ESTTA Tracking number: **ESTTA131228**
Filing date: **03/21/2007**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE TRADEMARK TRIAL AND APPEAL BOARD

Proceeding	92045000
Party	Plaintiff Bass Pro Trademarks, L.L.C.Bass Pro Trademarks, L.L.C. Bass Pro Trademarks, L.L.C. Bass Pro Trademarks, L.L.C. Bass Pro Trademarks, L.L.C. 2500 East Kearney Springfield, MO 65898 UNITED STATES
Correspondence Address	Dennis J.M. Donahue III Husch & Eppenberger, LLC 190 Carondelet Plaza Suite 600 St. Louis, MO 63105 UNITED STATES trademark@husch.com
Submission	Other Motions/Papers
Filer's Name	H. Frederick Rusche
Filer's e-mail	trademark@husch.com
Signature	/H. Frederick Rusche/
Date	03/21/2007
Attachments	Motion-Rebuttal Designations.PDF (4 pages)(935287 bytes) Rebuttal Designations pp1-50.pdf (50 pages)(2593335 bytes) Rebuttal Designations pp51-100.pdf (50 pages)(3657115 bytes) Rebuttal Designations pp101-150.pdf (50 pages)(3003498 bytes) Rebuttal Designations pp151-200.pdf (50 pages)(3063174 bytes) Rebuttal Designations pp201-250.pdf (50 pages)(2766608 bytes) Rebuttal Designations pp251-293.pdf (43 pages)(2227917 bytes)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE TRADEMARK TRIAL AND APPEAL BOARD

In the matter of Trademark Registration No. 2,390,988

For the mark SPORTSMAN'S WAREHOUSE HUNTING FISHING CAMPING RELOADING OUTERWEAR FOOTWEAR and Design

Date registered: October 3, 2000

Bass Pro Trademarks, L.L.C., PETITIONER, v. Sportsman's Warehouse, Inc., RESPONDENT	:	CANCELLATION NO. 92045000

PETITIONER'S OFFER OF DEPOSITION DESIGNATIONS INTO EVIDENCE AS REBUTTAL TESTIMONY

Pursuant to 37 C.F.R. §2.120(j)(3) and the stipulation of the parties in this matter, Petitioner, Bass Pro Trademarks, LLC, hereby submits the following designations of discovery deposition portions to be offered into evidence as rebuttal trial testimony in. Petitioner submits herewith copies of the deposition transcript pages listed herein, including all exhibits referenced in those page.

1. Deposition of Dale Smith Taken May 10, 2006:

   Page 9, line 19 to page 10, line 5

   Page 19, line 18 to page 20, line 6

   Page 28, line 18 to page 29, line 24

2. Deposition of Jamison Bryan Hensley Taken June 1, 2006:

Page 24, line 17 to page 25, line 3

3. Deposition of Stanley W. Lippelman Taken August 9, 2006:

Page 10, line 5 to page 11, line 10

Page 18, line 17 to page 20, line 23

Page 21, line 23 to page 24, line 20

4. Deposition of Gary T. Ford Taken September 15, 2006:

Petitioner also notes that Respondent Sportsman's Warehouse, Inc. has designated the entire depositions of Dale R. Smith, Scott Nielsen, Stuart Utgaard, Jim Hagale, Toni Miller, John David Hagel, April L. Henderson, Stanley W. Lippelman, Gary T. Ford, and Michael B. Mazis without identifying the particular portion(s) of such depositions upon which Respondent intends to rely as trial testimony. Consequently, while Petitioner has attempted to identify relevant portions of these depositions that may be relevant as rebuttal testimony, Petitioner is unable to determine the specific nature of the trial testimony that Respondent seeks to rely upon in this matter and is, therefore, not able to determine the full extent of rebuttal testimony that may be necessary. Therefore, Petitioner reserves the right to amend and supplement these rebuttal deposition designations and to request an opportunity to present additional live rebuttal testimony at such time as Respondent identifies the actual trial testimony upon which it seeks to rely in this matter.

Dated: March 21, 2007

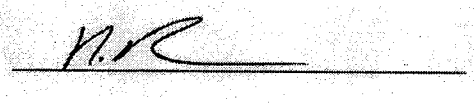
Respectfully submitted,

HUSCH & EPPENBERGER, LLC

Dennis J.M. Donahue III, Esq.
Dutro E. Campbell, II, Esq.
H. Frederick Rusche, Esq.
Husch & Eppenberger, LLC
190 Carondelet Plaza, Suite 600
St Louis, MO 63105
Phone: 314-480-1642
Fax: 314-290-5342
E-mail: trademark@husch.com

CERTIFICATE OF SERVICE

It is hereby certified that a true and correct copy of the foregoing document was served by first class mail, postage prepaid, on Christopher R. Smith, Lindquist & Vennum, PLLP, 80 South 8th Street, 4200 IDS Center, Minneapolis, MN 55402-2205, the attorney for Respondent, on this 21st day of March, 2007.



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE TRADEMARK TRIAL AND APPEAL BOARD

BASS PRO TRADEMARKS, L.L.C.,	)	
	)	Cancellation No. 92045000
Petitioner,	)	
	)	
vs.	)	Deposition of:
	)	DALE R. SMITH
SPORTSMAN'S WAREHOUSE, INC.,	)	
	)	
Respondent.	)	

May 10, 2006
10:00 a.m.

Sportsman's Warehouse
7035 High Tech Drive
Midvale, UT 84047

Sharon Morgan, CSR, RPR, CRR
Notary Public in and for the State of Utah

1 the attorneys, whoever, and we'll figure that out,
2 responded.
3 A. Okay.
4 Q. So the response part is under Answer, which
5 starts at like the middle of this page.
6 A. Okay.
7 Q. And if I'm reading this correctly, you show
8 up under B.
9 A. Yep. Yes.
10 Q. So if we go up to the top, it says, "Identify
11 the person or persons who are most knowledgeable about
12 (B) the selection and design of respondent's mark."
13 A. Okay.
14 Q. Respondent is Sportsman's Warehouse, and the
15 mark is the Sportsman's Warehouse logo with the
16 mountains and there's a registration. In fact, we'll
17 make that an exhibit later on.
18 A. Okay.
19 Q. So what I would like to know is what you know
20 about the selection and design of that mark.
21 A. At the time after Stu bought us, he asked my
22 wife, who happened to do all the advertising at that
23 time, to come up with a logo, basically, with our
24 printer. His name was Dave Egan. He passed away
25 about five years ago, but he was the printer at the

1 time. They designed it, still using the words and
2 then the departments underneath, hunting, fishing
3 camping, and the mountains in the background. They
4 presented it to Stu, he liked it, and it became the
5 mark.
6 (Exhibit No. 2 marked.)
7 Q. (By Mr. Upchurch) I've had the reporter mark
8 U.S. Trademark Registration No. 2,390,988.
9 A. Okay.
10 Q. Does that show the mark that we've been
11 talking about?
12 A. Yes.
13 Q. So that's the logo? What part did your wife
14 and the printer come up with?
15 A. Well, we already had the words "Sportsman's
16 Warehouse." They came up with the rest of it.
17 Q. The rest of it would be --
18 A. The style of the font and the mountains
19 behind, and then our categories underneath.
20 Q. You've said several times the categories
21 underneath. Are those the categories of products that
22 were sold in the store?
23 A. Yes.
24 Q. And is that still true today --
25 A. Yes.

1 Loveland.
2 Q. What did they say? What did the fellow in
3 Memphis say?
4 A. He just said at the time that it was causing
5 a little confusion, that they received a gift card. I
6 said, "Don't let the customers go away. Give them a
7 free hat, but we can't accept the gift card." And I
8 had a similar conversation with Brian Mackey in
9 Loveland.
10 Q. Have you had any other conversations with
11 store managers concerning Bass Pro and customers being
12 confused?
13 A. No.
14 Q. If you look on page 5, at the top -- and
15 you're welcome to look on the previous page to see
16 what the interrogatory was, if you want to.
17 A. Okay.
18 Q. What I'm interested in is the sentence that
19 starts, "Although Sportsman's Warehouse formally
20 changed the name on the store's sign to Sportsman's
21 Warehouse during 1995, respondent believes it had been
22 using the mark Sportsman's Warehouse prior to 1995."
23 Do you see that statement?
24 A. Uh-huh (affirmative).
25 Q. Do you know of any documents that show the

1 use of the mark Sportsman's Warehouse by your company
2 prior to 1995?
3 A. Documents, no.
4 Q. Do you know if Sportsman's Warehouse has
5 looked for such documents?
6 A. Yes, we have looked.
7 Q. Do you know what formed the basis of the
8 belief referred to in that sentence that Sportsman's
9 Warehouse had been using that mark prior to 1995?
10 A. Probably just that it was talked about. I
11 mean, we know it was spring of '95 sometime, but
12 trying to narrow down a date -- but it was talked
13 about prior to '95, yes, on the name change.
14 Q. Okay. And who would have talked about
15 changing the name prior to 1995?
16 A. Well, I remember talking with customers about
17 it on the sales floor.
18 Q. Okay. What did the customers say?
19 A. They would come in and say, how come you
20 don't carry baseball? That's a sport. How come you
21 don't carry golf? That's a sport. You ought to be
22 called Sportsman's Warehouse. I was very happy when
23 we changed the name.
24 Q. Because you didn't have to have that
25 conversation again?

1 Sportsman's Warehouse.
2 Q. (By Mr. Upchurch) Are you aware of this dba
3 Sportsman's Warehouse that your counsel just
4 mentioned?
5 A. Sports Warehouse dba Sportsman's Warehouse,
6 yes.
7 Q. And is that the previous company that you
8 talked about, the predecessor company, Sports
9 Warehouse?
10 A. I'm not sure why Stu set it up that way.
11 Q. But it's all one company?
12 A. It's all one company, yes. I think that's
13 how it was listed when we were incorporated.
14 Q. Do you own stock in Sportsman's Warehouse?
15 A. Yes.
16 Q. Is it publicly traded?
17 A. No.
18 Q. On pages 9 and 10 of Exhibit 1, there are a
19 list of stores. Are these all stores that are doing
20 business under the mark or under the name Sportsman's
21 Warehouse?
22 A. Yes.
23 MR. ALLGEYER: This is missing a store,
24 right?
25 THE WITNESS: Well, it's been relocated to --

1 MR. ALLGEYER: Okay. I see.
2 Q. (By Mr. Upchurch) So the date for store
3 No. 12 represents when that particular store moved to
4 its current location?
5 A. Yes. But it was prior to that that Pacific
6 Flyway relocated, because we made this building all
7 store.
8 Q. So the date that the first store in Midvale
9 opened would be much earlier than November of 2002?
10 A. The first store opened on Memorial weekend in
11 1996.
12 Q. And was the second store the one in Provo in
13 1998?
14 A. The second store being -- we relocated the
15 first store three times. So the first store in 1986
16 was the other half of Pacific Flyway wholesale which
17 opened up to retail. Then it moved to a building on
18 its own just down from Pacific Flyway, and then it
19 moved to this building in 1991.
20 Q. And then in 1998, another --
21 A. 1998, yes. Stu bought us in 1996 and in 1998
22 we added our second store under Stu's tenure.
23 Q. And that was the one in Provo?
24 A. Yes.
25 Q. Do you see the third column on pages 9 and 10

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE TRADEMARK TRIAL AND APPEAL BOARD

In the Matter of Trademark Registration No. 2,390,988

For the Mark SPORTSMAN'S WAREHOUSE HUNTING FISHING CAMPING RELOADING OUTERWEAR FOOTWEAR and DESIGN

Date registered: October 3, 2000

Bass Pro Trademarks, L.L.C.,

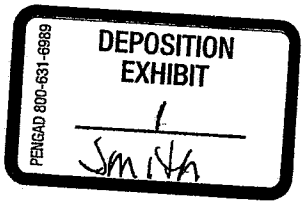
Petitioner,

Cancellation No. 92045000

v.

Sportsman's Warehouse, Inc.,

Respondent.



SPORTSMAN'S WAREHOUSE, INC.'S ANSWERS TO PETITIONER'S FIRST SET OF INTERROGATORIES AND REQUESTS FOR PRODUCTION OF DOCUMENTS

GENERAL RESPONSES

1. The General Responses and the General and Specific Objections apply to all answers of Respondent Sportsman's Warehouse, Inc. ("Sportsman's Warehouse") to Petitioner's interrogatories. In connection with these answers, Sportsman's Warehouse response is made without waiving or intending to waive, and expressly reserving: (a) the right to object on the grounds of competency, privilege, relevancy, materiality or any other proper ground to the use of the answers for any purpose, in whole or in part, in any subsequent step or proceeding in this action or any other action; (b) the right to object on any and all grounds at any time to other interrogatories or other discovery procedures involving or relating to the subject matter of the

CONFIDENTIAL

interrogatories; and (c) the right at any time to revise, correct add to or clarify any of the answers propounded herein.

2. The answers herein reflect the present state of Sportsman's Warehouse's knowledge and information following investigation regarding Petitioner's interrogatories. The investigation is continuing. Except as otherwise stated below, an objection to a interrogatory indicating that Sportsman's Warehouse will produce documents relating to the interrogatory does not necessarily mean that such documents exist. Sportsman's Warehouse reserves the right to supplement these answers, and to rely upon, at any time, including trial, subsequently discovered information or information omitted from these responses as a result of mistake, error, oversight, or inadvertence.

GENERAL OBJECTIONS

1. Sportsman's Warehouse objects to each interrogatory to the extent that it seeks information that is subject to the attorney-client privilege, work product privilege or other privilege on the ground that privileged matter is exempt from discovery.

2. Sportsman's Warehouse objects to any and all instructions or definitions on the grounds that they are overly broad and beyond the requirements imposed by the applicable rules of procedure.

3. Sportsman's Warehouse does not waive any of its general or particular objections in the event it furnishes information or documents coming within the scope of any such objections.

4. Any responsive documents to be produced will be produced at a time and place mutually agreeable to counsel.

CONFIDENTIAL

ANSWERS TO INTERROGATORIES

INTERROGATORY NO. 1: Identify the person or persons who are most knowledgeable about:

a. the selection of Respondent's Abandoned Mark;
b. the selection and design of Respondent's Mark;
c. the decision to apply for registration of Respondent's Mark in the United States Patent and Trademark Office;
d. any search reports, opinions and/or investigations prepared by or for Respondent which concern the combination of the terms SPORTSMAN'S and WAREHOUSE, or variants and synonyms thereof in whole or in part;
e. the preparation and filing of all U.S. applications by or for Respondent which include the phrase SPORTSMAN'S WAREHOUSE;
f. the use of trademarks and service marks in Respondent's advertising, marketing and promotional materials;
g. each of the services for which Respondent's Mark has been or are now being used, and the respective channels of trade and classes of purchasers for each of these services.

ANSWER: Sportsman's Warehouse objects to Interrogatory No. 1 and the grounds that it is overly broad and further objects to the extent that it seeks information protected by the attorney-client privilege or the attorney work product doctrine. Subject to and without waiving these objections or the foregoing General Objections, Sportsman's Warehouse responds as follows:

a. Bill Hayes, Scott Nielson, both former owners of Sportsman's Warehouse; Stu Utgaard, Chairman and Chief Executive Officer, Sportsman's Warehouse.
b. Dale Smith, President, Sportsman's Warehouse.
c. Stu Utgaard, Chairman and Chief Executive Officer, Sportsman's Warehouse, with assistance from counsel for Sportsman's Warehouse.
d. Stu Utgaard, Chairman and Chief Executive Officer, Sportsman's Warehouse, with assistance from counsel for Sportsman's Warehouse.
e. Stu Utgaard, Chairman and Chief Executive Officer, Sportsman's Warehouse, with assistance from counsel for Sportsman's Warehouse.
f. Stu Utgaard, Chairman and Chief Executive Officer, Sportsman's Warehouse.
g. Stu Utgaard, Chairman and Chief Executive Officer, Sportsman's Warehouse.

INTERROGATORY NO. 2: Describe all relevant facts and circumstances regarding Respondent's first awareness of Petitioner's Marks.

ANSWER: Respondent objects to Interrogatory No. 2 on the grounds that it vague and overly broad. Subject to and without waiving these objections and the foregoing General Objections, Sportsman's Warehouse responds as follows: During approximately 1997, Sportsman's Warehouse learned that Petitioner had opened a store near Atlanta, Georgia. This was the first time that Sportsman's Warehouse became aware that Petitioner was using the name SPORTSMAN'S WAREHOUSE. – no circumstances? who? how?

INTERROGATORY NO. 3: Describe all relevant facts and circumstances for each instance in which an individual was confused or mistaken, or otherwise uncertain, about a possible connection or affiliation between Petitioner's and Respondent's respective marks or services.

ANSWER: Sportsman's Warehouse objects to Interrogatory No. 3 to the extent that it is overly broad, unduly burdensome and seeks information that is not in Sportsman's Warehouse's possession. Subject to and without waiving this objection or the foregoing General Objections, Sportsman's Warehouse responds as follows: After Petitioner opened its store in St. Louis, Missouri, in 2000, customers attempted to redeem gift cards from Petitioner's store at Sportsman's Warehouse store locations, apparently because information on gift cards stated that the cards were redeemable at "Sportsman's Warehouse" stores. This same type of confusion occurred in 2005 after Petitioner issued similar gift cards from a store near Denver, Colorado.

INTERROGATORY NO. 4: Describe all relevant facts and circumstances regarding Respondent's first use of Respondent's Abandoned Mark in connection with each of the services identified in the abandoned application for Respondent's Abandoned Mark.

ANSWER: Respondent objects to Interrogatory No. 4 on the grounds that it is vague and overly broad. Subject to and without waiving these objections and the foregoing General Objections, Sportsman's Warehouse responds as follows: From 1986 through approximately 1995, Sportsman's Warehouse used the name SPORTS WAREHOUSE as the name of its retail sporting goods store in Midvale, Utah. During this time, customers often referred to the

operation as "Sportsman's Warehouse," which eventually led Respondent to change its name to SPORTSMAN'S WAREHOUSE. Sportsman's Warehouse is investigating the date on which it began using the mark SPORTSMAN'S WAREHOUSE in its retail stores. Although Sportsman's Warehouse formally changed the name on the store's sign to SPORTSMAN'S WAREHOUSE during 1995, Respondent believes it had been using the mark SPORTSMAN'S WAREHOUSE prior to 1995. Documents concerning the adoption of the mark SPORTSMAN'S WAREHOUSE will be produced to the extent they are available.

INTERROGATORY NO. 5: Describe all relevant facts and circumstances regarding Respondent's first use of Respondent's Mark in connection with each of the services identified in Reg. No. 2,390,988.

ANSWER: Respondent objects to Interrogatory No. 5 on the grounds that it is vague and overly broad. Subject to and without waiving these objections and the foregoing General Objections, Sportsman's Warehouse responds as follows: During 1995, after Sportsman's Warehouse had adopted the mark SPORTSMAN'S WAREHOUSE, it designed and began using the Respondent's Mark in connection with the sale and advertising of the services identified in Registration No. 2,390,988.

INTERROGATORY NO. 6: For each year, beginning with the date Respondent's Mark was first used in the United States, provide all relevant facts and circumstances regarding the Respondent's expenditures made in connection with Respondent's Mark for advertising, promoting and marketing each of the services identified in Reg. No. 2,390,988.

ANSWER: Respondent objects to Interrogatory No. 6 on the grounds that it is overly broad and unduly burdensome. Subject to and without waiving these objections and the foregoing General Objections, Sportsman's Warehouse responds as follows: All of Respondents' advertising is made in connection with Respondent's Mark. Sportsman's Warehouse is unable to compile information concerning every advertising expenditure because it would include thousands of advertisements and millions of dollars. Sportsman's Warehouse has

available and has compiled some general information, beginning with 1997. The available data is summarized below.

Sportsman's Warehouse
Advertising Expenditures
Year Ended October 31

1997	$	56,000
1998		107,000
1999		205,000
2000		264,000
2001		449,000
2002		563,000
2003		782,000
2004		1,792,000
2005		3,070,000
Total	**$**	**7,288,000**

Exterior Signage

1997	8,000 Est'd
1998	21,777
1999	31,394
2000	31,723
2001	113,675
2002	372,466
2003	555,596
2004	773,150
2005	1,154,720
Total	$3,062,501

Logo Merchandise
7/1/03 — 1/30/06
(Only Data Currently Available)

	Units	Dollars
Camping	1,032,034	$ 5,627,054
Clothing	472,718	6,718,164
Hunting	50,652	2,775,347
Fishing	138,690	1,584,647
Shoes	66,773	1,136,895
Gift Bar	63,132	348,412
	1,823,999	$ 18,190,519
Prior to 7/1/03	Est'd 200,000	Est'd 2,000,000
Grand Total	**2,023,999**	**$ 20,190,519**

CONFIDENTIAL

INTERROGATORY NO. 7: For each year, beginning with the date Respondent's Mark was first used in the United States, provide all relevant facts and circumstances regarding Respondent's total amount of sales made in connection with Respondent's Mark for each of the specific services identified in Reg. No. 2,390,988.

ANSWER: Respondent objects to Interrogatory No. 7 on the grounds that it vague and overly broad. Subject to and without waiving these objections and the foregoing General Objections, Sportsman's Warehouse responds as follows: Since 1995, Sportsman's Warehouse has completed millions of transactions under Respondent's Mark and the mark SPORTSMAN'S WAREHOUSE. Because of the enormous volume of transactions, Sportsman's Warehouse will provide a summary of its total sales by calendar year. The available data is summarized below:

All sales since late 1994 or early 1995 have been made under the Sportsman's Warehouse name. These sales exceed $1.4 billion.

Historical Sportsman's Warehouse Sales
Year Ending October 31
(Thousands of Dollars)

1995	$8,000,000 Est'd
1996	14,000,000
1997	20,519,000
1998	25,281,000
1999	39,726,000
2000	52,245,000
2001	77,484,000
2002	119,257,000
2003	118,129,000
2004	292,629,000
2005	415,592,000
2006 (3 Mos.)	155,717,000
Total	**$ 1,406,579,000**

CONFIDENTIAL

INTERROGATORY NO. 8: For each year, beginning with the date Respondent's Mark was first used in the United States, identify any and all businesses, retailers, or distributors to whom Respondent has sold goods at wholesale under Respondent's Mark.

ANSWER: Respondent objects to Interrogatory No. 8 on the grounds that it is vague and overly broad. Subject to and without waiving these objections and the foregoing General Objections, Sportsman's Warehouse responds that it sells goods only at retail and does not sell goods at a wholesale level.

INTERROGATORY NO. 9: Describe all relevant facts and circumstances regarding Respondent's relationship with any wholesale store or any other wholesale business, including but not limited to Pacific Flyway Wholesale, beginning with initiation of the relationship to the current state of the relationship and the development of the relationship therebetween.

ANSWER: Sportsman's Warehouse objects to Interrogatory No. 9 to the extent that it is overly broad, unduly burdensome, and not reasonably calculated to lead to the discovery of admissible evidence. Subject to and without waiving this objection or the foregoing General Objections, Sportsman's Warehouse responds as follows: The parent company of Sportsman's Warehouse, Sportsman's Warehouse Holdings, Inc., also owns Pacific Flyway. Pacific Flyway sells goods at wholesale to retailers. Pacific Flyway also acts as a distributor for goods sold at the retail stores of Sportsman's Warehouse.

INTERROGATORY NO. 10: Explain the differences between wholesale stores or other wholesale business, such as Pacific Flyway Wholesale, and Respondent's retail stores, including any differences between the respective channels of trade and classes of purchasers for each type of business.

ANSWER: Sportsman's Warehouse objects to Interrogatory No. 10 to the extent that it is vague, overly broad, unduly burdensome, and not reasonably calculated to lead to the discovery of admissible evidence. Subject to and without waiving this objection or the foregoing General Objections, Sportsman's Warehouse responds as follows: See Respondent's response to Interrogatory No. 9.

INTERROGATORY NO. 11: For each one of Respondent's stores that provides any service in connection with Respondent's Mark, identify the location of the store, the date that the store opened, and the scope of services currently provided by the store as pertaining to each one of the services identified in Reg. No. 2,390,988.

ANSWER: Respondent objects to Interrogatory No. 11 on the grounds that it is vague and overly broad. Subject to and without waiving these objections and the foregoing General Objections, Sportsman's Warehouse provides the following list of stores and the date they opened.

Location	Date Opened	Continuously Active States
1. Provo, UT	5/8/98	1 Utah
2. Riverdale, UT	4/26/99	
3. Meridian, ID	4/7/00	2 Idaho
4. Idaho Falls, ID	11/10/00	
5. Loveland, CO	5/18/01	3 Colorado
6. Spokane, WA	8/15/01	4 Washington
7. Grand Junction, CO	11/20/01	
8. Salem, OR	7/9/02	5 Oregon
9. Kennewick, WA	8/24/02	
10. St. Cloud, MN	8/30/02	6 Minnesota
11. Aurora, CO	10/4/02	
12. Midvale, UT	11/8/02	Successor Store to original Sportsman's Warehouse
13. Coon Rapids, MN	11/27/02	
14. Missoula, MT	2/26/03	7 Montana
15. Portland, OR	7/9/03	
16. Phoenix, AZ	8/21/03	8 Arizona
17. Twin Falls, ID	9/17/03	
18. Thornton, CO	10/2/03	
19. Ankeny, IA	1/26/03	9 Iowa
20. Anchorage, AK	2/9/04	10 Alaska
21. Silverdale, WA	5/14/04	
22. Lewisville, TX	5/21/04	11 Texas
23. Albuquerque, NM	8/19/04	12 New Mexico
24. Reno, NV	8/26/04	13 Nevada
25. New Berlin, WI	10/13/04	14 Wisconsin
26. Sioux Falls, SD	10/14/04	15 South Dakota
27. Round Pock, TX	10/18/04	
28. DePere, WI	10/28/04	
29. Goodyear, AZ	11/22/04	

Location	Date Opened	Continuously Active States
30. Fargo, ND	11/24/0416	16 North Dakota
31. Memphis, TN	6/15/05	17 Tennessee
32. Bend, OR	8/4/05	
33. Mesa, AZ	8/18/05	
34. San Antonio, TX	8/25/05	
35. San Antonio, TX	9/8/05	
36. Tucson, AZ	9/29/05	
37. Casper, WY	9/11/05	19 Wyoming
38. Oklahoma City, OK	9/29/05	18 Oklahoma
39. Southaven	10/5/05	20 Mississippi
40. Littleton, CO	10/7/05	
41. Fairbanks, AK	11/2/05	
42. Pittsburg, PA	11/23/05	21 Pennsylvania
43. Colorado Springs, CO	1/23/05	

Stores Under Construction

Location	Scheduled Opening Date	
44. St. George, UT	4/6/06	
45. Rodgers, AR	4/13/06	22 Arkansas
46. Wichita, KS	4/27/06	23 Kansas
47. West Bend, WI	5/4/06	
48. Woodbury, MN	5/1/06	
49. Chattanooga, TN	6/1/06	
50. Federal Way, WA	6/8/06	
51. Las Cruces, NM	8/3/06	
52. Lacey, WA	8/24/06	
53. Columbia, SC	8/31/06	24 South Carolina
54. Lafayette, IN	8/31/06	25 Indiana

Stores Scheduled to Start Construction

55. Rocklin, CA	9/21/06	26 California
56. Bozeman, MT	9/28/06	
57. Medford, OR	10/26/06	
58. Vancouver, WA	10/26/06	

INTERROGATORY NO. 12: Identify all persons or entities to whom Respondent has licensed or granted authority to use Respondent's Mark in connection with any goods or services.

ANSWER: Respondent objects to Interrogatory No. 12 on the grounds that it is vague and seeks information that is not reasonably calculated to lead to the discovery of admissible evidence. Subject to and without waiving these objections and the foregoing General Objections, Sportsman's Warehouse responds that it has granted a license to use the mark SPORTSMAN'S NEWS and DESIGN, U.S. Serial No.78/566,721, in connection with its news publication SPORTSMAN'S NEWS. Copies of the publication are available to customers at Sportsman's Warehouse store locations and by subscription. A copy of the License Agreement will be produced.

INTERROGATORY NO. 13: Describe all relevant facts and circumstances regarding Respondent's use of the phrase SPORTSMAN'S WAREHOUSE in the United States in combination with the registered trademark symbol, ®, but not in combination with the other words that comprise Respondent's Mark (i.e., HUNTING FISHING CAMPING RELOADING OUTERWEAR FOOTWEAR) or not in combination with the mountain range design element of the Respondent's Mark. If your answer is other than an unqualified negative, explain all relevant facts and circumstance' .s regarding Respondent's use in its advertisements, promotions, flyers and web site page of SPORTSMAN'S WAREHOUSE AMERICA'S PREMIER OUTFITTER in combination with the mountain range design and the registered trademark symbol but excluding the additional words that comprise Respondent's Mark.

ANSWER: Respondent objects to Interrogatory No. 13 on the grounds that it is vague, overly broad, and seeks information that it irrelevant and not reasonably calculated to lead to the discovery of admissible evidence. Subject to and without waiving these objections and the foregoing General Objections, Sportsman's Warehouse responds as follows: Sportsman's Warehouse is not aware of any such use. To extent there has been any use of the ® symbol not in connection with the Respondent's design, such use of the ® symbol has been inadvertent and *de minimis* in nature.

INTERROGATORY NO. 14: State all facts and identify all documents upon which Respondent relies in claiming that (a) "Respondent has superior rights over Petitioner in the mark SPORTSMAN'S WAREHOUSE" and (b) "Petitioner uses the phrase SPORTSMAN'S WAREHOUSE only in conjunction with the mark BASS PRO SHOPS" and (c) "Petitioner has abandoned any rights...in its registered BASS PRO SHOPS SPORTSMAN'S WAREHOUSE and DESIGN mark."

ANSWER: Respondent objects to Interrogatory No. 14 on the grounds that it is vague, overly broad, and seeks information protected by attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the foregoing General Objections, Sportsman's Warehouse responds as follows:

(a) The public associates the mark SPORTSMAN'S WAREHOUSE with Respondent's business. Respondent operates more than forty stores across the nation under the mark SPORTSMAN'S WAREHOUSE, and has been using the mark since at least 1995, and perhaps earlier. Respondent has conducted millions of transactions and generated hundreds of millions of dollars in sales under the mark SPORTSMAN'S WAREHOUSE. Because of the widespread success and consumer association of the mark SPORTSMAN'S WAREHOUSE with Respondent's retail sales, Respondent has acquired secondary meaning in the mark. Petitioner, by contrast, currently has only one location that uses the phrase "Sportsman's Warehouse" in connection with a retail store. Petitioner conducts all of its business under the BASS PRO SHOPS mark, and to Respondent's knowledge, does not use the phrase "Sportsman's Warehouse" apart from the BASS PRO SHOPS mark. Respondent has disclaimed exclusive rights to the phrase SPORTSMAN'S WAREHOUSE apart from the BASS PRO SHOPS mark, and has not acquired any secondary meaning in that phrase. Thus, Respondent's rights in the mark SPORTSMAN'S WAREHOUSE are superior to Petitioner.

Additionally, Respondent is investigating Petitioner's claim to first use of the phrase Sportsman's Warehouse. Petitioner apparently began use of the mark on March 1, 1995, not on January 2, 1995, as claimed in its registered mark. Respondent also has reason to believe that its date of first use of the mark SPORTSMAN'S WAREHOUSE is prior to Petitioner's date of first use. Respondent's investigation is continuing.

(b) Sportsman's Warehouse is not aware of any instance where Petitioner has used the phrase SPORTSMAN'S WAREHOUSE apart from the mark BASS PRO SHOPS. Petitioner disclaimed the phrase "Sportsman's Warehouse" in connection with its trademark registration.

(c) The manner in which Petitioner is using its BASS PRO SHOPS SPORTSMAN'S WAREHOUSE and DESIGN mark is different than the manner specified in the trademark registration for this mark. Specifically, the registered mark contains the stylized BASS PRO SHOPS mark above the phrase Sportsman's Warehouse, with that phrase in horizontal format and in lower case letters. Respondent's remaining BASS PRO SHOPS stores and advertising appear to have the Sportsman's Warehouse phrase in a different lettering and format than is depicted in the registration.

Additionally, only Petitioner's store in the St. Louis market currently uses the phrase "Sportsman's Warehouse" in connection with retail sales. The use in only one state may not satisfy the requirement of use in interstate commerce, and thus Petitioner may no longer have any federal trademark rights.

Finally, in the Memphis market, Petitioner changed the name of its retail store from BASS PRO SHOPS SPORTSMAN'S WAREHOUSE to BASS PRO SHOPS SPORTSMAN'S CENTER as a result of competition with Respondent's SPORTSMAN'S WAREHOUSE store. This decision to change the store name shows that Petitioner has not acted as the owner of the mark SPORTSMAN'S WAREHOUSE, and shows that any rights it claims in the mark have been abandoned.

INTERROGATORY NO. 15: Identify all third parties that Respondent knows or believes to be using (or to have used) any combination of the terms SPORTSMAN'S and WAREHOUSE, or variants and synonyms thereof in whole or in part, in connection with retail or wholesale store services.

ANSWER: Respondent objects to Interrogatory 15 on the grounds that it is vague, overly broad, and seeks information protected the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the foregoing General Objections, Sportsman's Warehouse states as follows:

Petitioner has taken action against companies that have infringed its rights in the mark SPORTSMAN'S WAREHOUSE. During early 2006, Respondent became aware that there was a store on www.ebay.com that was selling goods under the store name "Sportsman's Warehouse." Respondent sent the store operator a cease and desist letter and also contacted ebay to cease this infringement. As a result of Respondent's actions, ebay ordered the store owner to change its name. Respondent will produce responsive documents.

Additionally, several years ago Respondent exchanged correspondence with a company that was attempting to file state business and/or corporate names including the words Sportsman's Warehouse. As a result of correspondence from Respondent, that company ceased its efforts to file for such names. Responsive documents will be produced to the extent they exist.

INTERROGATORY NO. 16: Identify all third parties to whom Respondent has caused to be sent a cease and desist letter or other communication asserting Respondent's rights in Respondent's Mark or in any combination of the terms SPORTSMAN'S and WAREHOUSE, or variants and synonyms thereof in whole or in part. If no such party exists, describe all relevant facts and circumstances regarding Respondent's actions that protect Respondent's Mark against improper uses by third patties.

ANSWER: Sportsman's Warehouse objects to Interrogatory No. 16 to the extent that it seeks information protected by the attorney-client privilege or attorney work product doctrine. Subject to and without waiving this objection or the foregoing General Objections, Sportsman's Warehouse responds as follows: See response to Interrogatory 15. Respondent will produce responsive documents concerning cease and desist letters it has sent.

INTERROGATORY NO. 17: Identify each expert that Respondent or any person acting for or on behalf of Respondent has consulted with or whose services have been retained with respect to any of the issues involved in this action, and for each, identify all documents upon which said expert will base his/her expert opinion, and specify the subject matter on which he/she was consulted or retained.

ANSWER: Sportsman's Warehouse objects to Interrogatory No. 17 to the extent that it seeks information protected by the attorney-client privilege or attorney work product doctrine. Subject to and without waiving this objection or the foregoing General Objections, Sportsman's Warehouse responds as follows: Sportsman's Warehouse has not yet retained an expert to testify at the trial of this matter. Sportsman's Warehouse will supplement this response as necessary.

INTERROGATORY NO. 18: With regard to Respondent's policies, either written or implicitly followed:

a. Identify any policy or practice with respect to retention and/or destruction of documents used by Respondent. If your answer is other than an unqualified negative, describe that policy or practice in detail and identify all documents evidencing, recording, summarizing, referring or relating thereto.

b. Identify any formal, written policy with respect to the proper use of Respondent's Mark. If no formal, written policy exists, describe all relevant facts and circumstances regarding Respondent's practice that promotes the proper use of Respondent's Mark.

ANSWER: Respondent objects to Interrogatory No. 18 on the grounds it is vague and overly broad, and seeks information that is irrelevant and not reasonably calculated to lead to the discovery of admissible evidence. Subject to and without waiving these objections and the foregoing General Objections, Sportsman's Warehouse responds as follows:

a. None.

b. Sportsman's Warehouse does not have a formal written policy concerning the use of its trademarks, but takes all reasonable steps to ensure that its use of the marks is proper.

INTERROGATORY NO. 19: If any interrogatory asks for information which could at some time have been answered by producing, consulting or referring to documents which are no longer in existence, identify each interrogatory for which that is the case and, with respect to each interrogatory, identify such documents and all facts and circumstances regarding the creation, maintenance and destruction of such documents.

ANSWER: Subject to and without waiving the foregoing General Objections, Sportsman's Warehouse responds as follows: Sportsman's Warehouse has reason to believe that it used the mark SPORTSMAN'S WAREHOUSE in advertising and on in-store signage at least as early as 1994, but is still attempting to locate the necessary documents related to this time period. Responsive documents will be produced to the extent they are available.

INTERROGATORY NO. 20: Identify each document, if any, which is being withheld from document production pursuant to the requests for production of documents below on the grounds of privilege or otherwise and state for each document withheld:

a. The type of document;
b. The date(s);
c. The persons who prepared the document;
d. All persons who received a copy;
e. Reason for non-production; and
f. The subject matter

ANSWER: Subject to and without waiving the foregoing General Objections, Sportsman's Warehouse states as follows: Sportsman's Warehouse will provide a privilege log, to the extent such a log is necessary, in connection with its production of documents.

INTERROGATORY NO. 21: Identify each person who assisted in the preparation of the response to these interrogatories and state the number of each interrogatory in connection with which such person rendered such assistance.

ANSWER: Respondent objects to Interrogatory No. 21 on the grounds that it seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without these objections and the General Objections, Sportsman's Warehouse states that Stu Utgaard and Dale Smith assisted in gathering information for these responses, with the assistance of counsel.

INTERROGATORY NO. 22: Identify all persons who assisted in the collection of documents in response to the requests for production of documents below, including for each such person, each separate request for which he or she assisted in the collection of documents.

ANSWER: Respondent objects to Interrogatory No. 22 on the grounds that it seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without these objections and the General Objections, Sportsman's Warehouse states that Stu Utgaard and Dale Smith assisted in gathering documents, with the assistance of counsel.

RESPONSES TO REQUESTS FOR PRODUCTION

REQUEST NO. 1: All documents concerning Respondent's selection, adoption and design of Respondent's Abandoned Mark, Respondent's Mark, and any variants thereof.

RESPONSE: Respondent objects to Request No. 1 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent will product responsive documents to the extent they exist.

REQUEST NO. 2: All documents concerning when and how Respondent first became aware of Petitioner's use of Petitioner's Marks.

RESPONSE: Respondent objects to Request No. 2 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent will produce responsive documents to the extent they exist.

REQUEST NO. 3: All documents concerning any decisions and attempts by Respondent to apply for registration of Respondent's Abandoned Mark, Respondent's Mark, and any variants thereof in the United States Patent and Trademark Office.

RESPONSE: Respondent objects to Request No. 3 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections,

Respondent states as follows: Respondent will produce responsive documents to the extent they exist.

REQUEST NO. 4: All documents concerning Respondent's first use of Respondent's Abandoned Mark and Respondent's Mark in connection with each of the services identified in the respective applications for the marks.

RESPONSE: Respondent objects to Request No. 4 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent will produce responsive documents to the extent they exist.

REQUEST NO. 5: All documents concerning the development, distribution and publication of advertising, marketing and promotional materials that use Respondent's Mark.

RESPONSE: Respondent objects to Request No. 5 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent's mark is used in all of its stores, and all of its advertising and other materials. The volume of documents is so great that Respondent cannot possibly comply with the enormous scope of this request. Respondent will produce representative samples of some advertising and promotional materials.

REQUEST NO. 6: Documents sufficient to identify the amount of sales made by Respondent or any parties authorized by Respondent to sell goods in connection with Respondent's Mark.

RESPONSE: Respondent objects to Request No. 6 on the grounds that it is vague and overly broad, seeks information that confidential, proprietary, irrelevant and not reasonably calculated to lead to the discovery of admissible evidence, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these

objections and the General Objections, Respondent states that it has provided sufficient details of its sales in the interrogatory answers above.

REQUEST NO. 7: All documents concerning Respondent's marketing plans with respect to sales made in connection with Respondent's Mark.

RESPONSE: Respondent objects to Request No. 7 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states that it has no such marketing plans.

REQUEST NO. 8: Documents sufficient to identify the various channels of trade though which Respondent sells goods in connection with Respondent's Mark.

RESPONSE: Respondent objects to Request No. 8 on the grounds that it is vague and overly broad. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent sells goods in connection with Respondent's Mark only at retail.

REQUEST NO. 9: All documents concerning any market research, consumer tests or surveys, focus groups or related studies with respect to Respondent's Abandoned Mark, Respondent's Mark or the combination of terms SPORTSMAN'S and WAREHOUSE, or variants and synonyms thereof in whole or in part.

RESPONSE: Respondent objects to Request No. 9 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent will produce responsive documents, if any exist.

REQUEST NO. 10: All documents concerning any surveys analyzing the likelihood of confusion between Petitioner and Respondent or their respective marks.

RESPONSE: Respondent objects to Request No. 10 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work

product doctrine. Subject to and without waiving these objections and the General Objections, Respondent will produce responsive documents, if any exist.

REQUEST NO. 11: All documents concerning any statements, inquiries, comments or other communications by any person evidencing or concerning any confusion, mistake, suspicion, belief or doubt as to a possible connection between Respondent and Petitioner or their respective services or marks.

RESPONSE: Respondent objects to Request No. 11 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent will produce responsive documents, if any exist.

REQUEST NO. 12: All documents concerning any plans or proposals to expand Respondent's use of Respondent's Mark.

RESPONSE: Respondent objects to Request No. 12 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent will produce responsive documents to the extent they exist.

REQUEST NO. 13: All documents concerning or relating to any license agreements between Respondent and any and all third parties to whom Respondent has granted a license or the right to use Respondent's Mark, and all documents evidencing or otherwise concerning any promises, contracts and other mutual agreements between Respondent and Pacific Flyway Wholesale.

RESPONSE: Respondent objects to Request No. 13 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent will produce responsive documents to the extent they exist.

REQUEST NO. 14: Documents sufficient to identify each different manner in which Respondent's Mark has been and is being used in connection with each of the services identified in Reg. No. 2,390,988.

RESPONSE: Respondent objects to Request No. 14 on the grounds that the phrase "each different manner" is vague and overly broad. Respondent is unable to determine which documents are being requested.

REQUEST NO. 15: Documents sufficient to identify all retailers or distributors to whom Respondent has sold goods or products at wholesale under Respondent's Mark and the nature of all goods and products sold at wholesale under Respondent's Mark.

RESPONSE: Respondent objects to Request No. 15 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent does not sell at wholesale under Respondent's Mark, and therefore no responsive documents exist.

REQUEST NO. 16: For each year, beginning with the date of first use in the United States of the Respondent's Mark through the present, representative documents demonstrating the use of Respondent's Mark in connection with retail store services and wholesale store services.

RESPONSE: Respondent objects to Request No. 16 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent states that it will produce some samples of its use of Respondent's Mark in connection with retail store services.

REQUEST NO. 17: For each year, beginning with the date of first use in the United States of the Respondent's Mark through the present, representative documents sufficient to demonstrate the total amount of money spent on any advertisement of goods and/or services in connection with Respondent's Mark, including expenditures for promoting and marketing thereof.

RESPONSE: Respondent objects to Request No. 17 on the grounds that it is vague and overly broad, and seeks information that is confidential, proprietary, irrelevant, and not reasonably calculated to lead to the discovery of admissible evidence. Subject to and without

waiving these objections and the General Objections, Respondent states as follows: Respondent states that it has provided sufficient details of its advertising expenditures in its responses to the interrogatories above.

REQUEST NO. 18: For each year, beginning with the date of first use in the United States of the Respondent's Mark through the present, all documents sufficient to identify the person or persons who participated in the placement of advertising related to goods and/or services in connection with which such marks have been utilized by Respondent, including the account manager for Respondent at an advertising firm and the as well as Respondent.

RESPONSE: Respondent objects to Request No. 18 on the grounds that it is vague and overly broad, and seeks information that is confidential, proprietary, irrelevant, and not reasonably calculated to lead to the discovery of admissible evidence. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent states that it has provided sufficient details of its advertising personnel in its responses to the interrogatories above.

REQUEST NO. 19: Any and all documents otherwise supporting or controverting the allegations contained in Respondent's Affirmative Defenses, including by way of example but not limitation, all documents upon which Respondent intends to rely in arguing that (a) "Respondent has superior rights over Petitioner in the mark SPORTSMAN'S WAREHOUSE" and (b) "Petitioner uses the phrase SPORTSMAN'S WAREHOUSE only in conjunction with the mark BASS PRO SHOPS" and (c) "Petitioner has abandoned any rights ... in its registered BASS PRO SHOPS SPORTSMAN'S WAREHOUSE and DESIGN mark."

RESPONSE: Respondent objects to Request No. 19 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent will produce responsive documents to the extent they exist.

REQUEST NO. 20: All documents that may show any third party, not related to Respondent, using any combination of the terms SPORTSMAN'S and WAREHOUSE, or variants and synonyms thereof in whole or in part, in connection with retail or wholesale services.

RESPONSE: Respondent objects to Request No. 20 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent will produce responsive documents to the extent they exist.

REQUEST NO. 21: All documents, including cease and desist letters, correspondence or any other evidence of communications with a third party, in which Respondent asserts or has asserted any rights in Respondent's Mark or in any combination of the terms SPORTSMAN'S and WAREHOUSE, or variants and synonyms thereof in whole or in part.

RESPONSE: Respondent objects to Request No. 21 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent will produce responsive documents to the extent they exist.

REQUEST NO. 22: All documents relating to any and all consumer surveys, market research, focus groups or other related studies conducted by or on behalf of Respondent concerning whether Respondent's Mark is distinctive or whether Respondent's Mark has acquired secondary meaning.

RESPONSE: Respondent objects to Request No. 22 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent will produce responsive documents to the extent they exist.

REQUEST NO. 23: All documents referring or relating to any policy or practice with respect to Respondent's retention and/or destruction of documents.

RESPONSE: Respondent objects to Request No. 23 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work

product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: No responsive documents exist.

REQUEST NO. 24: All documents relating to any policy regarding the proper use of Respondent's Mark and any other use of SPORTSMAN'S WAREHOUSE and all documents showing the use of the phrase SPORTSMAN'S WAREHOUSE in the United States in combination with the registered trademark symbol, ®, but not in combination with the other words that to comprise Respondent's Mark (i.e., HUNTING FISHING CAMPING RELOADING OUTERWEAR FOOTWEAR) or not in combination with the mountain range design element of the Respondent's Mark, including all documents relating to any discovery of such use, any discontinuation or correction of use once discovered, and any actions taken to prevent such use, including policy changes that have been considered or implemented.

RESPONSE: Respondent objects to Request No. 24 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent will produce responsive documents to the extent they exist.

REQUEST NO. 25: Any and all documents that relate to answers provided to the interrogatories above, or which are required to be identified in response to these interrogatories, even if such documents are not otherwise specifically requested in these requests for production.

RESPONSE: Respondent incorporates all of its objections to Petitioner's interrogatories. Subject to and without waiving those objections, Respondent will produce responsive documents.

REQUEST NO. 26: Any and all documents concerning the legal and operational relationship between Respondent and any affiliated or related entities, including subsidiaries, parent companies, and licensees.

RESPONSE: Respondent objects to Request No. 26 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent will produce responsive documents to the extent they exist.

REQUEST NO. 27: Any and all documents concerning any legal opinions concerning the Respondent's Mark, including availability of the Respondent's Mark prior to Respondent's adoption thereof or likelihood of confusion of Respondent's Mark and any other mark.

RESPONSE: Respondent objects to Request No. 27 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. privilege log

REQUEST NO. 28: Any marketing or consumer surveys undertaken concerning the Respondent's Mark, including individual questionnaires, pilot studies and focus groups.

RESPONSE: Respondent objects to Request No. 28 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: No responsive documents exist.

REQUEST NO. 29: Any and all radio scripts, television videotapes, newspaper advertisements, point of sale displays, flyers and any other advertisements or promotional materials created or used at any time concerning Respondent's Mark, including layouts, preliminary drawings and artwork as well as all material available to Respondent's distributors or members of the general public.

RESPONSE: Respondent objects to Request No. 29 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent will produce responsive documents to the extent they exist.

REQUEST NO. 30: Any and all documents concerning publicity (both solicited and unsolicited) relating to Respondent's Mark, including press releases and articles of any kind.

RESPONSE: Respondent objects to Request No. 30 on the grounds that it is vague and overly broad. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent will produce responsive documents to the extent they exist.

REQUEST NO. 31: Two specimen samples of each advertisement and promotional piece concerning Respondent's Mark.

RESPONSE: Respondent objects to Request No. 31 on the grounds that it is vague and overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent's mark is used in all of its stores, and all of its advertising and other materials. The volume of documents is so great that Respondent cannot possibly comply with the enormous scope of this request. Respondent will produce representative samples of some advertising and promotional materials.

REQUEST NO. 32: Any and all documents concerning all past and current licenses, franchises, sublicenses, sub-franchises, distributorships and assignments and other agreements, together with related correspondence or other documents concerning Respondent's Mark.

RESPONSE: Respondent objects to Request No. 32 on the grounds that it is vague and overly broad. Subject to and without waiving these objections and the General Objections, Respondent states that no responsive documents exist.

REQUEST NO. 33: All co-existence agreements, or other agreements addressing the concurrent use of similar marks, concerning the Respondent's Mark, including memoranda and correspondence concerning the negotiation thereof.

RESPONSE: Respondent objects to Request No. 33 on the grounds that it is vague and overly broad. Subject to and without waiving these objections and the General Objections, Respondent states that no responsive documents exist.

REQUEST NO. 34: Any and all documents concerning the stores, retail outlets, wholesale businesses, offices or other locations at which or with whom Respondent markets, promotes or sells in connection with Respondent s Mark.

RESPONSE: Respondent objects to Request No. 34 on the grounds that it is vague and overly broad. Subject to and without waiving these objections and the General Objections,

Respondent already has provided sufficient information in the interrogatory responses listing its current store locations.

REQUEST NO. 35: Any and all documents concerning the manner of marketing (past, current and contemplated) of all goods sold in connection with Respondent's Mark.

RESPONSE: Respondent objects to Request No. 35 on the grounds that it is vague and overly broad. Subject to and without waiving these objections and the General Objections, Respondent states that it will produce some samples of its advertising in connection with Respondent's Mark.

REQUEST NO. 36: Any and all documents concerning the class of purchasers for goods sold in connection with Respondent's Mark, including past, current and projected class(es) of purchasers.

RESPONSE: Respondent objects to Request No. 36 on the grounds that it is vague and overly broad. Subject to and without waiving these objections and the General Objections, Respondent states that it offers retail stores services in connection with Respondent's Mark and cannot determine which documents are requested by the wording "concerning the class of purchasers."

REQUEST NO. 37: Documents evidencing or otherwise concerning the annual amount of expenditures for advertising and/or promotions involving Respondent's Mark.

RESPONSE: See Respondent's response to Request No. 17.

REQUEST NO. 38: Documents evidencing or otherwise concerning communications threatening or initiating litigation, arbitration or proceedings before the United States Trademark Trial and Appeal Board (including opposition and cancellation proceedings), or other adversarial proceedings concerning Respondent's Mark.

RESPONSE: See Respondent's response to Request No. 21 and Interrogatories Nos. 15 and 16.

REQUEST NO. 39: Documents evidencing or otherwise concerning any pending or concluded litigation, arbitration or proceeding before the United States Trademark Trial and Appeal Board (including opposition and cancellation proceedings), or other adversarial proceedings concerning Respondent's Mark.

RESPONSE: Respondent objects to Request No. 39 on the grounds that it is vague, overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states that there are no responsive documents, except for this proceeding.

REQUEST NO. 40: All documents related to Respondent's policing of Respondent's Mark or any combination of the terms SPORTSMAN'S and WAREHOUSE, or variants and synonyms thereof in whole or in part, including any trademark watch services.

RESPONSE: Respondent objects to Request No. 40 on the grounds that it is vague, overly broad, and seeks information that is irrelevant, not reasonably calculated to lead to the discovery of admissible evidence, and protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states that there are no responsive documents

REQUEST NO. 41: Any and all settlement agreements concerning Respondent's Mark.

RESPONSE: Respondent objects to Request No. 41 on the grounds that it is vague, overly broad, and seeks information protected by the attorney-client privilege and the work product doctrine. Subject to and without waiving these objections and the General Objections, Respondent states that there are no responsive documents.

REQUEST NO. 42: Any and all documents setting forth the management structure of Respondent, any related companies, and any subsidiaries during the period of time in which Respondent has used Respondent's Abandoned Mark, Respondent's Mark, or any combination of the terms SPORTSMAN'S and WAREHOUSE, or variants and synonyms thereof in whole or in part.

RESPONSE: Respondent objects to Request No. 42 on the grounds that it is vague and overly broad, and seeks information that is confidential, proprietary, irrelevant, and not reasonably calculated to lead to the discovery of admissible evidence. Subject to and without waiving these objections and the General Objections, Respondent states as follows: Respondent will produce responsive documents sufficient to show its corporate structure.

AS TO OBJECTIONS:

Dated: February 20, 2006

LINDQUIST & VENNUM P.L.L.P.

By: Garrett Weber

Garrett M. Weber
80 South 8th Street
4200 IDS Center
Minneapolis, MN 55402-2205
(612) 371-3211

ATTORNEYS FOR RESPONDENT SPORTSMAN'S WAREHOUSE, INC.

AS TO ANSWERS:

SPORTSMAN'S WAREHOUSE, INC.

By: _____
(Signature)

_____
(Printed Name)

Subscribed and sworn to before me
this _____ day of February, 2006.

_____
Notary Public

CONFIDENTIAL

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE TRADEMARK TRIAL AND APPEAL BOARD

BASS PRO TRADEMARKS, L.L.C.,	
PETITIONER,	
VS.	CANCELATION NO. 92045000
SPORTSMAN'S WAREHOUSE, INC.,	
RESPONDENT.	

DEPOSITION
OF
JAMISON BRYAN HENSLEY
THURSDAY, JUNE 1, 2006

ALPHA REPORTING CORPORATION
236 Adams Avenue
Memphis, TN 38103
901-523-8974
www.alphareporting.com

1 A. Yes. And on occasion we'll have something sent
2 out that does say Sportman's Center.
3 Q. What's an example of something that would be
4 sent out that says Sportman's Center?
5 A. We send out ads. We send out typically a
6 biweekly ad, or tabs as it's referred to, and on the
7 back when it gives directions, it will typically say,
8 nine times out of ten, it says Sportsman's
9 Warehouse.
10 Q. What are some examples of some marketing
11 material that says Sportman's Center?
12 A. Sometimes an ad will say Sportman's Center on
13 the back. But like I said, nine times out of ten it
14 does say Sportsman's Warehouse.
15 Q. Any other examples besides those ads?
16 A. Not to my knowledge.
17 Q. Do you have any idea why the marketing
18 materials were not changed to reflect the change in
19 the sign on the outside of the building?
20 A. Because we still refer to ourselves as
21 Sportsman's Warehouse, and our customers know us as
22 Sportsman's Warehouse.
23 Q. As a merchandising manager, did it strike you
24 that the difference between the store signage on the
25 outside of the building and the marketing materials

1 might confuse your customers?
2 A. No. What confused our customers was the
3 competitor having our name on their building.
4 Q. Did you ever have customers point out the
5 difference between the sign on the outside of the
6 building and the marketing materials?
7 A. No.
8 Q. Does your store or company have a policy about
9 keeping track of customer concerns or comments about
10 any confusion?
11 A. Yes.
12 Q. What is that policy?
13 A. It's not necessarily about any confusion, it's
14 about any type of issues inside the store. A
15 customer comes in and fills out a comment card and
16 they drop it in the box.
17 My administrative manager goes through those on
18 a weekly basis. We obviously call the customers back
19 and take care of any issues that are there that
20 arise.
21 Q. Are all your customer comments or complaints in
22 writing in that box?
23 A. Well, when a customer turns them in they go in
24 the box. Do they stay in the box indefinitely? No.
25 Q. Do you get oral complaints or concerns from

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE TRADEMARK TRIAL AND APPEAL BOARD

In the matter of Trademark Registration No. 2,390,988

For the mark SPORTSMAN'S WAREHOUSE HUNTING FISHING CAMPING RELOADING OUTERWEAR FOOTWEAR and Design

Date registered: October 3, 2000

BASS PRO TRADEMARKS, L.L.C.,	)	
	)	
Petitioner,	)	
	)	CANCELLATION
vs.	)	NO. 92045000
	)	
SPORTSMAN'S WAREHOUSE, INC.,	)	
	)	
Respondent.	)	

TELEPHONE DEPOSITION OF STANLEY W. LIPPELMAN

TAKEN ON BEHALF OF RESPONDENT

AUGUST 9, 2006

COPY

1 **Missouri, where my last job, I was the vice president**
2 **of marketing and research and development. And three**
3 **years ago approximately, a little less than three years**
4 **ago, came to Bass Pro as vice president of marketing.**
5 Q. Okay. So that would have put your starting
6 date somewhere in 2003?
7 **A. It was February of -- that would have**
8 **been -- it was either 2003 or 2004. I'm coming up on**
9 **my -- let's see. That would be 2003 -- 2004. I'm**
10 **sorry. February of 2007 will be my third year.**
11 Q. And you came in as -- what was your title
12 when you began in February of 2004 at Bass Pro Shops?
13 **A. Vice president of marketing.**
14 Q. Has your title changed since the time you
15 started?
16 **A. No.**
17 Q. And you are still vice president of
18 marketing?
19 **A. Yes.**
20 Q. Could you describe your duties as vice
21 president of marketing?
22 **A. I'm responsible for all of the -- let's just**
23 **say most of the advertising, promotions, branding,**
24 **analysis, direct mail. Our public relations group**
25 **reports to me. Our incentive group reports to me.**

1 **We have an in-house production group on the**
2 **broadcast side that does our television shows, of which**
3 **we produce three right now for the Outdoor Channel and**
4 **a syndicated radio show that reports up through me. I**
5 **have a total of around 50 to 60 people in my**
6 **department.**
7 Q. So it's safe to say if it has to do with
8 advertising at Bass Pro Shops, you are the top person?
9 **A. Yes. Well, short of Mr. Hagale, who I**
10 **report to, and obviously John Morris, our founder.**
11 Q. Okay. Does your job entail design or
12 creation of exterior or interior signage of the retail
13 stores?
14 **A. No, it does not.**
15 Q. Okay. Does your job entail the marketing
16 materials that are put out in the stores?
17 **A. Yes, it does.**
18 Q. Okay. Does that involve some signage?
19 **A. Not permanent signage. We do like sale**
20 **signage and that kind of thing, but not the -- not the**
21 **permanent signage, the wood fixtures that you see in**
22 **the store. My group handles everything from basically**
23 **8 feet down.**
24 Q. Okay.
25 **A. And our design and development group handles**

1 **A. To the best of my knowledge, two.**
2 Q. Okay. Could you tell me where those are?
3 **A. Uh-huh. In Memphis, Tennessee, and**
4 **St. Louis, Missouri.**
5 Q. And that is -- the St. Louis store is the
6 St. Charles store?
7 **A. Yes.**
8 Q. Do you have any idea what goes into the
9 decision to open, for instance, an Outdoor World versus
10 a Sportsman's Warehouse type retail store?
11 **A. Honestly, no. Those decisions are made by**
12 **someone other than me; so, honestly, no.**
13 Q. Are you aware of any discussions by Bass Pro
14 employees or officers about seeking to use the name
15 Bass Pro Shops Sportsman's Warehouse?
16 **A. No, not at all.**
17 Q. Let's talk about this store in Memphis. You
18 said that's a Sportsman's Warehouse store; is that
19 correct?
20 **A. Yes.**
21 Q. Have you ever visited that store?
22 **A. Yes, I have.**
23 Q. Do you know what the sign on the front of
24 that store says?
25 **A. Yes. Today it says Sportsman's Center.**

1 Q. Okay. Do you know why that is?
2 **A. It's my understanding that it was changed**
3 **roughly a year or so ago. It was changed by a then**
4 **head of retail and did not have full approval to be**
5 **changed, but was changed.**
6 Q. Do you know, who was that person?
7 **A. Kerry Harmon.**
8 Q. Kerry Harmon. And do you know why Kerry
9 Harmon decided to change the signage on the front of
10 the store?
11 **A. It was due to the Sportsman's Warehouse**
12 **opening in Memphis, I believe the one in Germantown,**
13 **and there was a lot of confusion from our customers in**
14 **terms of bringing in coupons and bringing in ads and**
15 **thinking that they were in the competitor's store.**
16 Q. Okay. Do you know if changing the sign
17 alleviated any of those problems?
18 **A. Honestly, no, I don't know that it did.**
19 Q. Do you have any personal knowledge about
20 those instances of confusion between the two stores in
21 the Memphis market?
22 **A. Only from discussions with Kerry and from**
23 **just comments from the store manager at the time that,**
24 **you know, customers were coming in with coupons and**
25 **that sort of thing. And then all the coupons that we**

1 **get come actually to -- examples of them come to**
2 **corporate, so we would have seen some in the office, as**
3 **well.**
4 Q. You say examples come to corporate. How
5 does that happen? How does that work?
6 **A. The stores send them in. They**
7 **periodically -- not always, but they will send in**
8 **examples of coupons that they take in, particularly**
9 **competitive coupons --**
10 Q. Okay.
11 **A. -- because we have a price-matching policy.**
12 Q. So this is unrelated to, for instance,
13 coupons that are taken at the register and kept? Those
14 aren't coupons you're talking about? You're talking
15 about somebody takes the initiative to send an example?
16 **A. Yes, to my department.**
17 Q. Okay.
18 **A. Yes.**
19 Q. Are you -- are you aware if coupons,
20 competitors' coupons that are taken, if they're taken
21 at the registers in Bass Pro stores, are kept?
22 **A. I honestly don't know. I'm not sure. I**
23 **know we take them, but I'm not sure if they're kept.**
24 Q. Do you have any idea what the costs were --
25 what costs were involved in changing the Memphis

1 store's signage to Sportsman's Center?
2 **A. No, I don't.**
3 Q. Are you aware if anyone in the company ever
4 discussed changing the sign on the Memphis store back
5 to Bass Pro Shops Sportsman's Warehouse?
6 **A. Yes, that was discussed.**
7 Q. Okay. And how do you -- how do you know
8 that that was discussed?
9 **A. Through conversations again with the retail**
10 **team, that there was some discussion about changing it**
11 **back; that it had not been approved to be changed, and**
12 **it was going to be changed back.**
13 Q. And what was the content of those
14 discussions?
15 **A. That we should not be changing any of the**
16 **other things we were doing because we were going to be**
17 **change the name back.**
18 Q. Okay. And was the name -- was the sign
19 changed back?
20 **A. To my knowledge, no.**
21 Q. And do you know why?
22 **A. Honestly, no.**
23 Q. Now, from your vantage point of overseeing
24 advertising for the company, did the sign change in
25 Memphis affect how you advertise that store and that

1 market?

2 **A. No, not initially. And from a -- from a**
3 **strategic perspective, no, it didn't change it at all**
4 **because, again, it was never fully -- it was never**
5 **approved to be changed. So from our perspective, it**
6 **never should have been changed, and we continued to**
7 **advertise under the Sportsman's Warehouse name.**

8 Q. Okay. You just said not initially. Has
9 something recently changed?

10 **A. There have been some ads that have changed**
11 **recently because of, again, confusion, both from a**
12 **customer's perspective and from a perspective of our**
13 **creative services group updated their creative**
14 **guidelines and listed the -- listed the Memphis store**
15 **as the Sportsman's Center. So there are -- there have**
16 **been a few ads recently that have denoted the Memphis**
17 **store as a Sportsman's Center and not as a Sportsman's**
18 **Warehouse.**

19 Q. Do you know if this is indicative of a trend
20 to shift all advertising to reflect the name
21 Sportsman's Center or not?

22 **A. Absolutely not. It was not something that**
23 **was approved by me to be changed.**

24 Q. So going forward, is it your understanding
25 or your direction that some of the advertising will say

1 Sportsman's Center and some of it will say Sportsman's
2 Warehouse in the Memphis market?
3 **A. I would say that the advertising would say**
4 **Sportsman's Warehouse.**
5 Q. So the recent change to Sportsman's Center,
6 if I understand you correctly, is going to be stopped
7 and shifted back to Sportsman's Warehouse?
8 **A. Yes.**
9 Q. Okay.
10 **A. And it hasn't been changed on everything.**
11 **It was only changed in a few areas. Particularly our**
12 **fliers is where it got changed, to my knowledge, the**
13 **circulars.**
14 Q. Do you have any sense of whether recently
15 there has been confusion between my client's store and
16 the Bass Pro Memphis Sportsman's Center store or
17 Sportsman's Warehouse store in Memphis?
18 **A. It would just be my opinion, but my opinion**
19 **would be yes.**
20 Q. "Yes" meaning there has been recent
21 confusion or --
22 **A. "Yes" meaning there has been recent**
23 **confusion.**
24 Q. Okay. Do you have any sense of if the
25 amount of customer confusion or customer concerns have

1 changed about the similarity between the two stores'
2 names since, for instance, 2005 to the present?
3 **A. No. Just, again, because I don't -- I'm not**
4 **really responsible for the customer service and those**
5 **comments and that sort of thing.**
6 Q. What causes you to say that you do believe
7 there had been some confusion in the recent past in
8 that market?
9 **A. Just from, again, talking to the folks in**
10 **the store and the fact that they're still getting**
11 **customers who come in with a Sportsman's Warehouse ad**
12 **and believe that it's our store.**
13 Q. Okay. Have you had any discussions where
14 particular instances of confusion have been discussed?
15 **A. Nothing I can remember off the top of my**
16 **head. Just more in general.**
17 Q. Okay. So right now, sitting here today, you
18 can't remember any particular instances of confusion
19 that have been reported to you; is that correct?
20 **A. Not a specific one. Just in general.**
21 Q. Okay. Let's go over one more time the types
22 of advertising that you're responsible for. You
23 mentioned there was a television production --
24 **A. Uh-huh.**
25 Q. -- correct? Some radio advertising?

1 IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
2 Before the Trademark Trial and Appeal Board
3
4 In the matter of Trademark Registration No. 2,390,988
5 For the mark SPORTSMAN'S WAREHOUSE HUNTING FISHING
CAMPING RELOADING OUTERWEAR FOOTWEAR and Design
6
7 Date Registered: October 3, 2000
8 BASS PRO TRADEMARKS, LLC, *
9 Petitioner, * CANCELLATION NO.
10 vs. * 92045000
11 SPORTSMAN'S WAREHOUSE, INC., *
12 Respondent. *
13 * * * * * * * * * * * * * * * * * *
14 Friday, September 15, 2006
15 Washington, D.C.
16
17 The telephone deposition of GARY T.
18 FORD, Ph.D., called for examination by counsel
19 for respondent, commencing 2:15 p.m., at the
20 office of Olender Reporting, Inc., 1522 K Street,
21 Northwest, Suite 720, Washington, D.C. 20005,
22 before E. Faye Grant, Court Reporter and Notary

1 Public for the District of Columbia, when were
2 present on behalf of the respective parties:
3 * * *
4 A P P E A R A N C E S (via telephone)
5 For the Petitioner:
6 GREGORY UPCHURCH, ESQUIRE
7 Husch & Eppenberger, LLC
8 190 Carondelet Plaza, Suite 600
9 St. Louis, MO 63105
10 (314) 480-1854
11
12 For the Respondent:
13 DAVID A. ALLGEYER, ESQUIRE
14 CHRISTOPHER R. SMITH, ESQUIRE
15 Lindquist & Vennum PLLP
16 80 South 8th Street
17 4200 IDS Center
18 Minneapolis, MN 55402-2205
19 (612) 371-3211
20 -0-
21
22

1 I N D E X
2 WITNESS: GARY T. FORD, Ph.D.
3
4 Examination by:
5 Mr. Allgeyer Mr. Upchurch
6 4 --
7 -0-
8
9 Exhibits: Page:
10 No. 1 - Ford Rebuttal Report 6
11 No. 2 - "Reference Guide on Survey Research" 41
12 No. 3 - "Reference Guide on Statistics" 48
13 No. 4 - Winkelman Report 63
14
15 -0-
16
17
18
19
20
21
22

1 Whereupon,
2 GARY T. FORD, Ph.D.,
3 Was called for examination and, after
4 having been duly sworn by the Notary Public, was
5 examined and testified as follows:
6 EXAMINATION BY COUNSEL FOR RESPONDENT
7 BY MR. ALLGEYER:
8 Q. Good afternoon. Do you go by Professor
9 Ford, Dr. Ford, or something else?
10 A. Either one. It doesn't matter.
11 Q. Okay. I'll use Dr. Ford because it's
12 shorter.
13 A. That's fine.
14 Q. All right. Could you please state your
15 full name for the record.
16 A. Gary Thomas Ford.
17 Q. Thank you. And your business address.
18 A. Kogod, K-o-g-o-d, School of Business,
19 American University, 4400 Massachusetts Avenue,
20 Northwest, Washington, D.C. 20016.
21 Q. Okay, great. Because this is a
22 telephone deposition, if we get disconnected for

1 any reason we'll immediately call back to this
2 number. And we will ask the Court Reporter to
3 include the fact of disconnection on the record
4 as we go ahead.
5 Have you been deposed before, Dr. Ford?
6 A. I have.
7 Q. On how many occasions?
8 A. I'd say 20, 25.
9 Q. How many of those were in connection
10 with your work as an expert witness?
11 A. All.
12 Q. Of those, how many were trademark cases
13 to the best of your recollection?
14 A. I'd say maybe a third to a half.
15 Q. So, something like seven to 10?
16 A. Yes. That's an estimate.
17 Q. Are you able to tell me, as best as you
18 can recall, as you sit there, the marks that were
19 involved in the various matters for which you
20 gave depositions?
21 A. I think that if you refer to my CV --
22 Q. Okay.

A. -- toward the end, I have the list of cases that I've been in in the last four years. And I think it's probably easiest just to go down through those and see.
Q. All right. That will be great. Do you have with you your rebuttal report?
A. Well, the Court Reporter has it. You want to mark it as an exhibit?
MR. ALLGEYER: That will be great. Madame Court Reporter, if you'd mark that as Exhibit 1. Let's make it Ford Exhibit 1, I think has been the way we've been doing it.
(Ford Deposition Exhibit Number 1 was marked for identification.)
BY MR. ALLGEYER:
Q. And then if you can have a look at that, Dr. Ford, confirm that's your expert report. And then once that's done, I will ask you to turn to Bates Number Ford 30.
A. You want me to -- okay, hang on. I am there.
Q. Okay. Well, first of all, do you

recognize this as the rebuttal report that you provided in the matter of Bass Pro Trademarks v. Sportsman's Warehouse?
A. I do.
Q. All right. I believe that this is the listing that you were referring to?
A. Yes.
Q. "Expert and Expert Witness Assignments Since 2002"?
A. Yes.
Q. Can you just go through there and go ahead and just tick off for me the ones that involve trademarks, and then we'll go back and see which trademarks.
A. Hispanic.
Q. Okay.
A. Diarama.
Q. Okay.
A. I think Pharmavite did, but it might have been deception. It was close.
GEICO, 3M, Polo, Omni, 9 Squared, Medi-Flex, Align, and American Century.

Q. Okay. Now, Hispanic Broadcasting, if you could let me know what the trademark was and perhaps the issue?
A. It was a radio call sign; KLOV, K-L-O-V. And the question was whether there was confusion between those call signs from those two different providers of radio signals.
Q. Did you provide testimony on behalf of the plaintiff or the defendant?
A. Educational Media was my client.
Q. The defendant?
A. Yes. I guess, yeah.
Q. And the question was whether there was confusion; is that correct?
A. Yes.
Q. And what did you determine?
A. That there was not.
Q. Did you do a survey?
A. Yes.
Q. What kind?
A. Telephone.
Q. Did you start with a list or some other

way? How did you get the names for the survey?
A. It was random-digit dialing.
Q. What's random-digit dialing?
A. There's no list. You start with just knowing the area codes and knowing the three-digit prefixes, and then randomly selecting -- then having a random number generator generate the last four numbers, and dialing from there, and then screening.
Q. How did you qualify people?
A. Oh, gee, I don't remember. I mean, that was three or four years ago. But, basically, we had to get people that were familiar with the call letters. As I recall, Educational Media Foundation is owned by a Christian -- or is a Christian company. And KLOV was a Christian broadcasting company. And Hispanic Broadcasting Corporation owned another radio station with the initials KLOV, and that one broadcast in Spanish, and the other one broadcast in English. And one went to Hispanic consumers of popular music, and the other one went to Christian -- people

Page 10

1 interested in Christian music stations in
2 English.
3 Q. Was that case tried to a conclusion?
4 A. No. It settled.
5 Q. Okay. Did you have your deposition
6 taken?
7 A. No.
8 Q. Just provided a report?
9 A. Yes.
10 Q. Okay. And the next one is Diarama
11 Trading Company v. J. Walter Thompson and others.
12 A. Yes.
13 Q. What was the issue there?
14 A. The initials D-T-C.
15 Q. Okay. What was the claim with that,
16 trademark infringement basically?
17 A. Yes.
18 Q. Any confusion?
19 A. Well, you guys may know it better as
20 lawyers than I do. I can describe the situation.
21 Q. Okay.
22 A. J. Walter Thompson represents DeBeers.

Page 11

1 Q. Yes.
2 A. And they had a company that sells
3 diamonds called the Diamond Trading Company and
4 they wanted to use the initials D-T-C. And the
5 client I was working for, Diarama Trading
6 Company, sells diamonds, and it was also using
7 the initials D-T-C.
8 Q. Okay. So, you were working for Diarama
9 Trading Company?
10 A. I was.
11 Q. And you did a report. And I assume your
12 report found there was a likelihood of confusion
13 there?
14 A. It did.
15 Q. Did that involve a secondary-meaning
16 inquiry or not? Was it just confusion?
17 A. Well, you know, you raise a good
18 question. I just don't remember. I remember
19 that there was very little awareness of the
20 initials D-T-C having to do with DeBeers. But I
21 don't remember what else the survey found.
22 Q. Okay. And you did a report and had your

Page 12

1 deposition taken. Was that case tried to a
2 conclusion?
3 A. I think it's still alive.
4 Q. Okay. Was your report presented in
5 connection with any court proceedings?
6 A. I don't know anything more than what
7 I've told you. I said I think it's still alive.
8 I have no idea. I have had no contact with
9 anyone really since the deposition.
10 Q. Okay. The next one is Pharmavite v.
11 Weider Nutrition Group. Maybe you can describe
12 the issue there?
13 A. It was a dietary supplement. And I
14 actually think that it was probably a deception
15 issue in the sense that Weider was using or
16 describing a product and advertising it in a way
17 that was similar to the Pharmavite product. And
18 I don't remember whether it was confusion or
19 deception.
20 Q. Okay. And who was your client?
21 A. Pharmavite was my client.
22 Q. Okay. And you gave a report together

Page 13

1 with Michael Mazis?
2 A. Yes, I did.
3 Q. Who's Michael Mazis?
4 A. Michael Mazis is a colleague of mine at
5 American University, a marketing professor.
6 Q. And you were deposed in that case?
7 A. I was.
8 Q. Was that case tried to conclusion?
9 A. I think it settled.
10 Q. Okay. Was your report ever used in a
11 court proceeding?
12 A. Not that I -- I think it settled right
13 soon after my deposition, so I doubt it.
14 Q. Okay. What kind of survey did you do in
15 that case?
16 A. It was a mall intercept.
17 Q. I don't know what that means. Explain
18 that to me would you please?
19 A. Yes. It was screening people in a
20 shopping mall to make sure they met the
21 eligibility criteria, and then if they did,
22 inviting them back to a room to be shown some

Page 14

1 materials and interviewed.
2 Q. Okay. Is there a reason that you would
3 choose that in some circumstances rather than
4 others?
5 A. I think when you need to show people
6 something, you pretty much are going to use --
7 the most common procedure is to use mall
8 intercepts.
9 Q. Okay. Oh, you said mall intercepts?
10 Okay. I thought you said mobile intercept.
11 A. No. Mall. M-a-l-l. Sorry.
12 Q. I do know what that is.
13 A. Yeah.
14 Q. Okay. By the way, have you ever done a
15 survey where you have to show people things where
16 you use the Internet?
17 A. I have.
18 Q. Is that one of these cases? Are we
19 going to get to that or --
20 A. No. It doesn't exist on here.
21 Q. Okay. Maybe you can tell me about that.
22 What was the circumstance and why did you decide

Page 15

1 to use the Internet?
2 A. The case took place in Sweden, and it
3 involved Swedish consumers. And they needed to
4 -- and the case involved recognition of various
5 logos. And in that situation, because the
6 respondents were dispersed throughout the country
7 of Sweden, it was most effective to do that with
8 an Internet panel.
9 Q. Okay. Was that Internet survey in
10 conjunction with a telephone survey?
11 A. No.
12 Q. Just Internet?
13 A. Yes.
14 Q. What was the response rate on something
15 like that?
16 A. I don't remember. But it was a panel,
17 so it was high.
18 Q. What do you mean by a panel? I'm
19 unfamiliar with that.
20 A. A panel is a group of consumers who have
21 been pre-recruited and who have agreed to respond
22 to interviews. Panels exist probably in every

Page 16

1 developed country in the world. For example, in
2 the United States, Market Facts has a panel I
3 think of 750,000 households. And the panels are
4 set up so that they're demographically,
5 socio-economically and geographically balanced to
6 replicate the United States population.
7 Originally these panels started out as mail
8 panels when using the U.S. Postal Service. As
9 telephones became more prevalent, they started
10 developing telephone panels. And over the last
11 ten years, they have been developing Internet
12 panels in the same criteria of being
13 demographically, socio-economically, and
14 geographically representative of Internet
15 households.
16 Q. I see. Thank you. Next is the GEICO
17 case?
18 A. Yes.
19 Q. What was the issue there?
20 A. Initial interest confusion.
21 Q. Okay. That involved Google and Overture
22 Services. Who was your client?

Page 17

1 A. GEICO.
2 Q. All right. And did you do a survey?
3 A. I did.
4 Q. What kind?
5 A. Mall intercept.
6 Q. What was the issue; what was confusing
7 initially?
8 A. The issue had to do with the sponsored
9 links that appear. This case involved both
10 Yahoo, which is owned by Overture Services, and
11 Google. And the issue was that when people
12 entered the letters g-e-i-c-o into, say, the
13 Google search engine, not only do they get
14 organic search results, which we might all think
15 of as sort of the real search results, but over
16 on the right-hand side and sometimes at the top,
17 there are sponsored links, which are paid
18 advertising that is associated with using
19 different search terms.
20 Q. Okay.
21 A. And GEICO was concerned that those
22 sponsored links mentioned GEICO in them, and they

Page 18

1 had no affiliation in any sense at all with
2 GEICO. And they were there to -- and that
3 potential consumers were being misled by those
4 sponsored links.
5 Q. And that case went to trial it looks
6 like?
7 A. It did.
8 Q. What jurisdiction was that?
9 A. It was in Alexandria, Virginia.
10 Q. Close to home, huh?
11 A. Yes.
12 Q. Okay. What was the result of that
13 trial, if you know?
14 A. They settled. They started a trial, and
15 then in the middle the two sides got together and
16 worked out a settlement.
17 Q. Okay. Next is Polo Ralph Lauren versus
18 United States Polo Association. Tell me about
19 that one. It looks like that one went to trial.
20 Who is your client?
21 A. Polo Ralph Lauren.
22 Q. And what was the issue?

Page 19

1 A. The United States Polo Association
2 wanted to use a logo that Polo Ralph Lauren
3 perceived was too close to their horse and rider
4 logo.
5 Q. Did you do a survey?
6 A. I did.
7 Q. What kind?
8 A. It was a mall intercept.
9 Q. Okay. And that was because you had to
10 show people things?
11 A. Yes. It was a post-confusion study.
12 Q. What do you mean by post-confusion?
13 A. Excuse me. I mean post-purchase study.
14 We showed people the shirts on mannequins as if
15 they were seeing someone on the street wearing a
16 Polo shirt.
17 Q. Oh, I see. Okay. Did the case get
18 tried to a verdict?
19 A. Yes.
20 Q. Oh. How did it turn out?
21 A. Won some, lost some.
22 Q. Okay. Except on the shirt, right?

Page 20

1 A. Yeah, I guess. Parts of it they won and
2 parts of it they lost. Or both sides, I guess.
3 Q. Okay. Maybe that's why so many of them
4 settle. Omni Pacific, Inc. versus OmniBrands.
5 A. Yes.
6 Q. Well, what was the issue there, please?
7 A. The issue was that Omni Pacific was
8 suing OmniBrands for similarity in the names and
9 confusion about the names.
10 Q. And who was your client?
11 A. OmniBrands.
12 Q. So, you were the defense expert?
13 A. Yes.
14 Q. Did you do a survey?
15 A. No. I did a rebuttal report.
16 Q. Okay. Suggesting that their methodology
17 was incorrect for the survey showing confusion,
18 I'm guessing?
19 A. There were a number of problems with it.
20 But I think the sample size was something like
21 30 people or, as I remember, it was very small
22 and unlikely to be representative of anything.

Page 21

1 Q. Okay. 9 Squared, Inc. versus Moviso,
2 LLC and InfoSpace. Maybe you could describe that
3 one to me.
4 A. 9 Squared was my client. And 9 Squared
5 and the other firms provide ringtones.
6 Q. Okay. What was the issue?
7 A. The issue, as I remember it, was whether
8 certain terms, such as ringtones, are generic
9 terms or are they branded words.
10 Q. And you did a report?
11 A. It was a rebuttal report.
12 Q. Okay. And what did you conclude, just
13 in general?
14 A. There were some flaws in the other guy's
15 study.
16 Q. Okay. Did that case go to trial or
17 anything?
18 A. No.
19 Q. Medi-Flex.
20 A. Medi-Flex, well --
21 Q. If I did the wrong one --
22 A. No. No. Nice-Pak was my client.

1 Q. Okay.
2 A. And it was a preliminary injunction
3 hearing, and I did a rebuttal report.
4 Q. What was the issue in the case?
5 A. Whether consumers were going to be
6 confused by the name of a new product that
7 Nice-Pak was introducing that allegedly was
8 similar to a product sold by Medi-Flex.
9 Q. Okay. I'm sorry. You said your client
10 was who?
11 A. Nice-Pak.
12 Q. Okay. What was the result of the
13 hearing, do you know?
14 A. There was no confusion.
15 Q. So, no injunction was --
16 A. Yeah, right. Exactly.
17 Q. Okay. Next is Align Technology versus
18 Orthoclear. Who was your client in that one?
19 A. Align.
20 Q. And what was the issue?
21 A. The issue is invisible braces.
22 Q. Okay. And what about invisible braces?

1 A. Orthoclear was using the same
2 advertising slogan in their advertising as had
3 been used many years by Align Technology.
4 Q. I see. And, so, that was sort of a
5 confusion based on ad situation?
6 A. Yes.
7 Q. All right. And did you do a survey or
8 just a rebuttal report?
9 A. I did an Internet survey.
10 Q. And how did you get your list of
11 surveyants or whatever they're called?
12 A. The participants were general
13 practitioner dentists and orthodontists.
14 Q. Did you start with a list or how
15 did you --
16 A. Yes. Started with an Internet panel.
17 Q. They have Internet panels for something
18 that specialized, huh?
19 A. Yes. I think in the health professions
20 area it's always difficult to get doctors and
21 nurses and other health professionals and, so,
22 the Internet turns out to be a convenient way to

1 reach people. They get paid an incentive, but
2 they can answer the thing on their own time. So,
3 it turns out to be a way to get information from
4 that group, from those groups of health
5 professionals.
6 Q. What was your conclusion as a result of
7 the survey you did?
8 A. Well, because Align had been -- I think
9 because Align had been using the advertising
10 slogan for many years, consumers, the dentists
11 associated that slogan with Align rather than
12 with Orthoclear.
13 Q. Okay. Did that case go to trial?
14 A. It's still going.
15 Q. Still going, okay. And finally, I
16 think, we have American Century Proprietary
17 Holdings versus American Century Casualty
18 Company. Who was your client in that case?
19 A. The first one, the plaintiff.
20 Q. What's the issue?
21 A. Similarity of the names.
22 Q. Oh, the American Century part?

1 A. Yes.
2 Q. Okay. Did you do a survey?
3 A. I did.
4 Q. What kind?
5 A. Mall intercept.
6 Q. What did you conclude?
7 A. That people perceived that the two
8 companies were associated in some way.
9 Q. And you did a report?
10 A. Yes, I did.
11 Q. It hasn't gone to trial yet?
12 A. No. It's just recent.
13 Q. Have I pretty well now exhausted your
14 trademark survey and expert work, at least in the
15 last three years with this line of inquiry?
16 A. Yes. And there wasn't anything in 2002,
17 so that's it.
18 Q. All right. What did you do to prepare
19 for today's deposition?
20 A. I was sent the revised Mr. Winkelman
21 report. I looked through that, and I talked to
22 Mr. Upchurch.

Page 26

1 Q. What did you talk to Mr. Upchurch about?
2 A. About the deposition.
3 Q. How long did you talk to Mr. Upchurch
4 for?
5 A. I think about 45 minutes.
6 Q. Did Mr. Upchurch indicate that there
7 were any areas where you should concern yourself
8 about your testimony or make sure you said one
9 thing or another?
10 A. No. It was mostly give me time to
11 object if I need to.
12 Q. Okay.
13 A. And tell the truth and be clear. That
14 type of stuff.
15 Q. Have you had a chance to review the
16 Winkelman report fully, the new one?
17 A. Pretty fully, yes.
18 Q. Okay. We'll get to that in a minute.
19 If you'd just quickly turn to Appendix B of your
20 report, which is 432 on the Bates numbers in the
21 lower right-hand corner.
22 A. Yes.

Page 27

1 Q. There's a listing of documents that said
2 you consulted or that you referred to.
3 A. Right.
4 Q. If we were to add the Winkelman, the
5 newer Winkelman report, the September one I
6 believe it is, would this in fact list all the
7 documents you've looked at in connection with
8 this matter or have you seen some later?
9 A. No. That's it.
10 Q. Okay. Apart from your discussions with
11 -- well, let me go back. Have you seen Professor
12 Mazis' report on this matter?
13 A. I have not.
14 Q. You weren't asked to review it or
15 criticize it?
16 A. I was not.
17 Q. Okay. Have you spoken to Professor
18 Mazis about his report?
19 A. Only that he did one. But we haven't
20 spoken -- I have no idea what he did.
21 Q. Okay. Did you have the assistance of
22 anyone else in preparing your rebuttal report,

Page 28

1 Exhibit 1?
2 A. Mr. McCarty -- Dr. McCarty.
3 Q. Who is Dr. McCarty?
4 A. John McCarty is a marketing professor at
5 The College of New Jersey, which is near Trenton.
6 And he helped me in terms of looking up the
7 cities. For example, one of my exhibits talks
8 about the cities in which Bass Pro and
9 Sportsman's Warehouse are located. And he, I
10 think, helped me prepare the material in Exhibit
11 -- one of the exhibits here. Appendix C. And --
12 Q. Well, let me catch up with you. Just a
13 second.
14 A. Sorry. I'm sorry. It's a little more
15 difficult with you not here or me not there.
16 Q. That's true. But think of all the
17 airfare we saved. [laughter]
18 A. Yeah, I know.
19 Q. Okay. You said appendix --
20 A. I said Appendix C. And then appendix --
21 the numbers, the I.D. numbers in Appendix C
22 continued, and then Appendix D.

Page 29

1 Q. Okay. So, Bates 33 through 36?
2 A. Yes.
3 Q. Okay. What precisely did he do? Let's
4 start with Bates Ford 33, Appendix C.
5 A. Well, he originally looked up for me
6 the cities in which there were -- for what
7 respondents mentioned. He identified whether
8 there was a Bass Pro in the city or a Sportsman's
9 Warehouse in the city or both or neither.
10 Q. Okay. Now, do you know what his
11 criteria were? For example, I'll give you an
12 example I know about here. We have Minneapolis,
13 and we also have I don't know how many suburbs;
14 ten or eleven suburbs. In his calculation did he
15 count suburbs --
16 A. Yes.
17 Q. -- as close enough, or did he count
18 cities as close enough for suburbs, or do you
19 know?
20 A. In general he attempted to count suburbs
21 as close enough.
22 Q. And how do you --

Page 30

1 A. Well, we had discussions about that.
2 Q. Okay. Do you know in which cases that
3 was done, in which cities and which suburbs?
4 A. No.
5 Q. Okay. How about the suburbs of
6 St. Louis; do you know if he counted St. Charles
7 the same as St. Louis?
8 A. I don't know off the top of my head, no.
9 Q. Where would you have to go to look to
10 see whether that was done or not?
11 A. I would have to go back to -- I would go
12 -- I presume it's in a database someplace.
13 Q. Okay. But I'm not going to be able to
14 find it in the report; correct?
15 A. No. That's correct. But we do have --
16 you do have the information in the report by
17 respondent I.D. number, to see how things were
18 categorized.
19 Q. Okay. Where is that located?
20 A. If you continue with Appendix C, on the
21 second page.
22 Q. Right. Okay. And the respondent I.D.,

Page 31

1 that's just a number that's assigned to each row
2 on a data file from the Winkelman survey?
3 A. Yes. In the order in which it appeared
4 in the file that Mr. Winkelman supplied.
5 Q. I'm guessing then each number
6 corresponds in some way to a person; right?
7 A. Yes.
8 Q. Is that how you take that?
9 A. That's correct.
10 Q. Okay. Appendix D, was that also
11 prepared by Professor McCarty?
12 A. Yes.
13 Q. Did Mr. McCarty do anything besides
14 prepare Appendixes C and D?
15 A. Not really.
16 Q. Okay. Did anyone, apart from Professor
17 McCarty, contribute to this report in any way?
18 A. No.
19 Q. Apart from any discussions you may have
20 had with Professor McCarty or Mr. Upchurch, have
21 you discussed this report with anybody else?
22 A. Mr. Rusche.

Page 32

1 Q. Fred Rusche?
2 A. Yes.
3 Q. He's with Mr. Upchurch's law firm?
4 A. Yes.
5 Q. Okay. What did you discuss with
6 Mr. Rusche?
7 A. Mr. Rusche was just my contact person.
8 Q. Okay. Did you discuss anything
9 substantive with him?
10 A. No. Not that I recall.
11 Q. How did you determine what exactly it
12 was that you were going to comment on in this
13 report? Did the lawyers for Bass Pro kind of say
14 we'd like you to comment on this, or did they
15 just give you the report and say comment any way
16 you'd like, or some combination of the two?
17 A. Well, I think that, based on my
18 experience over the years, I know what to look
19 for in terms of the research methodology used in
20 a survey. So, as I recall, the only instructions
21 I got were we'd like you to evaluate the
22 Winkelman survey.

Page 33

1 Q. Okay. Have you done any surveys for
2 Bass Pro Shops other than -- have you ever done
3 any surveys for Bass Pro Shops?
4 A. I have not.
5 Q. Have you done any work for Bass Pro
6 Shops apart from this rebuttal report that we see
7 here today?
8 A. I have not.
9 Q. Do you know anyone from Bass Pro Shops?
10 A. Do I know anyone?
11 Q. Yes.
12 A. No.
13 Q. Okay. Thinking back on trademark-
14 related surveys that you have done, have any of
15 those surveys ever been actually entered into
16 evidence in a court proceeding?
17 A. Well, I think the Polo.
18 Q. Polo. Any others?
19 A. And I think -- well, I don't know.
20 Let's go back and look. I'm just thinking about
21 what we went through.
22 Q. Okay. So we can go back --

Page 34

1 A. Yeah, to page 30.
2 Q. That's page 30 of your report.
3 A. Well, GEICO is and Polo. I don't know
4 where these other ones stand that are still going
5 on.
6 Q. But you haven't been put on the witness
7 stand and asked to explain your report and give
8 your opinion in the others; right?
9 A. That's correct.
10 Q. Okay. How many surveys would you say --
11 and I appreciate you probably won't be able to
12 give me to the penny -- but how many surveys
13 would you say you've done in your professional
14 career?
15 A. Including the ones I've taught and
16 courses I've taught and what students have done?
17 Q. No. Let's break those down. Let's do
18 the ones that you've been hired to do; hired to
19 do and done surveys or supervised.
20 A. Oh, I'd say at least probably 50.
21 Q. Fifty, okay. And how about surveys you
22 might have supervised in your teaching role?

Page 35

1 A. Probably over 100.
2 Q. And of those hundred, those are
3 typically done by students?
4 A. Oh, yeah. That's their assignment.
5 Q. Okay. And these are marketing classes
6 or --
7 A. Marketing research. I've taught
8 undergraduate, MBA, and Ph.D. courses in
9 research.
10 Q. Okay. What would you say the sort of --
11 if there is a way to characterize -- sort of the
12 focus of your research and work, let's say about
13 right now in your academic career?
14 A. I think it's on an area called economics
15 of information.
16 Q. Can you just give me the layman's
17 explanation of that?
18 A. Oh, sure. It really has to do with how
19 people understand when they should be skeptical
20 or not skeptical of advertising.
21 Q. Okay. That's sort of a research and
22 writing focus of yours?

Page 36

1 A. It has been over the years, yes.
2 Q. Anything else is a focus of your --
3 A. Well, I do consumer behavioral research
4 on a variety of areas. I've done a series of
5 studies on nutrition labels, things like that.
6 Q. Okay. You mentioned in your report a
7 term called convenience sample.
8 A. Yes.
9 Q. And you probably remember this, but I
10 see references to it on page 3, and I'm sure it's
11 in here a couple of other places. Can you define
12 for me what a convenience sample is?
13 A. Generally, a convenience sample is a
14 term that is widely used for any non-probability
15 sample.
16 Q. Okay.
17 A. That could include people because they
18 volunteer to be interviewed, people we interview
19 because they're close to us in terms of -- or
20 because we belong to a church group, and we can
21 interview some of the people in the church, or
22 other kinds of reasons like that.

Page 37

1 Q. Okay. Now, in your report -- and it's
2 on page 11 if you want to look at it, but I bet
3 you remember this -- you indicated that the
4 Survey Sampling International list that Mr.
5 Winkelman used is a "convenience sample."
6 A. Yes.
7 Q. Tell me all the reasons that you believe
8 that it's a convenience sample.
9 A. Well, it's a non-probability sample for
10 sure. Because in order to be a probability
11 sample, one would have to have a list of the
12 people who are eligible or potentially eligible
13 for participating in this survey, and that
14 doesn't exist. And then it's a matter of just
15 how the list was pulled together. My
16 understanding is, a lot of the names on that list
17 come from people who fill out, for example,
18 warranty cards for products that they might buy,
19 such as a DVD player. And then on the warranty
20 card, it also has the list of items and said
21 whether you have an interest in those particular
22 areas.

Page 38

1 Q. What's the basis of your understanding
2 of how SSI does its sampling?
3 A. Through information from SSI.
4 Q. Who'd you talk to there?
5 A. Mr. McCarty talked to SSI. That's
6 another thing he did.
7 Q. Well, that wasn't a discussion you had
8 with SSI?
9 A. No. It was a discussion that he had.
10 It was a task that I gave him, and then he
11 reported back to me.
12 Q. Do you know who he talked to at SSI?
13 A. No.
14 Q. Okay. Did you ask him that?
15 A. No.
16 Q. I notice on paragraph 43 of your report
17 you suggest that Mr. Winkelman should have
18 demonstrated how and why he believes the SSI
19 sample replicates the relevant population. Do
20 you see that?
21 A. Yes.
22 Q. Is it fair to say that if Mr. Winkelman

Page 39

1 did demonstrate that or articulated that, that
2 you would withdraw your objection that he used a
3 convenience sample?
4 A. Well, I think it's still a convenience
5 sample, but whether it's representative or not is
6 important.
7 Q. Okay.
8 A. I think if it is representative, then
9 it's better.
10 Q. Okay. So, is it fair to say that just
11 because something's a convenience sample doesn't
12 always mean that it's a flawed sample that's not
13 usable?
14 A. I agree with that with a caveat. And
15 the caveat is that I think in those situations
16 what you need to do is set up a quota sample.
17 For example -- and this is all hypothetical, but
18 it'll indicate my point. Suppose that with prior
19 research we learn that two-thirds of the people
20 who engage in these activities are males and
21 one-third are females. And let's suppose that
22 they tend to skew younger than older, such that

Page 40

1 40 percent of them are between 18 and 34, 20
2 percent are between 35 and 54, and 20 percent are
3 55 and older. I'm trying to get to 100 percent
4 there, so I hope I did that. In terms of quota
5 sample then, we might establish quotas such that
6 the ending sample we have is broken down
7 two-thirds males, one-third females, 60 percent
8 between the ages of 18 to 34 and 20 percent each
9 in those other two age groups.
10 Q. And then that would allow you to use a
11 convenience sample if you did it in that way; is
12 that what you're saying?
13 A. Yes. That's the best you could do.
14 Q. I note that in a few cases you talked
15 about doing mall intercept samples. Isn't that
16 by definition a convenience sample?
17 A. Yes. But they are with quotas. And
18 that's really the point I was explaining. We're
19 not going to have -- it's very difficult to have
20 a random probability sample of people who use a
21 particular type of product, because most times
22 lists don't exist. So, what you have to do is

Page 41

1 try to develop a representative sample by
2 establishing quotas that allow you to wind up
3 with an ending sample that has the same
4 characteristics of the user population or the
5 relevant population.
6 Q. I note that in your report you refer to
7 the work by Shari Seidman Diamond.
8 A. I do.
9 Q. "Reference Guide on Survey Research."
10 A. I do.
11 Q. All right. Do you have that report
12 there?
13 A. It is in the pile that you have. Do you
14 want us to mark it as an exhibit?
15 MR. ALLGEYER: Yes. Let's mark that as
16 Ford Exhibit 2.
17 (Ford Deposition Exhibit Number 2
18 was marked for identification.)
19 MR. ALLGEYER: Let me know when you've
20 got that, please.
21 THE WITNESS: It's ready.
22 BY MR. ALLGEYER:

Page 42

1 Q. Thank you. Now, my understanding is you
2 recognize this as an authoritative guide on
3 survey research?
4 A. I think it's one of the best out there.
5 I don't agree with every word in it, but it's
6 quite good.
7 Q. Okay. Just in the end there is
8 something about mall sampling. If you look at
9 page 247.
10 A. Yes.
11 Q. And you can read as much of this as you
12 need to get the context.
13 A. Where am I looking? I'm on 247, but I'm
14 not sure where I'm supposed to be on it.
15 Q. I'm about to put you there.
16 A. Okay, great. Thank you.
17 Q. I'm focusing on the second full sentence
18 where she talks about, "Training that reduces the
19 interviewer's discretion in selecting a potential
20 respondent is likely to reduce bias in selection,
21 as are instructions to approach every nth person
22 entering the facility through a particular door."

Page 43

1 Do you see that?
2 A. Yes.
3 Q. Okay. Now my understanding is what
4 Professor Diamond is telling us is that one way
5 to try to make a mall intercept survey a little
6 more representative, is to make sure that you
7 don't just take everybody that walks in the door
8 but that you take sort of a random sampling of
9 them; every third person or something like that.
10 First of all, is that your understanding of what
11 she's trying to say?
12 A. I think she's -- that is what I believe
13 is a pretty good summary of what she's trying to
14 say, yes.
15 Q. Okay. Then I note that in footnote 68
16 she says that, "In the end, even if malls are
17 randomly sampled and shoppers are randomly
18 selected within malls, results from mall surveys
19 technically can be used to generalize only to the
20 population of mall shoppers." Do you see that?
21 A. I do.
22 Q. Do you agree with that?

Page 44

1 A. I agree that technically that is
2 correct. What happens though is that my
3 experience has been -- for example, a typical
4 study I might use, if it's a national product, I
5 might use 12 malls or 16 malls scattered
6 throughout the four census regions of the U.S.
7 And we scatter them around because we want to
8 make sure that we got some people from the South,
9 from the North and the East and the West and so
10 on. And even though technically the results
11 can't be generalized, though they're only
12 empirical results we have and they result in a --
13 if the quotas are set up so that the sample is
14 representative of the larger population, in
15 practice those results are generalized as if it
16 would be the same had you taken a random sample
17 of consumers from the entire United States.
18 Q. But what if the population you were
19 interested in tended not to go to malls, like I
20 would? That would be me, by the way, if I could
21 help it.
22 A. Well, it depends. And let's suppose

Page 45

1 that some shoppers go to malls and some don't.
2 The question is whether the ones who go to the
3 malls or don't go to the malls have different
4 perceptions of materials that they may be shown.
5 And all you know is that you have a sample that
6 "looks like" the population. And "looks like" is
7 in quotes. And it's the best we're going to do
8 given the reality of budget, time, and so on.
9 Q. I suppose if you had an unlimited budget
10 you could probably do a better survey than if you
11 didn't. That's true, isn't it?
12 A. That's probably true. And unlimited
13 time.
14 Q. Unlimited. Okay, fair enough. And is
15 it also fair to say that most surveys have to
16 make some compromises on time and budget since we
17 have neither unlimited time nor unlimited funds;
18 true?
19 A. I think there's always tradeoffs that
20 one makes between time and budget, yes.
21 Q. Okay. I'll tell you, most -- I
22 shouldn't say most. Is it fair to say that many

Page 46

1 survey experts, people in that field, may draw
2 the line of what's permissible in terms of how
3 much time and money to spend as opposed to others
4 who might think you might spend more? In other
5 words, do folks all agree with each other on that
6 necessarily?
7 A. I really can't answer that question,
8 other than to say that I think in situations, for
9 example, where you need to show somebody
10 something, mall intercepts are in general the
11 best way to do that. And sometimes you can't;
12 for example, with a study of dentists or health
13 professionals, you may not be able to do that
14 because of the extraordinary expense that might
15 be. And you're not going to get them to come
16 into a shopping mall. So, I guess, it depends on
17 the situation.
18 Q. But is it a fair assumption on my part
19 that in this, like most areas, that not all
20 experts agree with each other on all aspects of
21 how surveys should be conducted and how much time
22 and money should be spent on them?

Page 47

1 A. I think that that's oftentimes a -- I
2 think that the client budget and the timing are
3 going to be influential. And the expert has to
4 be very forthright in what is being traded off,
5 and decide whether the survey is going to be
6 reliable when it gets done if it uses a certain
7 methodology -- in other words, you shouldn't use
8 a methodology just because it's cheap if it
9 doesn't provide reliable information. And then
10 that's up to the expert to make a decision about.
11 Q. All right. And it's fair to say that
12 not all experts agree on exactly the way that
13 that ought to be done; is that a fair conclusion
14 on my part?
15 A. Yes. And I mean if it's going to be
16 used in a litigation context or you want to use
17 it in a peer reviewed context, you have to
18 recognize if you're doing something that the
19 referees or the person who might be evaluating it
20 is going to hold you to pretty high standards.
21 Q. Okay. There's another document that
22 I've -- on my behalf it was sent to you out

Page 48

1 there, entitled, "Reference Manual on Scientific
2 Evidence" on the front. And there's another
3 article, as I understand it, out of the same
4 reference guide the Diamond article is out of.
5 Do you have that there?
6 A. I believe you're referring to the
7 "Reference Guide on Statistics."
8 Q. That's correct. By David Kaye and
9 David Freedman?
10 A. We have it. Should we mark it?
11 MR. ALLGEYER: Would you, please. That
12 would be 3.
13 (Ford Deposition Exhibit Number 3
14 was marked for identification.)
15 BY MR. ALLGEYER:
16 Q. Thank you. Could you have a look at
17 Exhibit 3 please, Dr. Ford, and tell me whether
18 that is something you're familiar with?
19 A. I am not familiar with it.
20 Q. Do you recognize that as being in the
21 same reference guide as the -- in the Reference
22 Manual on Scientific Evidence?

Page 49

1 A. I have seen it in the table of contents.
2 Q. You have not reviewed it?
3 A. If I did, I only skimmed it.
4 Q. Okay. Well, I'll just ask you, because
5 it won't help me to have you read this. In this
6 reference manual, it is my impression that the
7 authors refer to mall intercept or surveys done
8 of mall shoppers as convenience samples. Do you
9 agree that mall surveys rely on convenience
10 samples?
11 A. It depends on how they're done. They're
12 always convenience samples, so I agree with that.
13 The question is whether the representative
14 samples are not, given that they're convenience
15 samples. It's similar to the discussion we had a
16 bit earlier where I indicated, I hope, that with
17 a mall intercept sample, it's a non-probability
18 sample. And because of that, the only way that
19 you can have some confidence that your results
20 are reliable, is to attempt to develop quota
21 sampling procedures that allow you to develop a
22 sample that, as the words I used, "looks like"

Page 50

1 the larger population. And by "looks like" I
2 mean have the same demographic and socioeconomic
3 characteristics on whatever variables you think
4 are important.
5 Q. Let me ask you, perhaps, a little better
6 and more pointed question: If you look at page
7 162 of Exhibit 3 --
8 A. Yes.
9 Q. -- the authors define convenience
10 sample. And in the last sentence of that they
11 say mall samples are convenience samples. Do you
12 see that?
13 A. Yes. And I also see the prior language
14 that says such samples are -- and remember I said
15 this is a non-probability sample. And what
16 they've said in the first sentence is a
17 non-random sample of units. That means
18 non-probability. Those are synonyms. And it
19 says, such samples are easy to take but may
20 suffer from serious bias. The way you try to get
21 around the bias, or to minimize the probability
22 of the bias, is by setting up quotas so that the

Page 51

1 sample looks like, has the same characteristics
2 as the larger population. If you don't do that,
3 you do run the risk of having biases.
4 Q. Okay. But biased or not, do you agree
5 with the statement that mall samples are
6 convenience samples?
7 A. Oh, I agree that they are convenience
8 samples. And for that reason, you have to be
9 careful to make sure that they have the same
10 characteristics of the population you're trying
11 to represent.
12 Q. Okay. Apart from the inquiries that
13 Dr. McCarty made to SSI, have you ever made any
14 inquiries to that organization?
15 A. I have.
16 Q. Tell me about that.
17 A. I have made inquiries to it a few times.
18 I can't give a number. But primarily for samples
19 based on SIC Code of Organizations.
20 Q. Okay. Have you used SSI samples in your
21 survey work? In other words, you said you asked
22 about SIC samples. I'm just trying to follow

Page 52

1 that up. Does that mean you got those kinds of
2 samples or lists from SSI to use in your work?
3 A. Well, I know that I was intending to. I
4 can't ever remember whether I did or not.
5 Q. Okay. Have you, apart from that, had
6 any other contact with SSI or used any of their
7 lists, apart from the SIC Code samples?
8 A. I don't think so. I think it's only
9 been in situations where I was attempting to do
10 interviews with businesses or with executives or
11 representatives of various kinds of businesses.
12 My understanding is that SSI develops this
13 business list samples by getting every Yellow
14 Pages in the country and calling those companies
15 that they don't already have information about,
16 to identify the SIC Code that they're in. And
17 they assign a business to only one SIC Code, and
18 that's how they developed those lists. And those
19 are the lists that I'm talking about using.
20 Q. Okay. But not the other way I guess,
21 huh?
22 A. I can never remember using it for a

Page 53

1 consumer study.
2 Q. Do any of your colleagues use SSI for
3 consumer studies?
4 A. Not that I know of.
5 Q. Okay. I wanted to clarify something
6 in your report. And we can either look at -- you
7 can actually look at your report, or I can just
8 refer to it. You've indicated that in Mr.
9 Winkelman's work the lack of a don't know or no
10 opinion option is a problem. And then you say
11 that studies indicate that presentation of an
12 explicit don't know or no opinion alternative
13 commonly leads to a 20 to 25 percent increase in
14 the proportion of respondents selecting that
15 response.
16 A. I do.
17 Q. You recall that?
18 A. Yes. And I think that's a quote from
19 Diamond.
20 Q. Okay. Now, have you yourself done any
21 research or analysis of the effect of an explicit
22 don't know or no opinion alternative in survey

Page 54

1 research?
2 A. I can never remember running any
3 experiment where I looked at that factor.
4 Q. Okay. So, in this particular instance,
5 you're relying on Diamond's work in that regard?
6 A. Yes. I am relying on Diamond's work,
7 and I'm also -- and without giving any
8 citations -- I'm also familiar that people have
9 been criticized in litigation areas for not
10 having a don't know option.
11 Q. Okay. We're talking about the Coors
12 case?
13 A. I'm not sure what case I'm referring to.
14 I'm just saying that it's in the back of my mind
15 that it's something that's important to include.
16 Q. In your report you refer to a case,
17 which if you're real lucky I'm going to be able
18 to find. Bear with me just a second. I thought
19 I had it marked. On page 18 of your report -- I
20 believe that's right. I could be wrong. Well,
21 skip that. I'm sorry. That's the wrong
22 reference. It's page 16, actually. Coors

Page 55

1 Brewing versus Anheuser Busch is a case that
2 Professor Diamond cites?
3 A. Yes.
4 Q. Okay. So, you just relied on
5 Professor Diamond's citation of that case?
6 A. That's correct. In this case I am. But
7 I wasn't necessarily referring just to this case.
8 It's just things that I've read over the years,
9 and I can't necessarily give you a citation for
10 that. That's all.
11 Q. Okay. Have you read the Coors case?
12 A. I may have. But it doesn't come to
13 mind.
14 Q. Apart from there being a -- not having a
15 no response or no option, do you know whether the
16 Court attacked the survey on any other grounds?
17 A. I do not know, no.
18 Q. In your report you've indicated that a
19 -- well, actually, you've indicated that
20 Professor Diamond indicates that a don't know
21 alternative would commonly lead to a 20 to 25
22 percent increase in the proportion of respondents

Page 56

1 selecting a don't know response?
2 A. That's correct.
3 Q. So, is my math correct that if, for
4 example, the I don't knows were, let's say,
5 nine percent if you had a -- with a closed-end
6 question, if you had an explicit no opinion
7 alternative, the increase would be 20 percent of
8 nine percent, whatever that is?
9 A. Well, I think that's what that reads.
10 But when you have a situation where, for example,
11 50 percent of the people say don't know, then
12 that would lead it up to 60 percent.
13 Q. Okay. And if you had ten percent, it
14 would lead to 12 percent; correct?
15 A. That's correct.
16 Q. Okay. But the one thing we're clear on,
17 you would not add 20 percent, you would take 20
18 percent of the don't know responses?
19 A. That's how I understand what she's
20 saying there.
21 Q. Okay. I did, too. But I want to make
22 sure I understood.

Page 57

1 A. But see below in Table 1 on the same
2 page --
3 Q. Yes.
4 A. -- where we have 116 out of 219?
5 Q. Yes.
6 A. That could lead from 53 to 63 percent,
7 about.
8 Q. Yes. 20 percent of --
9 A. Yeah, the 53, right.
10 Q. Yes.
11 A. Or 25 percent, yes.
12 Q. Okay. I want to talk a little bit about
13 control groups.
14 A. Yes.
15 Q. Do you use a control group of some type
16 in every survey that you do?
17 A. I try to.
18 Q. Okay.
19 A. If it's a survey where I'm investigating
20 a causal proposition. And by causal I mean,
21 A causes B. Or exposure -- I'll take it out of a
22 trademark context and put it in another one.

1 Exposure to an ad causes people to think
2 something about a product. In those situations,
3 I attempt to use a control group if I can.
4 Sometimes you can't, and you should use a control
5 question. But I think whenever you're
6 investigating causal issues, it's good to -- it's
7 mandatory, not good -- it's mandatory to have a
8 control of some type.
9 MR. UPCHURCH: I'm sorry. We've been
10 going about 70 minutes. Could we take a short
11 break?
12 MR. ALLGEYER: Oh, absolutely. Should
13 we just leave the line open?
14 MR. UPCHURCH: Yes. I think so.
15 MR. ALLGEYER: Okay. You want to take
16 -- how long do you want to take?
17 MR. UPCHURCH: Five minutes is fine.
18 MR. ALLGEYER: Okay, great.
19 MR. UPCHURCH: Thank you.
20 THE WITNESS: Thank you, gentlemen.
21 MR. UPCHURCH: Thank you.
22 (Thereupon, a recess was taken.)

1 BY MR. ALLGEYER:
2 Q. Okay. I wanted to just clarify one
3 thing. I'm not quite sure I'm so fascinated
4 with these mall intercepts as an alternative
5 method. But since we've been talking about it,
6 you said there's a way you randomize and take
7 other steps -- I don't want to get this wrong --
8 but to make sure you've got the appropriate cross
9 sample of consumers in such a survey, or words to
10 that effect. Do you recall that?
11 A. Yes. And I didn't use the word
12 randomize. But I did say that you want to get a
13 representative sample.
14 Q. Okay. And somehow you used some kind
15 of -- could you just walk me through how you
16 actually do that? Let me put it that way.
17 A. Well, okay. I think what you're
18 referring to -- and if I'm wrong, let me know --
19 I think what you're referring to is setting up or
20 developing a quota sample --
21 Q. That was the word; quota sampling
22 procedures. Yes.

1 A. -- to get a representative sample.
2 Q. Yes.
3 A. Okay. The example I used was, let's
4 suppose we have a situation where two-thirds of
5 the consumers are males and one-third are
6 females, and that 60 percent are between the ages
7 of 18 and 34, 20 percent are between 35 and 54,
8 and 20 percent are 55 and older. What you would
9 do is, as people are entering the mall or walking
10 through the mall, you would approach them and ask
11 them some screening questions. And let's suppose
12 that you already had enough men, well, then, you
13 would only approach women because you filled out
14 your quota of men. But let's suppose that you
15 need some men -- stop and go back again -- that
16 we need some men, but what we need are men who
17 are 55 and older.
18 Then as men are walking through the
19 mall, we wouldn't intercept kids, we wouldn't
20 intercept my son, but we would look for somebody
21 like me who maybe might be 45 or older as best we
22 can tell. We would ask that person his age and

1 if he says 46, we'd say, sorry, thank you very
2 much. But if he said 55 or 56, then we're still
3 trying to close that quota, and we would then ask
4 the person other screening questions. So, you
5 really are trying to, at the end, build a sample
6 that replicates the relevant population in terms
7 of the, in this case, the variables of age and
8 gender.
9 Q. But what is your method you use to
10 determine or set the quotas? How do you do that?
11 A. That's a good question, a great
12 question. I think what you do there is you go --
13 I'll tell you what I do most frequently. I'll
14 ask the client what information do you have about
15 the relevant population from your own internal
16 marketing research. And usually the client will
17 have some information saying who their target
18 market is or the relevant market is. If not,
19 then I will go to published sources and do a
20 literature review to identify the characteristics
21 from surveys that have been published or data
22 that has been published about the characteristics

Page 62

1 of that population. And we'll estimate what the
2 population should look like from either of those
3 two sources.
4 Q. Okay, thank you. I'm going to move on
5 for a moment. I want to go back to something --
6 in your report on page Ford 11. This is
7 Exhibit 1.
8 A. Yes.
9 Q. You have indicated, in paragraph 42,
10 agreement with Mr. Winkelman's survey objectives,
11 where you say they appear appropriate for the
12 issues of this case. Do you see that?
13 A. Yes, I do.
14 Q. Okay. What do you understand to be the
15 issues in this case?
16 A. Well, I understand it to be that people
17 may or may not associate the term Sportsman's
18 Warehouse with both your client and with Bass
19 Pro. And whether they're confused about whether
20 those stores with names that include Sportsman's
21 Warehouse are associated with each other in some
22 particular way.

Page 63

1 Q. Okay. Now, one of the questions
2 Mr. Winkelman asked was designed to determine
3 whether consumers associated Sportsman's
4 Warehouse with a particular source. And you
5 might recall those questions?
6 A. Yes. And it would be helpful maybe if
7 we could mark the Winkelman report as Exhibit 4,
8 so I could refer to those -- make sure that my
9 understanding of what you're asking about is the
10 same as you're actually asking about.
11 MR. ALLGEYER: Sure. Let's go ahead and
12 mark that.
13 THE WITNESS: That's, by the way, all
14 the exhibits that we have here.
15 MR. ALLGEYER: Okay. We've got the
16 whole thing nailed then. Okay, great.
17 (Ford Deposition Exhibit Number 4
18 was marked for identification.)
19 THE WITNESS: Okay. Let me make sure
20 that what you are asking about is what I think it
21 is.
22 BY MR. ALLGEYER:

Page 64

1 Q. If you go to page Ford 4-4.
2 A. Yes. Where it's questions -- oh, wait a
3 second. 4-4?
4 Q. Well, I've got a particular one.
5 Question A-4.
6 A. Oh, okay. Sorry. I was down at the
7 bottom. Yeah, good. Sorry.
8 Q. Okay. Now, without -- first of all, do
9 you understand the meaning of the term secondary
10 meaning in trademark law?
11 A. I have a vague understanding of it, but
12 I would not characterize it as any more than
13 vague.
14 Q. Okay.
15 A. And certainly not a lawyer's
16 understanding of it.
17 Q. Right. Well, it depends on the lawyer,
18 actually. [laughter]
19 Is it your understanding that in order
20 to establish secondary meaning, that a customer
21 has to be able to identify the specific source of
22 the product or service?

Page 65

1 MR. UPCHURCH: Objection. Calls for a
2 legal conclusion on the part of the witness.
3 MR. ALLGEYER: And I don't want a legal
4 conclusion, I want the witness' understanding of
5 a legal concept. So, unless Mr. Upchurch says
6 you can't answer, please go ahead and tell me.
7 MR. UPCHURCH: Oh, I'm not instructing
8 the witness not to answer. I just want the
9 record to be clear that this would be his
10 understanding, not a statement of the law, which
11 would be binding on Bass Pro.
12 MR. ALLGEYER: Absolutely. I understand
13 and agree with that.
14 And, by the way, I'm pleased I was able
15 to get on this far without objection. Thank you.
16 MR. UPCHURCH: I can do it more often.
17 MR. ALLGEYER: No. Don't feel like you
18 have to. Just kidding. [laughter]
19 THE WITNESS: I guess I don't know quite
20 how to answer that question. And I think I've
21 forgotten it. Maybe you could read it back to me
22 again.

1 BY MR. ALLGEYER:
2 Q. Let me state it over.
3 A. I'm sorry. I just kind of lost it.
4 Q. And I'll concede that Counsel will have
5 the same objection to this question as the last
6 one.
7 Is it your understanding that to
8 establish secondary meaning in trademark, that
9 one must establish that the customer is able to
10 identify the source of the product or service?
11 A. It's my understanding that you'd have to
12 come as close as possible to doing that. And I'm
13 only thinking about that from a measurement
14 perspective. I want to be able to say what
15 people are telling me and interpret that as
16 unambiguously as possible.
17 Q. Okay. One of the things I note in your
18 report, and you can look at page 20, is that --
19 and I'm probably going to get the math wrong
20 here. But, basically, you didn't count in Table
21 2, if I understand this right, as someone
22 recognizing Sportsman's Warehouse or associating

1 that mark with them if they couldn't properly
2 name the city for a Sportsman's Warehouse. Have
3 I said that about right?
4 A. You've said that correctly.
5 Q. Okay. So, just so I understand, for you
6 to believe that someone associates Sportsman's
7 Warehouse with a particular source, they have to
8 know what that source is or at least where it is,
9 the way you have delivered your report here; is
10 that true?
11 A. In terms of my adjusted analyses, that
12 is correct. Now, the problem I was working with
13 is that there seemed to be so much guessing going
14 on with the responses that were going on with the
15 Winkelman Report, that we needed a way to see
16 whether people actually were expressing knowledge
17 or were just giving a guess. And given the data
18 that existed, this was the way I came up with.
19 Q. Okay. Well, assume with me for a second
20 that in fact one is able to establish secondary
21 meaning if consumers associate a particular mark
22 with a single source, although that source may be

1 anonymous. Under those circumstances, would you
2 agree with me that simply not being able to
3 recognize the city doesn't necessarily disqualify
4 a response from associating a mark with a
5 particular source?
6 A. I agree with that point. However, it
7 seems to me that what you want to be able to do
8 in that situation is net out the guessing.
9 Because let's -- they associate it with a source,
10 but they don't have a lot of information about
11 the source, and how do we know they're guessing
12 or not. That's where the subject of controls
13 comes in. Because with controls you are able to
14 assume that the amount of guessing that goes on
15 would go on in both groups at a similar level.
16 And, so, you can wind up subtracting the control
17 results from the results you get to have a
18 little more confidence that the number you get
19 is representative of what the true level of
20 perception is among all consumers.
21 Q. Okay. So your concern, as I understand
22 it, was not necessarily to conform to the issue

1 of their being a single anonymous source, but you
2 rejected those answers because you thought that
3 might represent guessing?
4 A. That's correct. There really did seem
5 to be a substantial amount of guessing.
6 Q. I take it that you would agree with me
7 that those people who said I don't know or no
8 response weren't guessing, they were just saying
9 they didn't know?
10 A. That's right.
11 Q. And there was a significant number of
12 people who did indicate don't know responses to
13 this and other questions in the survey, true?
14 A. That is true.
15 Q. One of the criticisms that I understand
16 you had of the first Winkelman Report, which I
17 probably didn't put before you and you probably
18 don't have with you, but I bet you'll remember
19 from your report, is that it included information
20 for ten vendors. Do you recall that?
21 A. I do.
22 Q. Okay. And have you looked at the new

1 Winkelman Report, Exhibit 4?
2 A. Well, I have looked at the report. I
3 don't know where Exhibit 4 is specifically, but
4 I've looked at the report.
5 Q. Okay. The new one.
6 A. Do you mean Figure 4?
7 Q. No, I mean Exhibit 4. I'm sorry. The
8 final report. That's the new one that you just
9 recently looked at, right?
10 A. Yeah.
11 Q. Okay. And you'll note that or did you
12 note that Mr. Winkelman simply eliminated the
13 vendors from the report?
14 A. Oh, I see. I'm sorry. When you said
15 Exhibit 4 I was not thinking of Ford Exhibit 4 of
16 my deposition. I was looking for Exhibit 4 in
17 his report. Sorry.
18 Q. Okay, I got you. I could have done that
19 on a Friday afternoon, too, I guess.
20 A. I understand from his new report that
21 the vendors have been eliminated.
22 Q. All right. Does that eliminate your

1 concern based on the vendor data?
2 A. Well, if they're not in there, yeah.
3 MR. ALLGEYER: Okay. What I'm going to
4 do, Mr. Upchurch, with your permission, is take
5 about two minutes. I think we're done, but you
6 never know. So, if I may have just a couple of
7 minutes to look things over, we'll be able to
8 wrap it up quickly.
9 MR. UPCHURCH: That's fine.
10 MR. ALLGEYER: All right, thank you.
11 (Thereupon, a recess was taken.)
12 MR. ALLGEYER: Thank you very much,
13 Dr. Ford, for your time. Those are all the
14 questions I had. And, Mr. Upchurch, will you be
15 reading and signing?
16 MR. UPCHURCH: Yes. And I have no
17 questions for the witness at this time.
18 MR. ALLGEYER: Okay, thank you.
19 THE WITNESS: Thank you, sir.
20 (Thereupon, reading and signing not
21 being waived, the above proceedings
22 were concluded at 3:45 p.m.)

1 CERTIFICATE FOR READING AND SIGNING
2 I hereby certify that I have read and
3 examined the within transcript and the same is a
4 true and accurate record of the testimony given
5 by me.
6 Any corrections I have listed on the
7 separate errata sheet enclosed, indicating the
8 page and the line number of each correction.
9
10 _____
11 GARY T. FORD, Ph.D.
12 _____
13 Date
14
15
16
17
18
19
20
21
22

1 CERTIFICATE OF NOTARY PUBLIC
2 I, ELVIA FAYE GRANT, the officer before
3 whom the foregoing deposition was taken do hereby
4 certify that the witness whose testimony appears
5 in the foregoing deposition was duly sworn by me;
6 that the testimony of the witness was taken by me
7 in stenographic notes and thereafter transcribed
8 under my supervision; that said deposition is a
9 true record of the testimony given by said
10 witness; that I am neither counsel for, related
11 to, nor employed by any of the parties to the
12 action in which this deposition was taken; and,
13 further, that I am not a relative or employee of
14 any attorney or counsel employed by the parties
15 thereto nor financially or otherwise interested
16 in the outcome of the action.
17
18 _____
19 Elvia Faye Grant, Notary Public
20 District of Columbia
21
22 My Commission Expires: February 28, 2008

DEPOSITION EXHIBIT
FORD
1 9/15/06

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE TRADEMARK TRIAL AND APPEAL BOARD

In the matter of trademark Registration No. 2,390,988

For the mark of SPORTSMAN'S WAREHOUSE HUNTING FISHING CAMPING RELOADING OUTERWEAR FOOTWEAR and Design

Date registered: October 3, 2000

Bass Pro Trademarks, L.L.C.	)	
	)	
v.	)	
	)	Cancellation No. 92045000
Sportsman's Warehouse, Inc.	)	

REBUTTAL REPORT OF
GARY T. FORD

Professor of Marketing
Kogod School of Business
American University

Qualifications:

1. I am Professor and Chair of Marketing in the Kogod School of Business, at American University. Formerly, I was Chair and Associate Professor of Marketing in the College of Business and Management at the University of Maryland at College Park and Visiting Professor in the Department of Applied Economics at Catholic University of Leuven, Belgium.
2. During my career I have taught undergraduate and MBA courses in Marketing Research and Doctoral Seminars on Research Methodology, as well as Marketing Management, Consumer Behavior and other courses. The Marketing Research and Research Methods courses I have taught include material on survey research design, questionnaire design, sampling, statistical methods and other topics. I have supervised the research design,

questionnaire design, sampling procedures, analysis and final reports of over a hundred surveys completed in the Marketing Research courses I have taught.

3. In addition to the surveys I have supervised as a professor, I have also designed and conducted surveys for both my academic research and litigation. I have published over forty academic articles and papers and the vast majority of these involve surveys that I and my co-authors designed, conducted and analyzed. In addition, in a litigation context, I have designed, conducted, analyzed prepared expert reports on dozens of surveys and have evaluated the survey research efforts of others in trademark, false advertising and anti-trust matters.
4. My research has been published in the *Journal of Consumer Research, Journal of Marketing, Journal of Public Policy and Marketing, Journal of Marketing Research* and other journals, books and proceedings. In 1997 I was listed as one of the "best researchers in marketing" in an article "The Best Researchers in Marketing" published in the *Marketing Educator* (Summer).
5. My most recent peer-reviewed article, which was published in the *Journal of Public Policy and Marketing* in November 2005, is titled, "The Impact of the *Daubert* Decision on Survey Research Used in Litigation." This article is a comprehensive analysis of the criteria that are being used to admit surveys into or exclude surveys from trials since the U.S. Supreme Court decision in Daubert v. Merrell Dow Pharmaceuticals Inc.
6. I served on the Board of Directors of the Association for Consumer Research, the largest academic organization in the field of consumer behavior, for four years and served as an Editorial Review Board member for the *Journal of Marketing* for over ten years. At present, I serve on the Editorial Review Board of the *Journal of Public Policy and Marketing* and frequently review manuscripts for the *Journal of Consumer Research, Journal of Marketing Research, Journal of Marketing* and other journals and conferences. Currently, I serve as the academic representative of the American Marketing Association, our largest professional association, to the U.S. Bureau of the Census. I have served as a Marketing Expert for the Federal Trade Commission in both consumer protection and antitrust matters. I have also served as a Marketing Expert for the Department of Justice and the Consumer Product Safety Commission.

7. A complete copy of my resume including a list of my publications over the last ten years and a list of the proceedings in which I have testified in the last four years is attached at Appendix A.
8. I am being compensated at my usual rate of $625 per hour for my work on this case. My compensation is not dependant on the outcome of this case or on the opinions I express.

Scope of Work:

9. At the request of counsel for Bass Pro Trademarks, L.L.C. (Bass) I reviewed and evaluated the research methodology and findings presented in the Expert Report filed by Mark R. Winkelman (Winkelman report) regarding a survey he conducted ("Sportsman's Warehouse Recognition Assessment" hereinafter referred to as "the Winkelman survey") in the case of Bass Pro Trademarks, L.L.C. v. Sportsman's Warehouse, Inc.
10. A list of the documents considered and/or relied upon is attached as Appendix B.

Summary of Opinions:

11. The Winkelman survey is fatally flawed and consequently, does not provide scientifically valid data and should be disregarded. The Winkelman survey has the following flaws:

    - The sample used in the consumer survey is a "convenience sample" and no evidence inducates it is representative of the relevant population.
    - No information is provided about "call dispositions" in the consumer survey and consequently, the survey response rate is unknown leading to the conclusion that nonresponse bias is very likely.
    - The population for the vendor survey is never described so no one knows what types of firms were included in the vendor survey.
    - The vendor survey sample of 10 respondents was too small to be meaningful.

- The key closed-ended question on the survey assessing respondent recognition of the "Sportsman's Warehouse" name was biased because two of the three response alternatives favored Sportsman's Warehouse.
- Contrary to accepted research practice the aforementioned question, did not include a "don't know" option which encouraged guessing.
- The internet logo survey suffered from severe demand effects because respondents had already either mentioned or been asked about "Sportsman's Warehouse." It should have been conducted as a separate survey.
- The key closed-ended question on the survey assessing respondent recognition of the "Sportsman's Warehouse" logo contained a meaningless and uninterpretable response alternative which accounted for over 30% of the responses.
- As in the telephone survey, a "don't know" option was not offered.
- When adjusted for guessing less than 20% of the respondents to the consumer survey recognized the "Sportsman's Warehouse" name.
- Despite the demand effects, only 41.2% of respondents in the consumer internet logo survey recognized the Sportsman's Warehouse logo.

12. In my opinion, the Winkelman "survey" is fatally flawed and unreliable and therefore, unusable for determining recognition levels for the Sportsman's Warehouse name or logo and what data do exist lead to the conclusion that neither the name nor logo have high levels of recognition.

Summary of Winkelman Survey Methodology and Results:

13. According to the Winkelman report, the purpose of the survey, "was to obtain information that will be used to determine consumer's recognition of the words/name Sportsman's Warehouse and the Sportsman's Warehouse logo." The Winkelman Report then states, "To do this the study focused on the following objectives: Assess the implied meaning of the words 'Sportsman's Warehouse,'" and "Assess the implied meaning of the 'Sportsman's Warehouse' sample logo." (Winkelman report, p. 3-1.)

14. In the attempt to realize these objectives, Mr. Winkelman conducted telephone and web-based interviews with two populations of potential respondents: 1.) "telephone households who have expressed an interest in fishing, hunting and/or camping and live in a state in which ... there is both a Bass Pro Shop and a Sportsman's Warehouse; and 2.) "vendors fishing, hunting, camping and/or other outdoor gear." (Winkelman report, p. 3-1.)
15. The telephone numbers for the consumer survey were obtained from Survey Sampling International. The 52 vendors on this list were obtained from Sportsman's Warehouse. (Winkelman report, p. 3-1.)
16. According to the Winkelman report, 474 interviews were completed with consumers and 10 were completed with vendors. (Winkelman report, p. 3-1.)
17. Contrary to accepted and normal practice, the Winkelman report does not provide data on survey response rate for either the consumer or vendor surveys, and no information is provided in the Winkelman report about call dispositions, so the response rate cannot be calculated for the consumer survey.[1]
18. The telephone survey questionnaire began with the following introduction:

> "Hi, my name is _____ from Winkelman Consulting in Fargo, North Dakota, and I'm not selling anything. I'm simply conducting a short survey about stores that sell hunting, fishing, camping and outdoor gear in your state. The survey will only take about 1 to 3 minutes, there are no right or wrong answers and your answer will be kept confidential. This survey is for people who have bought, hunting fishing or outdoor gear in the past twelve months. Using a scale of within the past 30 days, 1 to 6 months, 7 to 12 months, more than 12 months or never, how recently have you bought any ...
>
> Fishing gear?
> Hunting gear?
> Camping gear?
> Other outdoor gear?"

19. If the respondent had not bought any of these types of items, the interviewer asked whether anyone else in the household had purchased any of these types of items in the past twelve

[1]The survey response rate can be inferred for the vendor survey because 10 interviews were completed from an initial population of 52, which makes the response rate approximately 19.2% (10/52 = 19.2%). The survey response rate measures the proportion of respondents who are eligible for a survey who were actually interviewed.

months and if so, asked to speak to him or her. If no eligible respondent was found, the interview was terminated. If an eligible respondent was not available, the interviewer arranged for a call back. When an eligible respondent was reached, the interviewer then asked:

"Do you or any members of your household work for a store that sells mainly hunting, fishing, camping or outdoor gear?"

20. If the answer was "yes," "not sure" or "no response," the interview was terminated. If the answer was "no," the interviewer asked:

"A1. First, do you have a computer that you can use to access the Internet?"

21. Next the interviewer said:

"A2. I'd like you to tell me what words (sic) or words you use to describe a large store that sells mainly hunting, fishing, camping and outdoor gear. For example, people may use the words, Mexican restaurant" to describe a business that serves meals that include mainly Mexican foods. Also, people may use the word "Thermos" to describe a travel vacuum bottle or container a person can carry with them and use to keep cold beverages cold and warm beverages warm. So, what word or words would you use to describe a large store that sells mainly hunting, fishing, camping and outdoor gear?

22. There were no probe phrases used with this question which is very unusual.[2]

23. Next the interviewer asked respondents:

"A3a. Now I'd like you to think about the names of large stores that sell mainly hunting, fishing, camping and outdoor gear. When you think of names of large store that sell mainly hunting, fishing, camping and outdoor gear, what is the one store name that comes to mind first?"

24. Respondent were then asked the following probe question,

[2] A probe is phrase designed to get at other thoughts the respondent may have, such as, "What else?" Without a probe or probes the researcher does not know if he has captured the ways respondents describe a large store that sells mainly hunting, fishing, camping and outdoor gear.

"A3b-g. What other store names come to mind when you think of large stores that sell mainly hunting, fishing, camping and outdoor gear?"

25. Next, respondents were asked the key question on the telephone portion of the consumer interview:

"A4. When you hear the words 'Sportsman's Warehouse', do you think these words are most likely ...

3. The name of one specific store or chain of stores,
4. Words that describe one specific store or chain of stores, or
5. General words that can be used to describe any large store that sells mainly hunting, fishing, camping or outdoor gear.
91. *(Do not read)* Not sure, No response."

26. If respondents provided either of the first two answers to question A4, they were asked:

"A5. To the best of your knowledge, in what city or town is a 'Sportsman's Warehouse' located in your state?"

27. Respondents who indicated earlier that they had Internet access, were asked:

"B1. For this interview I will also want you to look at some sample logos, and then answer a few questions about the sample logos. To do this, I will need you to go to an Internet Web site that has the sample logos. Can you talk to me on this telephone and access the Internet at the same time?"

28. If respondents said "yes" or that they could be online and talk on a different number the interview continued. If they said "no" or were not sure, the interviewer asked to write down an IP address (66.223.28.200/40), a five digit "user name ID number" and a five digit "password number." They were then asked to go to their computers and go to the website associated with the IP address, and using their user name number and password access the site and the logos. The interviewer also gave them instructions to disable the function on their web browser that blocks "pop-ups" and then asked them for an email address so they could be sent a reminder with the IP address, username and passwords.

29. Respondents who were able to be interviewed and be online simultaneously were told:

"B2c. First I want you to open Internet Explorer or the program you use for going to the Internet. The first screen that pops up should be your web browser. In the address box at the top of your web browser, type 66.223.28.200/20 and then either hit the enter key or click on the 'go' button. When you see the page that says 'Winkelman Consulting – Welcome to our sample logo web site!,' let me know."

30. If respondents had trouble getting online or to the website they were asked to "see if they can figure out what is wrong" and scheduled for a callback time.

31. The respondents who could get to the website were told:

"B3a. Next, I'm going to give you the user name and password you need to go to the next page. Before I do that however, let me tell you what will happen once you enter this information. As soon as you click on the submit button, a sample log will appear for about ten seconds, then disappear. Once the sample logo disappears, let me know. First, click on the "user name" box, then type in this 5-digit 'user name' number: *(Username from list)*. Now, click on the 'password box,' then type in this 5-digit password number *(password from list)*. Last, get ready to look at the sample logo and then click the 'continue' button.

32. Once the logo disappeared, the interviewer asked:

"B3b. After seeing this sample logo, do you think it is most likely ...

1. An actual logo that is currently used by one specific store or chain of stores,
2. A logo that could be used to identify one specific store or chain of stores or
3. A general logo that could be used by any sporting goods store to let consumers know their stores sell mainly hunting fishing camping or outdoor gear."

33. Respondents were then shown the second logo, and after it disappeared were asked the same question.

34. The number of consumers who responded to the "logo" Internet questions was dramatically smaller than to the telephone interviews, i.e., 68 as compared to 474 (Bates No. SP 01012). Thus, only 14.3% of the consumers answered the Internet survey. Although 338 respondents had Internet access, only a small percentage of consumers who could be interviewed while on the Internet at the same time, i.e., 8.0% (38 consumers) and those who could not had to follow elaborate and complicated instructions for completing the interview at another time. Of the 338 respondents who had could access the Internet from home, only

an additional 30 consumers completed the Internet portion by going to their computers at some later point and following the elaborate instructions requiring an IP address, and ID user name and password.[3] Thus, the response rate among this group was 10.0%, i.e., 30/(338 – 38).

35. The last section of the consumer survey asked three demographic questions regarding the respondent's age, the number of years they had lived in their state and their gender.

36. The vendor survey used the same questions as the consumer survey except that it asked for the number of years the respondent had been selling hunting, fishing, camping or outdoor gear to retailers, instead of the number of years the respondent lived in the state. Surprisingly, there was no attempt in the Winkelman survey of vendors to ensure that the respondent was qualified to respond to the survey. For example, the survey did not ensure that the respondents employed by the various vendors had appropriate positions in their firms (Vice President of Sales, Manager of Sales, etc.) to be knowledgeable about the topic of the survey.

37. Based on his survey, Mr. Winkelman drew four conclusions: 1.) "Nearly one out of every two respondents felt the words 'Sportsman's Warehouse' are at least the name of specific store or chain that sells mainly hunting, fishing, camping or outdoor gear."[4] (Winkelman report, p. 2-1.) 2.) Nearly one out of twenty respondents named "Sportsman's Warehouse' first (top-of-mind awareness), while more than one of every ten named Sportman's Warehouse (unaided awareness, which includes top-of-mind awareness).[5] (Winkelman report, p. 2-3.) (3.) "The results indicate that only a minuscule proportion of respondents used the words 'sportsman(s)' or 'sportsman's warehouse' as a generic term for a large store that sells mainly hunting, fishing, camping and outdoor gear." (Winkelman report, p.

---

[3]Seven of the 10 vendors completed the Internet portion of the survey.

[4]Apparently, Mr. Winkelman combined the responses to questions A2, A3a, A3b-g, A4 and A5 to calculate this percentage which he estimated to be 48.3%. In Figure 1a of the Winkelman report it is the sum of "Top-of-mind-awareness" (3.7%), "Unaided awareness" (7.0%), "Named a correct city, metro area" (10.7%) and "Name of one, specific store or chain" (26.9%). These percentages represent the combined results for the consumer and vendor surveys. Additionally, the ""Name of one, specific store or chain" category is actually the sum of two categories "Name of one, specific store or chain" and "Words that describe one specific store or chain of stores."

[5]These percentages appear to be "Top-of-mind-awareness" (3.7%), and the sum "Top-of-mind-awareness" (3.7%) and "Unaided awareness" (7.0%), equals 10.7%. Rather than 3.7% meaning "nearly one out of twenty," more accurately it represents "less than 1 out of 25."

2-4.) 4.) "Of those who viewed the sample logos, nearly eight of every ten felt the Sportsman's warehouse logo is at least a logo that could be used to identify one specific store or chain of stores.[6] (Winkelman report, p. 2-5.)

Evaluation of the Winkelman Survey Methodology:

38. Every survey can be evaluated by examining the strengths and weaknesses of the various components of the survey:

- The relevance of the survey objectives to the issues of the case,
- The research design including:
  - Definition of the universe, sampling procedures and representativeness of the ending sample,
  - The clarity and degree of bias in the questions and procedures,
  - The ability to account for alternative explanations for the results,
- The appropriateness of the data analysis and interpretation of results, and
- The conclusions drawn from the survey.

39. Additionally, "The Reference Guide on Survey Research" by Shari Diamond published in the Second Edition of the *Reference Manual on Scientific Evidence* (Federal Judicial Center, 2000, pp. 229-276) provides a list of factors that need be considered when designing and/or evaluating a survey used in litigation.
40. My evaluation of the Winkelman survey draws on criteria from listed in Professor Diamond's work and from my own teaching and research experience.
41. An important point that should be emphasized is that the various components (e.g., the research design, sample, questionnaire, etc.) of a survey are non-compensatory. By that I mean that a strong sample does not compensate for a biased questionnaire or vice-versa. In other words one fatal flaw in a survey renders it unusable.

---

[6]This percentage represents the sum of two of the responses to question B3b: "After seeing this sample logo, do you think it is most likely: An actual logo that is currently used by one specific store or chain of stores" (46.7%) and "A logo that could be used to identify one specific store or chain of stores" (29.3%) which equals 76.0%.

42. Relevance of survey objectives. As noted earlier the objective of the Winkelman survey was to "obtain information that will be used to determine consumer's recognition of the words/name Sportsman's Warehouse and the Sportsman's Warehouse logo." (Winkelman report, p. 3-1.) In my opinion, the survey objectives appear appropriate for the issues of this case.

43. Definition of universe, sampling plan and representativeness of the ending sample for the consumer survey. The universe for the consumer survey was defined as consumers who had purchased hunting, fishing, camping or outdoor gear in the past twelve months. To approximate this population Mr. Winkelman obtained a list from Survey Sampling International (SSI) which compiles lists from people who express an interest in a topic by for example, completing warranty registration cards. Thus, the SSI sample is a "convenience sample," which is not projectable to the population of households who have purchased hunting, fishing, camping or outdoor gear in the past 12 months. Professor Diamond states, "when respondents are not selected randomly from the relevant population, the expert should be prepared to justify the method used to select respondents." (Diamond, 200, p. 244). What that means here, is that Mr. Winkelman should have demonstrated how and why he believes the SSI sample replicates the relevant population. Nothing of the sort was done.

44. Mr. Winkelman stated the list was randomly drawn from the SSI convenience sample and that "telephone interviews were conducted from May 31 to June 12, 2006." (Winkelman report, p. 3-1.) Contrary to accepted survey research practice, no information is provided about the procedures used in the telephone survey. For example, no information is provided about the number and timing of "callbacks," i.e., subsequent attempts to reach people who were not reached on the first call. More importantly, no information is provided about the survey response rate, which is the key initial indicator of whether the sample is likely to be biased.[7] The usual and accepted practice is to provide a "call disposition report," i.e., a report that summarizes the disposition of every call that was made

[7]The response rate to a survey varies between 0% and 100%, with numbers closer to 100% indicating little likelihood of nonresponse bias and numbers closer to 0% indicating that nonresponse bias is a real concern. Standardized approaches for calculating survey response rates have been proposed by the Council of American Survey Research Organizations and American Association of Public Opinion Research. See, for example, www.aapor.org/calculator.xls.

by categories such as, no answers, refusals, wrong number, nonworking numbers, completed interviews, etc. The call disposition report is required in order to compute the survey response rate, which is a key indicator of the likelihood of non-response bias.

45. Professor Diamond says, "The completeness of the survey report is one indicator of the trustworthiness of the survey and the professionalism of the expert who is presenting the results of the survey. Regarding call disposition she states a survey report generally should provide "a description of the results of the sample implementation including a.) the number of potential respondents contacted, b.) the number not reached, c.) the number of incomplete interviews or terminations, d.) the number of noneligibles, and f.) the number of completed interviews." (Diamond, 2000, p. 270.) The fact that the Winkelman survey does not report the response rate raises suspicions that the response rate is very low, and hence that the sample is not representative. Without knowing the response rate it is impossible to have confidence that the sample is representative of the population.

46. In sum, the SSI sample frame for the consumer survey was a non-projectable convenience sample, no details were provided about the procedures used to increase the response rate and the response rate to the survey was not reported, all of which raises a red flag about the representativeness of the sample used in the Winkelman survey. Essentially, there is no evidence in the Winkelman report that the sample used in the consumer survey is representative of the population of consumers who would or would not be confused by the marks at issue.

47. Definition of universe, sampling plan and representativeness of the ending sample for the vendor survey. According to Mr. Winkelman, 52 telephone numbers for the vendor survey were obtained from Sportsman's Warehouse (Winkelman report, p. 3-1.) Mr. Winkelman attempted to contact all 52 and actually completed 10 interviews with vendors. The Winkelman report provides no information about the universe of vendors. In fact, it is not even clear why any vendors were interviewed at all since the objective of the survey was to ""obtain information that will be used to determine consumer's recognition of the words/name Sportsman's Warehouse and the Sportsman's Warehouse logo" and not to determine vendor's recognition. Also, it is not clear whether the vendors are vendors that

Sportsman's Warehouse buys from, sells to or are vendors that sell products to retailers.[8] If the "universe of vendors" is vendors who sell to Sportsman's Warehouse the list should include all vendors who sell to retailers who offer products for hunting, fishing, camping and outdoor use. If the vendors universe used in the Winkelman survey are all vendors who sell to Sportsman's Warehouse, the sample is very likely to be biased in a way that favors the defendant because they will be more likely to be familiar with Sportsman's Warehouse.

48. Additionally, as was the case with the consumer survey, no information was provided about call dispositions, although the survey response rate can be estimated at 10/52 = 19.2%. It is surprising to me that Mr. Winkelman was only able to complete interviews from less than 20% of respondents when he began with a list supplied by his client. Obviously, one has to be concerned about non-response bias with this low of a response rate, yet there was no indication that Mr. Winkelman considered this to be a potential problem.

49. There was no attempt to screen vendor survey respondents for eligibility. Apparently, the interviewer just interviewed whoever answered the phone, whether it was an administrative assistant, a secretary, a sales person, bookkeeper, or any other person. The accepted approach in business surveys is to screen for eligible respondents, arrange callbacks to reach unavailable respondents, and offer a cash incentive.

50. Finally, a sample of 10 is too small to be usable. And, since vendors and consumers have different experiences, the responses from the two samples cannot be added together as was done in the Winkelman report. It is really an issue of adding apples and oranges; it is improper and should not be done.

51. In sum, the vendor survey sample is indefensible. It is not clear why any vendors were interviewed, the sample was drawn from an unknown universe, there were no eligibility criteria for respondents, the response rate was low and the ending sample is too small to be used.

52. Clarity and degree of bias in the survey questions. There are numerous biases in the Winkelman questionnaire. They are discussed below.

---

[8]Question D2 on the vendors survey asks, "How long have you personally sold hunting, fishing, camping or outdoor gear to retailers?," so it appears that the respondents were wholesale vendors of some type. Whether they also sold to Sportsman's Warehouse or were its competitors is not clear.

53. Inexplicably and inexcusably, the introduction to the Winkelman survey tells respondents what the eligibility criterion is for the survey when it says, "This survey is for people who have bought hunting, fishing, camping or outdoor gear in the past twelve months." I can never remember seeing a survey that told people before they were screened what to answer in order to be eligible to continue or not continue. Mr. Winkelman could have informed people about the topic of the survey and then asked, "May I please speak to the person in your household who has most recently purchased hunting, fishing, camping or outdoor gear?" When that person came to the phone, he or she could have been asked when they last purchased hunting, fishing, camping or outdoor gear. With this approach Mr. Winkelman would have been interviewing the person with the most current knowledge and would not have revealed the eligibility criterion.

54. Question A2 which attempts to explain the concept of a generic product and give two examples is virtually useless for determining whether people perceive "sportsman's warehouse" is perceived as a generic term, because it is so unfocussed. The open-ended question was, "So, what word or words would you use to describe a large store that sells mainly hunting, fishing, camping and outdoor gear?" Thus, with this vague question, it is not surprising that the respondent gave over 100 different words and phrases. Furthermore, apparently a large proportion of respondents did not understand the explanation because approximately one-quarter of the responses were the brand names of retailers, rather than descriptive phrases.

55. The next series of open-ended questions, asked respondents to give the names of stores that sell mainly hunting, fishing, camping and outdoor gear. Respondents who said "Sportsman's Warehouse" were assumed by Mr. Winkelman as perceiving that "Sportsman's Warehouse" was a brand name without any further checks to confirm his assumption. This is particularly troublesome because the phrase "Sportsman's Warehouse" is part of the Bass Pro name. Further, as is discussed below when asked to name a city in which a "Sportsman's Warehouse" was located, respondents were almost as likely to name a city in which a Bass Pro store was located as one in which a "Sportsman's Warehouse" was located.

56. If respondents did not mention "Sportsman's Warehouse" to these questions they were asked the following closed-ended question[9]:

"A4. When you hear the words 'Sportsman's Warehouse', do you think these words are most likely...

3. The name of one specific store or chain of stores,
4. Words that describe one specific store or chain of stores, or
5. General words that can be used to describe any large store that sells mainly hunting, fishing, camping or outdoor gear.
91. *(Do not read)* Not sure, No response."

57. Question A4 is fatally flawed on two dimensions. First, Mr. Winkelman interprets two of the three responses that were read to respondents in question A4 in favor of his client, Sportsman's Warehouse.[10] Closed-ended questions are often criticized because they may suggest response alternatives the respondent would not have thought of. Clearly, if the researcher only gives three response alternatives and two of them favor his client, there is a substantial likelihood of bias. In my opinion, question A4 is fatally flawed because it is so fundamentally biased.

58. Second, the Winkelman survey does not have a caution against guessing or offer a "don't know" option, which is the accepted and expected practice in litigation surveys. Professor

[9] A "closed-ended" question is one in which response alternatives are suggested. The response alternatives in a closed-ended question should be mutually exclusive, internally homogeneous and collectively exhaustive. An "open-ended" question is one in which response alternatives are not provided.

[10] That can be seen by examining the results in Figure 1a, which shows that 10.7% of respondents "named a correct city or metro area." This percentage = 52 respondents, i.e., 10.7% * 474 = 52. A cross tabulation of variable A4 and A5 shows that 40 respondent both said Sportsman's Warehouse was the name of a specific store or chain of stores and named a correct city/metro area. Eleven respondents said Sportsman's Warehouse described one specific store or chain of stores and gave a correct city or metro area. One respondent said Sportsman's Warehouse described one specific store or chain of stores and gave a correct city or metro area in a different state. The sum of 40 + 11 + 1 is 52, or 10.7%, all of whom were counted in Figure 1a as naming a correct city or metro area. (These responses included respondents to consumer survey and one vendor sample respondent, who should have been excluded because the samples should not have been mixed. Also, two responses were coded improperly by Mr. Winkelman because the respondents said the Sportsman's Warehouse was in "Grapevine," a city that has a Bass Pro store but no Sportsman's Warehouse store.) More importantly, two of the three answers to question A4 favored the defendant, a clear case of a biased question.

Diamond states, "Studies indicate that although the relative distribution of the respondents selecting the *listed* choices is unlikely to change dramatically, presentation of an explicit 'don't know' or 'no opinion' alternative commonly leads to a 20-25% increase in the proportion of respondents selecting that response." (Diamond, 2000, p. 250.) One acceptable way to diminish the effects of guessing is to have a filter in the form of an explicit "don't know" option. In fact, Professor Diamond cites a case (*Coors Brewing Company v. Anheuser-Busch Cos*, F. Supp 965, 972-73, (S.D.N.Y. 1992)) in which a question without a "no opinion" option was rejected by the court (Diamond, 2000, p. 251, fn. 86).

59. Overall, 431 respondents to the consumer survey were asked question A4 and approximately half (219) gave one of the first two responses, i.e., 170 said "name of a specific store" and 49 said "words that describe a specific store." Respondents who gave either of the first two answers above were then asked question A5, "in what city or town is a "Sportsman's Warehouse" located in your state."

60. The responses of the 219 respondents who were asked question A5, were distributed as follows:

Table 1*

	N	%
City/metro area with Sportsman's Warehouse only:	27	12.3%
City/metro area with Bass Pro only:	22	10.0%
City/metro area with both SW and BP:	24	11.0%
City/metro area with neither:	28	12.8%
Do not know:	116	53.0%
Not interpretable:	02	00.9%
Total	219	100.0%

*See Appendix C for details.

61. Thus, when respondents to question A4 were asked where a Sportsman's Warehouse was located in their state over half (53.0%) said they did not know. The remaining answers were scattered fairly evenly among respondents who identified a city with a Sportsman's Warehouse but no Bass Pro store (12.3%); a city with a Bass Pro store but no Sportsman's Warehouse (10.0%); a city with both a Sportsman's Warehouse and a Bass Pro store (11.0%) and a city with neither a Sportsman's Warehouse or a Bass Pro Store (12.9%). It

seems clear that a substantial number of respondents were guessing and confusing Sportsman's Warehouse with Bass Pro. Furthermore, because respondents were approximately just as likely to identify a city with only a Bass Pro store as one with only a Sportsman's Warehouse store, one cannot know whether they were thinking of a Bass Pro or Sportsman's Warehouse when they identified cities that had both stores.

62. Because it is appears obvious that there was a substantial amount of guessing by respondents to question A5, and because of the biased nature of the response categories to question A4, I have no confidence that the results shown in Figure 1a regarding the 10.7% who "named a correct city metro area" have any meaning whatsoever. In my opinion, question A4 shows that a substantial but unknown percentage of respondents guessed where a Sportsman's Warehouse might be located and frequently confused Bass Pro and Sportsman's Warehouse stores.

63. The logo internet survey. Respondents to the first part of the telephone survey were asked to go online and look at two logos and answer questions. As noted earlier, those consumers who could not simultaneously go online and talk with the interviewer were unlikely to complete the interview. There are more severe problems with the logo internet survey, however, than complicated instructions.

64. The most important problem is known as "demand effects," that is, the tendency for a survey to cue respondents regarding the answer the researcher desires. Respondents who had just been asked a series of questions and had either volunteered the words "Sportsman's Warehouse" or had been directly asked about those words, are now being asked whether they perceive the logo which says Sportsman's Warehouse is an actual logo or could be an actual logo. Of course, some percentage of respondents will think, "I'm supposed to answer it's a real logo or could be a real logo." The logo survey should have been completed separately from the telephone survey to minimize bias due to demand effects. This is a fatal flaw because the extent of bias cannot be estimated.

65. Another fatal flaw is in the response categories that were used:

> "B3b. After seeing this sample logo, do you think it is most likely ...
> 1. An actual logo that is currently used by one specific store or chain of stores,
> 2. A logo that could be used to identify one specific store or chain of stores or

3. A general logo that could be used by any sporting goods store to let consumers know their stores sell mainly hunting fishing camping or outdoor gear."

66. As with the closed-ended question that was used in the telephone survey, there was no "do not know" option so guessing was not discouraged and two of the three response categories favor Mr. Winkelman's client, a definite indication of bias. Additionally, the second response category is indefinite and hence cannot be used because the responses are meaningless. The phrase "a logo that could be used" is very much different than saying a logo that "is used." "Could" is like "might" or "maybe" and hence does not allow one to categorize respondents as perceiving the logo is used. Since 32.4% of the respondents in the consumer survey said it was a logo that "could" be used, the results to the logo survey are changed dramatically. That is, even with the demand effects, problem discussed earlier, only 41.2% of consumers perceived the Sportsman's Warehouse logo was currently being used (Winkelman report, SP01013).

67. In my opinion, the logo internet study conducted by Mr. Winkelman was fatally flawed from the get-go. It never should have been conducted with consumers who had just been asked a series of questions. Instead Mr. Winkelman should have conducted an entirely separate logo study with consumers not contaminated with prior questions. The closed-ended question was both biased in favor of his client and so poorly worded that one of the response alternatives is meaningless. Finally, the results do not support the fact that the Sportsman's Warehouse is well-known. In fact, while two-thirds (66.2%) of the respondents perceived the Bass Pro logo as an actual logo, only 41.2% of the respondents perceived the Sportsman's Warehouse logo in the same way. Clearly, the Bass Pro logo is more widely recognized than is the Sportsman's Warehouse logo.

68. The ability to account for alternative explanations for the results. It is fundamental in all research and especially in litigation research to be able to control for alternative explanations for the results.[11] The typical way in which this is done is with either a control group or control questions. The Winkelman survey used neither which makes it impossible

[11]Being able to control for alternative explanations is one of the criteria used for admissibility of scientific evidence in Federal trials. This point is discussed in my article, "The Impact of the *Daubert* Decision on Survey Research Used in Litigation," (Ford, 2005, p. 244).

to know the extent to which the results are due to biased questions, question order or other factors.

69. For example, a defensible survey could have used a separate control group which was given the name of an imaginary (control) store such as, "Everything Outdoors" and asked the same series of questions. Then the control results could have been subtracted from the Sportsman's Warehouse results to net out the percentage of consumers who actually know "Sportsman's Warehouse" from those who were purely guessing. Based on the results from my analysis of respondents' lack of knowledge of cities in which Sportsman's Warehouses are located, it seems clear that the Winkelman survey results are based in large part on guessing. If consumers were randomly assigned to the treatment (i.e., Sportsman's Warehouse" and control (i.e., "Everything Outdoors") the groups would be equivalent, and the results would have had more credibility. Alternatively, the Winkelman survey could have used a control question to estimate the degree of guessing.

70. The failure to use any controls whatsoever makes it impossible to know whether the results of the Winkelman survey are valid. All one knows for sure, is that the key question on the telephone survey (question A4 which asked whether Sportman's Warehouse was the name of a specific store) was biased in favor of Mr. Winkelman's client and that when the responses were checked to see whether the people who gave the answer most favorable to his client knew the location of the store, 53.0% did not know, 12.8% guessed wrong and the remainder were just as likely to identify the location of a Bass Pro store as a Sportsman's Warehouse store. Thus, the most obvious alternative explanation for the Winkelman survey results is that respondents were guessing and because of the biased nature of question A4 they guessed in the direction that favored Sportsman's Warehouse. Without a control there is no way to rule out that alternative explanation for the results.

71. <u>The appropriateness of the data analysis and interpretation of results.</u>[12] As just discussed the overriding issue with question A4 in the Winkelman survey is that it is clear that there

[12]This discussion is limited to the telephone portion of the consumer survey only because as explained earlier: a.) the vendor sample is too small to be usable and it is improper to combine consumers and vendors into one database and b.) the logo survey suffers from such severe demand effects and improper response alternatives it cannot produce valid findings. Furthermore, despite these weaknesses only 41.2% of respondents said the Sportsman's Warehouse logo was currently

was a substantial amount of guessing. The Winkelman survey report did not attempt to estimate the degree of guessing or adjust the results in any way. While it is not as accurate as having a control group analysis, one way the Winkelman results can be adjusted is to remove the respondents who had no idea where a Sportsman's Warehouse was, respondents who guessed a city with no Sportsman's Warehouse, respondents who said it was in a city where there was only a Bass Pro store, and one-half of the respondents who named a city that included both a Bass Pro store and a Sportsman's Warehouse. If the Winkelman survey data are adjusted in this way the adjusted results for the consumer survey are:

Table 2

	N	%	Adjusted Recognition Of Sportsman's Warehouse:
Top-of-mind awareness:	13	02.7%	02.7%
Unaided awareness:	30	06.3	06.3%
Named correct city for Sportsman's Warehouse:	27	05.7	05.7%
½ of those who named correct city for Sportsman's Warehouse and Bass Pro store:	13	02.7	02.7%
½ of those who named correct city for Sportsman's Warehouse and Bass Pro store:	12	02.5	
Unable to name correct city for Sportsman's Warehouse:	167	35.2	
General name for any large sporting goods store:	170	35.9	
Not sure:	42	08.9	
Total	474	99.9%	17.4%

72. When the Winkelman survey results are adjusted for guessing, the percentage of respondents who recognized the Sportsman's Warehouse name can be estimated at 17.4%, a result that indicates little recognition of that name. Even if all of respondents who named a city in which a Sportsman's Warehouse was located regardless of whether there was a Bass Pro store in that city also are included, the percentage who recognize Sportsm\an's Warehouse only increases by 2.5% to 19.9%.

73. It is also important to remember from Table 2 that almost just as many respondents named a city in which only a Bass Pro store was located (i.e. 22 respondents) as named a store in

---

being used as compared to two-thirds who said the Bass Pro logo was being used. Clearly, the Bass Pro logo is more widely recognized than is the Sportsman's Warehouse logo.

which a only a Sportsman's Warehouse was located (i.e., 28 respondents) indicating that confusion exists between these outlets.

74. It is also important to note as is shown in Table 3 that in questions 3a through 3g when people were asked the names of "large stores that sell mainly hunting, fishing, camping or outdoor gear" they were much more likely to name a Bass Pro store than to name a Sportsman's Warehouse store:

Table 3*

	N	%
Unaided mention of Sportsman's Warehouse but not Bass Pro:	21	04.4%
Unaided mention of both Sportsman's Warehouse and Bass Pro:	22	04.6%
Unaided mention of Bass Pro but not Sportsman's Warehouse:	216	45.6%
Neither mentioned:	215	45.4%
Total:	474	100.0%

*See Appendix D for details.

75. When asked to name stores "large stores that sell mainly hunting, fishing, camping or outdoor gear," the data in Table 3 show that 50.2% of consumers named Bass Pro (i.e. 45.6% + 4.6%) and 9.0% named Sportsman's Warehouse (i.e., 4.4% + 4.6%), an overwhelming result, not reported by Mr. Winkelman. Thus, the results from the unaided question portion of the Winkelman survey supports the conclusion that Bass Pro is much recognized as having "large stores that sell mainly hunting, fishing, camping or outdoor gear" than is Sportsman's Warehouse.

Conclusions:

76. For the reasons identified and explained above, (including the convenience sample, lack of data supporting the representativeness of the sample, biased response categories, lack of

controls, incomplete data analysis, etc.) I believe the Winkelman survey useless for determining whether the phrase "Sportman's Warehouse" and the "Sportsman's Warehouse" logo is recognized by consumers. If anything, when properly analyzed, the survey results support the conclusion that "Sportsman's Warehouse" has little name recognition compared to Bass Pro.

Gary T. Ford 8/23/06

Gary T. Ford Date

Appendix A

Curriculum Vitae of Gary T. Ford

GARY T. FORD

American University
4400 Massachusetts Ave., N.W.
Washington, D.C. 20016
Tel. (202) 885-1978
Fax (202) 885-2691
gtford@american.edu

EDUCATION

Ph.D., Marketing, STATE UNIVERSITY OF NEW YORK AT BUFFALO, Buffalo, NY, 1973.

M.B.A., Marketing, STATE UNIVERSITY OF NEW YORK AT BUFFALO, Buffalo, NY, 1968.

B.B.A., Accounting, CLARKSON COLLEGE, Potsdam, NY, 1966.

PRESENT POSITION

AMERICAN UNIVERSITY, Kogod School Business, Washington, D.C., 1985-present
Professor and Chair of Marketing

ACADEMIC/PROFESSIONAL EXPERIENCE

AMERICAN UNIVERSITY, Kogod School of Business, Washington, D.C.,

Professor of Marketing, 1985-present.

Chairman of Marketing, 1989-94, 1999-2001, 2004, 2006 to present.

CATHOLIC UNIVERSITY OF LEUVEN, Department of Applied Economics, Leuven, Belgium, 1991-1992
Visiting Professor of Marketing

UNIVERSITY OF MARYLAND, College of Business and Management
Chairman, Faculty of Marketing, 1980-1985
Associate Professor of Marketing, 1978-1985
Assistant Professor of Marketing, 1973-1978

FEDERAL TRADE COMMISSION, Bureau of Economics, 1979-1980
Visiting Marketing Professor, Division of Consumer Protection

PUBLICATIONS

Refereed Journal Publications

1) "Effects of Donor Recruitment Methods on Population Responses," with E.L. Wallace, *Transfusion*, pp. 159-164 (March-April 1975).
2) "A Study of Prices and Market Shares in the Computer Mainframe Industry," with B.T. Ratchford, *Journal of Business*, pp. 194-218 (April 1976).
3) "Some Relationships of States' Characteristics to the Passage of Consumer Legislation," *Journal of Consumer Affairs*, pp. 177-182 (Summer 1977).
4) "Perceptions of Uncertainty Within A Buying Task Group," with R.E. Spekman, *Industrial Marketing Management*, pp. 395-403 (December 1977).

5) "Adoption of Consumerism Policy by the States: Some Empirical Perspectives," *Journal of Marketing Research*," pp. 125-134 (February 1978).
6) "A Study of Prices and Market Shares in the Computer Mainframe Industry: Reply," with B.T. Ratchford, *Journal of Business*, pp. 125-135 (January 1979).
7) "Evaluation of Consumer Education Programs," with P.N. Bloom, *Journal of Consumer Research*, pp. 270-279 (December 1979).
8) "Marketing and Marketing Research for Information Scientists," with P. Wasserman, *Journal of Library Administration*, pp. 27-31 (Fall 1982).
9) "Viewer Miscomprehension of Televised Communications: A Comment," with R. Yalch, *Journal of Marketing*, pp. 27-31 (Fall 1982). Reprinted in *Mass Communication Review Yearbook Vol. 4*, E. Wartella, D.C. Whitney and S. Windall (eds.), Beverly Hills: Sage Publications, pp. 145-150 (1983).
10) "Unit Pricing Ten Years Later: A Replication," with D.A. Aaker, *Journal of Marketing*, pp. 118-122 (Winter 1983).
11) "Recent Developments in FTC Policy on Deception," with J.E. Calfee, *Journal of Marketing*, 82-103 (July 1986).
12) "Inferential Beliefs in Consumer Evaluations: An Assessment of Alternative Processing Strategies," with R.A. Smith, *Journal of Consumer Research*, pp. 363-371 (December 1987).
13) "Consumer Skepticism of Advertising Claims: Testing Hypotheses from Economics of Information," with D.B. Smith and J.L. Swasy, *Journal of Consumer Research*, pp. 433-441 (March 1990).
14) "Normative Values for the Beck Anxiety Inventory, Fear Questionnaire, Penn State Worry Questionnaire and Social Phobia an Anxiety Inventory," with M. Gillis, D. Haaga and A.F. Ford, *Psychological Assessment*, vol. 7, no. 4, pp. 450-455 (1995).
15) "Can Consumers Interpret Nutrition Information in the Presence of a Health Claim? A Laboratory Investigation," with M. Hastak, A. Mitra and D.J. Ringold, *Journal of Public Policy and Marketing*, vol. 15, no. 1, pp. 16-27 (1996).
16) "Informing Buyers of Risks: An Analysis of the Marketing and Regulation of All-Terrain Vehicles," with M.B. Mazis, *Journal of Consumer Affairs*, pp. 90-123 (Summer 1996).
17) "Can the Educationally Disadvantaged Interpret the FDA-Mandated Nutrition facts Panel in the Presence of an Implied Health Claim," with M. Hastak, A. Mitra and D. J. Ringold, *Journal of Public Policy and Marketing*, vol. 18, no. 1, pp. 106-117 (Summer 1999).
18) "Consumer Search for Information in the Digital Age: An Empirical Study of Pre-Purchase Search for Automobiles," with L. Klein, *Journal of Interactive Marketing*, vol. 17, no. 3, pp. 1-22 (Summer 2003).
19) "Application of Research on Consumer Complaint Rates to the Estimation of the Financial Impact of Prospective Product Defects," (with D. Scheffman and D Weiskopf), *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, pp. 130-141 (Fall 2004)
20) "The Impact of the *Daubert* Decision on Survey Research Used in Litigation," *Journal of Public Policy and Marketing*, pp. 234-252, Fall 2005.

Refereed and Special Session, Proceedings Publications

1) "A Multivariate Investigation of Market Structure," refereed, *Combined Proceedings of the American Marketing Association*, pp. 177-182 (1974).
2) "The Status of Consumer Behavior: Some Empirical Perspectives," with P.G. Kuehl and R.F. Dyer, refereed, *Advances in Consumer Research*, vol. 2, pp. 51-61 (1975).

3) "Classifying and Measuring Deceptive Advertising: An Experimental Approach," with P.G. Kuehl and O. Reksten, refereed, *Combined Proceedings of the American Marketing Association*, pp. 493-497 (1975).
4) "Public Policy, The Sherman Act and the IBM Antitrust Case," with B.T. Ratchford, refereed, *Combined Proceedings of the American Marketing Association*, pp. 593-596 (1975).
5) "A Functional Analysis of Macro and Micro Marketing Systems," with W. Nickels, referred, *Proceedings of the Southern Marketing Association*, pp. 76-79 (1975).
6) "Measuring the Impact of Consumer Survival Kit: Some Preliminary Results," with P.N. Bloom and J.W. Harvey, refereed, *Advances in Consumer Research*, vol. 3, pp. 388-391 (1976).
7) "Consumer Research and Public Policy Formation: The Case of Truth in Contributions," with P.G. Kuehl and P.N. Bloom, refereed, *Combined Proceedings of the American Marketing Association*, pp. 445-450 (1976).
8) "An Assessment of the Consumer Protection Act of 1975," refereed, *Combined Proceedings of the American Marketing Association*, pp. 209-212 (1976).
9) "A Multivariate Analysis of State Consumerism Policy," refereed, *Proceedings of the Annual Meeting of the American Institute for Decision Sciences*, pp. 211-213 (1976).
10) "The Promotion of Medical and Legal Services," with P.G. Kuehl, refereed, *Proceedings of the American Marketing Association*, pp. 39-44 (1977).
11) "Consumer Protection Agencies: Their Budgets and Activities," refereed, *Proceedings of the American Marketing Association*, pp. 93-96 (1978).
12) "Box-Jenkins Analysis of a Retail Sales Intervention," with F.B. Alt, refereed abstract, *Northeast Aids Proceedings*, pp. 28-32 (1979).
13) "The Industrial Marketing Implications of Organizational Hierarchy Within Purchasing Departments," with R.E. Spekman, refereed, *Proceedings of the American Marketing Association*, pp. 178-181 (1981).
14) "Consumer Research Issues at the Federal Trade Commission," with J. Calfee and T. Maronick, refereed, *Advances in Consumer Research*, vol. 19, pp. 263-267 (1983).
15) "Consumer Psychology Research Needs at the Federal Trade Commission," with J. Calfee, refereed, *Proceedings of the Division of Consumer Psychology*, American Psychological Association, pp. 118-122 (1984).
16) "Market Forces, Information and Reduced Flammability Cigarettes," with J. Calfee, Special Session, *Advances in Consumer Research*, vol. 14, pp. 274-278 (1987).
17) "An Empirical Test of the Search, Experience and Credence Attributes Framework," with D.B. Smith and J. Swasy, special session, *Advances in Consumer Research*, vol. 15, pp. 239-243 (1988).
18) "Economics, Information and Consumer Behavior," with J. Calfee, special session, *Advances in Consumer Research*, vol. 15, pp. 234-238 (1988).
19) "Cigarettes in the Popular Press, 1930-1960: Preliminary Research," with D.J. Ringold and M. Rogers, special session, *Advances in Consumer Research*, vol. 17, pp. 467-473 (1990).
20) "Regulation of Advertising in the European Economic Community: An Overview," special session, *European Advances in Consumer Research*, vol. 1, pp. 559-564 (1993).
21) "Consumer Search for Information in the Digital Age: an Empirical Study of Pre-Purchase Search for Automobiles" (with Lisa Klein) *Advances in Consumer Research*. (2001).

Articles in Books

1) "Problems in Education and Training in Marketing and Marketing Research in Information Science," with P. Wasserman, *Education and Training: Theory and Provision*, Federation

International De Documentation: The Hague, pp. 105-112 (1979) (a different version of the *Journal of Library Administration* article).

2) "Label Warnings in OTC Drug Advertising: Some Experimental Results," with P.G. Kuehl, *Current Issues and Research in Advertising*, J.H. Leigh and C.R. Martin (eds.), Univ. of Michigan Press: Ann Arbor, pp. 115-130 (1979).
3) "Using Marketing Techniques to Increase Immunization Levels: A Field Experiment," with R.E. Spekman, *Exploring and Developing Government Marketing*, S. Permut and M. Mokwa (eds.), New York, Praeger Press, pp. 304-317 (1981).
4) "The FTC's Product Defects Program and Consumer Perceptions of Product Quality," with J. Calfee, *Perceived Quality*, J. Jacoby and J. Olson (eds.), Lexington, Massachusetts, Lexington Books, pp. 175-191 (1985).
5) "The Economics of Information: Research Issues," with D.B. Smith and J.L. Swasy, *Marketing and Advertising Regulation: The Federal Trade Commission in the 1990s*, P. Murphy and W. Wilkie (eds.), pp. 300-312 (1990).

BOOKS EDITED

1) *Marketing and the Library*, New York, Haworth Press (1984).
2) *AMA Educators Proceedings*, co-edited with R.L. Lusch, G.L. Frazier, R.D. Howell, C.A. Ingene, M. Reilly and R.W. Stampf, Chicago, American Marketing Association, 403 pages (1985).
3) *AMA Educators Proceedings*, co-edited with S.P. Douglas, M.R. Solomon, V. Mahajan, M.I. Alpert, W.M. Pride, G.L. Frazier, J.C. Anderson and P. Doyle, Chicago, American Marketing Association, 287 pages (1987).

RESEARCH REPORTS

1) "A Study of Parks, Recreation and Open Space in Prince George's and Montgomery Counties, Maryland," with R.W. Janes and P.G. Kuehl, for Maryland National Capital Park and Planning Commission, 253 pages (1975).
2) "A Feasibility Study to Identify Methods to Increase the Levels of Immunization of Children Receiving Services from BCHS Funded Clinics," with R.E. Spekman, prepared for Bureau of Community Health Services, D.H.E.W., 97 pages (1979).
3) "The Effects of Reduced Flammability Cigarettes on Smoker Behavior," with J.P. Brown and J.E. Calfee, prepared for the National Bureau of Standards, 65 pages (October 1986).
4) "Final Report on Undercover Investigation of ATV Dealers," prepared for Consumer Product Safety Commission, 53 pages (1989).
5) Preliminary Report "ABA Digital Evidence Project Survey on Electronic Discovery Trends and Proposed Amendments to the Federal Rules of Civil Procedure," prepared for ABA Section of Science & Technology, 64 pages (2005).

PRESENTATIONS

"The Role of Dispute Mediation in Consumer Protection," presented at Meetings of the Practicing Justice Institute, Marymount College, New York City (1978).

"The Use of Consumer Research in the Bureau of Economics, FTC," presented at Association for Consumer Research Conference, San Francisco (1979).

"The FTC's 1983 Deception Policy Statement," presented at Southern Marketing Association (November 1984).

Proposed, organized and chaired special session on "FTC Policy Toward Deception," at Association for Consumer Research Conference, Washington, D.C. (1984).

As faculty member at AMA Doctoral Students Consortium at Notre Dame, presented "Economics of Information, Advertising and Public Policy (1986) (same session was repeated at 1987 Consortium at NYU).

Proposed, organized and chaired special session on "Cigarettes and Regulation: Unintended Consequences?" at Association for Consumer Research Conference, Toronto (1986).

"An Economics of Information Approach to the Regulation of Advertising," with J. Calfee, Winter Educators Conference of the American Marketing Association (1988).

"Signals in Advertising: Preliminary Results," with D.B. Smith and J.L. Swasy, special session, Winter Educators Conference of the American Marketing Association (1991).

"Content Analysis of Advertising for All-Terrain Vehicles, 1980-1987," presented to the marketing faculty at INSEAD, Fontainebleu, France and to marketing faculty at Catholic University at Leuven (1992).

"Can Consumers Interpret Nutrition Information in the Presence of a Health Claim? A Laboratory Investigation," with M. Hastak, A. Mitra and D.J. Ringold, presented at the Annual Meeting of the Association for Consumer Research (1993).

"Interpretation of Health Claims and Nutrition Information by Disadvantaged Consumers," with A. Mitra, M. Hastak and D.J. Ringold, presented at the Annual Meeting of the Association for Consumer Research (1994).

"The Effects of Health Claims on Consumer Interpretation of FDA-mandated Nutrition Disclosures: a Mall Intercept Study," with M. Hastak, A. Mitra and D.J. Ringold, presented at the Annual Marketing Association Public Policy Conference (1997).

"Regulation of Advertising on the Internet," with J. Calfee, presented at the Annual American Marketing Association Public Policy Conference (1997).

"Consumer Search on the Internet," with Lisa Klein, presented at the Annual American Marketing Association Public Policy Conference (1999).

"Consumer Search on the Internet: Predictions from the Economics of Information," with Lisa Klein, presented at the Annual Meeting of the Association for Consumer Research (1999).

"Philosophy of Science and the Supreme Court: The Impact of the *Daubert* Decision on Survey Research Used in Litigation," presented at Kenan-Flager School of Business, UNC at Chapel Hill (Fall 2002) and Marketing Faculty Consortium at Georgetown University (April 2003)

"Philosophy of Science and the Supreme Court: The Impact of the *Daubert* Decision on Survey Research Used in Litigation," presented at the Annual American Marketing Association Public Policy Conference (2003).

CURRENT RESEARCH INTERESTS

Ford, Hastak, and Mitra are completing third study of health claims in food advertising.

Ford is preparing a manuscript regarding the Dormant Commerce Clause.

PROFESSIONAL ACTIVITIES/MEMBERSHIPS

Manuscript reviewer for the American Marketing Association Educators' Conferences, 1976-present; Southern Marketing Association Conferences, 1977-1978; *Journal of Marketing*, 1979-1981, 1999-2001; *Journal of Business Research*, 1980; Association for Consumer Research Conferences, 1980, 1982, 1983, 1985-1990, 1999-2000 and *Journal of Consumer Research*, 1987-1992, 1995, 1997-2001, *Journal of Marketing Research*, 1997-2000, *Journal of Consumer Psychology*, 1999.

Reviewer for AMA Dissertation Competition, 1983, 1987, 1995. Proposal reviewer for the National Science Foundation, the Ford Foundation and the Department of Energy.

Discussant at AMA Consumerism Workshop, 1976; Southern Marketing Association Conference, 1977; American Marketing Association Educators' Conference, 1978-1980; Association for Consumer Research Conference, 1978-1980; AMA Professional Services Marketing Conference, 1981 and Public Policy Conference, 1993, 1994, 1995, 1997.

Member of Program Committee, Association for Consumer Research meeting, 1980, 1984, 2000.

Co-Chairman of AMA Doctoral Students Consortium, 1981.

Faculty participant at AMA Doctoral Students Consortium, 1980, 1986 and 1987.

Elected to Board of Directors, Association for Consumer Research, 1982-1985.

Editorial Review Board, *Journal of Marketing*, 1982-1997.

Editorial Review Board, *Journal of Public Policy and Marketing*, 1983-present.

Special Editor, Marketing and Information Science Issue, *Journal of Library Administration*, 1983-1984.

Public Policy Track Chairman, Educators' Conference of the American Marketing Association, 1985, 1987, 2001.

Book Review Editor Journal of Public Policy and Marketing, 2001 to 2004.

American Marketing Association representative to the U.S. Bureau of the Census, 2006- present.

GRANTS, CONTRACTS AND AWARDS RECEIVED

Received competitively-bid contract from the Maryland National Park and Planning Commission for *A Survey of Parks, Recreation and Open Space in Prince George's and Montgomery Counties, Maryland*, with R.W. Janes and P.G. Kuehl, $33,878 (Spring and Summer 1975).

Received contract from National Institute of Health to develop curriculum for a two-day Cancer Communications Marketing Seminar, $8,000 (Summer 1978).

Received contract for "A Feasibility Study to Identify Methods to Increase the Levels of Immunization of Children and Adolescents Receiving Services from BCHS Funded Clinics," with Robert Spekman, from Bureau of Community Health Services, DHEW, $9,972 (Fall 1998).

Received contract for "The Effects of Reduced Flammability Cigarettes on Smoker Behavior," with John P. Brown, from Consumer Product Safety Commission and National Bureau of Standards, $19,925.

Course Release, Senate Research Committee, American University (Spring 1987 and Spring 1988).

Summer Research Grants, Kogod College of Business Administration, American University (1986 and 1987).

Received award for "The Effects of New Food Labels on Disadvantaged Consumers," with M. Hastak, A. Mitra and D. Ringold, from Marketing Science Institute, $26,000 (1993) (proposal was one of six funded out of 45 entries in MSI "Using Research to Help Society Competition").

Listed as one of "The Best Researchers in Marketing," *Marketing Educator*, p. 5 (Summer 1997).

CONSULTING WORK

American Automobile Association
Organization of American States
Insituto De Investigaciones Electricas, Mexico
Dames and Moore, Inc.

Public Broadcasting System

EXPERT AND EXPERT WITNESS ASSIGNMENTS SINCE 2002

National Association of Optometrists & Opticians v. Lockyer
Morrison & Foerster (2003)
Expert report, deposition

Hispanic Broadcasting Corp. v. Educational Media Foundation
Kenyon & Kenyon (2003)
Expert Report

Cass. v. AmeriDebt et al.; FTC v. Debtworks
Collier Shannon (2004/5)
Expert Report, deposition (Cass 2004); (FTC 2005)

Diarama Trading Company v. J.Walter Thompson and others
Kenyon and Kenyon (2004)
Expert Report, deposition

Gardner and Blangeres v. Stimson Lumber Company
Steven H. Gurnee & Associates (2004)
Deposition, declaration, trial testimony

Pharmavite , LLC v. Weider Nutrition Group, Inc. et al.
Blakely, Sokoloff, Taylor & Zafman (2004)
Expert Report (with Michael Mazis), deposition

Government Employees Insurance Company v. Google, Inc. and Overture Services, Inc.
Arnold & Porter, LLP (2004)
Expert Report, deposition, trial testimony

Minnesota Mining & Manufacturing Company v. Saint Gobain
Arnold & Porter, LLP (2005)
Expert Report, deposition, testimony

Polo Ralph Lauren v. United States Polo Association
Paul Weiss (2005)
Expert Report, Rebuttal Report, deposition, testimony

Schick Manufacturing, Inc. v. The Gillette Company
Ropes & Gray (2005)
Expert Report, Rebuttal Report, deposition

Omni Pacific, Inc. v. OmniBrands, Inc.
DLA Piper (2005)
Expert Report

USA v. QVC, Inc.
Baker & Hostetler (2005)
Expert Report, deposition

9 Squared, Inc. v. Moviso, LLC and InfoSpace, Inc.
Holland & Hart (2006)
Expert Report, deposition

Medi-Flex, Inc. v. Nice-Pak Products, Inc and Professional Disposables, Inc.
Lerner, David, Littenberg, Krumholz & Mentlik (2006)
Declaration, deposition

Align Techology, Inc. v. Orthoclear, Inc. and Orthoclear Holdings, Inc.
Paul, Hastings, Janofsky and Walker LLP (2006)

Expert report, deposition
Shuttlesworth et al. v. Carleton Sheets and American Marketing Systems, Inc.
Sachnoff & Weaver (2006)
Expert report
American Century Proprietary Holdings, Inc. v. American Century Casualty Company and American Century Claims Service, Inc.
Leydig, Voit and Mayer (2006)
Expert report

THESIS COMMITTEES

Chairman of dissertation committees for George Coan, Dennis Pitta, Debra Ringold and Darlene Smith.

Member of dissertation committee for Bill Grazer, Michael McGinnis, Dennis McDonald, Frank Franzak, Ronald Hill and Dennis McDonnell.

COMMITTEE SERVICE AT AMERICAN UNIVERSITY

Kogod Rank and Tenure Committee (1987-1988, 1995-1998, 2004-2005)
Faculty Senate Research Committee (1985-1989)
Marketing Department Faculty Recruiting Committee (1985-present)
Ad hoc Research committee, KCBA (1986-1989)
Committee on Faculty Relations (1988-1989)
Deans Cabinet (1989-1994, 1999-2001, 2004, 2006-present)
Executive Committee of the AU Chairs (1993-1994)
AU Presidential Search Committee (1993-1994)
President's Committee on Strategic Planning (1995-1997)
AU Provost's Committee on Academic Programs (1995-1997)
Director of MBA Field Studies (1995-1997)
Chair of Executive Education Committee (1998)
Women's Varsity Soccer Coach Search Committee (2000)
Faculty Advisor Men's Soccer (2002 to present)
Associate Dean Search Committee, KSB, (2003)

UNDERGRADUATE COURSES TAUGHT

"Principles of Marketing"
"Marketing Research Methods"
"Fundamentals of Marketing and Business for Communications"

GRADUATE COURSES TAUGHT

"Marketing Research Methods"
"Doctoral Seminar in Marketing and Public Policy"
"Research Methodology for Doctoral Students"
"Consumer Behavior"
"Marketing Management"

August 2006

Appendix B

Documents Consulted or Referred To

"Amended Petition to Cancel," Bass Pro Trademarks, L.L.C. v. Sportsman's Warehouse, Inc., Cancellation No. 92045000, March 29, 2006.

Diamond, Shari (2000) "Reference Guide of Survey Research," *Reference Manual on Scientific Evidence, 2nd Ed.*, Federal Judicial Center, pp. 229-276.

Ford, Gary, T. (2005) "The Impact of the *Daubert* Decision on Survey Research Used in Litigation," *Journal of Public Policy and Marketing,* Vol. 24 (Fall), pp. 234-252.

"Registrant's Answer and Affirmative Defenses to Amended Petition to Cancel," Bass Pro Trademarks, L.L.C. v. Sportsman's Warehouse, Inc., Cancellation No. 92045000, June 12, 2006.

Winkelman Consulting (2006) "Sportsman's Warehouse Recognition Assessment," (June)

Appendix C

Responses to Question A5 Categorized
By Whether There Is a Sportsman's Warehouse or Bass Pro
In City Named by Respondents

Response to Question A 5	N	%
Mentioned Sportsman's Warehouse city	27	12.3
Mentioned Bass Pro city	22	10.0
Mentioned City/Metro Area with Sportsman's Warehouse and Bass Pro	24	11.0
Mentioned City/Metro Area without Sportsman's Warehouse or Bass Pro	28	12.8
City mentioned could not be coded	2	0.9
Don't Know/No Response	116	53.0
Total	219	100.0

Appendix C (cont.)

Responses to Question A5 Categorized
By Whether There Is a Sportsman's Warehouse or Bass Pro
In City Named by Respondents and by Respondent ID*

Response to Question A 5	Respondent ID*
Mentioned Sportsman's Warehouse city	8, 16, 21, 26, 27, 30, 37, 38, 45, 53, 56, 67, 78, 90, 97, 101, 102, 110, 152, 155, 162, 170, 179, 350, 393, 449, 453
Mentioned Bass Pro city	116, 131, 140, 157, 164, 171, 176, 180, 225, 233, 256, 257, 276, 285, 290, 322, 327, 330, 344, 394, 458, 471
Mentioned City/Metro Area with Sportsman's Warehouse and Bass Pro	22, 35, 47, 49, 50, 62, 123, 125, 197, 200, 201, 202, 205, 212, 305, 312, 314, 324, 331, 374, 405, 427, 431, 479
Mentioned City/Metro Area without Sportsman's Warehouse or Bass Pro	25, 63, 87, 127, 129, 135, 143, 169, 186, 192, 195, 221, 227, 230, 231, 243, 255, 258, 262, 291, 316, 404, 411, 439, 442, 447, 450, 466
City mentioned could not be coded	39, 320
Don't Know/No Response	15, 28, 34, 36, 46, 55, 68, 74, 76, 79, 83, 84, 86, 93, 94, 95, 103, 105, 106, 107, 113, 114, 117, 120, 126, 128, 130, 133, 136, 142, 146, 147, 159, 160, 166, 172, 177, 182, 184, 189, 190, 193, 194, 207, 209, 213, 216, 222, 223, 235, 237, 244, 245, 247, 249, 252, 253, 261, 264, 266, 272, 273, 275, 279, 281, 282, 293, 294, 295, 297, 301, 302, 306, 309, 310, 311, 321, 326, 329, 336, 339, 341, 346, 349, 351, 354, 365, 366, 369, 375, 376, 378, 381, 383, 384, 385, 387, 397, 407, 408, 409, 425, 428, 435, 445, 454, 455, 465, 468, 469, 470, 473, 476, 480, 481, 482

*Respondent ID numbers were assigned to each row of the SPSS data file from the Winkelman survey.

Appendix D
Responses to Questions A3a-g Categorized
By Whether They Mentioned Sportsman's Warehouse,
Or Bass Pro or Both or Neither

Responses to Question A 3 a-g	N	%
Mentioned Sportsman's Warehouse but not Bass Pro	21	4.4
Mentioned Sportsman's Warehouse, and Bass Pro	22	4.6
Mentioned Bass Pro, but not Sportsman's Warehouse	216	45.6
Did not mention Sportsman's Warehouse or Bass Pro	215	45.4
Total	474	100.0

Appendix D (cont.)
Responses to Questions A3a-g Categorized
By Whether They Mentioned Sportsman's Warehouse,
Or Bass Pro or Both or Neither And by Respondent ID*

Responses to Question A 3	Respondent ID*
Mentioned Sportsman's Warehouse, but not Bass Pro	19, 24, 33, 42, 48, 54, 58, 59, 61, 77, 163 167, 181, 218, 356, 360, 361, 396, 400, 403, 457
Mentioned Sportsman's Warehouse and Bass Pro	1, 20, 29, 32, 40, 41, 44, 60, 64, 65, 89, 111, 121, 153, 156, 300, 347, 380, 390, 395, 415, 421
Mentioned Bass Pro, but not Sportsman's Warehouse	8, 17, 21, 26, 35, 37, 45, 47, 49, 51, 63, 67, 69, 71, 76, 78, 84, 85, 90, 93, 95, 99, 101, 105, 107, 110, 114, 115, 116, 117, 119, 120, 122, 124, 125, 127, 131, 133, 135, 137, 139, 141, 142, 146, 147, 148, 149, 150, 151, 154, 155, 157, 159, 160, 162, 164, 165, 170, 172, 174, 176, 180, 182, 184, 185, 186, 187, 188, 189, 190, 193, 194, 195, 197, 198, 199, 200, 201, 203, 204, 205, 206, 207, 208, 209, 210, 211, 224, 226, 231, 233, 239, 242, 248, 249, 256, 257, 258, 261, 264, 265, 267, 269, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 283, 284, 285, 287, 288, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 302, 304, 305, 306, 309, 310, 313, 314, 320, 323, 326, 328, 330, 331, 336, 338, 339, 340, 344, 346, 348, 349, 350, 351, 355, 363, 366, 367, 368, 369, 370, 372, 373, 376, 379, 383, 384, 385, 386, 387, 388, 392, 397, 398, 401, 402, 404, 405, 408, 409, 411, 413, 414, 417, 418, 419, 420, 423, 424, 425, 427, 428, 431, 432, 433, 434, 435, 436, 440, 444, 445, 449, 452, 453, 455, 458, 460, 461, 465, 467, 468, 471, 473, 474, 477, 478, 479, 480
Did not mention Sportsman's Warehouse or Bass Pro	10, 14, 15, 16, 18, 22, 23, 25, 27, 28, 30, 31, 34, 36, 38, 39, 43, 46, 50, 52, 53, 55, 56, 57, 62, 66, 68, 70, 72, 73, 74, 75, 79, 80, 81, 82, 83, 86, 87, 88, 91, 92, 94, 96, 97, 98, 100, 102, 103, 104, 106, 108, 109, 112, 113, 118, 123, 126, 128, 129, 130, 132, 134, 136, 138, 140, 143, 144, 145, 152, 158, 161, 166, 168, 169, 171, 173, 175, 177, 178, 179, 183, 191, 192, 196, 202, 212, 213, 214, 215, 216, 217, 219, 220, 221, 222, 223, 225, 227, 228, 229, 230, 232, 234, 235, 236, 237, 238, 240, 241, 243, 244, 245, 246, 247, 250, 251, 252, 253, 254, 255, 259, 260, 262, 263, 266, 268, 270, 271, 282, 286, 289, 301, 303, 307, 308, 311, 312, 315, 316, 317, 318, 319, 321, 322, 324, 325, 327, 329, 332, 333, 334, 335, 337, 341, 342, 343, 345, 352, 353, 354, 357, 358, 359, 362, 364, 365, 371, 374, 375, 377, 378, 381, 382, 389, 391, 393, 394, 399, 406, 407, 410, 412, 416, 422, 426, 429, 430, 437, 438, 439, 441, 442, 443, 446, 447, 448, 450, 451, 454, 456, 459, 462, 463, 464, 466, 469, 470, 472, 475, 476, 481, 482, 483, 484

*Respondent ID numbers were assigned to each row of the SPSS data file from the Winkelman survey.

GARY T. FORD

American University
4400 Massachusetts Ave., N.W.
Washington, D.C. 20016
Tel. (202) 885-1978
Fax (202) 885-2691
gtford@american.edu

EDUCATION

Ph.D., Marketing, STATE UNIVERSITY OF NEW YORK AT BUFFALO, Buffalo, NY, 1973.

M.B.A., Marketing, STATE UNIVERSITY OF NEW YORK AT BUFFALO, Buffalo, NY, 1968.

B.B.A., Accounting, CLARKSON COLLEGE, Potsdam, NY, 1966.

PRESENT POSITION

AMERICAN UNIVERSITY, Kogod School Business, Washington, D.C., 1985-present
Professor and Chair of Marketing

ACADEMIC/PROFESSIONAL EXPERIENCE

AMERICAN UNIVERSITY, Kogod School of Business, Washington, D.C.,

Professor of Marketing, 1985-present.

Chairman of Marketing, 1989-94, 1999-2001, 2004, 2006 to present.

CATHOLIC UNIVERSITY OF LEUVEN, Department of Applied Economics, Leuven, Belgium, 1991-1992
Visiting Professor of Marketing

UNIVERSITY OF MARYLAND, College of Business and Management
Chairman, Faculty of Marketing, 1980-1985
Associate Professor of Marketing, 1978-1985
Assistant Professor of Marketing, 1973-1978

FEDERAL TRADE COMMISSION, Bureau of Economics, 1979-1980
Visiting Marketing Professor, Division of Consumer Protection

PUBLICATIONS

Refereed Journal Publications

1) "Effects of Donor Recruitment Methods on Population Responses," with E.L. Wallace, *Transfusion*, pp. 159-164 (March-April 1975).
2) "A Study of Prices and Market Shares in the Computer Mainframe Industry," with B.T. Ratchford, *Journal of Business*, pp. 194-218 (April 1976).
3) "Some Relationships of States' Characteristics to the Passage of Consumer Legislation," *Journal of Consumer Affairs*, pp. 177-182 (Summer 1977).
4) "Perceptions of Uncertainty Within A Buying Task Group," with R.E. Spekman, *Industrial Marketing Management*, pp. 395-403 (December 1977).
5) "Adoption of Consumerism Policy by the States: Some Empirical Perspectives," *Journal of Marketing Research*," pp. 125-134 (February 1978).
6) "A Study of Prices and Market Shares in the Computer Mainframe Industry: Reply," with B.T. Ratchford, *Journal of Business*, pp. 125-135 (January 1979).
7) "Evaluation of Consumer Education Programs," with P.N. Bloom, *Journal of Consumer Research*, pp. 270-279 (December 1979).
8) "Marketing and Marketing Research for Information Scientists," with P. Wasserman, *Journal of Library Administration*, pp. 27-31 (Fall 1982).
9) "Viewer Miscomprehension of Televised Communications: A Comment," with R. Yalch, *Journal of Marketing*, pp. 27-31 (Fall 1982). Reprinted in *Mass Communication Review Yearbook Vol. 4*, E. Wartella, D.C. Whitney and S. Windall (eds.), Beverly Hills: Sage Publications, pp. 145-150 (1983).
10) "Unit Pricing Ten Years Later: A Replication," with D.A. Aaker, *Journal of Marketing*, pp. 118-122 (Winter 1983).
11) "Recent Developments in FTC Policy on Deception," with J.E. Calfee, *Journal of Marketing*, 82-103 (July 1986).
12) "Inferential Beliefs in Consumer Evaluations: An Assessment of Alternative Processing Strategies," with R.A. Smith, *Journal of Consumer Research*, pp. 363-371 (December 1987).
13) "Consumer Skepticism of Advertising Claims: Testing Hypotheses from Economics of Information," with D.B. Smith and J.L. Swasy, *Journal of Consumer Research*, pp. 433-441 (March 1990).
14) "Normative Values for the Beck Anxiety Inventory, Fear Questionnaire, Penn State Worry Questionnaire and Social Phobia an Anxiety Inventory," with M. Gillis, D. Haaga and A.F. Ford, *Psychological Assessment*, vol. 7, no. 4, pp. 450-455 (1995).
15) "Can Consumers Interpret Nutrition Information in the Presence of a Health Claim? A Laboratory Investigation," with M. Hastak, A. Mitra and D.J. Ringold, *Journal of Public Policy and Marketing*, vol. 15, no. 1, pp. 16-27 (1996).

16) "Informing Buyers of Risks: An Analysis of the Marketing and Regulation of All-Terrain Vehicles," with M.B. Mazis, *Journal of Consumer Affairs*, pp. 90-123 (Summer 1996).

17) "Can the Educationally Disadvantaged Interpret the FDA-Mandated Nutrition facts Panel in the Presence of an Implied Health Claim," with M. Hastak, A. Mitra and D. J. Ringold, *Journal of Public Policy and Marketing*, vol. 18, no. 1, pp. 106-117 (Summer 1999).

18) "Consumer Search for Information in the Digital Age: An Empirical Study of Pre-Purchase Search for Automobiles," with L. Klein, *Journal of Interactive Marketing*, vol. 17, no. 3, pp. 1-22 (Summer 2003).

19) "Application of Research on Consumer Complaint Rates to the Estimation of the Financial Impact of Prospective Product Defects," (with D. Scheffman and D Weiskopf), *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, pp. 130-141 (Fall 2004)

20) "The Impact of the *Daubert* Decision on Survey Research Used in Litigation," *Journal of Public Policy and Marketing*, pp. 234-252, Fall 2005.

Refereed and Special Session, Proceedings Publications

1) "A Multivariate Investigation of Market Structure," refereed, *Combined Proceedings of the American Marketing Association*, pp. 177-182 (1974).

2) "The Status of Consumer Behavior: Some Empirical Perspectives," with P.G. Kuehl and R.F. Dyer, refereed, *Advances in Consumer Research*, vol. 2, pp. 51-61 (1975).

3) "Classifying and Measuring Deceptive Advertising: An Experimental Approach," with P.G. Kuehl and O. Reksten, refereed, *Combined Proceedings of the American Marketing Association*, pp. 493-497 (1975).

4) "Public Policy, The Sherman Act and the IBM Antitrust Case," with B.T. Ratchford, refereed, *Combined Proceedings of the American Marketing Association*, pp. 593-596 (1975).

5) "A Functional Analysis of Macro and Micro Marketing Systems," with W. Nickels, referred, *Proceedings of the Southern Marketing Association*, pp. 76-79 (1975).

6) "Measuring the Impact of Consumer Survival Kit: Some Preliminary Results," with P.N. Bloom and J.W. Harvey, refereed, *Advances in Consumer Research*, vol. 3, pp. 388-391 (1976).

7) "Consumer Research and Public Policy Formation: The Case of Truth in Contributions," with P.G. Kuehl and P.N. Bloom, refereed, *Combined Proceedings of the American Marketing Association*, pp. 445-450 (1976).

8) "An Assessment of the Consumer Protection Act of 1975," refereed, *Combined Proceedings of the American Marketing Association*, pp. 209-212 (1976).

9) "A Multivariate Analysis of State Consumerism Policy," refereed, *Proceedings of the Annual Meeting of the American Institute for Decision Sciences*, pp. 211-213 (1976).

10) "The Promotion of Medical and Legal Services," with P.G. Kuehl, refereed, *Proceedings of the American Marketing Association*, pp. 39-44 (1977).

11) "Consumer Protection Agencies: Their Budgets and Activities," refereed, *Proceedings of the American Marketing Association*, pp. 93-96 (1978).

12) "Box-Jenkins Analysis of a Retail Sales Intervention," with F.B. Alt, refereed abstract, *Northeast Aids Proceedings*, pp. 28-32 (1979).

13) "The Industrial Marketing Implications of Organizational Hierarchy Within Purchasing Departments," with R.E. Spekman, refereed, *Proceedings of the American Marketing Association*, pp. 178-181 (1981).

14) "Consumer Research Issues at the Federal Trade Commission," with J. Calfee and T. Maronick, refereed, *Advances in Consumer Research*, vol. 19, pp. 263-267 (1983).

15) "Consumer Psychology Research Needs at the Federal Trade Commission," with J. Calfee, refereed, *Proceedings of the Division of Consumer Psychology*, American Psychological Association, pp. 118-122 (1984).

16) "Market Forces, Information and Reduced Flammability Cigarettes," with J. Calfee, Special Session, *Advances in Consumer Research*, vol. 14, pp. 274-278 (1987).

17) "An Empirical Test of the Search, Experience and Credence Attributes Framework," with D.B. Smith and J. Swasy, special session, *Advances in Consumer Research*, vol. 15, pp. 239-243 (1988).

18) "Economics, Information and Consumer Behavior," with J. Calfee, special session, *Advances in Consumer Research*, vol. 15, pp. 234-238 (1988).

19) "Cigarettes in the Popular Press, 1930-1960: Preliminary Research," with D.J. Ringold and M. Rogers, special session, *Advances in Consumer Research*, vol. 17, pp. 467-473 (1990).

20) "Regulation of Advertising in the European Economic Community: An Overview," special session, *European Advances in Consumer Research*, vol. 1, pp. 559-564 (1993).

21) "Consumer Search for Information in the Digital Age: an Empirical Study of Pre-Purchase Search for Automobiles" (with Lisa Klein) *Advances in Consumer Research*. (2001).

Articles in Books

1) "Problems in Education and Training in Marketing and Marketing Research in Information Science," with P. Wasserman, *Education and Training: Theory and Provision*, Federation International De Documentation: The Hague, pp. 105-112 (1979) (a different version of the *Journal of Library Administration* article).

2) "Label Warnings in OTC Drug Advertising: Some Experimental Results," with P.G. Kuehl, *Current Issues and Research in Advertising*, J.H. Leigh and C.R. Martin (eds.), Univ. of Michigan Press: Ann Arbor, pp. 115-130 (1979).

3) "Using Marketing Techniques to Increase Immunization Levels: A Field Experiment," with R.E. Spekman, *Exploring and Developing Government Marketing*, S. Permut and M. Mokwa (eds.), New York, Praeger Press, pp. 304-317 (1981).

4) "The FTC's Product Defects Program and Consumer Perceptions of Product Quality," with J. Calfee, *Perceived Quality*, J. Jacoby and J. Olson (eds.), Lexington, Massachusetts, Lexington Books, pp. 175-191 (1985).
5) "The Economics of Information: Research Issues," with D.B. Smith and J.L. Swasy, *Marketing and Advertising Regulation: The Federal Trade Commission in the 1990s*, P. Murphy and W. Wilkie (eds.), pp. 300-312 (1990).

Books Edited

1) *Marketing and the Library*, New York, Haworth Press (1984).
2) *AMA Educators Proceedings*, co-edited with R.L. Lusch, G.L. Frazier, R.D. Howell, C.A. Ingene, M. Reilly and R.W. Stampf, Chicago, American Marketing Association, 403 pages (1985).
3) *AMA Educators Proceedings*, co-edited with S.P. Douglas, M.R. Solomon, V. Mahajan, M.I. Alpert, W.M. Pride, G.L. Frazier, J.C. Anderson and P. Doyle, Chicago, American Marketing Association, 287 pages (1987).

Research Reports

1) "A Study of Parks, Recreation and Open Space in Prince George's and Montgomery Counties, Maryland," with R.W. Janes and P.G. Kuehl, for Maryland National Capital Park and Planning Commission, 253 pages (1975).
2) "A Feasibility Study to Identify Methods to Increase the Levels of Immunization of Children Receiving Services from BCHS Funded Clinics," with R.E. Spekman, prepared for Bureau of Community Health Services, D.H.E.W., 97 pages (1979).
3) "The Effects of Reduced Flammability Cigarettes on Smoker Behavior," with J.P. Brown and J.E. Calfee, prepared for the National Bureau of Standards, 65 pages (October 1986).
4) "Final Report on Undercover Investigation of ATV Dealers," prepared for Consumer Product Safety Commission, 53 pages (1989).
5) Preliminary Report "ABA Digital Evidence Project Survey on Electronic Discovery Trends and Proposed Amendments to the Federal Rules of Civil Procedure," prepared for ABA Section of Science & Technology, 64 pages (2005).

PRESENTATIONS

"The Role of Dispute Mediation in Consumer Protection," presented at Meetings of the Practicing Justice Institute, Marymount College, New York City (1978).

"The Use of Consumer Research in the Bureau of Economics, FTC," presented at Association for Consumer Research Conference, San Francisco (1979).

"The FTC's 1983 Deception Policy Statement," presented at Southern Marketing Association (November 1984).

Proposed, organized and chaired special session on "FTC Policy Toward Deception," at Association for Consumer Research Conference, Washington, D.C. (1984).

As faculty member at AMA Doctoral Students Consortium at Notre Dame, presented "Economics of Information, Advertising and Public Policy (1986) (same session was repeated at 1987 Consortium at NYU).

Proposed, organized and chaired special session on "Cigarettes and Regulation: Unintended Consequences?" at Association for Consumer Research Conference, Toronto (1986).

"An Economics of Information Approach to the Regulation of Advertising," with J. Calfee, Winter Educators Conference of the American Marketing Association (1988).

"Signals in Advertising: Preliminary Results," with D.B. Smith and J.L. Swasy, special session, Winter Educators Conference of the American Marketing Association (1991).

"Content Analysis of Advertising for All-Terrain Vehicles, 1980-1987," presented to the marketing faculty at INSEAD, Fontainebleu, France and to marketing faculty at Catholic University at Leuven (1992).

"Can Consumers Interpret Nutrition Information in the Presence of a Health Claim? A Laboratory Investigation," with M. Hastak, A. Mitra and D.J. Ringold, presented at the Annual Meeting of the Association for Consumer Research (1993).

"Interpretation of Health Claims and Nutrition Information by Disadvantaged Consumers," with A. Mitra, M. Hastak and D.J. Ringold, presented at the Annual Meeting of the Association for Consumer Research (1994).

"The Effects of Health Claims on Consumer Interpretation of FDA-mandated Nutrition Disclosures: a Mall Intercept Study," with M. Hastak, A. Mitra and D.J. Ringold, presented at the Annual Marketing Association Public Policy Conference (1997).

"Regulation of Advertising on the Internet," with J. Calfee, presented at the Annual American Marketing Association Public Policy Conference (1997).

"Consumer Search on the Internet," with Lisa Klein, presented at the Annual American Marketing Association Public Policy Conference (1999).

"Consumer Search on the Internet: Predictions from the Economics of Information," with Lisa Klein, presented at the Annual Meeting of the Association for Consumer Research (1999).

"Philosophy of Science and the Supreme Court: The Impact of the *Daubert* Decision on Survey Research Used in Litigation," presented at Kenan-Flager School of Business, UNC at Chapel Hill (Fall 2002) and Marketing Faculty Consortium at Georgetown University (April 2003)

"Philosophy of Science and the Supreme Court: The Impact of the *Daubert* Decision on Survey Research Used in Litigation," presented at the Annual American Marketing Association Public Policy Conference (2003).

CURRENT RESEARCH INTERESTS

Ford, Hastak, and Mitra are completing third study of health claims in food advertising.

Ford is preparing a manuscript regarding the Dormant Commerce Clause.

PROFESSIONAL ACTIVITIES/MEMBERSHIPS

Manuscript reviewer for the American Marketing Association Educators' Conferences, 1976-present; Southern Marketing Association Conferences, 1977-1978; *Journal of Marketing*, 1979-1981, 1999-2001; *Journal of Business Research*, 1980; Association for Consumer Research Conferences, 1980, 1982, 1983, 1985-1990, 1999-2000 and *Journal of Consumer Research*, 1987-1992, 1995, 1997-2001, *Journal of Marketing Research*, 1997-2000, *Journal of Consumer Psychology*, 1999.

Reviewer for AMA Dissertation Competition, 1983, 1987, 1995. Proposal reviewer for the National Science Foundation, the Ford Foundation and the Department of Energy.

Discussant at AMA Consumerism Workshop, 1976; Southern Marketing Association Conference, 1977; American Marketing Association Educators' Conference, 1978-1980; Association for Consumer Research Conference, 1978-1980; AMA Professional Services Marketing Conference, 1981 and Public Policy Conference, 1993, 1994, 1995, 1997.

Member of Program Committee, Association for Consumer Research meeting, 1980, 1984, 2000.

Co-Chairman of AMA Doctoral Students Consortium, 1981.

Faculty participant at AMA Doctoral Students Consortium, 1980, 1986 and 1987.

Elected to Board of Directors, Association for Consumer Research, 1982-1985.

Editorial Review Board, *Journal of Marketing*, 1982-1997.

Editorial Review Board, *Journal of Public Policy and Marketing*, 1983-present.

Special Editor, Marketing and Information Science Issue, *Journal of Library Administration*, 1983-1984.

Public Policy Track Chairman, Educators' Conference of the American Marketing Association, 1985, 1987, 2001.

Book Review Editor Journal of Public Policy and Marketing, 2001 to 2004.

American Marketing Association representative to the U.S. Bureau of the Census, 2006-present.

GRANTS, CONTRACTS AND AWARDS RECEIVED

Received competitively-bid contract from the Maryland National Park and Planning Commission for *A Survey of Parks, Recreation and Open Space in Prince George's and*

Montgomery Counties, Maryland, with R.W. Janes and P.G. Kuehl, $33,878 (Spring and Summer 1975).

Received contract from National Institute of Health to develop curriculum for a two-day Cancer Communications Marketing Seminar, $8,000 (Summer 1978).

Received contract for "A Feasibility Study to Identify Methods to Increase the Levels of Immunization of Children and Adolescents Receiving Services from BCHS Funded Clinics," with Robert Spekman, from Bureau of Community Health Services, DHEW, $9,972 (Fall 1998).

Received contract for "The Effects of Reduced Flammability Cigarettes on Smoker Behavior," with John P. Brown, from Consumer Product Safety Commission and National Bureau of Standards, $19,925.

Course Release, Senate Research Committee, American University (Spring 1987 and Spring 1988).

Summer Research Grants, Kogod College of Business Administration, American University (1986 and 1987).

Received award for "The Effects of New Food Labels on Disadvantaged Consumers," with M. Hastak, A. Mitra and D. Ringold, from Marketing Science Institute, $26,000 (1993) (proposal was one of six funded out of 45 entries in MSI "Using Research to Help Society Competition").

Listed as one of "The Best Researchers in Marketing," *Marketing Educator*, p. 5 (Summer 1997).

CONSULTING WORK

American Automobile Association
Organization of American States
Insituto De Investigaciones Electricas, Mexico
Dames and Moore, Inc.
Public Broadcasting System

EXPERT AND EXPERT WITNESS ASSIGNMENTS SINCE 2002

National Association of Optometrists & Opticians v. Lockyer
Morrison & Foerster (2003)
Expert report, deposition

Hispanic Broadcasting Corp. v. Educational Media Foundation
Kenyon & Kenyon (2003)
Expert Report

Cass. v. AmeriDebt et al.; FTC v. Debtworks
Collier Shannon (2004/5)
Expert Report, deposition (Cass 2004); (FTC 2005)

Diarama Trading Company v. J.Walter Thompson and others
Kenyon and Kenyon (2004)
Expert Report, deposition

Gardner and Blangeres v. Stimson Lumber Company
Steven H. Gurnee & Associates (2004)
Deposition, declaration, trial testimony
Pharmavite , LLC v. Weider Nutrition Group, Inc. et al.
Blakely, Sokoloff, Taylor & Zafman (2004)
Expert Report (with Michael Mazis), deposition
Government Employees Insurance Company v. Google, Inc. and Overture Services, Inc.
Arnold & Porter, LLP (2004)
Expert Report, deposition, trial testimony
Minnesota Mining & Manufacturing Company v. Saint Gobain
Arnold & Porter, LLP (2005)
Expert Report, deposition, testimony
Polo Ralph Lauren v. United States Polo Association
Paul Weiss (2005)
Expert Report, Rebuttal Report, deposition, testimony
Schick Manufacturing, Inc. v. The Gillette Company
Ropes & Gray (2005)
Expert Report, Rebuttal Report, deposition
Omni Pacific, Inc. v. OmniBrands, Inc.
DLA Piper (2005)
Expert Report
USA v. QVC, Inc.
Baker & Hostetler (2005)
Expert Report, deposition
9 Squared, Inc. v. Moviso, LLC and InfoSpace, Inc.
Holland & Hart (2006)
Expert Report, deposition
Medi-Flex, Inc. v. Nice-Pak Products, Inc and Professional Disposables, Inc.
Lerner, David, Littenberg, Krumholz & Mentlik (2006)
Declaration, deposition
Align Techology, Inc. v. Orthoclear, Inc. and Orthoclear Holdings, Inc.
Paul, Hastings, Janofsky and Walker LLP (2006)
Expert report, deposition
Shuttlesworth et al. v. Carleton Sheets and American Marketing Systems, Inc.
Sachnoff & Weaver (2006)
Expert report
American Century Proprietary Holdings, Inc. v. American Century Casualty Company and American Century Claims Service, Inc.
Leydig, Voit and Mayer (2006)
Expert report

THESIS COMMITTEES

Chairman of dissertation committees for George Coan, Dennis Pitta, Debra Ringold and Darlene Smith.

Member of dissertation committee for Bill Grazer, Michael McGinnis, Dennis McDonald, Frank Franzak, Ronald Hill and Dennis McDonnell.

COMMITTEE SERVICE AT AMERICAN UNIVERSITY

Kogod Rank and Tenure Committee (1987-1988, 1995-1998, 2004-2005)
Faculty Senate Research Committee (1985-1989)
Marketing Department Faculty Recruiting Committee (1985-present)
Ad hoc Research committee, KCBA (1986-1989)
Committee on Faculty Relations (1988-1989)
Deans Cabinet (1989-1994, 1999-2001, 2004, 2006-present)
Executive Committee of the AU Chairs (1993-1994)
AU Presidential Search Committee (1993-1994)
President's Committee on Strategic Planning (1995-1997)
AU Provost's Committee on Academic Programs (1995-1997)
Director of MBA Field Studies (1995-1997)
Chair of Executive Education Committee (1998)
Women's Varsity Soccer Coach Search Committee (2000)
Faculty Advisor Men's Soccer (2002 to present)
Associate Dean Search Committee, KSB, (2003)

UNDERGRADUATE COURSES TAUGHT

"Principles of Marketing"
"Marketing Research Methods"
"Fundamentals of Marketing and Business for Communications"

GRADUATE COURSES TAUGHT

"Marketing Research Methods"
"Doctoral Seminar in Marketing and Public Policy"
"Research Methodology for Doctoral Students"
"Consumer Behavior"
"Marketing Management"

August 2006

DEPOSITION EXHIBIT
FORD 2
PB 9-15-06

Reference Guide on Survey Research

SHARI SEIDMAN DIAMOND

Shari Seidman Diamond, J.D., Ph.D., is Professor of Law and Psychology, Northwestern University, Evanston, Illinois, and Senior Research Fellow, American Bar Foundation, Chicago, Illinois.

CONTENTS

I. Introduction, 231
- A. Use of Surveys in Court, 233
- B. A Comparison of Survey Evidence and Individual Testimony, 235

II. Purpose and Design of the Survey, 236
- A. Was the Survey Designed to Address Relevant Questions? 236
- B. Was Participation in the Design, Administration, and Interpretation of the Survey Appropriately Controlled to Ensure the Objectivity of the Survey? 237
- C. Are the Experts Who Designed, Conducted, or Analyzed the Survey Appropriately Skilled and Experienced? 238
- D. Are the Experts Who Will Testify About Surveys Conducted by Others Appropriately Skilled and Experienced? 239

III. Population Definition and Sampling, 239
- A. Was an Appropriate Universe or Population Identified? 239
- B. Did the Sampling Frame Approximate the Population? 240
- C. How Was the Sample Selected to Approximate the Relevant Characteristics of the Population? 242
- D. Was the Level of Nonresponse Sufficient to Raise Questions About the Representativeness of the Sample? If So, What Is the Evidence That Nonresponse Did Not Bias the Results of the Survey? 245
- E. What Procedures Were Used to Reduce the Likelihood of a Biased Sample? 246
- F. What Precautions Were Taken to Ensure That Only Qualified Respondents Were Included in the Survey? 247

IV. Survey Questions and Structure, 248
- A. Were Questions on the Survey Framed to Be Clear, Precise, and Unbiased? 248
- B. Were Filter Questions Provided to Reduce Guessing? 249
- C. Did the Survey Use Open-Ended or Closed-Ended Questions? How Was the Choice in Each Instance Justified? 251
- D. If Probes Were Used to Clarify Ambiguous or Incomplete Answers, What Steps Were Taken to Ensure That the Probes Were Not Leading and Were Administered in a Consistent Fashion? 253

E. What Approach Was Used to Avoid or Measure Potential Order or Context Effects? 254
F. If the Survey Was Designed to Test a Causal Proposition, Did the Survey Include an Appropriate Control Group or Question? 256
G. What Limitations Are Associated with the Mode of Data Collection Used in the Survey? 260
1. In-person interviews, 260
2. Telephone surveys, 261
3. Mail surveys, 263
4. Internet surveys, 264
V. Surveys Involving Interviewers, 264
A. Were the Interviewers Appropriately Selected and Trained? 264
B. What Did the Interviewers Know About the Survey and Its Sponsorship? 266
C. What Procedures Were Used to Ensure and Determine That the Survey Was Administered to Minimize Error and Bias? 267
VI. Data Entry and Grouping of Responses, 268
A. What Was Done to Ensure That the Data Were Recorded Accurately? 268
B. What Was Done to Ensure That the Grouped Data Were Classified Consistently and Accurately? 268
VII. Disclosure and Reporting, 269
A. When Was Information About the Survey Methodology and Results Disclosed? 269
B. Does the Survey Report Include Complete and Detailed Information on All Relevant Characteristics? 270
C. In Surveys of Individuals, What Measures Were Taken to Protect the Identities of Individual Respondents? 271
Glossary of Terms, 273
References on Survey Research, 276

I. Introduction

Surveys are used to describe or enumerate objects or the beliefs, attitudes, or behavior of persons or other social units.[1] Surveys typically are offered in legal proceedings to establish or refute claims about the characteristics of those objects, individuals, or social units. Although surveys may count or measure every member of the relevant population (e.g., all plaintiffs eligible to join in a suit, all employees currently working for a corporation, all trees in a forest), sample surveys count or measure only a portion of the objects, individuals, or social units that the survey is intended to describe.[2]

Some statistical and sampling experts apply the phrase "sample survey" only to a survey in which probability sampling techniques are used to select the sample.[3] Although probability sampling offers important advantages over nonprobability sampling,[4] experts in some fields (e.g., marketing) regularly rely on various forms of nonprobability sampling when conducting surveys. Consistent with Federal Rule of Evidence 703, courts generally have accepted such evidence.[5] Thus, in this reference guide, both the probability sample and the nonprobability sample are discussed. The strengths of probability sampling and the weaknesses of various types of nonprobability sampling are described so that the trier of fact can consider these features in deciding what weight to give to a particular sample survey.

As a method of data collection, surveys have several crucial potential advantages over less systematic approaches.[6] When properly designed, executed, and

1. Social scientists describe surveys as "conducted for the purpose of collecting data from individuals about themselves, about their households, or about other larger social units." Peter H. Rossi et al., *Sample Surveys: History, Current Practice, and Future Prospects*, *in* Handbook of Survey Research 1, 2 (Peter H. Rossi et al. eds., 1983). Used in its broader sense, however, the term *survey* applies to any description or enumeration, whether or not a person is the source of this information. Thus, a report on the number of trees destroyed in a forest fire might require a survey of the trees and stumps in the damaged area.

2. In *J.H. Miles & Co. v. Brown*, 910 F. Supp. 1138 (E.D. Va. 1995), clam processors and fishing vessel owners sued the Secretary of Commerce for failing to use the unexpectedly high results from 1994 survey data on the size of the clam population to determine clam fishing quotas for 1995. The estimate of clam abundance is obtained from surveys of the amount of fishing time the research survey vessels require to collect a specified yield of clams in major fishing areas over a period of several weeks. *Id.* at 1144–45.

3. *E.g.*, Leslie Kish, Survey Sampling 26 (1965).

4. *See infra* § III.C.

5. Fed. R. Evid. 703 recognizes facts or data "of a type reasonably relied upon by experts in the particular field . . . ."

6. This does not mean that surveys can be relied on to address all types of questions. For example, some respondents may not be able to predict accurately whether they would volunteer for military service if Washington, D.C., were to be bombed. Their inaccuracy may arise not because they are unwilling to answer the question or to say they don't know, but because they believe they can predict accurately, and they are simply wrong. Thus, the availability of a "don't know" option cannot cure the inaccuracy. Although such a survey is suitable for assessing their predictions, it may not provide a very accurate estimate of what their actual responses would be.

described, surveys (1) economically present the characteristics of a large group of objects or respondents and (2) permit an assessment of the extent to which the measured objects or respondents are likely to adequately represent a relevant group of objects, individuals, or social units.[7] All questions asked of respondents and all other measuring devices used can be examined by the court and the opposing party for objectivity, clarity, and relevance, and all answers or other measures obtained can be analyzed for completeness and consistency. To make it possible for the court and the opposing party to closely scrutinize the survey so that its relevance, objectivity, and representativeness can be evaluated, the party proposing to offer the survey as evidence should describe in detail the design and execution of the survey.

The questions listed in this reference guide are intended to assist judges in identifying, narrowing, and addressing issues bearing on the adequacy of surveys either offered as evidence or proposed as a method for developing information.[8] These questions can be (1) raised from the bench during a pretrial proceeding to determine the admissibility of the survey evidence; (2) presented to the contending experts before trial for their joint identification of disputed and undisputed issues; (3) presented to counsel with the expectation that the issues will be addressed during the examination of the experts at trial; or (4) raised in bench trials when a motion for a preliminary injunction is made to help the judge evaluate what weight, if any, the survey should be given.[9] These questions are intended to improve the utility of cross-examination by counsel, where appropriate, not to replace it.

All sample surveys, whether they measure objects, individuals, or other social units, should address the issues concerning purpose and design (section II), population definition and sampling (section III), accuracy of data entry (section VI), and disclosure and reporting (section VII). Questionnaire and interview surveys raise methodological issues involving survey questions and structure (section IV) and confidentiality (section VII.C), and interview surveys introduce additional issues (e.g., interviewer training and qualifications) (section V). The sections of this reference guide are labeled to direct the reader to those topics that are relevant to the type of survey being considered. The scope of this reference guide is necessarily limited, and additional issues might arise in particular cases.

7. The ability to quantitatively assess the limits of the likely margin of error is unique to probability sample surveys.

8. *See infra* text accompanying note 27.

9. Lanham Act cases involving trademark infringement or deceptive advertising frequently require expedited hearings that request injunctive relief, so judges may need to be more familiar with survey methodology when considering the weight to accord a survey in these cases than when presiding over cases being submitted to a jury. Even in a case being decided by a jury, however, the court must be prepared to evaluate the methodology of the survey evidence in order to rule on admissibility. *See* Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 589 (1993).

A. Use of Surveys in Court

Forty years ago the question whether surveys constituted acceptable evidence still was unsettled.[10] Early doubts about the admissibility of surveys centered on their use of sampling techniques[11] and their status as hearsay evidence.[12] Federal Rule of Evidence 703 settled both matters for surveys by redirecting attention to the "validity of the techniques employed."[13] The inquiry under Rule 703 focuses on whether facts or data are "of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject."[14] For a survey, the question becomes, "Was the poll or survey conducted in accordance with generally accepted survey principles, and were the results used in a

10. Hans Zeisel, *The Uniqueness of Survey Evidence*, 45 Cornell L.Q. 322, 345 (1960).

11. In an early use of sampling, Sears, Roebuck & Co. claimed a tax refund based on sales made to individuals living outside city limits. Sears randomly sampled 33 of the 826 working days in the relevant working period, computed the proportion of sales to out-of-city individuals during those days, and projected the sample result to the entire period. The court refused to accept the estimate based on the sample. When a complete audit was made, the result was almost identical to that obtained from the sample. *Sears, Roebuck & Co. v. City of Inglewood*, tried in Los Angeles Superior Court in 1955, is described in R. Clay Sprowls, *The Admissibility of Sample Data into a Court of Law: A Case History*, 4 UCLA L. Rev. 222, 226–29 (1956–1957).

12. Judge Wilfred Feinberg's thoughtful analysis in *Zippo Manufacturing Co. v. Rogers Imports, Inc.*, 216 F. Supp. 670, 682–83 (S.D.N.Y. 1963), provides two alternative grounds for admitting opinion surveys: (1) surveys are not hearsay because they are not offered in evidence to prove the truth of the matter asserted; and (2) even if they are hearsay, they fall under one of the exceptions as a "present sense impression." In *Schering Corp. v. Pfizer Inc.*, 189 F.3d 218 (2d Cir. 1999), the Second Circuit distinguished between perception surveys designed to reflect the present sense impressions of respondents and "memory" surveys designed to collect information about a past occurrence based on the recollections of the survey respondents. The court in *Schering* suggested that if a survey is offered to prove the existence of a specific idea in the public mind, then the survey does constitute hearsay evidence. As the court observed, Federal Rule of Evidence 803(3), creating "an exception to the hearsay rule for such statements [i.e., state of mind expressions] rather than excluding the statements from the definition of hearsay, makes sense only in this light." *Id.* at 230 n.3.

Two additional exceptions to the hearsay exclusion can be applied to surveys. First, surveys may constitute a hearsay exception if the survey data were collected in the normal course of a regularly conducted business activity, unless "the source of information or the method or circumstances of preparation indicate lack of trustworthiness." Fed. R. Evid. 803(6); *see also* Ortho Pharm. Corp. v. Cosprophar, Inc., 828 F. Supp. 1114, 1119–20 (S.D.N.Y. 1993) (marketing surveys prepared in the course of business were properly excluded due to lack of foundation from a person who saw the original data or knew what steps were taken in preparing the report), *aff'd*, 32 F.3d 690 (2d Cir. 1994). In addition, if a survey shows guarantees of trustworthiness equivalent to those in other hearsay exceptions, it can be admitted if the court determines that the statement is offered as evidence of a material fact, it is more probative on the point for which it is offered than any other evidence which the proponent can procure through reasonable efforts, and admissibility serves the interests of justice. Fed. R. Evid. 807; *e.g.*, Keith v. Volpe, 618 F. Supp. 1132 (C.D. Cal. 1985); *Schering*, 189 F.3d at 232. Admissibility as an exception to the hearsay exclusion thus depends on the trustworthiness of the survey.

13. Fed. R. Evid. 703 advisory committee's note.

14. Fed. R. Evid. 703.

statistically correct way?"[15] This focus on the adequacy of the methodology used in conducting and analyzing results from a survey is also consistent with the Supreme Court's discussion of admissible scientific evidence in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[16]

Because the survey method provides an economical and systematic way to gather information about a large number of individuals or social units, surveys are used widely in business, government, and, increasingly, administrative settings and judicial proceedings. Both federal and state courts have accepted survey evidence on a variety of issues. In a case involving allegations of discrimination in jury panel composition, the defense team surveyed prospective jurors to obtain age, race, education, ethnicity, and income distribution.[17] Surveys of employees or prospective employees are used to support or refute claims of employment discrimination.[18] In ruling on the admissibility of scientific claims, courts have examined surveys of scientific experts to assess the extent to which the theory or technique has received widespread acceptance.[19] Some courts have admitted surveys in obscenity cases to provide evidence about community standards.[20] Requests for a change of venue on grounds of jury pool bias often are backed by evidence from a survey of jury-eligible respondents in the area of the original venue.[21] The plaintiff in an antitrust suit conducted a survey to assess what characteristics, including price, affected consumers' preferences. The sur-

15. Manual for Complex Litigation § 2.712 (1982). Survey research also is addressed in the Manual for Complex Litigation, Second § 21.484 (1985) [hereinafter MCL 2d] and the Manual for Complex Litigation, Third § 21.493 (1995) [hereinafter MCL 3d]. Note, however, that experts who collect survey data, along with the professions that rely on those surveys, may differ in some of their methodological standards and principles. An assessment of the precision of sample estimates and an evaluation of the sources and magnitude of likely bias are required to distinguish methods that are acceptable from methods that are not.

16. 509 U.S. 579 (1993). *See also* General Elec. Co. v. Joiner, 522 U.S. 136, 147 (1997).

17. People v. Harris, 679 P.2d 433 (Cal.), *cert. denied*, 469 U.S. 965 (1984).

18. EEOC v. Sears, Roebuck & Co., 628 F. Supp. 1264, 1308 (N.D. Ill. 1986), *aff'd*, 839 F.2d 302 (7th Cir. 1988); Stender v. Lucky Stores, Inc., 803 F. Supp. 259, 326 (N.D. Cal. 1992); Richardson v. Quik Trip Corp., 591 F. Supp. 1151, 1153 (S.D. Iowa 1984).

19. United States v. Scheffer, 523 U.S. 303, 309 (1998); Meyers v. Arcudi, 947 F. Supp. 581, 588 (D. Conn. 1996); United States v. Varoudakis, No. 97-10158, 1998 WL 151238 (D. Mass. Mar. 27, 1998); United States v. Bishop, 64 F. Supp. 2d 1149 (D. Utah 1999); United States v. Orians, 9 F. Supp. 2d 1168, 1174 (D. Ariz. 1998) (all cases in which courts determined, based on the inconsistent reactions revealed in several surveys, that the polygraph test has failed to achieve general acceptance in the scientific community).

20. *E.g.*, People v. Page Books, Inc., 601 N.E.2d 273, 279–80 (Ill. App. Ct. 1992); People v. Nelson, 410 N.E.2d 476, 477–79 (Ill. App. Ct. 1980); State v. Williams, 598 N.E.2d 1250, 1256–58 (Ohio Ct. App. 1991).

21. *E.g.*, United States v. Eagle, 586 F.2d 1193, 1195 (8th Cir. 1978); United States v. Tokars, 839 F. Supp. 1578, 1583 (D. Ga. 1993), *aff'd*, 95 F.3d 1520 (11th Cir. 1996); Powell v. Superior Court, 283 Cal. Rptr. 777, 783 (Ct. App. 1991).

vey was offered as one way to estimate damages.[22] A routine use of surveys in federal courts occurs in Lanham Act[23] cases, where the plaintiff alleges trademark infringement[24] or claims that false advertising[25] has confused or deceived consumers. The pivotal legal question in such cases virtually demands survey research because it centers on consumer perception and memory (i.e., is the consumer likely to be confused about the source of a product, or does the advertisement imply an inaccurate message?).[26] In addition, survey methodology has been used creatively to assist federal courts in managing mass torts litigation. Faced with the prospect of conducting discovery concerning 10,000 plaintiffs, the plaintiffs and defendants in *Wilhoite v. Olin Corp.*[27] jointly drafted a discovery survey that was administered in person by neutral third parties, thus replacing interrogatories and depositions. It resulted in substantial savings in both time and cost.

B. A Comparison of Survey Evidence and Individual Testimony

To illustrate the value of a survey, it is useful to compare the information that can be obtained from a competently done survey with the information obtained

22. Dolphin Tours, Inc. v. Pacifico Creative Servs., Inc., 773 F.2d 1506, 1508 (9th Cir. 1985). *See also* SMS Sys. Maintenance Servs., Inc. v. Digital Equip. Corp., 188 F.3d 11 (1st Cir. 1999); Benjamin F. King, *Statistics in Antitrust Litigation*, *in* Statistics and the Law 49 (Morris H. DeGroot et al. eds., 1986). Surveys also are used in litigation to help define relevant markets. In *United States v. E.I. DuPont de Nemours & Co.*, 118 F. Supp. 41, 60 (D. Del. 1953), *aff'd*, 351 U.S. 377 (1956), a survey was used to develop the "market setting" for the sale of cellophane. In *Mukand, Ltd. v. United States*, 937 F. Supp. 910 (Ct. Int'l Trade 1996), a survey of purchasers of stainless steel wire rods was conducted to support a determination of competition and fungibility between domestic and Indian wire rod.

23. Lanham Act § 43(a), 15 U.S.C. § 1125(a) (1946) (amended 1992).

24. *E.g.*, Union Carbide Corp. v. Ever-Ready, Inc., 531 F.2d 366 (7th Cir.), *cert. denied*, 429 U.S. 830 (1976); Qualitex Co. v. Jacobson Prods. Co., No. CIV-90-1183HLH, 1991 U.S. Dist. LEXIS 21172 (C.D. Cal. Sept. 3, 1991), *aff'd in part & rev'd on other grounds*, 13 F.3d 1297 (9th. Cir. 1994), *rev'd on other grounds*, 514 U.S. 159 (1995). According to Neal Miller, *Facts, Expert Facts, and Statistics: Descriptive and Experimental Research Methods in Litigation*, 40 Rutgers L. Rev. 101, 137 (1987), trademark law has relied on the institutionalized use of statistical evidence more than any other area of the law.

25. *E.g.*, Southland Sod Farms v. Stover Seed Co., 108 F.3d 1134, 1142–43 (9th Cir. 1997); American Home Prods. Corp. v. Johnson & Johnson, 577 F.2d 160 (2d Cir. 1978).

26. Courts have observed that "the court's reaction is at best not determinative and at worst irrelevant. The question in such cases is, what does the person to whom the advertisement is addressed find to be the message?" American Brands, Inc. v. R.J. Reynolds Tobacco Co., 413 F. Supp. 1352, 1357 (S.D.N.Y. 1976). The wide use of surveys in recent years was foreshadowed in *Triangle Publications, Inc. v. Rohrlich*, 167 F.2d 969, 974 (2d Cir. 1948) (Frank, J., dissenting). Called on to determine whether a manufacturer of girdles labeled "Miss Seventeen" infringed the trademark of the magazine *Seventeen*, Judge Frank suggested that, in the absence of a test of the reactions of "numerous girls and women," the trial court judge's finding as to what was likely to confuse was "nothing but a surmise, a conjecture, a guess," noting that "neither the trial judge nor any member of this court is (or resembles) a teen-age girl or the mother or sister of such a girl." *Id.* at 976–77.

27. No. CV-83-C-5021-NE (N.D. Ala. filed Jan. 11, 1983). The case ultimately settled before trial. *See* Francis E. McGovern & E. Allan Lind, *The Discovery Survey*, Law & Contemp. Probs., Autumn 1988, at 41.

by other means. A survey is presented by a survey expert who testifies about the responses of a substantial number of individuals who have been selected according to an explicit sampling plan and asked the same set of questions by interviewers who were not told who sponsored the survey or what answers were predicted or preferred. Although parties presumably are not obliged to present a survey conducted in anticipation of litigation by a nontestifying expert if it produced unfavorable results,[28] the court can and should scrutinize the method of respondent selection for any survey that is presented.

A party using a nonsurvey method generally identifies several witnesses who testify about their own characteristics, experiences, or impressions. While the party has no obligation to select these witnesses in any particular way or to report on how they were chosen, the party is not likely to select witnesses whose attributes conflict with the party's interests. The witnesses who testify are aware of the parties involved in the case and have discussed the case before testifying.

Although surveys are not the only means of demonstrating particular facts, presenting the results of a well-done survey through the testimony of an expert is an efficient way to inform the trier of fact about a large and representative group of potential witnesses. In some cases, courts have described surveys as the most direct form of evidence that can be offered.[29] Indeed, several courts have drawn negative inferences from the absence of a survey, taking the position that failure to undertake a survey may strongly suggest that a properly done survey would not support the plaintiff's position.[30]

II. Purpose and Design of the Survey

A. Was the Survey Designed to Address Relevant Questions?

The report describing the results of a survey should include a statement describing the purpose or purposes of the survey. One indication that a survey offers probative evidence is that it was designed to collect information relevant to the legal controversy (e.g., to estimate damages in an antitrust suit or to assess con-

28. Loctite Corp. v. National Starch & Chem. Corp., 516 F. Supp. 190, 205 (S.D.N.Y. 1981) (distinguishing between surveys conducted in anticipation of litigation and surveys conducted for nonlitigation purposes which cannot be reproduced because of the passage of time, concluding that parties should not be compelled to introduce the former at trial, but may be required to provide the latter).

29. *E.g.*, Charles Jacquin et Cie, Inc. v. Destileria Serralles, Inc., 921 F.2d 467, 475 (3d Cir. 1990). *See also* Brunswick Corp. v. Spinit Reel Co., 832 F.2d 513, 522 (10th Cir. 1987).

30. E.S. Originals, Inc. v. Stride Rite Corp., 656 F. Supp. 484, 490 (S.D.N.Y. 1987); *see also* Ortho Pharm. Corp. v. Cosprophar, Inc., 32 F.3d 690, 695 (2d Cir. 1994); Henri's Food Prods. Co. v. Kraft, Inc., 717 F.2d 352, 357 (7th Cir. 1983); Information Clearing House, Inc. v. Find Magazine, 492 F. Supp. 147, 160 (S.D.N.Y. 1980).

sumer confusion in a trademark case). Surveys not conducted specifically in preparation for, or in response to, litigation may provide important information,[31] but they frequently ask irrelevant questions[32] or select inappropriate samples of respondents for study.[33] Nonetheless, surveys do not always achieve their stated goals. Thus, the content and execution of a survey must be scrutinized even if the survey was designed to provide relevant data on the issue before the court. Moreover, if a survey was not designed for purposes of litigation, one source of bias is less likely: The party presenting the survey is less likely to have designed and constructed the survey to prove its side of the issue in controversy.

B. *Was Participation in the Design, Administration, and Interpretation of the Survey Appropriately Controlled to Ensure the Objectivity of the Survey?*

An early handbook for judges recommended that survey interviews be "conducted independently of the attorneys in the case."[34] Some courts have interpreted this to mean that any evidence of attorney participation is objectionable.[35] A better interpretation is that the attorney should have no part in carrying out the survey.[36] However, some attorney involvement in the survey design is

31. *See, e.g.*, Wright v. Jeep Corp., 547 F. Supp. 871, 874 (E.D. Mich. 1982). Indeed, as courts increasingly have been faced with scientific issues, parties have requested in a number of recent cases that the courts compel production of research data and testimony by unretained experts. The circumstances under which an unretained expert can be compelled to testify or to disclose research data and opinions, as well as the extent of disclosure that can be required when the research conducted by the expert has a bearing on the issues in the case, are the subject of considerable current debate. *See, e.g.*, Richard L. Marcus, *Discovery Along the Litigation/Science Interface*, 57 Brook. L. Rev. 381, 393–428 (1991); Joe S. Cecil, *Judicially Compelled Disclosure of Research Data*, 1 Cts. Health Sci. & L. 434 (1991); *see also* Symposium, *Court-Ordered Disclosure of Academic Research: A Clash of Values of Science and Law*, Law & Contemp. Probs., Summer 1996, at 1.

32. Loctite Corp. v. National Starch & Chem. Corp., 516 F. Supp. 190, 206 (S.D.N.Y. 1981) (marketing surveys conducted before litigation were designed to test for brand awareness, whereas the "single issue at hand . . . [was] whether consumers understood the term 'Super Glue' to designate glue from a single source").

33. In *Craig v. Boren*, 429 U.S. 190 (1976), the state unsuccessfully attempted to use its annual roadside survey of the blood alcohol level, drinking habits, and preferences of drivers to justify prohibiting the sale of 3.2% beer to males under the age of 21 and to females under the age of 18. The data were biased because it was likely that the male would be driving if both the male and female occupants of the car had been drinking. As pointed out in 2 Joseph L. Gastwirth, Statistical Reasoning in Law and Public Policy: Tort Law, Evidence, and Health 527 (1988), the roadside survey would have provided more relevant data if all occupants of the cars had been included in the survey (and if the type and amount of alcohol most recently consumed had been requested so that the consumption of 3.2% beer could have been isolated).

34. Judicial Conference of the U.S., Handbook of Recommended Procedures for the Trial of Protracted Cases 75 (1960).

35. *E.g.*, Boehringer Ingelheim G.m.b.H. v. Pharmadyne Lab., 532 F. Supp. 1040, 1058 (D.N.J. 1980).

36. Upjohn Co. v. American Home Prods. Corp., No. 1-95-CV-237, 1996 U.S. Dist. LEXIS 8049, at *42 (W.D. Mich. Apr. 5, 1996) (objection that "counsel reviewed the design of the survey

necessary to ensure that relevant questions are directed to a relevant population.[37] The trier of fact evaluates the objectivity and relevance of the questions on the survey and the appropriateness of the definition of the population used to guide sample selection. These aspects of the survey are visible to the trier of fact and can be judged on their quality, irrespective of who suggested them. In contrast, the interviews themselves are not directly visible, and any potential bias is minimized by having interviewers and respondents blind to the purpose and sponsorship of the survey and by excluding attorneys from any part in conducting interviews and tabulating results.

C. Are the Experts Who Designed, Conducted, or Analyzed the Survey Appropriately Skilled and Experienced?

Experts prepared to design, conduct, and analyze a survey generally should have graduate training in psychology (especially social, cognitive, or consumer psychology), sociology, marketing, communication sciences, statistics, or a related discipline; that training should include courses in survey research methods, sampling, measurement, interviewing, and statistics. In some cases, professional experience in conducting and publishing survey research may provide the requisite background. In all cases, the expert must demonstrate an understanding of survey methodology, including sampling,[38] instrument design (questionnaire and interview construction), and statistical analysis.[39] Publication in peer-reviewed journals, authored books, membership in professional organizations, faculty appointments, consulting experience, research grants, and membership on scientific advisory panels for government agencies or private foundations are indications of a professional's area and level of expertise. In addition, if the survey involves highly technical subject matter (e.g., the particular preferences of electrical engineers for various pieces of electrical equipment and the bases for those preferences) or involves a special population (e.g., developmentally disabled adults with limited cognitive skills), the survey expert also should be able to demonstrate sufficient familiarity with the topic or population (or assistance from an individual on the research team with suitable expertise) to design a survey instrument that will communicate clearly with relevant respondents.

carries little force with this Court because [opposing party] has not identified any flaw in the survey that might be attributed to counsel's assistance").

37. 3 J. Thomas McCarthy, McCarthy on Trademarks and Unfair Competition § 32:166 (4th ed. 1996).

38. The one exception is that sampling expertise is unnecessary if the survey is administered to all members of the relevant population. *See, e.g.*, McGovern & Lind, *supra* note 27.

39. If survey expertise is being provided by several experts, a single expert may have general familiarity but not special expertise in all these areas.

D. Are the Experts Who Will Testify About Surveys Conducted by Others Appropriately Skilled and Experienced?

Parties often call on an expert to testify about a survey conducted by someone else. The secondary expert's role is to offer support for a survey commissioned by the party who calls the expert, to critique a survey presented by the opposing party, or to introduce findings or conclusions from a survey not conducted in preparation for litigation or by any of the parties to the litigation. The trial court should take into account the exact issue that the expert seeks to testify about and the nature of the expert's field of expertise.[40] The secondary expert who gives an opinion about the adequacy and interpretation of a survey not only should have general skills and experience with surveys and be familiar with all of the issues addressed in this reference guide, but also should demonstrate familiarity with the following properties of the survey being discussed:

1. the purpose of the survey;
2. the survey methodology, including
   a. the target population,
   b. the sampling design used in conducting the survey,
   c. the survey instrument (questionnaire or interview schedule), and
   d. (for interview surveys) interviewer training and instruction;
3. the results, including rates and patterns of missing data; and
4. the statistical analyses used to interpret the results.

III. Population Definition and Sampling

A. Was an Appropriate Universe or Population Identified?

One of the first steps in designing a survey or in deciding whether an existing survey is relevant is to identify the target population (or universe).[41] The target population consists of all elements (i.e., objects, individuals, or other social units) whose characteristics or perceptions the survey is intended to represent. Thus, in trademark litigation, the relevant population in some disputes may include all prospective and actual purchasers of the plaintiff's goods or services and all prospective and actual purchasers of the defendant's goods or services. Similarly, the population for a discovery survey may include all potential plaintiffs or all em-

40. Margaret A. Berger, The Supreme Court's Triology on the Admissibility of Expert Testimony § IV.C, in this manual.

41. Identification of the proper universe is recognized uniformly as a key element in the development of a survey. *See, e.g.*, Judicial Conference of the U.S., *supra* note 34; MCL 3d, *supra* note 15, § 21.493. *See also* 3 McCarthy, *supra* note 37, § 32:166; Council of Am. Survey Res. Orgs., Code of Standards and Ethics for Survey Research § III.B.4 (1997).

ployees who worked for Company A between two specific dates. In a community survey designed to provide evidence for a motion for a change of venue, the relevant population consists of all jury-eligible citizens in the community in which the trial is to take place.[42] The definition of the relevant population is crucial because there may be systematic differences in the responses of members of the population and nonmembers. (For example, consumers who are prospective purchasers may know more about the product category than consumers who are not considering making a purchase.)

The universe must be defined carefully. For example, a commercial for a toy or breakfast cereal may be aimed at children, who in turn influence their parents' purchases. If a survey assessing the commercial's tendency to mislead were conducted based on the universe of prospective and actual adult purchasers, it would exclude a crucial group of eligible respondents. Thus, the appropriate population in this instance would include children as well as parents.[43]

B. Did the Sampling Frame Approximate the Population?

The target population consists of all the individuals or units that the researcher would like to study. The sampling frame is the source (or sources) from which the sample actually is drawn. The surveyor's job generally is easier if a complete list of every eligible member of the population is available (e.g., all plaintiffs in a discovery survey), so that the sampling frame lists the identity of all members of the target population. Frequently, however, the target population includes members who are inaccessible or who cannot be identified in advance. As a result, compromises are sometimes required in developing the sampling frame. The survey report should contain a description of the target population, a description of the survey population actually sampled, a discussion of the difference between the two populations, and an evaluation of the likely consequences of that difference.

42. A second relevant population may consist of jury-eligible citizens in the community where the party would like to see the trial moved. By questioning citizens in both communities, the survey can test whether moving the trial is likely to reduce the level of animosity toward the party requesting the change of venue. *See* United States v. Haldeman, 559 F.2d 31, 140, 151, app. A at 176–79 (D.C. Cir. 1976) (court denied change of venue over the strong objection of Judge MacKinnon, who cited survey evidence that Washington, D.C., residents were substantially more likely to conclude, before trial, that the defendants were guilty), *cert. denied*, 431 U.S. 933 (1977); *see also* People v. Venegas, 31 Cal. Rptr. 2d 114, 117 (Ct. App. 1994) (change of venue denied because defendant failed to show that the defendant would face a less hostile jury in a different court).

43. Children and some other populations create special challenges for researchers. For example, very young children should not be asked about sponsorship or licensing, concepts that are foreign to them. Concepts, as well as wording, should be age-appropriate.

A survey that provides information about a wholly irrelevant universe of respondents is itself irrelevant.[44] Courts are likely to exclude the survey or accord it little weight. Thus, when the plaintiff submitted the results of a survey to prove that the green color of its fishing rod had acquired a secondary meaning, the court gave the survey little weight in part because the survey solicited the views of fishing rod dealers rather than consumers.[45] More commonly, however, the sampling frame is either underinclusive or overinclusive relative to the target population. If it is underinclusive, the survey's value depends on the extent to which the excluded population is likely to react differently from the included population. Thus, a survey of spectators and participants at running events would be sampling a sophisticated subset of those likely to purchase running shoes. Because this subset probably would consist of the consumers most knowledgeable about the trade dress used by companies that sell running shoes, a survey based on this population would be likely to substantially overrepresent the strength of a particular design as a trademark, and the extent of that overrepresentation would be unknown and not susceptible to any reasonable estimation.[46]

Similarly, in a survey designed to project demand for cellular phones, the assumption that businesses would be the primary users of cellular service led surveyors to exclude potential nonbusiness users from the survey. The Federal Communications Commission (FCC) found the assumption unwarranted and concluded that the research was flawed, in part because of this underinclusive universe.[47]

44. A survey aimed at assessing how persons in the trade respond to an advertisement should be conducted on a sample of persons in the trade and not on a sample of consumers. Home Box Office v. Showtime/The Movie Channel, 665 F. Supp. 1079, 1083 (S.D.N.Y.), *aff'd in part & vacated in part*, 832 F.2d 1311 (2d Cir. 1987). *But see* Lon Tai Shing Co. v. Koch + Lowy, No. 90-C4464, 1990 U.S. Dist. LEXIS 19123, at *50 (S.D.N.Y. Dec. 14, 1990), in which the judge was willing to find likelihood of consumer confusion from a survey of lighting store salespersons questioned by a survey researcher posing as a customer. The court was persuaded that the salespersons who were misstating the source of the lamp, whether consciously or not, must have believed reasonably that the consuming public would be misled by the salespersons' inaccurate statements about the name of the company that manufactured the lamp they were selling.

45. R.L. Winston Rod Co. v. Sage Mfg. Co., 838 F. Supp. 1396, 1401-02 (D. Mont. 1993).

46. Brooks Shoe Mfg. Co. v. Suave Shoe Corp., 533 F. Supp. 75, 80 (S.D. Fla. 1981), *aff'd*, 716 F.2d 854 (11th Cir. 1983). *See also* Winning Ways, Inc. v. Holloway Sportswear, Inc., 913 F. Supp. 1454, 1467 (D. Kan. 1996) (survey flawed in failing to include sporting goods customers who constituted a major portion of customers). *But see* Thomas & Betts Corp. v. Panduit Corp., 138 F.3d 277, 294–95 (7th Cir. 1998) (survey of store personnel admissible because relevant market included both distributors and ultimate purchasers).

47. Gencom, Inc., 56 Rad. Reg. 2d (P&F) 1597, 1604 (1984). This position was affirmed on appeal. *See* Gencom, Inc. v. FCC, 832 F.2d 171, 186 (D.C. Cir. 1987).

In some cases, it is difficult to determine whether an underinclusive universe distorts the results of the survey and, if so, the extent and likely direction of the bias. For example, a trademark survey was designed to test the likelihood of confusing an analgesic currently on the market with a new product that was similar in appearance.[48] The plaintiff's survey included only respondents who had used the plaintiff's analgesic, and the court found that the universe should have included users of other analgesics, "so that the full range of potential customers for whom plaintiff and defendants would compete could be studied."[49] In this instance, it is unclear whether users of the plaintiff's product would be more or less likely to be confused than users of the defendant's product or users of a third analgesic.[50]

An overinclusive universe generally presents less of a problem in interpretation than does an underinclusive universe. If the survey expert can demonstrate that a sufficiently large (and representative) subset of respondents in the survey was drawn from the appropriate universe, the responses obtained from that subset can be examined, and inferences about the relevant universe can be drawn based on that subset.[51] If the relevant subset cannot be identified, however, an overbroad universe will reduce the value of the survey.[52] If the sample is drawn from an underinclusive universe, there is generally no way to know how the unrepresented members would have responded.[53]

C. How Was the Sample Selected to Approximate the Relevant Characteristics of the Population?

Identification of a survey population must be followed by selection of a sample that accurately represents that population.[54] The use of probability sampling techniques maximizes both the representativeness of the survey results and the ability to assess the accuracy of estimates obtained from the survey.

Probability samples range from simple random samples to complex multistage sampling designs that use stratification, clustering of population elements into various groupings, or both. In simple random sampling, the most basic type

48. American Home Prods. Corp. v. Barr Lab., Inc., 656 F. Supp. 1058 (D.N.J.), *aff'd*, 834 F.2d 368 (3d Cir. 1987).

49. *Id.* at 1070.

50. *See also* Craig v. Boren, 429 U.S. 190 (1976).

51. This occurred in *National Football League Properties, Inc. v. Wichita Falls Sportswear, Inc.*, 532 F. Supp. 651, 657–58 (W.D. Wash. 1982).

52. Schieffelin & Co. v. Jack Co. of Boca, 850 F. Supp. 232, 246 (S.D.N.Y. 1994).

53. *See, e.g.*, Amstar Corp. v. Domino's Pizza, Inc., 615 F.2d 252, 263–64 (5th Cir.) (court found both plaintiff's and defendant's surveys substantially defective for a systematic failure to include parts of the relevant population), *cert. denied*, 449 U.S. 899 (1980).

54. MCL 3d, *supra* note 15, § 21.493. *See also* David H. Kaye & David A. Freedman, Reference Guide on Statistics § II.B, in this manual.

of probability sampling, every element in the population has a known, equal probability of being included in the sample, and all possible samples of a given size are equally likely to be selected.[55] In all forms of probability sampling, each element in the relevant population has a known, nonzero probability of being included in the sample.[56]

Probability sampling offers two important advantages over other types of sampling. First, the sample can provide an unbiased estimate of the responses of all persons in the population from which the sample was drawn; that is, the expected value of the sample estimate is the population value being estimated. Second, the researcher can calculate a confidence interval that describes explicitly how reliable the sample estimate of the population is likely to be. Thus, suppose a survey tested a sample of 400 dentists randomly selected from the population of all dentists licensed to practice in the United States and found that 80, or 20%, of them mistakenly believed that a new toothpaste, Goldgate, was manufactured by the makers of Colgate. A survey expert could properly compute a confidence interval around the 20% estimate obtained from this sample. If the survey was repeated a large number of times, and a 95% confidence interval was computed each time, 95% of the confidence intervals would include the actual percentage of dentists in the entire population who would believe that Goldgate was manufactured by the makers of Colgate.[57] In this example, the confidence interval, or margin of error, is the estimate (20%) plus or minus 4%, or the distance between 16% and 24%.

All sample surveys produce estimates of population values, not exact measures of those values. Strictly speaking, the margin of sampling error associated with the sample estimate assumes probability sampling. Assuming a probability sample, a confidence interval describes how stable the mean response in the sample is likely to be. The width of the confidence interval depends on three characteristics:

55. Systematic sampling, in which every *n*th unit in the population is sampled and the starting point is selected randomly, fulfills the first of these conditions. It does not fulfill the second, because no systematic sample can include elements adjacent to one another on the list of population members from which the sample is drawn. Except in very unusual situations when periodicities occur, systematic samples and simple random samples generally produce the same results. Seymour Sudman, *Applied Sampling*, *in* Handbook of Survey Research, *supra* note 1, at 145, 169.

56. Other probability sampling techniques include (1) stratified random sampling, in which the researcher subdivides the population into mutually exclusive and exhaustive subpopulations, or strata, and then randomly selects samples from within these strata; and (2) cluster sampling, in which elements are sampled in groups or clusters, rather than on an individual basis. Martin Frankel, *Sampling Theory*, *in* Handbook of Survey Research, *supra* note 1, at 21, 37, 47.

57. Actually, since survey interviewers would be unable to locate some dentists and some dentists would be unwilling to participate in the survey, technically the population to which this sample would be projectable would be all dentists with current addresses who would be willing to participate in the survey if they were asked.

1. the size of the sample (the larger the sample, the narrower the interval);
2. the variability of the response being measured; and
3. the confidence level the researcher wants to have.

Traditionally, scientists adopt the 95% level of confidence, which means that if 100 samples of the same size were drawn, the confidence interval expected for at least 95 of the samples would be expected to include the true population value.[58]

Although probability sample surveys often are conducted in organizational settings and are the recommended sampling approach in academic and government publications on surveys, probability sample surveys can be expensive when in-person interviews are required, the target population is dispersed widely, or qualified respondents are scarce. A majority of the consumer surveys conducted for Lanham Act litigation present results from nonprobability convenience samples.[59] They are admitted into evidence based on the argument that nonprobability sampling is used widely in marketing research and that "results of these studies are used by major American companies in making decisions of considerable consequence."[60] Nonetheless, when respondents are not selected randomly from the relevant population, the expert should be prepared to justify the method used to select respondents. Special precautions are required to reduce the likelihood of biased samples.[61] In addition, quantitative values computed from such samples (e.g., percentage of respondents indicating confusion) should be viewed as rough indicators rather than as precise quantitative estimates. Confidence intervals should not be computed.

58. To increase the likelihood that the confidence interval contains the actual population value (e.g., from 95% to 99%), the width of the confidence interval can be expanded. An increase in the confidence interval brings an increase in the confidence level. For further discussion of confidence intervals, see David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.A, in this manual.

59. Jacob Jacoby & Amy H. Handlin, *Non-Probability Sampling Designs for Litigation Surveys*, 81 Trademark Rep. 169, 173 (1991). For probability surveys conducted in trademark cases, see *National Football League Properties, Inc. v. Wichita Falls Sportswear, Inc.*, 532 F. Supp. 651 (W.D. Wash. 1982); James Burrough, Ltd. v. Sign of Beefeater, Inc., 540 F.2d 266 (7th Cir. 1976).

60. National Football League Properties, Inc. v. New Jersey Giants, Inc., 637 F. Supp. 507, 515 (D.N.J. 1986). A survey of members of the Council of American Survey Research Organizations, the national trade association for commercial survey research firms in the United States, revealed that 95% of the in-person independent contacts in studies done in 1985 took place in malls or shopping centers. Jacoby & Handlin, *supra* note 59, at 172–73, 176.

D. Was the Level of Nonresponse Sufficient to Raise Questions About the Representativeness of the Sample? If So, What Is the Evidence That Nonresponse Did Not Bias the Results of the Survey?

Even when a sample is drawn randomly from a complete list of elements in the target population, responses or measures may be obtained on only part of the selected sample. If this lack of response were distributed randomly, valid inferences about the population could be drawn from the characteristics of the available elements in the sample. The difficulty is that nonresponse often is not random, so that, for example, persons who are single typically have three times the "not at home" rate in U.S. Census Bureau surveys as do family members.[62] Efforts to increase response rates include making several attempts to contact potential respondents and providing financial incentives for participating in the survey.

One suggested formula for quantifying a tolerable level of nonresponse in a probability sample is based on the guidelines for statistical surveys issued by the former U.S. Office of Statistical Standards.[63] According to these guidelines, response rates of 90% or more are reliable and generally can be treated as random samples of the overall population. Response rates between 75% and 90% usually yield reliable results, but the researcher should conduct some check on the representativeness of the sample. Potential bias should receive greater scrutiny when the response rate drops below 75%. If the response rate drops below 50%, the survey should be regarded with significant caution as a basis for precise quantitative statements about the population from which the sample was drawn.[64]

Determining whether the level of nonresponse in a survey is critical generally requires an analysis of the determinants of nonresponse. For example, even a survey with a high response rate may seriously underrepresent some portions of the population, such as the unemployed or the poor. If a general population sample was used to chart changes in the proportion of the population that knows someone with HIV, the survey would underestimate the population value if some groups more likely to know someone with HIV (e.g., intravenous drug users) were underrepresented in the sample. The survey expert should be prepared to provide evidence on the potential impact of nonresponse on the survey results.

61. *See infra* § III.E.

62. 2 Gastwirth, *supra* note 33, at 501. This volume contains a useful discussion of sampling, along with a set of examples. *Id.* at 467.

63. This standard is cited with approval by Gastwirth. *Id.* at 502.

64. For thoughtful examples of judges closely scrutinizing potential sample bias when response rates were below 75%, see *Vuyanich v. Republic National Bank*, 505 F. Supp. 224 (N.D. Tex. 1980); Rosado v. Wyman, 322 F. Supp. 1173 (E.D.N.Y.), *aff'd*, 437 F.2d 619 (2d Cir. 1970), *aff'd*, 402 U.S. 991 (1971).

In surveys that include sensitive or difficult questions, particularly surveys that are self-administered, some respondents may refuse to provide answers or may provide incomplete answers. To assess the impact of nonresponse to a particular question, the survey expert should analyze the differences between those who answered and those who did not answer. Procedures to address the problem of missing data include recontacting respondents to obtain the missing answers and using the respondent's other answers to predict the missing response.[65]

E. What Procedures Were Used to Reduce the Likelihood of a Biased Sample?

If it is impractical for a survey researcher to sample randomly from the entire target population, the researcher still can apply probability sampling to some aspects of respondent selection to reduce the likelihood of biased selection. For example, in many studies the target population consists of all consumers or purchasers of a product. Because it is impractical to randomly sample from that population, research is conducted in shopping malls where some members of the target population may not shop. Mall locations, however, can be sampled randomly from a list of possible mall sites. By administering the survey at several different malls, the expert can test for and report on any differences observed across sites. To the extent that similar results are obtained in different locations using different on-site interview operations, it is less likely that idiosyncrasies of sample selection or administration can account for the results.[66] Similarly, since the characteristics of persons visiting a shopping center vary by day of the week and time of day, bias in sampling can be reduced if the survey design calls for sampling time segments as well as mall locations.[67]

In mall intercept surveys, the organization that manages the on-site interview facility generally employs recruiters who approach potential survey respondents in the mall and ascertain if they are qualified and willing to participate in the survey. If a potential respondent agrees to answer the questions and meets the specified criteria, he or she is escorted to the facility where the survey interview takes place. If recruiters are free to approach potential respondents without controls on how an individual is to be selected for screening, shoppers who spend more time in the mall are more likely to be approached than shoppers who visit the mall only briefly. Moreover, recruiters naturally prefer to approach friendly-

65. Andy B. Anderson et al., *Missing Data: A Review of the Literature*, *in* Handbook of Survey Research, *supra* note 1, at 415.

66. Note, however, that differences in results across sites may be due to genuine differences in respondents across geographic locations or to a failure to administer the survey consistently across sites.

67. Seymour Sudman, *Improving the Quality of Shopping Center Sampling*, 17 J. Marketing Res. 423 (1980).

looking potential respondents, so that it is more likely that certain types of individuals will be selected. These potential biases in selection can be reduced by providing appropriate selection instructions and training recruiters effectively. Training that reduces the interviewer's discretion in selecting a potential respondent is likely to reduce bias in selection, as are instructions to approach every *n*th person entering the facility through a particular door.[68]

F. What Precautions Were Taken to Ensure That Only Qualified Respondents Were Included in the Survey?

In a carefully executed survey, each potential respondent is questioned or measured on the attributes that determine his or her eligibility to participate in the survey. Thus, the initial questions screen potential respondents to determine if they are within the target population of the survey (e.g., Is she at least fourteen years old? Does she own a dog? Does she live within ten miles?). The screening questions must be drafted so that they do not convey information that will influence the respondent's answers on the main survey. For example, if respondents must be prospective and recent purchasers of Sunshine orange juice in a trademark survey designed to assess consumer confusion with Sun Time orange juice, potential respondents might be asked to name the brands of orange juice they have purchased recently or expect to purchase in the next six months. They should not be asked specifically if they recently have purchased, or expect to purchase, Sunshine orange juice, because this may affect their responses on the survey either by implying who is conducting the survey or by supplying them with a brand name that otherwise would not occur to them.

The content of a screening questionnaire (or screener) can also set the context for the questions that follow. In *Pfizer, Inc. v. Astra Pharmaceutical Products, Inc.*,[69] physicians were asked a screening question to determine whether they prescribed particular drugs. The court found that the screener conditioned the physicians to respond with the name of a drug rather than a condition.[70]

The criteria for determining whether to include a potential respondent in the survey should be objective and clearly conveyed, preferably using written instructions addressed to those who administer the screening questions. These instructions and the completed screening questionnaire should be made avail-

68. In the end, even if malls are randomly sampled and shoppers are randomly selected within malls, results from mall surveys technically can be used to generalize only to the population of mall shoppers. The ability of the mall sample to describe the likely response pattern of the broader relevant population will depend on the extent to which a substantial segment of the relevant population (1) is not found in malls and (2) would respond differently to the interview.

69. 858 F. Supp. 1305, 1321 & n.13 (S.D.N.Y. 1994).

70. *Id.* at 1321.

able to the court and the opposing party along with the interview form for each respondent.

IV. Survey Questions and Structure

A. Were Questions on the Survey Framed to Be Clear, Precise, and Unbiased?

Although it seems obvious that questions on a survey should be clear and precise, phrasing questions to reach that goal is often difficult. Even questions that appear clear can convey unexpected meanings and ambiguities to potential respondents. For example, the question "What is the average number of days each week you have butter?" appears to be straightforward. Yet some respondents wondered whether margarine counted as butter, and when the question was revised to include the introductory phrase "not including margarine," the reported frequency of butter use dropped dramatically.[71]

When unclear questions are included in a survey, they may threaten the validity of the survey by systematically distorting responses if respondents are misled in a particular direction, or by inflating random error if respondents guess because they do not understand the question.[72] If the crucial question is sufficiently ambiguous or unclear, it may be the basis for rejecting the survey. For example, a survey was designed to assess community sentiment that would warrant a change of venue in trying a case for damages sustained when a hotel skywalk collapsed.[73] The court found that the question "Based on what you have heard, read or seen, do you believe that in the current compensatory damage trials, the defendants, such as the contractors, designers, owners, and operators of the Hyatt Hotel, should be punished?" could neither be correctly understood nor easily answered.[74] The court noted that the phrase "compensatory damages," although well-defined for attorneys, was unlikely to be meaningful for laypersons.[75]

Texts on survey research generally recommend pretests as a way to increase the likelihood that questions are clear and unambiguous,[76] and some courts have

71. Floyd J. Fowler, Jr., *How Unclear Terms Affect Survey Data*, 56 Pub. Opinion Q. 218, 225–26 (1992).

72. *Id.* at 219.

73. Firestone v. Crown Ctr. Redevelopment Corp., 693 S.W.2d 99 (Mo. 1985) (en banc).

74. *Id.* at 102, 103.

75. *Id.* at 103. When there is any question about whether some respondent will understand a particular term or phrase, the term or phrase should be defined explicitly.

76. For a thorough treatment of pretesting methods, see Jean M. Converse & Stanley Presser, Survey Questions: Handcrafting the Standardized Questionnaire 51 (1986). *See also* Fred W. Morgan, *Judicial Standards for Survey Research: An Update and Guidelines*, 54 J. Marketing 59, 64 (1990).

recognized the value of pretests.[77] In a pretest or pilot test,[78] the proposed survey is administered to a small sample (usually between twenty-five and seventy-five)[79] of the same type of respondents who would be eligible to participate in the full-scale survey. The interviewers observe the respondents for any difficulties they may have with the questions and probe for the source of any such difficulties so that the questions can be rephrased if confusion or other difficulties arise. Attorneys who commission surveys for litigation sometimes are reluctant to approve pilot work or to reveal that pilot work has taken place because they are concerned that if a pretest leads to revised wording of the questions, the trier of fact may believe that the survey has been manipulated and is biased or unfair. A more appropriate reaction is to recognize that pilot work can improve the quality of a survey and to anticipate that it often results in word changes that increase clarity and correct misunderstandings. Thus, changes may indicate informed survey construction rather than flawed survey design.[80]

B. Were Filter Questions Provided to Reduce Guessing?

Some survey respondents may have no opinion on an issue under investigation, either because they have never thought about it before or because the question mistakenly assumes a familiarity with the issue. For example, survey respondents may not have noticed that the commercial they are being questioned about guaranteed the quality of the product being advertised and thus they may have no opinion on the kind of guarantee it indicated. Likewise, in an employee survey, respondents may not be familiar with the parental leave policy at their company and thus may have no opinion on whether they would consider taking advantage of the parental leave policy if they became parents. The following three alternative question structures will affect how those respondents answer and how their responses are counted.

First, the survey can ask all respondents to answer the question (e.g., "Did you understand the guarantee offered by Clover to be a one-year guarantee, a sixty-day guarantee, or a thirty-day guarantee?"). Faced with a direct question, particularly one that provides response alternatives, the respondent obligingly may supply an answer even if (in this example) the respondent did not notice the guarantee (or is unfamiliar with the parental leave policy). Such answers will

77. *E.g.*, Zippo Mfg. Co. v. Rogers Imports, Inc., 216 F. Supp. 670 (S.D.N.Y. 1963).

78. The terms *pretest* and *pilot test* are sometimes used interchangeably to describe pilot work done in the planning stages of research. When they are distinguished, the difference is that a pretest tests the questionnaire, whereas a pilot test generally tests proposed collection procedures as well.

79. Converse & Presser, *supra* note 76, at 69. Converse and Presser suggest that a pretest with twenty-five respondents is appropriate when the survey uses professional interviewers.

80. *See infra* § VII.B for a discussion of obligations to disclose pilot work.

reflect only what the respondent can glean from the question, or they may reflect pure guessing. The imprecision introduced by this approach will increase with the proportion of respondents who are unfamiliar with the topic at issue.

Second, the survey can use a quasi-filter question to reduce guessing by providing "don't know" or "no opinion" options as part of the question (e.g., "Did you understand the guarantee offered by Clover to be for more than a year, a year, or less than a year, or don't you have an opinion?").[81] By signaling to the respondent that it is appropriate not to have an opinion, the question reduces the demand for an answer and, as a result, the inclination to hazard a guess just to comply. Respondents are more likely to choose a "no opinion" option if it is mentioned explicitly by the interviewer than if it is merely accepted when the respondent spontaneously offers it as a response. The consequence of this change in format is substantial. Studies indicate that, although the relative distribution of the respondents selecting the *listed* choices is unlikely to change dramatically, presentation of an explicit "don't know" or "no opinion" alternative commonly leads to a 20%–25% increase in the proportion of respondents selecting that response.[82]

Finally, the survey can include full-filter questions, that is, questions that lay the groundwork for the substantive question by first asking the respondent if he or she has an opinion about the issue or happened to notice the feature that the interviewer is preparing to ask about (e.g., "Based on the commercial you just saw, do you have an opinion about how long Clover stated or implied that its guarantee lasts?"). The interviewer then asks the substantive question only of those respondents who have indicated that they have an opinion on the issue.

Which of these three approaches is used and the way it is used can affect the rate of "no opinion" responses that the substantive question will evoke.[83] Respondents are more likely to say they do not have an opinion on an issue if a full filter is used than if a quasi-filter is used.[84] However, in maximizing respondent expressions of "no opinion," full filters may produce an underreporting of opinions. There is some evidence that full-filter questions discourage respondents who actually have opinions from offering them by conveying the implicit suggestion that respondents can avoid difficult follow-up questions by saying that they have no opinion.[85]

81. Norbert Schwarz & Hans-Jürgen Hippler, *Response Alternatives: The Impact of Their Choice and Presentation Order*, *in* Measurement Errors in Surveys 41, 45–46 (Paul P. Biemer et al. eds., 1991).

82. Howard Schuman & Stanley Presser, Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context 113–46 (1981).

83. Considerable research has been conducted on the effects of filters. For a review, see George F. Bishop et al., *Effects of Filter Questions in Public Opinion Surveys*, 47 Pub. Opinion Q. 528 (1983).

84. Schwarz & Hippler, *supra* note 81, at 45–46.

85. *Id.* at 46.

In general, then, a survey that uses full filters tends to provide a conservative estimate of the number of respondents holding an opinion, whereas a survey that uses neither full filters nor quasi-filters tends to overestimate the number of respondents with opinions, because some respondents offering opinions are guessing. The strategy of including a "no opinion" or "don't know" response as a quasi-filter avoids both of these extremes. Thus, rather than asking, "Based on the commercial, do you believe that the two products are made in the same way, or are they made differently?"[86] or prefacing the question with a preliminary, "Do you have an opinion, based on the commercial, concerning the way that the two products are made?" the question could be phrased, "Based on the commercial, do you believe that the two products are made in the same way, or that they are made differently, or don't you have an opinion about the way they are made?"

C. Did the Survey Use Open-Ended or Closed-Ended Questions? How Was the Choice in Each Instance Justified?

The questions that make up a survey instrument may be open-ended, closed-ended, or a combination of both. Open-ended questions require the respondent to formulate and express an answer in his or her own words (e.g., "What was the main point of the commercial?" "Where did you catch the fish you caught in these waters?"[87]). Closed-ended questions provide the respondent with an explicit set of responses from which to choose; the choices may be as simple as yes or no (e.g., "Is Colby College coeducational?"[88]) or as complex as a range of alternatives (e.g., "The two pain relievers have (1) the same likelihood of causing gastric ulcers; (2) about the same likelihood of causing gastric ulcers; (3) a somewhat different likelihood of causing gastric ulcers; (4) a very different likelihood of causing gastric ulcers; or (5) none of the above."[89]).

Open-ended and closed-ended questions may elicit very different responses.[90]

86. The question in the example without the "no opinion" alternative was based on a question rejected by the court in *Coors Brewing Co. v. Anheuser-Busch Cos.*, 802 F. Supp. 965, 972–73 (S.D.N.Y. 1992).

87. A relevant example from *Wilhoite v. Olin Corp.* is described in McGovern & Lind, *supra* note 27, at 76.

88. Presidents & Trustees of Colby College v. Colby College–N.H., 508 F.2d 804, 809 (1st Cir. 1975).

89. This question is based on one asked in *American Home Products Corp. v. Johnson & Johnson*, 654 F. Supp. 568, 581 (S.D.N.Y. 1987), that was found to be a leading question by the court, primarily because the choices suggested that the respondent had learned about aspirin's and ibuprofen's relative likelihood of causing gastric ulcers. In contrast, in *McNeilab, Inc. v. American Home Products Corp.*, 501 F. Supp. 517, 525 (S.D.N.Y. 1980), the court accepted as nonleading the question, "Based only on what the commercial said, would Maximum Strength Anacin contain more pain reliever, the same amount of pain reliever, or less pain reliever than the brand you, yourself, currently use most often?"

90. Howard Schuman & Stanley Presser, *Question Wording as an Independent Variable in Survey Analysis*,

Most responses are less likely to be volunteered by respondents who are asked an open-ended question than they are to be chosen by respondents who are presented with a closed-ended question. The response alternatives in a closed-ended question may remind respondents of options that they would not otherwise consider or which simply do not come to mind as easily.[91]

The advantage of open-ended questions is that they give the respondent fewer hints about the answer that is expected or preferred. Precoded responses on a closed-ended question, in addition to reminding respondents of options that they might not otherwise consider,[92] may direct the respondent away from or toward a particular response. For example, a commercial reported that in shampoo tests with more than 900 women, the sponsor's product received higher ratings than other brands.[93] According to a competitor, the commercial deceptively implied that each woman in the test rated more than one shampoo, when in fact each woman rated only one. To test consumer impressions, a survey might have shown the commercial and asked an open-ended question: "How many different brands mentioned in the commercial did each of the 900 women try?"[94] Instead, the survey asked a closed-ended question; respondents were given the choice of "one," "two," "three," "four," or "five or more." The fact that four of the five choices in the closed-ended question provided a response that was greater than one implied that the correct answer was probably more than one.[95] Note, however, that the open-ended question also may suggest that the answer is more than one. By asking "how many different brands," the question suggests (1) that the viewer should have received some message from the commercial about the number of brands each woman tried and (2) that different brands were tried. Thus, the wording of a question, open-ended or closed-ended, can be leading, and the degree of suggestiveness of each question must be considered in evaluating the objectivity of a survey.

6 Soc. Methods & Res. 151 (1977); Schuman & Presser, *supra* note 82, at 79–112; Converse & Presser, *supra* note 76, at 33.

91. For example, when respondents in one survey were asked, "What is the most important thing for children to learn to prepare them for life?", 62% picked "to think for themselves" from a list of five options, but only 5% spontaneously offered that answer when the question was open-ended. Schuman & Presser, *supra* note 82, at 104–07. An open-ended question presents the respondent with a free-recall task, whereas a closed-ended question is a recognition task. Recognition tasks in general reveal higher performance levels than recall tasks. Mary M. Smyth et al., Cognition in Action 25 (1987). In addition, there is evidence that respondents answering open-ended questions may be less likely to report some information that they would reveal in response to a closed-ended question when that information seems self-evident or irrelevant.

92. Schwarz & Hippler, *supra* note 81, at 43.

93. *See* Vidal Sassoon, Inc. v. Bristol-Myers Co., 661 F.2d 272, 273 (2d Cir. 1981).

94. This was the wording of the stem of the closed-ended question in the survey discussed in *Vidal Sassoon*, 661 F.2d at 275–76.

95. Ninety-five percent of the respondents who answered the closed-ended question in the plaintiff's survey said that each woman had tried two or more brands. The open-ended question was never asked.

Closed-ended questions have some additional potential weaknesses that arise if the choices are not constructed properly. If the respondent is asked to choose one response from among several choices, the response chosen will be meaningful only if the list of choices is exhaustive, that is, if the choices cover all possible answers a respondent might give to the question. If the list of possible choices is incomplete, a respondent may be forced to choose one that does not express his or her opinion.[96] Moreover, if respondents are told explicitly that they are not limited to the choices presented, most respondents nevertheless will select an answer from among the listed ones.[97]

Although many courts prefer open-ended questions on the grounds that they tend to be less leading, the value of any open-ended or closed-ended question depends on the information it is intended to elicit. Open-ended questions are more appropriate when the survey is attempting to gauge what comes first to a respondent's mind, but closed-ended questions are more suitable for assessing choices between well-identified options or obtaining ratings on a clear set of alternatives.

D. If Probes Were Used to Clarify Ambiguous or Incomplete Answers, What Steps Were Taken to Ensure That the Probes Were Not Leading and Were Administered in a Consistent Fashion?

When questions allow respondents to express their opinions in their own words, some of the respondents may give ambiguous or incomplete answers. In such instances, interviewers may be instructed to record any answer that the respondent gives and move on to the next question, or they may be instructed to probe to obtain a more complete response or clarify the meaning of the ambiguous response. In either situation, interviewers should record verbatim both what the respondent says and what the interviewer says in the attempt to get clarification. Failure to record every part of the exchange in the order in which it occurs raises questions about the reliability of the survey, because neither the court nor the opposing party can evaluate whether the probe affected the views expressed by the respondent.

Vidal Sassoon, 661 F.2d at 276. Norbert Schwarz, *Assessing Frequency Reports of Mundane Behaviors: Contributions of Cognitive Psychology to Questionnaire Construction*, *in* Research Methods in Personality and Social Psychology 98 (Clyde Hendrick & Margaret S. Clark eds., 1990), suggests that respondents often rely on the range of response alternatives as a frame of reference when they are asked for frequency judgments. *See, e.g.*, Roger Tourangeau & Tom W. Smith, *Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context*, 60 Pub. Opinion Q. 275, 292 (1996).

96. *See, e.g.*, American Home Prods. Corp. v. Johnson & Johnson, 654 F. Supp. 568, 581 (S.D.N.Y. 1987).

97. *See* Howard Schuman, *Ordinary Questions, Survey Questions, and Policy Questions*, 50 Pub. Opinion Q. 432, 435–36 (1986).

If the survey is designed to allow for probes, interviewers must be given explicit instructions on when they should probe and what they should say in probing. Standard probes used to draw out all that the respondent has to say (e.g., "Any further thoughts?" "Anything else?" "Can you explain that a little more?") are relatively innocuous and noncontroversial in content, but persistent continued requests for further responses to the same or nearly identical questions may convey the idea to the respondent that he or she has not yet produced the "right" answer.[98] Interviewers should be trained in delivering probes to maintain a professional and neutral relationship with the respondent (as they should during the rest of the interview), which minimizes any sense of passing judgment on the content of the answers offered. Moreover, interviewers should be given explicit instructions on when to probe, so that probes are administered consistently.

A more difficult type of probe to construct and deliver reliably is one that requires a substantive question tailored to the answer given by the respondent. The survey designer must provide sufficient instruction to interviewers so that they avoid giving directive probes that suggest one answer over another. Those instructions, along with all other aspects of interviewer training, should be made available for evaluation by the court and the opposing party.

E. What Approach Was Used to Avoid or Measure Potential Order or Context Effects?

The order in which questions are asked on a survey and the order in which response alternatives are provided in a closed-ended question can influence the answers.[99] Thus, although asking a general question before a more specific question on the same topic is unlikely to affect the response to the specific question, reversing the order of the questions may influence responses to the general question. As a rule, then, surveys are less likely to be subject to order effects if

98. *See, e.g.*, Johnson & Johnson-Merck Consumer Pharms. Co. v. Rhone-Poulenc Rorer Pharms., Inc., 19 F.3d 125, 135 (3d Cir. 1994); American Home Prods. Corp. v. Procter & Gamble Co., 871 F. Supp. 739, 748 (D.N.J. 1994).

99. *See* Schuman & Presser, *supra* note 82, at 23, 56–74; Norman M. Bradburn, *Response Effects*, *in* Handbook of Survey Research, *supra* note 1, at 289, 302. In *R.J. Reynolds Tobacco Co. v. Loew's Theatres, Inc.*, 511 F. Supp. 867, 875 (S.D.N.Y. 1980), the court recognized the biased structure of a survey which disclosed the tar content of the cigarettes being compared before questioning respondents about their cigarette preferences. Not surprisingly, respondents expressed a preference for the lower tar product. *See also* E. & J. Gallo Winery v. Pasatiempos Gallo, S.A., 905 F. Supp. 1403, 1409–10 (E.D. Cal. 1994) (court recognized that earlier questions referring to playing cards, board or table games, or party supplies, such as confetti, increased the likelihood that respondents would include these items in answers to the questions that followed).

the questions go from the general (e.g., "What do you recall being discussed in the advertisement?") to the specific (e.g., "Based on your reading of the advertisement, what companies do you think the ad is referring to when it talks about rental trucks that average five miles per gallon?").[100]

The mode of questioning can influence the form that an order effect takes. In mail surveys, respondents are more likely to select the first choice offered (a primacy effect); in telephone surveys, respondents are more likely to choose the last choice offered (a recency effect). Although these effects are typically small, no general formula is available that can adjust values to correct for order effects, because the size and even the direction of the order effects may depend on the nature of the question being asked and the choices being offered. Moreover, it may be unclear which order is most appropriate. For example, if the respondent is asked to choose between two different products, and there is a tendency for respondents to choose the first product mentioned,[101] which order of presentation will produce the more accurate response?[102]

To control for order effects, the order of the questions and the order of the response choices in a survey should be rotated,[103] so that, for example, one-third of the respondents have Product A listed first, one-third of the respondents have Product B listed first, and one-third of the respondents have Product C listed first. If the three different orders[104] are distributed randomly among respondents, no response alternative will have an inflated chance of being selected because of its position, and the average of the three will provide a reasonable estimate of response level.[105]

100. This question was accepted by the court in *U-Haul International, Inc. v. Jartran, Inc.*, 522 F. Supp. 1238, 1249 (D. Ariz. 1981), *aff'd*, 681 F.2d 1159 (9th Cir. 1982).

101. Similarly, candidates in the first position on the ballot tend to attract extra votes when the candidates are not well known. Henry M. Bain & Donald S. Hecock, Ballot Position and Voter's Choice: The Arrangement of Names on the Ballot and Its Effect on the Voter (1973).

102. *See* Rust Env't & Infrastructure, Inc. v. Teunissen, 131 F.3d 1210, 1218 (7th Cir. 1997) (survey did not pass muster in part because of failure to incorporate random rotation of corporate names that were the subject of a trademark dispute).

103. *See, e.g.*, Stouffer Foods Corp., 118 F.T.C. 746, No. 9250, 1994 FTC LEXIS 196, at *24–25 (Sept. 26, 1994); *cf.* Winning Ways, Inc. v. Holloway Sportswear, Inc., 913 F. Supp. 1454, 1465–67 (D. Kan. 1996) (failure to rotate the order in which the jackets were shown to the consumers led to reduced weight for the survey).

104. Actually, there are six possible orders of the three alternatives: ABC, ACB, BAC, BCA, CAB, and CBA. Thus, the optimal survey design would allocate equal numbers of respondents to each of the six possible orders.

105. Although rotation is desirable, many surveys are conducted with no attention to this potential bias. Since it is impossible to know in the abstract whether a particular question suffers much, little, or not at all from an order bias, lack of rotation should not preclude reliance on the answer to the question, but it should reduce the weight given to that answer.

F. If the Survey Was Designed to Test a Causal Proposition, Did the Survey Include an Appropriate Control Group or Question?

Most surveys that are designed to provide evidence of trademark infringement or deceptive advertising are not conducted to describe consumer beliefs. Instead, they are intended to show how a trademark or the content of a commercial influences respondents' perceptions or understanding of a product or commercial. Thus, the question is whether the commercial misleads the consumer into thinking that Product A is a superior pain reliever, not whether consumers hold inaccurate beliefs about the product. Yet if consumers already believe, before viewing the commercial, that Product A is a superior pain reliever, a survey that records consumers' impressions after they view the commercial may reflect those preexisting beliefs rather than impressions produced by the commercial.

Surveys that record consumer impressions have a limited ability to answer questions about the origins of those impressions. The difficulty is that the consumer's response to any question on the survey may be the result of information or misinformation from sources other than the trademark the respondent is being shown or the commercial he or she has just watched. In a trademark survey attempting to show secondary meaning, for example, respondents were shown a picture of the stripes used on Mennen stick deodorant and asked, "[W]hich [brand] would you say uses these stripes on their package?"[106] The court recognized that the high percentage of respondents selecting "Mennen" from an array of brand names may have represented "merely a playback of brand share"[107]; that is, respondents asked to give a brand name may guess the one that is most familiar, generally the brand with the largest market share.[108]

Some surveys attempt to reduce the impact of preexisting impressions on respondents' answers by instructing respondents to focus solely on the stimulus as a basis for their answers. Thus, the survey includes a preface (e.g., "based on the commercial you just saw") or directs the respondent's attention to the mark at issue (e.g., "these stripes on the package"). Such efforts are likely to be only partially successful. It is often difficult for respondents to identify accurately the source of their impressions.[109] The more routine the idea being examined in the survey (e.g., that the advertised pain reliever is more effective than others on the

106. Mennen Co. v. Gillette Co., 565 F. Supp. 648, 652 (S.D.N.Y. 1983), *aff'd*, 742 F.2d 1437 (2d Cir. 1984). To demonstrate secondary meaning, "the [c]ourt must determine whether the mark has been so associated in the mind of consumers with the entity that it identifies that the goods sold by that entity are distinguished by the mark or symbol from goods sold by others." *Id.*

107. *Id.*

108. *See also* Upjohn Co. v. American Home Prods. Corp., No. 1-95-CV-237, 1996 U.S. Dist. LEXIS 8049, at *42-44 (W.D. Mich. Apr. 5, 1996).

109. *See* Richard E. Nisbett & Timothy D. Wilson, *Telling More Than We Can Know: Verbal Reports on Mental Processes*, 84 Psychol. Rev. 231 (1977).

market; that the mark belongs to the brand with the largest market share), the more likely it is that the respondent's answer is influenced by preexisting impressions, by expectations about what commercials generally say (e.g., the product being advertised is better than its competitors), or by guessing, rather than by the actual content of the commercial message or trademark being evaluated.

It is possible to adjust many survey designs so that causal inferences about the effect of a trademark or an allegedly deceptive commercial become clear and unambiguous. By adding an appropriate control group, the survey expert can test directly the influence of the stimulus.[110] In the simplest version of a survey experiment, respondents are assigned randomly to one of two conditions.[111] For example, respondents assigned to the experimental condition view an allegedly deceptive commercial, and respondents assigned to the control condition either view a commercial that does not contain the allegedly deceptive material or do not view any commercial.[112] Respondents in both the experimental and control groups answer the same set of questions. The effect of the allegedly deceptive message is evaluated by comparing the responses made by the experimental group members with those of the control group members. If 40% of the respondents in the experimental group responded with the deceptive message (e.g., the advertised product has fewer calories than its competitor), whereas only 8% of the respondents in the control group gave that response, the difference between 40% and 8% (within the limits of sampling error[113]) can be attributed only to the allegedly deceptive commercial. Without the control group, it is not possible to determine how much of the 40% is due to respondents' preexisting beliefs or other background noise (e.g., respondents who misunderstand the question or misstate their responses). Both preexisting beliefs and other background noise should have produced similar response levels in the experimental

110. *See* Shari S. Diamond, *Using Psychology to Control Law: From Deceptive Advertising to Criminal Sentencing*, 13 Law & Hum. Behav. 239, 244–46 (1989); Shari S. Diamond & Linda Dimitropoulos, Deception and Puffery in Advertising: Behavioral Science Implications for Regulation (American Bar Found. Working Paper Series No. 9105, 1994); Jacob Jacoby & Constance Small, *Applied Marketing: The FDA Approach to Defining Misleading Advertising*, 39 J. Marketing 65, 68 (1975). For a more general discussion of the role of control groups, see David H. Kaye & David A. Freedman, Reference Guide on Statistics, § II.A, in this manual.

111. Random assignment should not be confused with random selection. When respondents are assigned randomly to different treatment groups (e.g., respondents in each group watch a different commercial), the procedure ensures that within the limits of sampling error the two groups of respondents will be equivalent except for the different treatments they receive. Respondents selected for a mall intercept study, and not from a probability sample, may be assigned randomly to different treatment groups. Random selection, in contrast, describes the method of selecting a sample of respondents in a probability sample. *See supra* § III.C.

112. This alternative commercial could be a "tombstone" advertisement that includes only the name of the product or a more elaborate commercial that does not include the claim at issue.

113. For a discussion of sampling error, see David H. Kaye & David A. Freedman, Reference Guide on Statistics, § IV, in this manual.

and control groups. In addition, if respondents who viewed the allegedly deceptive commercial respond differently than respondents who viewed the control commercial, the difference cannot be the result of a leading question, because both groups answered the same question. The ability to evaluate the effect of the wording of a particular question makes the control group design particularly useful in assessing responses to closed-ended questions,[114] which may encourage guessing or particular responses. Thus, the focus on the response level in a control group design is not on the absolute response level, but on the difference between the response level of the experimental group and that of the control group.

In designing a control group study, the expert should select a stimulus for the control group that shares as many characteristics with the experimental stimulus as possible, with the key exception of the characteristic whose influence is being assessed. A survey with an imperfect control group generally provides better information than a survey with no control group at all, but the choice of the specific control group requires some care and should influence the weight that the survey receives. For example, a control stimulus should not be less attractive than the experimental stimulus if the survey is designed to measure how familiar the experimental stimulus is to respondents, since attractiveness may affect perceived familiarity.[115] Nor should the control stimulus share with the experimental stimulus the feature whose impact is being assessed. If, for example, the control stimulus in a case of alleged trademark infringement is itself a likely source of consumer confusion, reactions to the experimental and control stimuli may not differ because both cause respondents to express the same level of confusion.[116]

Explicit attention to the value of control groups in trademark and deceptive-advertising litigation is a recent phenomenon, but it is becoming more common.[117] A LEXIS search using *Lanham Act* and *control group* revealed fourteen

114. The Federal Trade Commission has long recognized the need for some kind of control for closed-ended questions, although it has not specified the type of control that is necessary. Stouffer Foods Corp., 118 F.T.C. 746, No. 9250, 1994 FTC LEXIS 196, at *31 (Sept. 26, 1994).

115. *See, e.g.*, Indianapolis Colts, Inc. v. Metropolitan Baltimore Football Club Ltd. Partnership, 34 F.3d 410, 415–16 (7th Cir. 1994) (The court recognized that the name "Baltimore Horses" was less attractive for a sports team than the name "Baltimore Colts."). *See also* Reed-Union Corp. v. Turtle Wax, Inc., 77 F.3d 909, 912 (7th Cir. 1996) (court noted that one expert's choice of a control brand with a well-known corporate source was less appropriate than the opposing expert's choice of a control brand whose name did not indicate a specific corporate source).

116. *See, e.g.*, Western Publ'g Co. v. Publications Int'l, Ltd., No. 94-C-6803, 1995 U.S. Dist. LEXIS 5917, at *45 (N.D. Ill. May 2, 1995) (court noted that the control product was "arguably more infringing than" the defendant's product) (emphasis omitted).

117. *See, e.g.*, American Home Prods. Corp. v. Procter & Gamble Co., 871 F. Supp. 739, 749 (D.N.J. 1994) (discounting survey results based on failure to control for participants' preconceived notions); ConAgra, Inc. v. Geo. A. Hormel & Co., 784 F. Supp. 700, 728 (D. Neb. 1992) ("Since no control was used, the . . . study, standing alone, must be significantly discounted."), *aff'd*, 990 F.2d 368 (8th Cir. 1993).

district court cases in the six years since the first edition of this manual in 1994,[118] five district court cases in the seven years from 1987 to 1993,[119] and only one case before 1987[120] in which surveys with control groups were discussed. Other cases, however, have described or considered surveys using control group designs without labeling the comparison group a control group.[121] Indeed, one reason why cases involving surveys with control groups may be underrepresented in reported cases is that a survey with a control group produces less ambiguous findings, which may lead to a resolution before a preliminary injunction hearing or trial occurs.[122]

Another more common use of control methodology is a control question. Rather than administering a control stimulus to a separate group of respondents,

118. National Football League Properties, Inc. v. Prostyle, Inc., 57 F. Supp. 2d 665 (E.D. Wis. 1999); Nabisco, Inc. v. PF Brands, Inc., 50 F. Supp. 2d 188 (S.D.N.Y. 1999); Proctor & Gamble Co. v. Colgate-Palmolive Co., No. 96 Civ. 9123, 1998 U.S. Dist. LEXIS 17773 (S.D.N.Y. Nov. 5, 1998); Mattel, Inc. v. MCA Records, Inc., 28 F. Supp. 2d 1120 (C.D. Cal. 1998); Westchester Media Co. v. PRL USA Holdings, No. H-97-3278, 1998 U.S. Dist. LEXIS 11737 (S.D. Tex. July 2, 1998); Time Inc. v. Petersen Publ'g Co., 976 F. Supp. 263 (S.D.N.Y. 1997), *aff'd*, 173 F.3d 113 (2d Cir. 1999); Adjusters Int'l, Inc. v. Public Adjusters Int'l, Inc., No. 92-CV-1426, 1996 U.S. Dist. LEXIS 12604 (N.D.N.Y. Aug. 27, 1996); Upjohn Co. v. American Home Prods. Corp., No. 1-95-CV-237, 1996 U.S. Dist. LEXIS 8049 (W.D. Mich. Apr. 5, 1996); Copy Cop, Inc. v. Task Printing, Inc., 908 F. Supp. 37 (D. Mass. 1995); Volkswagen Aktiengesellschaft v. Uptown Motors, No. 91-CIV-3447, 1995 U.S. Dist. LEXIS 13869 (S.D.N.Y. July 13, 1995); Western Publ'g Co. v. Publications Int'l, Ltd., No. 94-C-6803, 1995 U.S. Dist. LEXIS 5917 (N.D. Ill. May 2, 1995); Dogloo, Inc. v. Doskocil Mfg. Co., 893 F. Supp. 911 (C.D. Cal. 1995); Reed-Union Corp. v. Turtle Wax, Inc., 869 F. Supp. 1304 (N.D. Ill. 1994), *aff'd*, 77 F.3d 909 (7th Cir. 1996); Pfizer, Inc. v. Miles, Inc., 868 F. Supp. 437 (D. Conn. 1994).

119. ConAgra, Inc. v. Geo. A. Hormel & Co., 784 F. Supp. 700 (D. Neb. 1992), *aff'd*, 990 F.2d 368 (8th Cir. 1993); Johnson & Johnson-Merck Consumer Pharms. Co. v. Smithkline Beecham Corp., No. 91 Civ. 0960, 1991 U.S. Dist. LEXIS 13689 (S.D.N.Y. Sept. 30, 1991), *aff'd*, 960 F.2d 294 (2d Cir. 1992); Goya Foods, Inc. v. Condal Distribs., Inc., 732 F. Supp. 453 (S.D.N.Y. 1990); Sturm, Ruger & Co. v. Arcadia Mach. & Tool, Inc., No. 85-8459, 1988 U.S. Dist. LEXIS 16451 (C.D. Cal. Nov. 7, 1988); Frisch's Restaurant, Inc. v. Elby's Big Boy, Inc., 661 F. Supp. 971 (S.D. Ohio 1987), *aff'd*, 849 F.2d 1012 (6th Cir. 1988).

120. American Basketball Ass'n v. AMF Voit, Inc., 358 F. Supp. 981 (S.D.N.Y.), *aff'd*, 487 F.2d 1393 (2d Cir. 1973).

121. Indianapolis Colts, Inc. v. Metropolitan Baltimore Football Club Ltd. Partnership, No. 94-727-C, 1994 U.S. Dist. LEXIS 19277, at *10–11 (S.D. Ind. June 27, 1994), *aff'd*, 34 F.3d 410 (7th Cir. 1994). In *Indianapolis Colts*, the district court described a survey conducted by the plaintiff's expert in which half of the interviewees were shown a shirt with the name "Baltimore CFL Colts" on it and half were shown a shirt on which the word "Horses" had been substituted for the word "Colts." *Id.* The court noted that the comparison of reactions to the horse and colt versions of the shirt made it possible "to determine the impact from the use of the word 'Colts.'" *Id.* at *11. *See also* Quality Inns Int'l, Inc. v. McDonald's Corp., 695 F. Supp. 198, 218 (D. Md. 1988) (survey revealed confusion between McDonald's and McSleep, but control survey revealed no confusion between McDonald's and McTavish).

122. The relatively infrequent mention of control groups in surveys discussed in federal cases is not confined to Lanham Act litigation. A LEXIS search using *survey* and *control group* revealed thirty district court cases in the six years from 1994 in which *control group* was used to refer to a methodological feature: the fourteen Lanham Act cases cited *supra* note 118; nine that referred to medical, physiological, or pharmacological experiments; and seven others.

the survey asks all respondents one or more control questions along with the question about the product or service. In a trademark dispute, for example, a survey indicated that 7.2% of respondents believed that "The Mart" and "K-Mart" were owned by the same individuals. The court found no likelihood of confusion based on survey evidence that 5.7% of the respondents also thought that "The Mart" and "King's Department Store" were owned by the same source.[123]

Similarly, a standard technique used to evaluate whether a brand name is generic is to present survey respondents with a series of product or service names and ask them to indicate in each instance whether they believe the name is a brand name or a common name. By showing that 68% of respondents considered Teflon a brand name (a proportion similar to the 75% of respondents who recognized the acknowledged trademark Jell-O as a brand name, and markedly different from the 13% who thought aspirin was a brand name), the makers of Teflon retained their trademark.[124]

Every measure of opinion or belief in a survey reflects some degree of error. Control groups and control questions are the most reliable means for assessing response levels against the baseline level of error associated with a particular question.

G. What Limitations Are Associated with the Mode of Data Collection Used in the Survey?

Three primary methods are used to collect survey data: (1) in-person interviews, (2) telephone surveys, and (3) mail surveys.[125] The choice of a data collection method for a survey should be justified by its strengths and weaknesses.

1. In-person interviews

Although costly, in-person interviews generally are the preferred method of data collection, especially when visual materials must be shown to the respondent under controlled conditions.[126] When the questions are complex and the interviewers are skilled, in-person interviewing provides the maximum oppor-

123. S.S. Kresge Co. v. United Factory Outlet, Inc., 598 F.2d 694, 697 (1st Cir. 1979). Note that the aggregate percentages reported here do not reveal how many of the same respondents were confused by both names, an issue that may be relevant in some situations. *See* Joseph L. Gastwirth, *Reference Guide on Survey Research*, 36 Jurimetrics J. 181, 187–88 (1996) (review essay).

124. E.I. DuPont de Nemours & Co. v. Yoshida Int'l, Inc., 393 F. Supp. 502, 526-27 & n.54 (E.D.N.Y. 1975).

125. Methods also may be combined, as when the telephone is used to "screen" for eligible respondents, who then are invited to participate in an in-person interview.

126. A mail survey also can include limited visual materials but cannot exercise control over when and how the respondent views them.

tunity to clarify or probe. Unlike a mail survey, both in-person and telephone interviews have the capability to implement complex skip sequences (in which the respondent's answer determines which question will be asked next) and the power to control the order in which the respondent answers the questions. As described in section V.A, appropriate interviewer training is necessary if these potential benefits are to be realized. Objections to the use of in-person interviews arise primarily from their high cost or, on occasion, from evidence of inept or biased interviewers.

2. *Telephone surveys*

Telephone surveys offer a comparatively fast and low-cost alternative to in-person surveys and are particularly useful when the population is large and geographically dispersed. Telephone interviews (unless supplemented with mailed materials) can be used only when it is unnecessary to show the respondent any visual materials. Thus, an attorney may present the results of a telephone survey of jury-eligible citizens in a motion for a change of venue in order to provide evidence that community prejudice raises a reasonable suspicion of potential jury bias.[127] Similarly, potential confusion between a restaurant called McBagel's and the McDonald's fast-food chain was established in a telephone survey. Over objections from defendant McBagel's that the survey did not show respondents the defendant's print advertisements, the court found likelihood of confusion based on the survey, noting that "by soliciting audio responses [, the telephone survey] was closely related to the radio advertising involved in the case."[128] In contrast, when words are not sufficient because, for example, the survey is assessing reactions to the trade dress or packaging of a product that is alleged to promote confusion, a telephone survey alone does not offer a suitable vehicle for questioning respondents.[129]

In evaluating the sampling used in a telephone survey, the trier of fact should consider

- (when prospective respondents are not business personnel) whether some form of random-digit dialing[130] was used instead of or to supplement tele-

127. United States v. Partin, 320 F. Supp. 275, 279-80 (E.D. La. 1970). For a discussion of surveys used in motions for change of venue, see Neal Miller, *Facts, Expert Facts, and Statistics: Descriptive and Experimental Research Methods in Litigation, Part II*, 40 Rutgers L. Rev. 467, 470–74 (1988); National Jury Project, Jurywork: Systematic Techniques (Elissa Krauss & Beth Bonora eds., 2d ed. 1983).

128. McDonald's Corp. v. McBagel's, Inc., 649 F. Supp. 1268, 1278 (S.D.N.Y. 1986).

129. Thompson Med. Co. v. Pfizer Inc., 753 F.2d 208 (2d Cir. 1985); Incorporated Publ'g Corp. v. Manhattan Magazine, Inc., 616 F. Supp. 370 (S.D.N.Y. 1985), *aff'd without op.*, 788 F.2d 3 (2d Cir. 1986).

130. Random digit dialing provides coverage of households with both listed and unlisted telephone numbers by generating numbers at random from the frame of all possible telephone numbers. James M. Lepkowski, *Telephone Sampling Methods in the United States*, *in* Telephone Survey Methodology 81–91 (Robert M. Groves et al. eds., 1988).

phone numbers obtained from telephone directories, because up to 65% of all residential telephone numbers in some areas may be unlisted;[131]

- whether the sampling procedures required the interviewer to sample within the household or business, instead of allowing the interviewer to administer the survey to any qualified individual who answered the telephone;[132] and
- whether interviewers were required to call back at several different times of the day and on different days to increase the likelihood of contacting individuals or businesses with different schedules.

Telephone surveys that do not include these procedures may, like other nonprobability sampling approaches, be adequate for providing rough approximations. The vulnerability of the survey depends on the information being gathered. More elaborate procedures for achieving a representative sample of respondents are advisable if the survey instrument requests information that is likely to differ for individuals with listed telephone numbers and individuals with unlisted telephone numbers, or individuals rarely at home and those usually at home.

The report submitted by a survey expert who conducts a telephone survey should specify

1. the procedures that were used to identify potential respondents;
2. the number of telephone numbers for which no contact was made; and
3. the number of contacted potential respondents who refused to participate in the survey.

Computer-assisted telephone interviewing, or CATI, is increasingly used in the administration and data entry of large-scale surveys.[133] A computer protocol may be used to generate telephone numbers and dial them as well as to guide the interviewer. The interviewer conducting a computer-assisted interview (CAI), whether by telephone or in a face-to-face setting, follows the script for the interview generated by the computer program and types in the respondent's answers as the interview proceeds. A primary advantage of CATI and other CAI procedures is that skip patterns can be built into the program so that, for example, if the respondent is asked whether she has ever been the victim of a burglary and she says yes, the computer will generate further questions about

131. In 1992, the percentage of households with unlisted numbers reached 65% in Las Vegas and 62% in Los Angeles. Survey Sampling, Inc., The Frame 2 (March 1993). Studies comparing listed and unlisted household characteristics show some important differences. Lepkowski, *supra* note 130, at 76.

132. This is a consideration only if the survey is sampling individuals. If the survey is seeking information on the household, more than one individual may be able to answer questions on behalf of the household.

133. William L. Nicholls II & R.M. Groves, *The Status of Computer-Assisted Telephone Interviewing*, 2 J. Official Stat. 93 (1986); Mary A. Spaeth, *CATI Facilities at Academic Research Organizations*, 21 Surv. Res. 11 (1990); William E. Saris, Computer-Assisted Interviewing (1991).

the burglary, but if she says no, the program will automatically skip the follow-up burglary questions. Interviewer errors in following the skip patterns are therefore avoided, making CAI procedures particularly valuable when the survey involves complex branching and skip patterns.[134] CAI procedures can also be used to control for order effects by having the program rotate the order in which questions or choices are presented.[135] CAI procedures, however, require additional planning to take advantage of the potential for improvements in data quality. When a CAI protocol is used in a survey presented in litigation, the party offering the survey should supply for inspection the computer program that was used to generate the interviews. Moreover, CAI procedures do not eliminate the need for close monitoring of interviews to ensure that interviewers are accurately reading the questions in the interview protocol and accurately entering the answers that the respondent is giving to those questions.

3. Mail surveys

In general, mail surveys tend to be substantially less costly than both in-person and telephone surveys.[136] Although response rates for mail surveys are often low, researchers have obtained 70% response rates in some general public surveys and response rates of over 90% with certain specialized populations.[137] Procedures that encourage high response rates include multiple mailings, highly personalized communications, prepaid return envelopes and incentives or gratuities, assurances of confidentiality, and first-class outgoing postage.[138]

A mail survey will not produce a high rate of return unless it begins with an accurate and up-to-date list of names and addresses for the target population. Even if the sampling frame is adequate, the sample may be unrepresentative if some individuals are more likely to respond than others. For example, if a survey targets a population that includes individuals with literacy problems, these individuals will tend to be underrepresented. Open-ended questions are generally of limited value on a mail survey because they depend entirely on the respondent to answer fully and do not provide the opportunity to probe or clarify

134. Saris, *supra* note 133, at 20, 27.

135. *See, e.g.*, Intel Corp. v. Advanced Micro Devices, Inc., 756 F. Supp. 1292, 1296–97 (N.D. Cal. 1991) (survey designed to test whether the term *386* as applied to a microprocessor was generic used a CATI protocol that tested reactions to five terms presented in rotated order).

136. Don A. Dillman, *Mail and Other Self-Administered Questionnaires*, *in* Handbook of Survey Research, *supra* note 1, at 359, 373.

137. *Id.* at 360.

138. *See, e.g.*, Richard J. Fox et al., *Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response*, 52 Pub. Opinion Q. 467, 482 (1988); Eleanor Singer et al., *Confidentiality Assurances and Response: A Quantitative Review of the Experimental Literature*, 59 Pub. Opinion Q. 66, 71 (1995); Kenneth D. Hopkins & Arlen R. Gullickson, *Response Rates in Survey Research: A Meta-Analysis of the Effects of Monetary Gratuities*, 61 J. Experimental Educ. 52, 54–57, 59 (1992).

unclear answers. Similarly, if eligibility to answer some questions depends on the respondent's answers to previous questions, such skip sequences may be difficult for some respondents to follow. Finally, because respondents complete mail surveys without supervision, survey personnel are unable to prevent respondents from discussing the questions and answers with others before completing the survey and to control the order in which respondents answer the questions. If it is crucial to have respondents answer questions in a particular order, a mail survey cannot be depended on to provide adequate data.[139]

4. *Internet surveys*

A more recent innovation in survey technology is the Internet survey in which potential respondents are contacted and their responses are collected over the Internet. Internet surveys can substantially reduce the cost of reaching potential respondents and offer some of the advantages of in-person interviews by allowing the computer to show the respondent pictures or lists of response choices in the course of asking the respondent questions. The key limitation is that the respondents accessible over the Internet must fairly represent the relevant population whose responses the survey was designed to measure. Thus, a litigant presenting the results of a web-based survey should be prepared to provide evidence on the potential bias in sampling that the web-based survey is likely to introduce. If the target population consists of computer users, the bias may be minimal. If the target population consists of owners of television sets, significant bias is likely.

V. Surveys Involving Interviewers

A. Were the Interviewers Appropriately Selected and Trained?

A properly defined population or universe, a representative sample, and clear and precise questions can be depended on to produce trustworthy survey results only if "sound interview procedures were followed by competent interviewers."[140] Properly trained interviewers receive detailed written instructions on everything they are to say to respondents, any stimulus materials they are to use in the survey, and how they are to complete the interview form. These instructions should be made available to the opposing party and to the trier of fact. Thus, interviewers should be told, and the interview form on which answers are recorded should indicate, which responses, if any, are to be read to the respondent. Interviewers also should be instructed to record verbatim the respondent's

139. Dillman, *supra* note 136, at 368–70.
140. Toys "R" Us, Inc. v. Canarsie Kiddie Shop, Inc., 559 F. Supp. 1189, 1205 (E.D.N.Y. 1983).

answers, to indicate explicitly whenever they repeat a question to the respondent, and to record any statements they make to or supplementary questions they ask the respondent.

Interviewers require training to ensure that they are able to follow directions in administering the survey questions. Some training in general interviewing techniques is required for most interviews (e.g., practice in pausing to give the respondent enough time to answer and practice in resisting invitations to express the interviewer's beliefs or opinions). Although procedures vary, one treatise recommends at least five hours of training in general interviewing skills and techniques for new interviewers.[141]

The more complicated the survey instrument is, the more training and experience the interviewers require. Thus, if the interview includes a skip pattern (where, e.g., Questions 4–6 are asked only if the respondent says yes to Question 3, and Questions 8–10 are asked only if the respondent says no to Question 3), interviewers must be trained to follow the pattern. Similarly, if the questions require specific probes to clarify ambiguous responses, interviewers must receive instruction on when to use the probes and what to say. In some surveys, the interviewer is responsible for last-stage sampling (i.e., selecting the particular respondents to be interviewed), and training is especially crucial to avoid interviewer bias in selecting respondents who are easiest to approach or easiest to find.

Training and instruction of interviewers should include directions on the circumstances under which interviews are to take place (e.g., question only one respondent at a time out of the hearing of any other respondent). The trustworthiness of a survey is questionable if there is evidence that some interviews were conducted in a setting in which respondents were likely to have been distracted or in which others were present and could overhear. Such evidence of careless administration of the survey was one ground used by a court to reject as inadmissible a survey that purported to demonstrate consumer confusion.[142]

Some compromises may be accepted when surveys must be conducted swiftly. In trademark and deceptive advertising cases, the plaintiff's usual request is for a preliminary injunction, because a delay means irreparable harm. Nonetheless, careful instruction and training of interviewers who administer the survey and complete disclosure of the methods used for instruction and training are crucial elements that, if compromised, seriously undermine the trustworthiness of any survey.

141. Eve Weinberg, *Data Collection: Planning and Management*, *in* Handbook of Survey Research, *supra* note 1, at 329, 332.

142. *Toys "R" Us*, 559 F. Supp. at 1204 (some interviews apparently were conducted in a bowling alley; some interviewees waiting to be interviewed overheard the substance of the interview while they were waiting).

B. *What Did the Interviewers Know About the Survey and Its Sponsorship?*

One way to protect the objectivity of survey administration is to avoid telling interviewers who is sponsoring the survey. Interviewers who know the identity of the survey's sponsor may affect results inadvertently by communicating to respondents their expectations or what they believe are the preferred responses of the survey's sponsor. To ensure objectivity in the administration of the survey, it is standard interview practice to conduct double-blind research whenever possible: both the interviewer and the respondent are blind to the sponsor of the survey and its purpose. Thus, the survey instrument should provide no explicit clues (e.g., a sponsor's letterhead appearing on the survey) and no implicit clues (e.g., reversing the usual order of the yes and no response boxes on the interviewer's form next to a crucial question, thereby potentially increasing the likelihood that *no* will be checked[143]) about the sponsorship of the survey or the expected responses.

Nonetheless, in some surveys (e.g., some government surveys), disclosure of the survey's sponsor to respondents (and thus to interviewers) is required. Such surveys call for an evaluation of the likely biases introduced by interviewer or respondent awareness of the survey's sponsorship. In evaluating the consequences of sponsorship awareness, it is important to consider (1) whether the sponsor has views and expectations that are apparent and (2) whether awareness is confined to the interviewers or involves the respondents. For example, if a survey concerning attitudes toward gun control is sponsored by the National Rifle Association, it is clear that responses opposing gun control are likely to be preferred. In contrast, if the survey on gun control attitudes is sponsored by the Department of Justice, the identity of the sponsor may not suggest the kind of responses the sponsor expects or would find acceptable.[144] When interviewers are well trained, their awareness of sponsorship may be a less serious threat than respondents' awareness. The empirical evidence for the effects of interviewers' prior expectations on respondents' answers generally reveals modest effects when the interviewers are well trained.[145]

143. Centaur Communications, Ltd. v. A/S/M Communications, Inc., 652 F. Supp. 1105, 1111 n.3 (S.D.N.Y.) (pointing out that reversing the usual order of response choices, yes or no, to no or yes may confuse interviewers as well as introduce bias), *aff'd*, 830 F.2d 1217 (2d Cir. 1987).

144. *See, e.g.*, Stanley Presser et al., *Survey Sponsorship, Response Rates, and Response Effects*, 73 Soc. Sci. Q. 699, 701 (1992) (different responses to a university-sponsored telephone survey and a newspaper-sponsored survey for questions concerning attitudes toward the mayoral primary, an issue on which the newspaper had taken a position).

145. *See, e.g.*, Seymour Sudman et al., *Modest Expectations: The Effects of Interviewers' Prior Expectations on Responses*, 6 Soc. Methods & Res. 171, 181 (1977).

C. What Procedures Were Used to Ensure and Determine That the Survey Was Administered to Minimize Error and Bias?

Three methods are used to ensure that the survey instrument was implemented in an unbiased fashion and according to instructions. The first, monitoring the interviews as they occur, is done most easily when telephone surveys are used. A supervisor listens to a sample of interviews for each interviewer. Field settings make monitoring more difficult, but evidence that monitoring has occurred provides an additional indication that the survey has been reliably implemented.

Second, validation of interviews occurs when respondents in a sample are recontacted to ask whether the initial interviews took place and to determine whether the respondents were qualified to participate in the survey. The standard procedure for validation of in-person interviews is to telephone a random sample of about 10% to 15% of the respondents.[146] Some attempts to reach the respondent will be unsuccessful, and occasionally a respondent will deny that the interview took place even though it did. Because the information checked is limited to whether the interview took place and whether the respondent was qualified, this validation procedure does not determine whether the initial interview as a whole was conducted properly. Nonetheless, this standard validation technique warns interviewers that their work is being checked and can detect gross failures in the administration of the survey.

A third way to verify that the interviews were conducted properly is to compare the work done by each individual interviewer. By reviewing the interviews and individual responses recorded by each interviewer, researchers can identify any response patterns or inconsistencies for further investigation.

When a survey is conducted at the request of a party for litigation rather than in the normal course of business, a heightened standard for validation checks may be appropriate. Thus, independent validation of at least 50% of interviews by a third party rather than by the field service that conducted the interviews increases the trustworthiness of the survey results.[147]

146. *See, e.g.*, National Football League Properties, Inc. v. New Jersey Giants, Inc., 637 F. Supp. 507, 515 (D.N.J. 1986); Davis v. Southern Bell Tel. & Tel. Co., No. 89-2839, 1994 U.S. Dist. LEXIS 13257, at *16 (S.D. Fla. Feb. 1, 1994).

147. In *Rust Environment & Infrastructure, Inc. v. Teunissen*, 131 F.3d 1210, 1218 (7th Cir. 1997), the court criticized a survey in part because it "did not comport with accepted practice for independent validation of the results."

VI. Data Entry and Grouping of Responses

A. What Was Done to Ensure That the Data Were Recorded Accurately?

Analyzing the results of a survey requires that the data obtained on each sampled element be recorded, edited, and often coded before the results can be tabulated and processed. Procedures for data entry should include checks for completeness, checks for reliability and accuracy, and rules for resolving inconsistencies. Accurate data entry is maximized when responses are verified by duplicate entry and comparison, and when data entry personnel are unaware of the purposes of the survey.

B. What Was Done to Ensure That the Grouped Data Were Classified Consistently and Accurately?

Coding of answers to open-ended questions requires a detailed set of instructions so that decision standards are clear and responses can be scored consistently and accurately. Two trained coders should independently score the same responses to check for the level of consistency in classifying responses. When the criteria used to categorize verbatim responses are controversial or allegedly inappropriate, those criteria should be sufficiently clear to reveal the source of disagreements. In all cases, the verbatim responses should be available so that they can be recoded using alternative criteria.[148]

148. *See, e.g.*, Coca-Cola Co. v. Tropicana Prods., Inc., 538 F. Supp. 1091, 1094–96 (S.D.N.Y.) (plaintiff's expert stated that respondents' answers to the several open-ended questions revealed that 43% of respondents thought Tropicana was portrayed as fresh squeezed; the court's own tabulation found no more than 15% believed this was true), *rev'd on other grounds*, 690 F.2d 312 (2d Cir. 1982). *See also* McNeilab, Inc. v. American Home Prods. Corp., 501 F. Supp. 517 (S.D.N.Y. 1980); Rock v. Zimmerman, 959 F.2d 1237, 1253 n.9 (3d Cir. 1992) (court found that responses on a change of venue survey incorrectly categorized respondents who believed the defendant was insane as believing he was guilty); Revlon Consumer Prods. Corp. v. Jennifer Leather Broadway, Inc., 858 F. Supp. 1268, 1276 (S.D.N.Y. 1994) (inconsistent scoring and subjective coding led court to find survey so unreliable that it was entitled to no weight), *aff'd*, 57 F.3d 1062 (2d Cir. 1995).

VII. Disclosure and Reporting

A. When Was Information About the Survey Methodology and Results Disclosed?

Objections to the definition of the relevant population, the method of selecting the sample, and the wording of questions generally are raised for the first time when the results of the survey are presented. By that time it is too late to correct methodological deficiencies that could have been addressed in the planning stages of the survey. The plaintiff in a trademark case[149] submitted a set of proposed survey questions to the trial judge, who ruled that the survey results would be admissible at trial while reserving the question of the weight the evidence would be given.[150] The court of appeals called this approach a commendable procedure and suggested that it would have been even more desirable if the parties had "attempt[ed] in good faith to agree upon the questions to be in such a survey."[151]

The *Manual for Complex Litigation, Second,* recommended that parties be required, "before conducting any poll, to provide other parties with an outline of the proposed form and methodology, including the particular questions that will be asked, the introductory statements or instructions that will be given, and other controls to be used in the interrogation process."[152] The parties then were encouraged to attempt to resolve any methodological disagreements before the survey was conducted.[153] Although this passage in the second edition of the manual has been cited with apparent approval,[154] the prior agreement the manual recommends has occurred rarely and the *Manual for Complex Litigation, Third,* recommends, but does not advocate requiring, prior disclosure and discussion of survey plans.[155]

Rule 26 of the Federal Rules of Civil Procedure requires extensive disclosure of the basis of opinions offered by testifying experts. However, these provisions may not produce disclosure of all survey materials, because parties are not obli-

149. Union Carbide Corp. v. Ever-Ready, Inc., 392 F. Supp. 280 (N.D. Ill. 1975), *rev'd*, 531 F.2d 366 (7th Cir.), *cert. denied*, 429 U.S. 830 (1976).

150. Before trial, the presiding judge was appointed to the court of appeals, so the case was tried by another district court judge.

151. *Union Carbide*, 531 F.2d at 386. More recently, the Seventh Circuit recommended the filing of a motion *in limine*, asking the district court to determine the admissibility of a survey based on an examination of the survey questions and the results of a preliminary survey before the party undertakes the expense of conducting the actual survey. Piper Aircraft Corp. v. Wag-Aero, Inc., 741 F.2d 925, 929 (7th Cir. 1984).

152. MCL 2d, *supra* note 15, § 21.484.

153. *Id.*

154. *E.g.*, National Football League Properties, Inc. v. New Jersey Giants, Inc., 637 F. Supp. 507, 514 n.3 (D.N.J. 1986).

155. MCL 3d, *supra* note 15, § 21.493.

gated to disclose information about nontestifying experts. Parties considering whether to commission or use a survey for litigation are not obligated to present a survey that produces unfavorable results. Prior disclosure of a proposed survey instrument places the party that ultimately would prefer not to present the survey in the position of presenting damaging results or leaving the impression that the results are not being presented because they were unfavorable. Anticipating such a situation, parties do not decide whether an expert will testify until after the results of the survey are available.

Nonetheless, courts are in a position to encourage early disclosure and discussion even if they do not lead to agreement between the parties. In *McNeilab, Inc. v. American Home Products Corp.*,[156] Judge William C. Conner encouraged the parties to submit their survey plans for court approval to ensure their evidentiary value; the plaintiff did so and altered its research plan based on Judge Conner's recommendations. Parties can anticipate that changes consistent with a judicial suggestion are likely to increase the weight given to, or at least the prospects of admissibility of, the survey.[157]

B. Does the Survey Report Include Complete and Detailed Information on All Relevant Characteristics?

The completeness of the survey report is one indicator of the trustworthiness of the survey and the professionalism of the expert who is presenting the results of the survey. A survey report generally should provide in detail

1. the purpose of the survey;
2. a definition of the target population and a description of the population that was actually sampled;
3. a description of the sample design, including the method of selecting respondents, the method of interview, the number of callbacks, respondent eligibility or screening criteria, and other pertinent information;
4. a description of the results of sample implementation, including (a) the number of potential respondents contacted, (b) the number not reached, (c) the number of refusals, (d) the number of incomplete interviews or terminations, (e) the number of noneligibles, and (f) the number of completed interviews;
5. the exact wording of the questions used, including a copy of each version of the actual questionnaire, interviewer instructions, and visual exhibits;
6. a description of any special scoring (e.g., grouping of verbatim responses into broader categories);

156. 848 F.2d 34, 36 (2d Cir. 1988) (discussing with approval the actions of the district court).

157. Larry C. Jones, *Developing and Using Survey Evidence in Trademark Litigation*, 19 Memphis St. U. L. Rev. 471, 481 (1989).

7. estimates of the sampling error, where appropriate (i.e., in probability samples);
8. statistical tables clearly labeled and identified as to source of data, including the number of raw cases forming the base for each table, row, or column; and
9. copies of interviewer instructions, validation results, and code books.[158]

A description of the procedures and results of pilot testing is not included on this list. Survey professionals generally do not describe pilot testing in their reports. The Federal Rules of Civil Procedure, however, may require that a testifying expert disclose pilot work that serves as a basis for the expert's opinion. The situation is more complicated when a nontestifying expert conducts the pilot work and the testifying expert learns about the pilot testing only indirectly through the attorney's advice about the relevant issues in the case. Some commentators suggest that attorneys are obligated to disclose such pilot work.[159]

C. In Surveys of Individuals, What Measures Were Taken to Protect the Identities of Individual Respondents?

The respondents questioned in a survey generally do not testify in legal proceedings and are unavailable for cross-examination. Indeed, one of the advantages of a survey is that it avoids a repetitious and unrepresentative parade of witnesses. To verify that interviews occurred with qualified respondents, standard survey practice includes validation procedures,[160] the results of which should be included in the survey report.

Conflicts may arise when an opposing party asks for survey respondents' names and addresses in order to reinterview some respondents. The party introducing the survey or the survey organization that conducted the research generally resists supplying such information.[161] Professional surveyors as a rule guarantee

158. These criteria were adapted from the Council of Am. Survey Res. Orgs., *supra* note 41, § III. B. Failure to supply this information substantially impairs a court's ability to evaluate a survey. *In re* Prudential Ins. Co. of Am. Sales Practices Litig., 962 F. Supp. 450, 532 (D.N.J. 1997) (citing the first edition of this manual). *But see Florida Bar v. Went for It, Inc.*, 515 U.S. 618, 626–28 (1995), in which a majority of the Supreme Court relied on a summary of results prepared by the Florida Bar from a consumer survey purporting to show consumer objections to attorney solicitation by mail. In a strong dissent, Justice Kennedy, joined by three of his colleagues, found the survey inadequate based on the document available to the court, pointing out that the summary included "no actual surveys, few indications of sample size or selection procedures, no explanations of methodology, and no discussion of excluded results . . . no description of the statistical universe or scientific framework that permits any productive use of the information the so-called Summary of Record contains." *Id.* at 640.

159. Yvonne C. Schroeder, *Pretesting Survey Questions*, 11 Am. J. Trial Advoc. 195, 197–201 (1987).

160. *See supra* § V.C.

161. *See, e.g.*, Alpo Petfoods, Inc. v. Ralston Purina Co., 720 F. Supp. 194 (D.D.C. 1989), *aff'd in part & vacated in part*, 913 F.2d 958 (D.C. Cir. 1990).

confidentiality in an effort to increase participation rates and to encourage candid responses. Because failure to extend confidentiality may bias both the willingness of potential respondents to participate in a survey and their responses, the professional standards for survey researchers generally prohibit disclosure of respondents' identities. "The use of survey results in a legal proceeding does not relieve the Survey Research Organization of its ethical obligation to maintain in confidence all Respondent-identifiable information or lessen the importance of Respondent anonymity."[162] Although no surveyor–respondent privilege currently is recognized, the need for surveys and the availability of other means to examine and ensure their trustworthiness argue for deference to legitimate claims for confidentiality in order to avoid seriously compromising the ability of surveys to produce accurate information.[163]

Copies of all questionnaires should be made available upon request so that the opposing party has an opportunity to evaluate the raw data. All identifying information, such as the respondent's name, address, and telephone number, should be removed to ensure respondent confidentiality.

162. Council of Am. Survey Res. Orgs., *supra* note 41, § I.A.3.f. Similar provisions are contained in the By-Laws of the American Association for Public Opinion Research.

163. Litton Indus., Inc., No. 9123, 1979 FTC LEXIS 311, at *13 & n.12 (June 19, 1979) (Order Concerning the Identification of Individual Survey-Respondents with Their Questionnaires) (citing Frederick H. Boness & John F. Cordes, Note, *The Researcher–Subject Relationship: The Need for Protection and a Model Statute*, 62 Geo. L.J. 243, 253 (1973)). *See also* Lampshire v. Procter & Gamble Co., 94 F.R.D. 58, 60 (N.D. Ga. 1982) (defendant denied access to personal identifying information about women involved in studies by the Centers for Disease Control based on Fed. R. Civ. P. 26(c) giving court the authority to enter "any order which justice requires to protect a party or persons from annoyance, embarrassment, oppression, or undue burden or expense.") (citation omitted).

Glossary of Terms

The following terms and definitions were adapted from a variety of sources, including Handbook of Survey Research (Peter H. Rossi et al. eds., 1983); 1 Environmental Protection Agency, Survey Management Handbook (1983); Measurement Errors in Surveys (Paul P. Biemer et al. eds., 1991); William E. Saris, Computer-Assisted Interviewing (1991); Seymour Sudman, Applied Sampling (1976).

branching. A questionnaire structure that uses the answers to earlier questions to determine which set of additional questions should be asked (e.g., citizens who report having served as jurors on a criminal case are asked different questions about their experiences than citizens who report having served as jurors on a civil case).

CAI (computer-assisted interviewing). A method of conducting interviews in which an interviewer asks questions and records the respondent's answer by following a computer-generated protocol.

CATI (computer-assisted telephone interviewing). A method of conducting telephone interviews in which an interviewer asks questions and records the respondent's answer by following a computer-generated protocol.

closed-ended question. A question that provides the respondent with a list of choices and asks the respondent to choose from among them.

cluster sampling. A sampling technique allowing for the selection of sample elements in groups or clusters, rather than on an individual basis; it may significantly reduce field costs and may increase sampling error if elements in the same cluster are more similar to one another than are elements in different clusters.

confidence interval. An indication of the probable range of error associated with a sample value obtained from a probability sample. Also, margin of error.

convenience sample. A sample of elements selected because they were readily available.

double-blind research. Research in which the respondent and the interviewer are not given information that will alert them to the anticipated or preferred pattern of response.

error score. The degree of measurement error in an observed score (see true score).

full-filter question. A question asked of respondents to screen out those who do not have an opinion on the issue under investigation before asking them the question proper.

mall intercept survey. A survey conducted in a mall or shopping center in which potential respondents are approached by a recruiter (intercepted) and invited to participate in the survey.

multistage sampling design. A sampling design in which sampling takes place in several stages, beginning with larger units (e.g., cities) and then proceeding with smaller units (e.g., households or individuals within these units).

nonprobability sample. Any sample that does not qualify as a probability sample.

open-ended question. A question that requires the respondent to formulate his or her own response.

order effect. A tendency of respondents to choose an item based in part on the order in which it appears in the question, questionnaire, or interview (see primacy effect and recency effect); also referred to as a context effect because the context of the question influences the way the respondent perceives and answers it.

parameter. A summary measure of a characteristic of a population (e.g., average age, proportion of households in an area owning a computer). Statistics are estimates of parameters.

pilot test. A small field test replicating the field procedures planned for the full-scale survey; although the terms *pilot test* and *pretest* are sometimes used interchangeably, a pretest tests the questionnaire, whereas a pilot test generally tests proposed collection procedures as well.

population. The totality of elements (objects, individuals, or other social units) that have some common property of interest; the target population is the collection of elements that the researcher would like to study; the survey population is the population that is actually sampled and for which data may be obtained. Also, universe.

population value, population parameter. The actual value of some characteristic in the population (e.g., the average age); the population value is estimated by taking a random sample from the population and computing the corresponding sample value.

pretest. A small preliminary test of a survey questionnaire. See pilot test.

primacy effect. A tendency of respondents to choose early items from a list of choices; the opposite of a recency effect.

probability sample. A type of sample selected so that every element in the population has a known nonzero probability of being included in the sample; a simple random sample is a probability sample.

probe. A follow-up question that an interviewer asks to obtain a more complete answer from a respondent (e.g., "Anything else?" "What kind of medical problem do you mean?").

quasi-filter question. A question that offers a "don't know" or "no opinion" option to respondents as part of a set of response alternatives; used to screen out respondents who may not have an opinion on the issue under investigation.

random sample. See simple random sample.

recency effect. A tendency of respondents to choose later items from a list of choices; the opposite of a primacy effect.

sample. A subset of a population or universe selected so as to yield information about the population as a whole.

sampling error. The estimated size of the difference between the result obtained from a sample study and the result that would be obtained by attempting a complete study of all units in the sampling frame from which the sample was selected in the same manner and with the same care.

sampling frame. The source or sources from which the objects, individuals, or other social units in a sample are drawn.

secondary meaning. A descriptive term that becomes protectable as a trademark if it signifies to the purchasing public that the product comes from a single producer or source.

simple random sample. The most basic type of probability sample; each unit in the population has an equal probability of being in the sample, and all possible samples of a given size are equally likely to be selected.

skip pattern, skip sequence. A sequence of questions in which some should not be asked (should be skipped) based on the respondent's answer to a previous question (e.g., if the respondent indicates that he does not own a car, he should not be asked what brand of car he owns).

stratified sampling. A sampling technique that permits the researcher to subdivide the population into mutually exclusive and exhaustive subpopulations, or strata; within these strata, separate samples are selected; results can be combined to form overall population estimates or used to report separate within-stratum estimates.

survey population. See population.

systematic sampling. A sampling technique that consists of a random starting point and the selection of every *n*th member of the population; it generally produces the same results as simple random sampling.

target population. See population.

trade dress. A distinctive and nonfunctional design of a package or product protected under state unfair competition law and the federal Lanham Act §43(a), 15 U.S.C. §1125(a) (1946) (amended 1992).

true score. The underlying true value, which is unobservable because there is always some error in measurement; the observed score = true score + error score.

universe. See population.

References on Survey Research

William G. Cochran, Sampling Techniques (3d ed. 1977).

Jean M. Converse & Stanley Presser, Survey Questions: Handcrafting the Standardized Questionnaire (1986).

Thomas D. Cook & Donald T. Campbell, Quasi-Experimentation: Design and Analysis Issues for Field Settings (1979).

Shari S. Diamond, *Methods for the Empirical Study of Law, in* Law and the Social Sciences (Leon Lipson & Stanton Wheeler eds., 1986).

Floyd J. Fowler, Survey Research Methods (2d ed. 1984).

Robert M. Groves & Robert L. Kahn, Surveys by Telephone: A National Comparison with Personal Interviews (1979).

Handbook of Survey Research (Peter H. Rossi et al. eds., 1983).

Leslie Kish, Survey Sampling (1965).

Measurement Errors in Surveys (Paul P. Biemer et al. eds., 1991).

Questions About Questions: Inquiries into the Cognitive Bases of Surveys (Judith M. Tanur ed., 1992).

Howard Schuman & Stanley Presser, Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context (1981).

Seymour Sudman, Applied Sampling (1976).

Seymour Sudman & Norman M. Bradburn, Response Effects in Surveys: A Review and Synthesis (1974).

Telephone Survey Methodology (Robert M. Groves et al. eds., 1988).

DEPOSITION EXHIBIT

Reference Guide on Statistics

DAVID H. KAYE AND DAVID A. FREEDMAN

David H. Kaye, M.A., J.D., is Regents' Professor, Arizona State University College of Law, and Fellow, Center for the Study of Law, Science, and Technology, Tempe, Arizona

David A. Freedman, Ph.D., is Professor of Statistics, University of California, Berkeley, California

CONTENTS

I. Introduction, 85
- A. Admissibility and Weight of Statistical Studies, 86
- B. Varieties and Limits of Statistical Expertise, 86
- C. Procedures that Enhance Statistical Testimony, 88
  1. Maintaining Professional Autonomy, 88
  2. Disclosing Other Analyses, 88
  3. Disclosing Data and Analytical Methods Before Trial, 89
  4. Presenting Expert Statistical Testimony, 89

II. How Have the Data Been Collected? 90
- A. Is the Study Properly Designed to Investigate Causation? 90
  1. Types of Studies, 90
  2. Randomized Controlled Experiments, 93
  3. Observational Studies, 94
  4. Can the Results Be Generalized? 96
- B. Descriptive Surveys and Censuses, 98
  1. What Method Is Used to Select the Units? 98
  2. Of the Units Selected, Which Are Measured? 101
- C. Individual Measurements, 102
  1. Is the Measurement Process Reliable? 102
  2. Is the Measurement Process Valid? 103
  3. Are the Measurements Recorded Correctly? 104

III. How Have the Data Been Presented? 104
- A. Are Rates or Percentages Properly Interpreted? 105
  1. Have Appropriate Benchmarks Been Provided? 105
  2. Have the Data-Collection Procedures Changed? 105
  3. Are the Categories Appropriate? 106
  4. How Big Is the Base of a Percentage? 107
  5. What Comparisons Are Made? 107
- B. Is an Appropriate Measure of Association Used? 108
- C. Does a Graph Portray Data Fairly? 110
  1. How Are Trends Displayed? 110
  2. How Are Distributions Displayed? 112
- D. Is an Appropriate Measure Used for the Center of a Distribution? 113

E. Is an Appropriate Measure of Variability Used? 114
IV. What Inferences Can Be Drawn from the Data? 115
A. Estimation, 117
1. What Estimator Should Be Used? 117
2. What Is the Standard Error? The Confidence Interval? 117
B. Significance Levels and Hypothesis Tests, 121
1. What Is the *p*-value? 121
2. Is a Difference Statistically Significant? 123
C. Evaluating Hypothesis Tests, 125
1. What Is the Power of the Test? 125
2. One- or Two-tailed Tests? 126
3. How Many Tests Have Been Performed? 127
4. Tests or Interval Estimates? 128
5. What Are the Rival Hypotheses? 129
D. Posterior Probabilities, 131
V. Correlation and Regression, 133
A. Scatter Diagrams, 134
B. Correlation Coefficients, 135
1. Is the Association Linear? 137
2. Do Outliers Influence the Correlation Coefficient? 137
3. Does a Confounding Variable Influence the Coefficient? 138
C. Regression Lines, 139
1. What Are the Slope and Intercept? 140
2. What Is the Unit of Analysis? 141
D. Statistical Models, 143
1. A Social Science Example, 145
2. Standard Errors, *t*-statistics, and Statistical Significance, 148
3. Summary, 148
Appendix, 151
A. Probability and Statistical Inference, 151
B. Technical Details on the Standard Error, the Normal Curve, and Significance Levels, 153
Glossary of Terms, 160
References on Statistics, 178

I. Introduction

Statistics, broadly defined, is the art and science of gaining information from data. For statistical purposes, data mean observations or measurements, expressed as numbers. A statistic may refer to a particular numerical value, derived from the data. Baseball statistics, for example, is the study of data about the game; a player's batting average is a statistic. The field of statistics includes methods for (1) collecting data, (2) analyzing data, and (3) drawing inferences from data.

Statistical assessments are prominent in many kinds of cases, ranging from antitrust to voting rights. Statistical reasoning can be crucial to the interpretation of psychological tests, toxicological and epidemiological studies, disparate treatment of employees, and DNA fingerprinting; this list could easily be extended.[1]

This reference guide describes the elements of statistical thinking. We hope that the explanations provided will permit judges and lawyers who deal with statistical evidence to understand the terminology, place the evidence in context, appreciate its strengths and weaknesses, and apply legal doctrine governing the use of such evidence. The reference guide is organized as follows:

- Section I provides an overview of the field, discusses the admissibility of statistical studies, and offers some suggestions about procedures that encourage the best use of statistical expertise in litigation.
- Section II addresses data collection. The design of a study is the most important determinant of its quality. The section reviews controlled experiments, observational studies, and surveys, indicating when these designs are likely to give useful data.
- Section III discusses the art of describing and summarizing data. The section considers the mean, median, and standard deviation. These are basic descriptive statistics, and most statistical analyses seen in court use them as building blocks. Section III also discusses trends and associations in data as summarized by graphs, percentages, and tables.
- Section IV describes the logic of statistical inference, emphasizing its foundations and limitations. In particular, this section explains statistical estimation, standard errors, confidence intervals, p-values, and hypothesis tests.
- Section V shows how relationships between two variables can be described by means of scatter diagrams, correlation coefficients, and regression lines. Statisticians often use regression techniques in an attempt to infer causation

1. *See generally* Statistics and the Law (Morris H. DeGroot et al. eds., 1986); Panel on Statistical Assessments as Evidence in the Courts, National Research Council, The Evolving Role of Statistical Assessments as Evidence in the Courts (Stephen E. Fienberg ed., 1989) [hereinafter The Evolving Role of Statistical Assessments as Evidence in the Courts]; Michael O. Finkelstein & Bruce Levin, Statistics for Lawyers (1990); 1 & 2 Joseph L. Gastwirth, Statistical Reasoning in Law and Public Policy (1988); Hans Zeisel & David Kaye, Prove It with Figures: Empirical Methods in Law and Litigation (1997).

from association; section V briefly explains the techniques and some of their limitations.

- An appendix presents certain technical details, and the glossary defines many statistical terms that might be encountered in litigation.

A. Admissibility and Weight of Statistical Studies

Statistical studies suitably designed to address a material issue generally will be admissible under the Federal Rules of Evidence. The hearsay rule rarely is a serious barrier to the presentation of statistical studies, since such studies may be offered to explain the basis for an expert's opinion or may be admissible under the learned treatise exception to the hearsay rule.[2] Likewise, since most statistical methods relied on in court are described in textbooks and journal articles and are capable of producing useful results when carefully and appropriately applied, such methods generally satisfy important aspects of the "scientific knowledge" requirement articulated in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[3] Of course, a particular study may use a method that is entirely appropriate, but so poorly executed that it should be inadmissible under Federal Rules of Evidence 403 and 702.[4] Or, the method may be inappropriate for the problem at hand and thus lacks the "fit" spoken of in *Daubert*.[5] Or, the study may rest on data of the type not reasonably relied on by statisticians or substantive experts, and hence run afoul of Federal Rule of Evidence 703. Often, however, the battle over statistical evidence concerns weight or sufficiency rather than admissibility.

B. Varieties and Limits of Statistical Expertise

For convenience, the field of statistics may be divided into three subfields: probability, theoretical statistics, and applied statistics. Theoretical statistics is the study of the mathematical properties of statistical procedures, such as error rates; probability theory plays a key role in this endeavor. Results may be used by

2. *See generally* 2 McCormick on Evidence §§ 321, 324.3 (John W. Strong ed., 5th ed. 1999). Studies published by government agencies also may be admissible as public records. *Id.* § 296. *See also* United States v. Esquivel, 88 F.3d 722, 727 (9th Cir. 1996) (taking judicial notice of 1990 census data showing the number of Hispanics eligible for jury service).

3. 509 U.S. 579, 589–90 (1993). For a discussion of the implications and scope of *Daubert* generally, see 1 Modern Scientific Evidence: The Law and Science of Expert Testimony § 1–3.0 (David L. Faigman et al. eds., 1997).

4. *See, e.g.*, Sheehan v. Daily Racing Form, Inc., 104 F.3d 940, 942 (7th Cir. 1997) ("failure to exercise the degree of care that a statistician would use in his scientific work, outside of the context of litigation" renders analysis inadmissible under *Daubert*).

5. 509 U.S. at 591; *cf.* People Who Care v. Rockford Bd. of Educ., 111 F.3d 528, 537–38 (7th Cir. 1997) ("a statistical study that fails to correct for salient explanatory variables, or even to make the most elementary comparisons, has no value as causal explanation and is therefore inadmissible in a federal court"); *Sheehan*, 104 F.3d at 942 (holding that expert's "failure to correct for any potential explanatory variables other than age" made the analyst's finding that "there was a significant correlation between age and retention" inadmissible).

applied statisticians who specialize in particular types of data collection, such as survey research, or in particular types of analysis, such as multivariate methods.

Statistical expertise is not confined to those with degrees in statistics. Because statistical reasoning underlies all empirical research, researchers in many fields are exposed to statistical ideas. Experts with advanced degrees in the physical, medical, and social sciences—and some of the humanities—may receive formal training in statistics. Such specializations as biostatistics, epidemiology, econometrics, and psychometrics are primarily statistical, with an emphasis on methods and problems most important to the related substantive discipline.

Individuals who specialize in using statistical methods—and whose professional careers demonstrate this orientation—are most likely to apply appropriate procedures and correctly interpret the results. On the other hand, forensic scientists and technicians often testify to probabilities or statistics derived from studies or databases compiled by others, even though some of these testifying experts lack the training or knowledge required to understand and apply the information. *State v. Garrison*[6] illustrates the problem. In a murder prosecution involving bite-mark evidence, a dentist was allowed to testify that "the probability factor of two sets of teeth being identical in a case similar to this is, approximately, eight in one million," even though "he was unaware of the formula utilized to arrive at that figure other than that it was 'computerized.'"[7]

At the same time, the choice of which data to examine, or how best to model a particular process, could require subject matter expertise that a statistician might lack. Statisticians often advise experts in substantive fields on the procedures for collecting data and often analyze data collected by others. As a result, cases involving statistical evidence often are (or should be) "two-expert" cases of interlocking testimony.[8] A labor economist, for example, may supply a definition of the relevant labor market from which an employer draws its employees, and the statistical expert may contrast the racial makeup of those hired to the racial composition of the labor market. Naturally, the value of the statistical analysis depends on the substantive economic knowledge that informs it.[9]

6. 585 P.2d 563 (Ariz. 1978).

7. *Id.* at 566, 568.

8. Sometimes a single witness presents both the substantive underpinnings and the statistical analysis. Ideally, such a witness has extensive expertise in both fields, although less may suffice to qualify the witness under Fed. R. Evid. 702. In deciding whether a witness who clearly is qualified in one field may testify in a related area, courts should recognize that qualifications in one field do not necessarily imply qualifications in the other.

9. In *Vuyanich v. Republic National Bank*, 505 F. Supp. 224, 319 (N.D. Tex. 1980), *vacated*, 723 F.2d 1195 (5th Cir. 1984), defendant's statistical expert criticized the plaintiffs' statistical model for an implicit, but restrictive, assumption about male and female salaries. The district court trying the case accepted the model because the plaintiffs' expert had a "very strong guess" about the assumption, and her expertise included labor economics as well as statistics. *Id.* It is doubtful, however, that economic knowledge sheds much light on the assumption, and it would have been simple to perform a less restrictive analysis. In this case, the court may have been overly impressed with a single expert who

C. Procedures that Enhance Statistical Testimony

1. Maintaining Professional Autonomy

Ideally, experts who conduct research in the context of litigation should proceed with the same objectivity that they would apply in other contexts. Thus, experts who testify (or who supply results that are used in testimony by others) should be free to do whatever analysis is required to address in a professionally responsible fashion the issues posed by the litigation.[10] Questions about the freedom of inquiry accorded to testifying experts, as well as the scope and depth of their investigations, may reveal some of the limitations to the analysis being presented.

2. Disclosing Other Analyses

Statisticians analyze data using a variety of statistical models and methods. There is much to be said for looking at the data in a variety of ways. To permit a fair evaluation of the analysis that the statistician does settle on, however, the testifying expert may explain the history behind the development of the final statistical approach.[11] Indeed, some commentators have urged that counsel who know of other data sets or analyses that do not support the client's position should reveal this fact to the court, rather than attempt to mislead the court by presenting only favorable results.[12]

combined substantive and statistical expertise. Once the issue is defined by legal and substantive knowledge, some aspects of the statistical analysis will turn on statistical considerations alone, and expertise in another subject will not be pertinent.

10. *See* The Evolving Role of Statistical Assessments as Evidence in the Courts, *supra* note 1, at 164 (recommending that the expert be free to consult with colleagues who have not been retained by any party to the litigation and that the expert receive a letter of engagement providing for these and other safeguards).

11. *See, e.g.*, Mikel Aickin, *Issues and Methods in Discrimination Statistics*, *in* Statistical Methods in Discrimination Litigation 159 (David H. Kaye & Mikel Aickin eds., 1986).

12. The Evolving Role of Statistical Assessments as Evidence in the Courts, *supra* note 1, at 167; *cf.* William W Schwarzer, *In Defense of "Automatic Disclosure in Discovery,"* 27 Ga. L. Rev. 655, 658–59 (1993) ("[T]he lawyer owes a duty to the court to make disclosure of core information."). The Panel on Statistical Assessments as Evidence in the Courts also recommends that "if a party gives statistical data to different experts for competing analyses, that fact be disclosed to the testifying expert, if any." The Evolving Role of Statistical Assessments as Evidence in the Courts, *supra* note 1, at 167. Whether and under what circumstances a particular statistical analysis might be so imbued with counsel's thoughts and theories of the case that it should receive protection as the attorney's work product is an issue beyond the scope of this reference guide.

3. Disclosing Data and Analytical Methods Before Trial

The collection of data often is expensive, and data sets typically contain at least some minor errors or omissions. Careful exploration of alternative modes of analysis also can be expensive and time consuming. To minimize the occurrence of distracting debates at trial over the accuracy of data and the choice of analytical techniques, and to permit informed expert discussions of method, pretrial procedures should be used, particularly with respect to the accuracy and scope of the data, and to discover the methods of analysis. Suggested procedures along these lines are available elsewhere.[13]

4. Presenting Expert Statistical Testimony

The most common format for the presentation of evidence at trial is sequential. The plaintiff's witnesses are called first, one by one, without interruption except for cross-examination, and testimony is in response to specific questions rather than by an extended narration. Although traditional, this structure is not compelled by the Federal Rules of Evidence.[14] Some alternatives have been proposed that might be more effective in cases involving substantial statistical testimony. For example, when the reports of witnesses go together, the judge might allow their presentations to be combined and the witnesses to be questioned as a panel rather than sequentially. More narrative testimony might be allowed, and the expert might be permitted to give a brief tutorial on statistics as a preliminary to some testimony. Instead of allowing the parties to present their experts in the midst of all the other evidence, the judge might call for the experts for opposing sides to testify at about the same time. Some courts, particularly in bench trials, may have both experts placed under oath and, in effect, permit them to engage in a dialogue. In such a format, experts are able to say whether they agree or disagree on specific issues. The judge and counsel can interject questions. Such practices may improve the judge's understanding and reduce the tensions associated with the experts' adversarial role.[15]

13. *See* The Special Comm. on Empirical Data in Legal Decision Making, Recommendations on Pretrial Proceedings in Cases with Voluminous Data, *reprinted in* The Evolving Role of Statistical Assessments as Evidence in the Courts, *supra* note 1, app. F. *See also* David H. Kaye, *Improving Legal Statistics*, 24 L. & Soc'y Rev. 1255 (1990).

14. *See* Fed. R. Evid. 611.

15. The Evolving Role of Statistical Assessments as Evidence in the Courts, *supra* note 1, at 174.

II. How Have the Data Been Collected?

An analysis is only as good as the data on which it rests.[16] To a large extent, the design of a study determines the quality of the data. Therefore, the proper interpretation of data and their implications begins with an understanding of study design, and different designs help answer different questions. In many cases, statistics are introduced to show causation. Would additional information in a securities prospectus disclosure have caused potential investors to behave in some other way? Does capital punishment deter crime? Do food additives cause cancer? The design of studies intended to prove causation is the first and perhaps the most important topic of this section.

Another issue is the use of sample data to characterize a population: the population is the whole class of units that are of interest; the sample is a set of units chosen for detailed study. Inferences from the part to the whole are justified only when the sample is representative, and that is the second topic of this section.

Finally, it is important to verify the accuracy of the data collection. Errors can arise in the process of making and recording measurements on individual units. This aspect of data quality is the third topic in this section.

A. Is the Study Properly Designed to Investigate Causation?

1. Types of Studies

When causation is at issue, advocates have relied on three major types of information: anecdotal evidence, observational studies, and controlled experiments.[17] As we shall see, anecdotal reports can provide some information, but they are

16. For introductory treatments of data collection, see, e.g., David Freedman et al., Statistics (3d ed. 1998); Darrell Huff, How to Lie with Statistics (1954); David S. Moore, Statistics: Concepts and Controversies (3d ed. 1991); Hans Zeisel, Say It with Figures (6th ed. 1985); and Zeisel & Kaye, *supra* note 1.

17. When relevant studies exist before the commencement of the litigation, it becomes the task of the lawyer and appropriate experts to explain this research to the court. Examples of such "off-the-shelf" research are experiments pinpointing conditions under which eyewitnesses tend to err in identifying criminals and studies of how sex stereotyping affects perceptions of women in the workplace. *See, e.g.*, State v. Chapple, 660 P.2d 1208, 1223–24 (Ariz. 1983) (reversing a conviction for excluding expert testimony about scientific research on eyewitness accuracy); Price Waterhouse v. Hopkins, 490 U.S. 228, 235 (1989). Some psychologists have questioned the applicability of these experiments to litigation. *See, e.g.*, Gerald V. Barrett & Scott B. Morris, *The American Psychological Association's Amicus Curiae Brief in* Price Waterhouse v. Hopkins*: The Values of Science Versus the Values of the Law*, 17 Law & Hum. Behav. 201 (1993). For a rejoinder, see Susan T. Fiske et al., *What Constitutes a Scientific Review?: A Majority Retort to Barrett and Morris*, 17 Law & Hum. Behav. 217 (1993).

If no preexisting studies are available, a case-specific one may be devised. *E.g.*, United States v. Youritan Constr. Co., 370 F. Supp. 643, 647 (N.D. Cal. 1973) (investigating racial discrimination in the rental-housing market by using "testers"—who should differ only in their race—to rent a property),

more useful as a stimulus for further inquiry than as a basis for establishing association. Observational studies can establish that one factor is associated with another, but considerable analysis may be necessary to bridge the gap from association to causation.[18] Controlled experiments are ideal for ascertaining causation, but they can be difficult to undertake.

"Anecdotal evidence" means reports of one kind of event following another. Typically, the reports are obtained haphazardly or selectively, and the logic of "post hoc, ergo propter hoc" does not suffice to demonstrate that the first event causes the second. Consequently, while anecdotal evidence can be suggestive,[19] it can also be quite misleading.[20] For instance, some children who live near power lines develop leukemia; but does exposure to electrical and magnetic fields cause this disease? The anecdotal evidence is not compelling because leukemia also occurs among children who have minimal exposure to such fields.[21] It is necessary to compare disease rates among those who are exposed and those who are not. If exposure causes the disease, the rate should be higher among the exposed, lower among the unexposed. Of course, the two groups may differ in crucial ways other than the exposure. For example, children who live near power

aff'd in part, 509 F.2d 623 (9th Cir. 1975). For a critical review of studies using testers, see James J. Heckman & Peter Siegelman, *The Urban Institute Audit Studies: Their Methods and Findings*, *in* Clear and Convincing Evidence: Measurement of Discrimination in America 187 (Michael Fix & Raymond J. Struyk eds., 1993) (including commentary).

18. For example, smokers have higher rates of lung cancer than nonsmokers; thus smoking and lung cancer are associated.

19. In medicine, evidence from clinical practice is often the starting point for the demonstration of a causal effect. One famous example involves exposure of mothers to German measles during pregnancy, followed by blindness in their babies. N. McAlister Gregg, *Congenital Cataract Following German Measles in the Mother*, 3 Transactions Ophthalmological Soc'y Austl. 35 (1941), *reprinted in* The Challenge of Epidemiology 426 (Carol Buck et al. eds., 1988).

20. Indeed, some courts have suggested that attempts to infer causation from anecdotal reports are inadmissible as unsound methodology under *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993). *See, e.g.*, Haggerty v. Upjohn Co., 950 F. Supp. 1160, 1163–64 (S.D. Fla. 1996) (holding that reports to the Food and Drug Administration of "adverse medical events" involving the drug Halcion and "anecdotal case reports appearing in medical literature . . . can be used to generate hypotheses about causation, but not causation conclusions" because "scientifically valid cause and effect determinations depend on controlled clinical trials and epidemiological studies"); Cartwright v. Home Depot U.S.A., Inc., 936 F. Supp. 900, 905 (M.D. Fla. 1996) (excluding an expert's opinion that latex paint caused plaintiff's asthma, in part because "case reports . . . are no substitute for a scientifically designed and conducted inquiry").

21. *See* Committee on the Possible Effects of Electromagnetic Fields on Biologic Sys., National Research Council, Possible Health Effects of Exposure to Residential Electric and Magnetic Fields (1997); Zeisel & Kaye, *supra* note 1, at 66–67. There are serious problems in measuring exposure to electromagnetic fields, and results are somewhat inconsistent from one study to another. For such reasons, the epidemiologic evidence for an effect on health is quite inconclusive. *Id.*; Martha S. Linet et al., *Residential Exposure to Magnetic Fields and Acute Lymphoblastic Leukemia in Children*, 337 New Eng. J. Med. 1 (1997); Edward W. Campion, *Power Lines, Cancer, and Fear*, 337 New Eng. J. Med. 44 (1997) (editorial); Gary Taubes, *Magnetic Field-Cancer Link: Will It Rest in Peace?*, 277 Science 29 (1997) (quoting various epidemiologists).

lines could come from poorer families and be exposed to other environmental hazards. These differences could create the appearance of a cause-and-effect relationship, or they can mask a real relationship. Cause-and-effect relationships often are quite subtle, and carefully designed studies are needed to draw valid conclusions.[22]

Typically, a well-designed study will compare outcomes for subjects who are exposed to some factor—the treatment group—and other subjects who are not so exposed—the control group. A distinction must then be made between controlled experiments and observational studies. In a controlled experiment, the investigators decide which subjects are exposed to the factor of interest and which subjects go into the control group. In most observational studies, the subjects themselves choose their exposures. Because of this self-selection, the treatment and control groups are likely to differ with respect to important factors other than the one of primary interest.[23] (These other factors are called confounding variables or lurking variables.[24]) With studies on the health effects of power lines, family background is a possible confounder; so is exposure to other hazards.[25]

22. Here is a classic example from epidemiology. At one time, it was thought that lung cancer was caused by fumes from tarring the roads, because many lung cancer patients lived near roads that had recently been paved. This is anecdotal evidence. But the logic is quite incomplete, because many people without lung cancer were exposed to asphalt fumes. A comparison of rates is needed. Careful study showed that lung cancer patients had similar rates of exposure to tar fumes as other people; the real difference was in exposure to cigarette smoke. Richard Doll & A. Bradford Hill, *A Study of the Aetiology of Carcinoma of the Lung*, 2 Brit. Med. J. 1271 (1952).

23. For present purposes, a variable is a numerical characteristic of units in a study. For instance, in a survey of people, the unit of analysis is the person, and variables might include income (in dollars per year) and educational level (years of schooling completed). In a study of school districts, the unit of analysis is the district, and variables might include average family income of residents and average test scores of students. When investigating a possible cause-and-effect relationship, the variable that characterizes the effect is called the dependent variable, since it may depend on the causes; dependent variables also are called response variables. In contrast, the variables that represent the causes are called independent variables; independent variables also are called factors or explanatory variables.

24. A confounding variable is correlated with the independent variables and with the dependent variable. If the units being studied differ on the independent variables, they are also likely to differ on the confounder. Therefore, the confounder—not the independent variables—could be responsible for differences seen on the dependent variable.

25. Confounding is a problem even in careful epidemiologic studies. For example, women with herpes are more likely to develop cervical cancer than women who have not been exposed to the virus. It was concluded that herpes caused cancer; in other words, the association was thought to be causal. Later research suggests that herpes is only a marker of sexual activity. Women who have had multiple sexual partners are more likely to be exposed not only to herpes but also to human papilloma virus. Certain strains of papilloma virus seem to cause cervical cancer, while herpes does not. Apparently, the association between herpes and cervical cancer is not causal but is due to the effect of other variables. *See* Viral Etiology of Cervical Cancer (Richard Peto & Harald zur Hausen eds., 1986); The Epidemiology of Cervical Cancer and Human Papillomavirus (N. Muñoz et al. eds. 1992). For additional examples and discussion, see Freedman et al., *supra* note 16, at 12–27, 150–52; David Freedman, *From Association to Causation: Some Remarks on the History of Statistics*, 14 Stat. Sci. 243 (1999).

2. *Randomized Controlled Experiments*

In randomized controlled experiments, investigators assign subjects to treatment or control groups at random. The groups are therefore likely to be quite comparable—except for the treatment. Choosing at random tends to balance the groups with respect to possible confounders, and the effect of remaining imbalances can be assessed by statistical techniques.[26] Consequently, inferences based on well-executed randomized experiments are more secure than inferences based on observational studies.[27]

The following illustration brings together the points made thus far. Many doctors think that taking aspirin helps prevent heart attacks, but there is some controversy. Most people who take aspirin do not have heart attacks; this is anecdotal evidence for the protective effect, but proves very little. After all, most people do not suffer heart attacks—whether or not they take aspirin regularly. A careful study must compare heart attack rates for two groups: persons who take aspirin (the treatment group) and persons who do not (the controls). An observational study would be easy to do, but then the aspirin-takers are likely to be different from the controls. If, for instance, the controls are healthier to begin with, the study would be biased against the drug. Randomized experiments with aspirin are harder to do, but they provide much better evidence. It is the experiments that demonstrate a protective effect.

To summarize: First, outcome figures from a treatment group without a control group generally reveal very little and can be misleading. Comparisons are essential. Second, if the control group was obtained through random assignment before treatment, a difference in the outcomes between treatment and control groups may be accepted, within the limits of statistical error, as the true measure of the treatment effect.[28] However, if the control group was created in any

26. *See infra* § IV.

27. Experiments, however, are often impractical, as in the power-line example. Even when randomized controlled experiments are feasible, true randomization can be difficult to achieve. *See, e.g.*, Kenneth F. Schulz, *Subverting Randomization in Controlled Trials*, 274 JAMA 1456 (1995); Rachel Nowak, *Problems in Clinical Trials Go Far Beyond Misconduct*, 264 Science 1538 (1994). For statistical purposes, randomization should be accomplished using some definite, objective method (like a random number generator on a computer); haphazard assignment may not be sufficient.

28. Of course, the possibility that the two groups will not be comparable in some unrecognized way can never be eliminated. Random assignment, however, allows the researcher to compute the probability of seeing a large difference in the outcomes when the treatment actually has no effect. When this probability is small, the difference in the response is said to be "statistically significant." *See infra* § IV.B.2. Randomization of subjects to treatment or control groups puts statistical tests of significance on a secure footing. Freedman et al., *supra* note 16, at 503–24, 547–78.

Even more important, randomization also ensures that the assignment of subjects to treatment and control groups is free from conscious or unconscious manipulation by investigators or subjects. Randomization may not be the only way to ensure such protection, but "it is the simplest and best understood way to certify that one has done so." Philip W. Lavori et al., *Designs for Experiments—Parallel Comparisons of Treatment*, *in* Medical Uses of Statistics 61, 66 (John C. Bailar III & Frederick Mosteller

other way, differences in the groups that existed before treatment may contribute to differences in the outcomes, or mask differences that otherwise would be observed. Thus, observational studies succeed to the extent that their treatment and control groups are comparable—apart from the treatment.

3. Observational Studies

The bulk of the statistical studies seen in court are observational, not experimental. Take the question of whether capital punishment deters murder. To do a randomized controlled experiment, people would have to be assigned randomly to a control group and a treatment group. The controls would know that they could not receive the death penalty for murder, while those in the treatment group would know they could be executed. The rate of subsequent murders by the subjects in these groups would be observed. Such an experiment is unacceptable—politically, ethically, and legally.[29]

Nevertheless, many studies of the deterrent effect of the death penalty have been conducted, all observational, and some have attracted judicial attention.[30] Researchers have catalogued differences in the incidence of murder in states with and without the death penalty, and they have analyzed changes in homicide rates and execution rates over the years. In such observational studies, investigators may speak of control groups (such as the states without capital punishment) and of controlling for potentially confounding variables (e.g., worsening economic conditions).[31] However, association is not causation, and the causal inferences that can be drawn from such analyses rest on a less secure foundation than that provided by a randomized controlled experiment.[32]

eds., 2d ed. 1992). To avoid ambiguity, the researcher should be explicit "about how the randomization was done (e.g., table of random numbers) and executed (e.g., by sealed envelopes prepared in advance)." *Id. See also* Colin Begg et al., *Improving the Quality of Reporting of Randomized Controlled Trials: The CONSORT Statement*, 276 JAMA 637 (1996).

29. *Cf.* Experimentation in the Law: Report of the Federal Judicial Center Advisory Committee on Experimentation in the Law (Federal Judicial Center 1981) [hereinafter Experimentation in the Law] (study of ethical issues raised by controlled experimentation in the evaluation of innovations in the justice system).

30. *See generally* Hans Zeisel, *The Deterrent Effect of the Death Penalty: Facts v. Faith*, 1976 Sup. Ct. Rev. 317.

31. A procedure often used to control for confounding in observational studies is regression analysis. The underlying logic is described *infra* § V.D and in Daniel L. Rubinfeld, Reference Guide on Multiple Regression, § II, in this manual. The early enthusiasm for using multiple regression analysis to study the death penalty was not shared by reviewers. *Compare* Isaac Ehrlich, *The Deterrent Effect of Capital Punishment: A Question of Life and Death*, 65 Am. Econ. Rev. 397 (1975), *with, e.g.*, Lawrence R. Klein et al., *The Deterrent Effect of Capital Punishment: An Assessment of the Estimates*, *in* Panel on Research on Deterrent and Incapacitative Effects, National Research Council, Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates 336 (Alfred Blumstein et al. eds., 1978); Edward Leamer, *Let's Take the Con Out of Econometrics*, 73 Am. Econ. Rev. 31 (1983).

32. *See, e.g.*, Experimentation in the Law, *supra* note 29, at 18:

> [G]roups selected without randomization will [almost] always differ in some systematic way other than exposure to the experimental program. Statistical techniques can eliminate chance as a feasible explanation for the

Of course, observational studies can be very useful. The evidence that smoking causes lung cancer in humans, although largely observational, is compelling. In general, observational studies provide powerful evidence in the following circumstances:

- The association is seen in studies of different types among different groups. This reduces the chance that the observed association is due to a defect in one type of study or a peculiarity in one group of subjects.
- The association holds when the effects of plausible confounding variables are taken into account by appropriate statistical techniques, such as comparing smaller groups that are relatively homogeneous with respect to the factor.[33]
- There is a plausible explanation for the effect of the independent variables; thus, the causal link does not depend on the observed association alone. Other explanations linking the response to confounding variables should be less plausible.[34]

When these criteria are not fulfilled, observational studies may produce legitimate disagreement among experts, and there is no mechanical procedure for ascertaining who is correct. In the end, deciding whether associations are causal is not a matter of statistics, but a matter of good scientific judgment, and the questions that should be asked with respect to data offered on the question of causation can be summarized as follows:

- Was there a control group? If not, the study has little to say about causation.
- If there was a control group, how were subjects assigned to treatment or control: through a process under the control of the investigator (a controlled experiment) or a process outside the control of the investigator (an observational study)?

differences, . . . [b]ut without randomization there are no certain methods for determining that observed differences between groups are not related to the preexisting, systematic difference. . . . [C]omparison between systematically different groups will yield ambiguous implications whenever the systematic difference affords a plausible explanation for apparent effects of the experimental program.

33. The idea is to control for the influence of a confounder by making comparisons separately within groups for which the confounding variable is nearly constant and therefore has little influence over the variables of primary interest. For example, smokers are more likely to get lung cancer than nonsmokers. Age, gender, social class, and region of residence are all confounders, but controlling for such variables does not really change the relationship between smoking and cancer rates. Furthermore, many different studies—of different types and on different populations—confirm the causal link. That is why most experts believe that smoking causes lung cancer and many other diseases. For a review of the literature, see 38 International Agency for Research on Cancer (IARC), World Health Org., IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans: Tobacco Smoking (1986).

34. A. Bradford Hill, *The Environment and Disease: Association or Causation?*, 58 Proc. Royal Soc'y Med. 295 (1965); Alfred S. Evans, Causation and Disease: A Chronological Journey 187 (1993).

- If the study was a controlled experiment, was the assignment made using a chance mechanism (randomization), or did it depend on the judgment of the investigator?
- If the data came from an observational study or a nonrandomized controlled experiment, how did the subjects come to be in treatment or in control groups? Are the groups comparable? What factors are confounded with treatment? What adjustments were made to take care of confounding? Were they sensible?[35]

4. Can the Results Be Generalized?

Any study must be conducted on a certain group of subjects, at certain times and places, using certain treatments. With respect to these subjects, the study may be persuasive. There may be adequate control over confounding variables, and there may be an unequivocally large difference between the treatment and control groups. If so, the study's internal validity will not be disputed: for the subjects in the study, the treatment had an effect. But an issue of external validity remains. To extrapolate from the conditions of a study to more general circumstances always raises questions. For example, studies suggest that definitions of insanity given to jurors influence decisions in cases of incest;[36] would the definitions have a similar effect in cases of murder? Other studies indicate that recidivism rates for ex-convicts are not affected by temporary financial support after release.[37] Would the same results be obtained with different conditions in the labor market?

Confidence in the appropriateness of an extrapolation cannot come from the experiment itself.[38] It must come from knowledge about which outside factors

35. These questions are adapted from Freedman et al., *supra* note 16, at 28. For discussions of the admissibility or weight of studies that overlook obvious possible confounders, see *People Who Care v. Rockford Board of Education*, 111 F.3d 528, 537–38 (7th Cir. 1997) ("The social scientific literature on educational achievement identifies a number of other variables besides poverty and discrimination that explain differences in scholastic achievement, such as the educational attainments of the student's parents and the extent of their involvement in their children's schooling. . . . These variables cannot be assumed to be either randomly distributed across the different racial and ethnic groups in Rockford or perfectly correlated with poverty. . . ."); cases cited *supra* note 5 and *infra* note 230.

36. *See* Rita James Simon, The Jury and the Defense of Insanity 58–59 (1967).

37. For an experiment on income support and recidivism, see Peter H. Rossi et al., Money, Work, and Crime: Experimental Evidence (1980). The interpretation of the data has proved controversial. *See* Hans Zeisel, *Disagreement over the Evaluation of a Controlled Experiment*, 88 Am. J. Soc. 378 (1982) (with commentary).

38. Suppose an epidemiologic study is conducted on the relationship between a toxic substance and a disease. The rate of occurrence of the disease in a group of persons exposed to the substance is compared to the rate in a control group, and the rate in the exposed group turns out to be more than double the rate in the control group. (More technically, the relative risk exceeds two.) Do these data imply that a plaintiff who was exposed to the toxic substance and contracted the disease probably would not have contracted the disease but for the exposure? If we assume that the substance causes the disease and all confounding has been properly accounted for (a judgment that might not be easy to defend),

would or would not affect the outcome.[39] Sometimes, several experiments or other studies, each having different limitations, all point in the same direction. This is the case, for example, with eight studies indicating that jurors who approve of the death penalty are more likely to convict in a capital case.[40] Such convergent results strongly suggest the validity of the generalization.

then we can conclude that over half the cases of disease in the exposed group would not be there but for the exposure. Applying this arithmetic to a specific person, however, is problematic. For instance, the relative risk is an average over all the subjects included in the study. The exposures and susceptibilities almost certainly are not uniform, and the plaintiff's exposure and susceptibility cannot be known from the study. Nevertheless, several courts and commentators have stated that a relative risk of more than two demonstrates specific causation, or, conversely, that a relative risk of two or less precludes a finding of specific causation. *E.g.*, DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 958–59 (3d Cir. 1990); Marder v. G.D. Searle & Co., 630 F. Supp. 1087, 1092 (D. Md. 1986) ("a two-fold increased risk is . . . the equivalent of the required legal burden of proof—a showing of causation by the preponderance of the evidence or, in other words, a probability of greater than 50%"), *aff'd sub nom.* Wheelahan v. G.D. Searle & Co., 814 F.2d 655 (4th Cir. 1987); Bert Black & David E. Lilienfeld, *Epidemiologic Proof in Toxic Tort Litigation*, 52 Fordham L. Rev. 732, 769 (1984); Michael D. Green et al., Reference Guide on Epidemiology, § VII, in this manual. A few commentators have sharply criticized this reasoning. Steven E. Fienberg et al., *Understanding and Evaluating Statistical Evidence in Litigation*, 36 Jurimetrics J. 1, 9 (1995); Diana B. Petitti, *Reference Guide on Epidemiology*, 36 Jurimetrics J. 159, 168 (1996) (review essay); D.A. Freedman & Philip B. Stark, *The Swine Flu Vaccine and Guillain-Barré Syndrome: A Case Study in Relative Risk and Specific Causation*, 23 Evaluation Rev. 619 (1999); James Robins & Sander Greenland, *The Probability of Causation Under a Stochastic Model for Individual Risk*, 45 Biometrics 1125, 1126 (1989); Melissa Moore Thompson, Comment, *Causal Inference in Epidemiology: Implications for Toxic Tort Litigation*, 71 N.C. L. Rev. 247 (1992).

39. Such judgments are easiest in the physical and life sciences, but even here, there are problems. For example, it may be difficult to infer human reactions to substances that affect animals. First, there are often inconsistencies across test species: A chemical may be carcinogenic in mice but not in rats. Extrapolation from rodents to humans is even more problematic. Second, to get measurable effects in animal experiments, chemicals are administered at very high doses. Results are extrapolated—using mathematical models—to the very low doses of concern in humans. However, there are many dose-response models to use and few grounds for choosing among them. Generally, different models produce radically different estimates of the "virtually safe dose" in humans. David A. Freedman & Hans Zeisel, *From Mouse to Man: The Quantitative Assessment of Cancer Risks*, 3 Stat. Sci. 3 (1988). For these reasons, many experts—and some courts in toxic tort cases—have concluded that evidence from animal experiments is generally insufficient by itself to establish causation. *See generally* Bruce N. Ames et al., *The Causes and Prevention of Cancer*, 92 Proc. Nat'l Acad. Sci. USA 5258 (1995); Susan R. Poulter, *Science and Toxic Torts: Is There a Rational Solution to the Problem of Causation?*, 7 High Tech. L.J. 189 (1993) (epidemiological evidence on humans is needed). *See also* Committee on Comparative Toxicity of Naturally Occurring Carcinogens, National Research Council, Carcinogens and Anticarcinogens in the Human Diet: A Comparison of Naturally Occurring and Synthetic Substances (1996); Committee on Risk Assessment of Hazardous Air Pollutants, National Research Council, Science and Judgment in Risk Assessment 59 (1994) ("There are reasons based on both biologic principles and empirical observations to support the hypothesis that many forms of biologic responses, including toxic responses, can be extrapolated across mammalian species, including *Homo sapiens*, but the scientific basis of such extrapolation is not established with sufficient rigor to allow broad and definitive generalizations to be made.").

40. Phoebe C. Ellsworth, *Some Steps Between Attitudes and Verdicts*, *in* Inside the Juror 42, 46 (Reid Hastie ed., 1993). Nevertheless, in *Lockhart v. McCree*, 476 U.S. 162 (1986), the Supreme Court held that the exclusion of opponents of the death penalty in the guilt phase of a capital trial does not violate the constitutional requirement of an impartial jury.

B. Descriptive Surveys and Censuses

Having discussed the statistical logic of studies to investigate causation, we now turn to a second topic—sampling, that is, choosing units for study. A census tries to measure some characteristic of every unit in a population of individuals or objects. A survey, alternatively, measures characteristics only in part of a population. The accuracy of the information collected in a census or survey depends on how the units are selected, which units are actually measured, and how the measurements are made.[41]

1. What Method Is Used to Select the Units?

By definition, a census seeks to measure some characteristic of every unit in a whole population. It may fall short of this goal, in which case the question must be asked whether the missing data are likely to differ in some systematic way from the data that are collected. The U.S. Bureau of the Census estimates that the past six censuses failed to count everyone, and there is evidence that the undercount is greater in certain subgroups of the population.[42] Supplemental studies may enable statisticians to adjust for such omissions, but the adjustments may rest on uncertain assumptions.[43]

The methodological framework of a scientific survey is more complicated than that of a census. In surveys that use probability sampling methods, a sampling frame (that is, an explicit list of units in the population) is created. Individual units then are selected by a kind of lottery procedure, and measurements are made on these sampled units. For example, a defendant charged with a notorious crime who seeks a change of venue may commission an opinion poll to show that popular opinion is so adverse and deep-rooted that it will be difficult

41. For more extended treatment of these issues, see Shari Seidman Diamond, Reference Guide on Survey Research, § III, in this manual.

42. *See generally* Harvey M. Choldin, Looking for the Last Percent: The Controversy Over Census Undercounts 42–43 (1994).

43. For conflicting views on proposed adjustments to the 1990 census, see the exchanges of papers at 9 Stat. Sci. 458 (1994), 18 Surv. Methodology No. 1 (1992), 88 J. Am. Stat. Ass'n 1044 (1993), and 34 Jurimetrics J. 65 (1993). In *Wisconsin v. City of New York*, 517 U.S. 1 (1996), the Supreme Court resolved the conflict among the circuits over the legal standard governing claims that adjustment is compelled by statute or the Constitution. The Court unanimously determined that the exacting requirements of the equal protection clause, as explicated in congressional redistricting and state reapportionment cases, do not "translate into a requirement that the Federal Government conduct a census that is as accurate as possible" and do not provide any basis for "preferring numerical accuracy to distributive accuracy." *Id.* at 17, 18. The Court therefore applied a much less demanding standard to the Secretary's decision. Concluding that the government had shown "a reasonable relationship" between the decision not to make post hoc adjustments and "the accomplishment of an actual enumeration of the population, keeping in mind the constitutional purpose of the census . . . to determine the apportionment of the Representatives among the States," the Court held that the decision satisfied the Constitution. Indeed, having rejected the argument that the Constitution compelled statistical adjustment, the Court noted that the Constitution might prohibit such adjustment. *Id.* at 19 n.9, 20.

to impanel an unbiased jury. The population consists of all persons in the jurisdiction who might be called for jury duty. A sampling frame here could be the list of these persons as maintained by appropriate officials.[44] In this case, the fit between the sampling frame and the population would be excellent.[45]

In other situations, the sampling frame may cover less of the population. In an obscenity case, for example, the defendant's opinion poll about community standards[46] should identify the population as all adults in the legally relevant community, but obtaining a full list of all such people may not be possible. If names from a telephone directory are used, people with unlisted numbers are excluded from the sampling frame. If these people, as a group, hold different opinions from those included in the sampling frame, the poll will not reflect this difference, no matter how many individuals are polled and no matter how well their opinions are elicited.[47] The poll's measurement of community opinion will be biased, although the magnitude of this bias may not be great.

44. If the jury list is not compiled properly from appropriate sources, it might be subject to challenge. *See* David Kairys et al., *Jury Representativeness: A Mandate for Multiple Source Lists*, 65 Cal. L. Rev. 776 (1977).

45. Likewise, in drug investigations the sampling frame for testing the contents of vials, bags, or packets seized by police easily can be devised to match the population of all the items seized in a single case. Because testing each and every item can be quite time-consuming and expensive, chemists often draw a probability sample, analyze the material that is sampled, and use the percentage of illicit drugs found in the sample to determine the total quantity of illicit drugs in all the items seized. *E.g.*, United States v. Shonubi, 895 F. Supp. 460, 470 (E.D.N.Y. 1995) (citing cases), *rev'd on other grounds*, 103 F.3d 1085 (2d Cir. 1997). For discussions of statistical estimation in such cases, see C.G.G. Aitken et al., *Estimation of Quantities of Drugs Handled and the Burden of Proof*, 160 J. Royal Stat. Soc'y 333 (1997); Dov Tzidony & Mark Ravreby, *A Statistical Approach to Drug Sampling: A Case Study*, 37 J. Forensic Sci. 1541 (1992); Johan Bring & Colin Aitken, *Burden of Proof and Estimation of Drug Quantities Under the Federal Sentencing Guidelines*, 18 Cardozo L. Rev. 1987 (1997).

46. On the admissibility of such polls, compare *Saliba v. State*, 475 N.E.2d 1181, 1187 (Ind. Ct. App. 1985) ("Although the poll did not . . . [ask] the interviewees . . . whether the particular film was obscene, the poll was relevant to an application of community standards"), with *United States v. Pryba*, 900 F.2d 748, 757 (4th Cir. 1990) ("Asking a person in a telephone interview as to whether one is offended by nudity, is a far cry from showing the materials . . . and then asking if they are offensive," so exclusion of the survey results was proper).

47. A classic example of selection bias is the 1936 *Literary Digest* poll. After successfully predicting the winner of every U.S. presidential election since 1916, the *Digest* used the replies from 2.4 million respondents to predict that Alf Landon would win 57% to 43%. In fact, Franklin Roosevelt won by a landslide vote of 62% to 38%. *See* Freedman et al., *supra* note 16, at 334–35. The *Digest* was so far off, in part, because it chose names from telephone books, rosters of clubs and associations, city directories, lists of registered voters, and mail order listings. *Id.* at 335, A-20 n.6. In 1936, when only one household in four had a telephone, the people whose names appeared on such lists tended to be more affluent. Lists that overrepresented the affluent had worked well in earlier elections, when rich and poor voted along similar lines, but the bias in the sampling frame proved fatal when the Great Depression made economics a salient consideration for voters. *See* Judith M. Tanur, *Samples and Surveys*, *in* Perspectives on Contemporary Statistics 55, 57 (David C. Hoaglin & David S. Moore eds., 1992). Today, survey organizations conduct polls by telephone, but most voters have telephones, and these organizations select the numbers to call at random rather than sampling names from telephone books.

Not all surveys use random selection. In some commercial disputes involving trademarks or advertising, the population of all potential purchasers of the products is difficult to identify. Some surveyors may resort to an easily accessible subgroup of the population, such as shoppers in a mall.[48] Such convenience samples may be biased by the interviewer's discretion in deciding whom to interview—a form of selection bias—and the refusal of some of those approached to participate—nonresponse bias.[49] Selection bias is acute when constituents write their representatives, listeners call into radio talk shows, interest groups collect information from their members,[50] or attorneys choose cases for trial.[51] Selection bias also affects data from jury-reporting services that gather information from readily available sources.

Various procedures are available to cope with selection bias. In quota sampling, the interviewer is instructed to interview so many women, so many older men, so many ethnic minorities, or the like. But quotas alone still leave too much discretion to the interviewers in selecting among the members of each category, and therefore do not solve the problem of selection bias.

Probability sampling methods, in contrast, ideally are suited to avoid selection bias. Once the conceptual population is reduced to a tangible sampling frame, the units to be measured are selected by some kind of lottery that gives each unit in the sampling frame a known, nonzero probability of being chosen. Selection according to a table of random digits or the like[52] leaves no room for selection bias. These procedures are used routinely to select individuals for jury

48. *E.g.*, R.J. Reynolds Tobacco Co. v. Loew's Theatres, Inc., 511 F. Supp. 867, 876 (S.D.N.Y. 1980) (questioning the propriety of basing a "nationally projectable statistical percentage" on a suburban mall intercept study).

49. Nonresponse bias is discussed *infra* § II.B.2.

50. *E.g.*, Pittsburgh Press Club v. United States, 579 F.2d 751, 759 (3d Cir. 1978) (tax-exempt club's mail survey of its members to show little sponsorship of income-producing uses of facilities was held to be inadmissible hearsay because it "was neither objective, scientific, nor impartial"), *rev'd on other grounds*, 615 F.2d 600 (3d Cir. 1980).

51. *See In re* Chevron U.S.A., Inc., 109 F.3d 1016 (5th Cir. 1997). In that case, the district court decided to try 30 cases to resolve common issues or to ascertain damages in 3,000 claims arising from Chevron's allegedly improper disposal of hazardous substances. The court asked the opposing parties to select 15 cases each. Selecting 30 extreme cases, however, is quite different from drawing a random sample of 30 cases. Thus, the court of appeals wrote that although random sampling would have been acceptable, the trial court could not use the results in the 30 extreme cases to resolve issues of fact or ascertain damages in the untried cases. *Id.* at 1020. Those cases, it warned, were "not cases calculated to represent the group of 3,000 claimants." *Id.*

52. In simple random sampling, units are drawn at random without replacement. In particular, each unit has the same probability of being chosen for the sample. More complicated methods, such as stratified sampling and cluster sampling, have advantages in certain applications. In systematic sampling, every fifth, tenth, or hundredth (in mathematical jargon, every *n*th) unit in the sampling frame is selected. If the starting point is selected at random and the units are not in any special order, then this procedure is comparable to simple random sampling.

duty;[53] they also have been used to choose "bellwether" cases for representative trials to resolve issues in all similar cases.[54]

2. Of the Units Selected, Which Are Measured?

Although probability sampling ensures that, within the limits of chance, the sample will be representative of the sampling frame, the question remains as to which units actually get measured. When objects like receipts are sampled for an audit, or vegetation is sampled for a study of the ecology of a region, all the selected units can be examined. Human beings are more troublesome. Some may refuse to respond, and the survey should report the nonresponse rate. A large nonresponse rate warns of bias,[55] although supplemental study may establish that the nonrespondents do not differ systematically from the respondents with respect to the characteristics of interest[56] or may permit the missing data to

53. Before 1968, most federal districts used the "key man" system for compiling lists of eligible jurors. Individuals believed to have extensive contacts in the community would suggest names of prospective jurors, and the qualified jury wheel would be made up from those names. To reduce the risk of discrimination associated with this system, the Jury Selection and Service Act of 1968, 28 U.S.C. §§ 1861–1878 (1988), substituted the principle of "random selection of juror names from the voter lists of the district or division in which court is held." S. Rep. No. 891, 90th Cong., 1st Sess. 10 (1967), *reprinted in* 1968 U.S.C.C.A.N. 1792, 1793.

54. Hilao v. Estate of Marcos, 103 F.3d 767 (9th Cir. 1996); Cimino v. Raymark Indus., Inc., 751 F. Supp. 649 (E.D. Tex. 1990); *cf. In re* Chevron U.S.A., Inc., 109 F.3d 1016 (5th Cir. 1997) (discussed *supra* note 51). Although trials in a suitable random sample of cases can produce reasonable estimates of average damages, the propriety of precluding individual trials has been debated. *Compare* Michael J. Saks & Peter David Blanck, *Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts*, 44 Stan. L. Rev. 815 (1992), *with Chevron*, 109 F.3d at 1021 (Jones, J., concurring); Robert G. Bone, *Statistical Adjudication: Rights, Justice, and Utility in a World of Process Scarcity*, 46 Vand. L. Rev. 561 (1993).

55. The 1936 *Literary Digest* election poll (see *supra* note 47) illustrates the danger. Only 24% of the 10 million people who received questionnaires returned them. Most of the respondents probably had strong views on the candidates, and most of them probably objected to President Roosevelt's economic program. This self-selection is likely to have biased the poll. Maurice C. Bryson, *The* Literary Digest *Poll: Making of a Statistical Myth*, 30 Am. Statistician 184 (1976); Freedman et al., *supra* note 16, at 335–36.

In *United States v. Cometz*, 730 F.2d 475, 478 (7th Cir. 1984) (en banc), the Seventh Circuit recognized that "a low rate of response to juror questionnaires could lead to the underrepresentation of a group that is entitled to be represented on the qualified jury wheel." Nevertheless, the court held that under the Jury Selection and Service Act of 1968, 28 U.S.C. §§ 1861–1878 (1988), the clerk did not abuse his discretion by failing to take steps to increase a response rate of 30%. According to the court, "Congress wanted to make it possible for all qualified persons to serve on juries, which is different from forcing all qualified persons to be available for jury service." *Cometz*, 730 F.2d at 480. Although it might "be a good thing to follow up on persons who do not respond to a jury questionnaire," the court concluded that Congress "was not concerned with anything so esoteric as nonresponse bias." *Id.* at 479, 482.

56. Even when demographic characteristics of the sample match those of the population, however, caution still is indicated. In the 1980s, a behavioral researcher sent out 100,000 questionnaires to explore how women viewed their relationships with men. Shere Hite, Women and Love: A Cultural Revolution in Progress (1987). She amassed a huge collection of anonymous letters from thousands of women disillusioned with love and marriage, and she wrote that these responses established that the

be imputed.[57]

In short, a good survey defines an appropriate population, uses an unbiased method for selecting the sample, has a high response rate, and gathers accurate information on the sample units. When these goals are met, the sample tends to be representative of the population: the measurements within the sample describe fairly the characteristics in the population. It remains possible, however, that despite every precaution, the sample, being less than exhaustive, is not representative; proper statistical analysis helps address the magnitude of this risk, at least for probability samples.[58] Of course, surveys may be useful even if they fail to meet all of the criteria given above; but then, additional arguments are needed to justify the inferences.

C. Individual Measurements

1. Is the Measurement Process Reliable?

There are two main aspects to the accuracy of measurements—reliability and validity. In science, "reliability" refers to reproducibility of results.[59] A reliable measuring instrument returns consistent measurements of the same quantity. A scale, for example, is reliable if it reports the same weight for the same object time and again. It may not be accurate—it may always report a weight that is too high or one that is too low—but the perfectly reliable scale always reports the

"outcry" of feminists "against the many injustices of marriage—exploitation of women financially, physically, sexually, and emotionally" is "just and accurate." *Id.* at 344. The outcry may indeed be justified, but this research does little to prove the point. About 95% of the 100,000 inquiries did not produce responses. The nonrespondents may have had less distressing experiences with men and therefore did not see the need to write autobiographical letters. Furthermore, this systematic difference would be expected within every demographic and occupational class. Therefore, the argument that the sample responses are representative because "those participating according to age, occupation, religion, and other variables known for the U.S. population at large in most cases quite closely mirrors that of the U.S. female population" is far from convincing. *Id.* at 777. In fact, the results of this nonrandom sample differ dramatically from those of polls with better response rates. *See* Chamont Wang, Sense and Nonsense of Statistical Inference: Controversy, Misuse, and Subtlety 174–76 (1993). For further criticism of this study, see David Streitfeld, *Shere Hite and the Trouble with Numbers*, 1 Chance 26 (1988).

57. Methods for "imputing" missing data are discussed in, e.g., Tanur, *supra* note 47, at 66 and Howard Wainer, *Eelworms, Bullet Holes, and Geraldine Ferraro: Some Problems with Statistical Adjustment and Some Solutions*, 14 J. Educ. Stat. 121 (1989) (with commentary). The easy case is one in which the response rate is so high that even if all nonrespondents had responded in a way adverse to the proponent of the survey, the substantive conclusion would be unaltered. Otherwise, imputation can be problematic.

58. *See infra* § IV.

59. Courts often use "reliable" to mean "that which can be relied on" for some purpose, such as establishing probable cause or crediting a hearsay statement when the declarant is not produced for confrontation. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 590 n.9 (1993), for instance, distinguishes "evidentiary reliability" from reliability in the technical sense of giving consistent results. We use "reliability" to denote the latter.

same weight for the same object. Its errors, if any, are systematic; they always point in the same direction.

Reliability can be ascertained by measuring the same quantity several times. For instance, one method of DNA identification requires a laboratory to determine the lengths of fragments of DNA. By making duplicate measurements of DNA fragments, a laboratory can determine the likelihood that two measurements will differ by specified amounts.[60] Such results are needed when deciding whether an observed discrepancy between a crime sample and a suspect sample is sufficient to exclude the suspect.[61]

In many studies, descriptive information is obtained on the subjects. For statistical purposes, the information may have to be reduced to numbers, a process called "coding." The reliability of the coding process should be considered. For instance, in a study of death sentencing in Georgia, legally trained evaluators examined short summaries of cases and ranked them according to the defendant's culpability.[62] Two different aspects of reliability are worth considering. First, the "within-observer" variability of judgments should be small—the same evaluator should rate essentially identical cases the same way. Second, the "between-observer" variability should be small—different evaluators should rate the same cases the same way.

2. Is the Measurement Process Valid?

Reliability is necessary, but not sufficient, to ensure accuracy. In addition to reliability, "validity" is needed. A valid measuring instrument measures what it is supposed to. Thus, a polygraph measures certain physiological responses to stimuli. It may accomplish this task reliably. Nevertheless, it is not valid as a lie detector unless increases in pulse rate, blood pressure, and the like are well correlated with conscious deception. Another example involves the MMPI (Minnesota Multiphasic Personality Inventory), a pencil and paper test that, many psychologists agree, measures aspects of personality or psychological functioning. Its reliability can be quantified. But this does not make it a valid test of sexual deviancy.[63]

When an independent and reasonably accurate way of measuring the variable of interest is available, it may be used to validate the measuring system in ques-

60. *See* Committee on DNA Forensic Science: An Update, National Research Council, The Evaluation of Forensic DNA Evidence 139–41 (1996).

61. *Id.*; Committee on DNA Tech. in Forensic Science, National Research Council, DNA Technology in Forensic Science 61–62 (1992); David H. Kaye & George F. Sensabaugh, Jr., Reference Guide on DNA Evidence, § VII, in this manual.

62. David C. Baldus et al., Equal Justice and the Death Penalty: A Legal and Empirical Analysis 49–50 (1990).

63. *See* People v. John W., 229 Cal. Rptr. 783, 785 (Ct. App. 1986) (holding that because the use of the MMPI to diagnose sexual deviancy was not shown to be generally accepted as valid in the scientific community, a diagnosis based in part on the MMPI was inadmissible).

tion. Breathalyzer readings may be validated against alcohol levels found in blood samples. Employment test scores may be validated against job performance. A common measure of validity is the correlation coefficient between the criterion (job performance) and the predictor (the test score).[64]

3. Are the Measurements Recorded Correctly?

Judging the adequacy of data collection may involve examining the process by which measurements are recorded and preserved. Are responses to interviews coded and logged correctly? Are all the responses to a survey included? If gaps or mistakes are present, do they distort the results?[65]

III. How Have the Data Been Presented?

After data have been collected, they should be presented in a way that makes them intelligible. Data can be summarized with a few numbers or with graphical displays. However, the wrong summary can mislead.[66] Section III.A discusses rates or percentages, and gives some cautionary examples of misleading summaries, indicating the sorts of questions that might be considered when numerical summaries are presented in court. Percentages are often used to demonstrate statistical association, which is the topic of section III.B. Section III.C

64. *E.g.*, Washington v. Davis, 426 U.S. 229, 252 (1976); Albemarle Paper Co. v. Moody, 422 U.S. 405, 430–32 (1975). As the discussion of the correlation coefficient indicates, *infra* § V.B, the closer the coefficient is to 1, the greater the validity. Various statistics are used to characterize the reliability of laboratory instruments, psychological tests, or human judgments. These include the standard deviation as well as the correlation coefficient. *See infra* §§ III, V.

65. *See, e.g.*, McCleskey v. Kemp, 753 F.2d 877, 914–15 (11th Cir. 1985) (district court was unpersuaded by a statistical analysis of capital sentencing, in part because of various imperfections in the study, including discrepancies in the data and missing data; concurring and dissenting opinion concludes that the district court's findings on missing and misrecorded data were clearly erroneous because the possible errors were not large enough to affect the overall results; for an exposition of the study and response to such criticisms, see Baldus et al., *supra* note 62), *aff'd*, 481 U.S. 279 (1987); G. Heileman Brewing Co. v. Anheuser-Busch, Inc., 676 F. Supp. 1436, 1486 (E.D. Wis. 1987) ("many coding errors . . . affected the results of the survey"); EEOC v. Sears, Roebuck & Co., 628 F. Supp. 1264, 1304, 1305 (N.D. Ill. 1986) ("[E]rrors in EEOC's mechanical coding of information from applications in its hired and nonhired samples also make EEOC's statistical analysis based on this data less reliable." The EEOC "consistently coded prior experience in such a way that less experienced women are considered to have the same experience as more experienced men" and "has made so many general coding errors that its data base does not fairly reflect the characteristics of applicants for commission sales positions at Sears."), *aff'd*, 839 F.2d 302 (7th Cir. 1988); Dalley v. Michigan Blue Cross-Blue Shield, Inc., 612 F. Supp. 1444, 1456 (E.D. Mich. 1985) ("although plaintiffs show that there were some mistakes in coding, plaintiffs still fail to demonstrate that these errors were so generalized and so pervasive that the entire study is invalid").

66. *See generally* Freedman et al., *supra* note 16; Huff, *supra* note 16; Moore, *supra* note 16; Zeisel, *supra* note 16.

considers graphical summaries of data, while sections III.D and III.E discuss some of the basic descriptive statistics that are likely to be encountered in litigation, including the mean, median and standard deviation.

A. Are Rates or Percentages Properly Interpreted?

1. Have Appropriate Benchmarks Been Provided?

Selective presentation of numerical information is like quoting someone out of context. A television commercial for the Investment Company Institute (the mutual fund trade association) said that a $10,000 investment made in 1950 in an average common stock mutual fund would have increased to $113,500 by the end of 1972. On the other hand, according to the *Wall Street Journal,* the same investment spread over all the stocks making up the New York Stock Exchange Composite Index would have grown to $151,427. Mutual funds performed worse than the stock market as a whole.[67] In this example, and in many other situations, it is helpful to look beyond a single number to some benchmark that places the isolated figure into perspective.

2. Have the Data-Collection Procedures Changed?

Changes in the process of collecting data can create problems of interpretation. Statistics on crime provide many examples. The number of petty larcenies reported in Chicago more than doubled between 1959 and 1960—not because of an abrupt crime wave, but because a new police commissioner introduced an improved reporting system.[68] During the 1970s, police officials in Washington, D.C., "demonstrated" the success of President Nixon's law-and-order campaign by valuing stolen goods at $49, just below the $50 threshold for inclusion in the Federal Bureau of Investigation's (FBI) Uniform Crime Reports.[69]

Changes in data-collection procedures are by no means limited to crime statistics.[70] Indeed, almost all series of numbers that cover many years are affected by changes in definitions and collection methods. When a study includes such time series data, it is useful to inquire about changes and to look for any sudden jumps, which may signal such changes.[71]

67. Moore, *supra* note 16, at 161.

68. *Id.* at 162.

69. James P. Levine et al., Criminal Justice in America: Law in Action 99 (1986).

70. For example, improved survival rates for cancer patients may result from improvements in therapy. Or, the change may simply mean that cancers now are detected earlier, due to improvements in diagnostic techniques, so that patients with these cancers merely appear to live longer. *See* Richard Doll & Richard Peto, The Causes of Cancer: Quantitative Estimates of Avoidable Risks of Cancer in the United States Today app. C at 1278–79 (1981).

71. Moore, *supra* note 16, at 162.

3. Are the Categories Appropriate?

Misleading summaries also can be produced by choice of categories for comparison. In *Philip Morris, Inc. v. Loew's Theatres, Inc.*,[72] and *R.J. Reynolds Tobacco Co. v. Loew's Theatres, Inc.*,[73] Philip Morris and R.J. Reynolds sought an injunction to stop the maker of Triumph low-tar cigarettes from running advertisements claiming that participants in a national taste test preferred Triumph to other brands. Plaintiffs alleged that claims that Triumph was a "national taste test winner" or Triumph "beats" other brands were false and misleading. An exhibit introduced by the defendant contained the data shown in Table 1.[74]

Table 1. Data used by defendant to refute plaintiffs' false advertising claim

	Triumph much better than Merit	Triumph somewhat better than Merit	Triumph about the same as Merit	Triumph somewhat worse than Merit	Triumph much worse than Merit
Number	45	73	77	93	36
Percentage	14%	22%	24%	29%	11%

Only 14% + 22% = 36% of the sample preferred Triumph to Merit, while 29% + 11% = 40% preferred Merit to Triumph.[75] By selectively combining categories, however, defendant attempted to create a different impression. Since 24% found the brands about the same, and 36% preferred Triumph, defendant claimed that a clear majority (36% + 24% = 60%) found Triumph "as good or better than Merit."[76] The court correctly resisted this chicanery, finding that defendant's test results did not support the advertising claims.[77]

There was a similar distortion in claims for the accuracy of a home pregnancy test.[78] The manufacturer advertised the test as 99.5% accurate under laboratory conditions. The data underlying this claim are summarized in Table 2.

Table 2. Home pregnancy test results

	Actually pregnant	Actually not pregnant
Test says pregnant	197	0
Test says not pregnant	1	2
Total	198	2

72. 511 F. Supp. 855 (S.D.N.Y. 1980).
73. 511 F. Supp. 867 (S.D.N.Y. 1980).
74. 511 F. Supp. at 866.
75. *Id.* at 856.
76. *Id.* at 866.
77. *Id.* at 856–57. The statistical issues in these cases are discussed more fully in 2 Gastwirth, *supra* note 1, at 633–39.
78. This incident is reported in Arnold Barnett, *How Numbers Can Trick You*, Tech. Rev., Oct. 1994, at 38, 44–45.

Table 2 does indicate only one error in 200 assessments, or 99.5% overall accuracy. But the table also shows that the test can make two types of errors—it can tell a pregnant woman that she is not pregnant (a false negative), and it can tell a woman who is not pregnant that she is (a false positive). The reported 99.5% accuracy rate conceals a crucial fact—the company had virtually no data with which to measure the rate of false positives.[79]

4. How Big Is the Base of a Percentage?

Rates and percentages often provide effective summaries of data, but these statistics can be misinterpreted. A percentage makes a comparison between two numbers: one number is the base, and the other number is compared to that base. When the base is small, actual numbers may be more revealing than percentages. Media accounts in 1982 of a crime wave by the elderly give an example. The annual Uniform Crime Reports showed a near tripling of the crime rate by older people since 1964, while crimes by younger people only doubled. But people over 65 years of age account for less than 1% of all arrests. In 1980, for instance, there were only 151 arrests of the elderly for robbery out of 139,476 total robbery arrests.[80]

5. What Comparisons Are Made?

Finally, there is the issue of which numbers to compare. Researchers sometimes choose among alternative comparisons. It may be worthwhile to ask why they chose the one they did. Would another comparison give a different view? A government agency, for example, may want to compare the amount of service now being given with that of earlier years—but what earlier year ought to be the baseline? If the first year of operation is used, a large percentage increase should be expected because of start-up problems.[81] If last year is used as the base, was it also part of the trend, or was it an unusually poor year? If the base year is not representative of other years, then the percentage may not portray the trend fairly.[82] No single question can be formulated to detect such distortions, but it may help to ask for the numbers from which the percentages were obtained;

79. Only two women in the sample were not pregnant; the test gave correct results for both of them. Although a false-positive rate of zero is ideal, an estimate based on a sample of only two women is not.

80. Mark H. Maier, The Data Game: Controversies in Social Science Statistics 83 (1991). *See also* Alfred Blumstein & Jacqueline Cohen, *Characterizing Criminal Careers*, 237 Science 985 (1987).

81. *Cf.* Michael J. Saks, *Do We Really Know Anything About the Behavior of the Tort Litigation System—And Why Not?*, 140 U. Pa. L. Rev. 1147, 1203 (1992) (using 1974 as the base year for computing the growth of federal product liability filings exaggerates growth because "1974 was the first year that product liability cases had their own separate listing on the cover sheets. . . . The count for 1974 is almost certainly an understatement . . . .").

82. Jeffrey Katzer et al., Evaluating Information: A Guide for Users of Social Science Research 106 (2d ed. 1982).

asking about the base can also be helpful. Ultimately, however, recognizing which numbers are related to which issues requires a species of clear thinking not easily reducible to a checklist.[83]

B. Is an Appropriate Measure of Association Used?

Many cases involve statistical association. Does a test for employee promotion have an exclusionary effect that depends on race or gender? Does the incidence of murder vary with the rate of executions for convicted murderers? Do consumer purchases of a product depend on the presence or absence of a product warning? This section discusses tables and percentage-based statistics that are frequently presented to answer such questions.[84]

Percentages often are used to describe the association between two variables. Suppose that a university alleged to discriminate against women in admitting students consists of only two colleges, engineering and business. The university admits 350 out of 800 male applicants; by comparison, it admits only 200 out of 600 female applicants. Such data commonly are displayed as in Table 3.[85]

Table 3. Admissions by gender

Decision	Male	Female	Total
Admit	350	200	550
Deny	450	400	850
Total	800	600	1,400

As Table 3 indicates, 350/800 = 44% of the males are admitted, compared with only 200/600 = 33% of the females. One way to express the disparity is to subtract the two percentages: 44% − 33% = 11 percentage points. Although such subtraction is commonly seen in jury discrimination cases,[86] the difference is inevitably small when the two percentages are both close to zero. If the selection rate for males is 5% and that for females is 1%, the difference is only 4 percentage points. Yet, females have only 1/5 the chance of males of being admitted, and that may be of real concern.[87]

83. For assistance in coping with percentages, see Zeisel, *supra* note 16, at 1–24.

84. Correlation and regression are discussed *infra* § V.

85. A table of this sort is called a "cross-tab" or a "contingency table." Table 3 is "two-by-two" because it has two rows and two columns, not counting rows or columns containing totals.

86. *See, e.g.*, D.H. Kaye, *Statistical Evidence of Discrimination in Jury Selection*, *in* Statistical Methods in Discrimination Litigation, *supra* note 11, at 13.

87. *Cf.* United States v. Jackman, 46 F.3d 1240, 1246–47 (2d Cir. 1995) (holding that the small percentage of minorities in the population makes it "inappropriate" to use an "absolute numbers" or "absolute impact" approach for measuring underrepresentation of these minorities in the list of potential jurors).

For Table 3, the selection ratio (used by the Equal Employment Opportunity Commission (EEOC) in its "80% rule")[88] is 33/44 = 75%, meaning that, on average, women have 75% the chance of admission that men have.[89] However, the selection ratio has its own problems. In the last example, if the selection rates are 5% and 1%, then the exclusion rates are 95% and 99%. The corresponding ratio is 99/95 = 104%, meaning that females have, on average, 104% the risk of males of being rejected. The underlying facts are the same, of course, but this formulation sounds much less disturbing.[90]

The odds ratio is more symmetric. If 5% of male applicants are admitted, the odds on a man being admitted are 5/95 = 1/19; the odds on a woman being admitted are 1/99. The odds ratio is (1/99)/(1/19) = 19/99. The odds ratio for rejection instead of acceptance is the same, except that the order is reversed.[91] Although the odds ratio has desirable mathematical properties, its meaning may be less clear than that of the selection ratio or the simple difference.

Data showing disparate impact are generally obtained by aggregating—putting together—statistics from a variety of sources. Unless the source material is fairly homogenous, aggregation can distort patterns in the data. We illustrate the problem with the hypothetical admission data in Table 3. Applicants can be classified not only by gender and admission but also by the college to which they applied, as in Table 4:

Table 4. Admissions by gender and college

	Engineering		Business	
Decision	Male	Female	Male	Female
Admit	300	100	50	100
Deny	300	100	150	300

The entries in Table 4 add up to the entries in Table 3; said more technically, Table 3 is obtained by aggregating the data in Table 4. Yet, there is no association between gender and admission in either college; men and women are ad-

88. The EEOC generally regards any procedure that selects candidates from the least successful group at a rate less than 80% of the rate for the most successful group as having an adverse impact. EEOC Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.4(D) (1993). The rule is designed to help spot instances of substantially discriminatory practices, and the commission usually asks employers to justify any procedures that produce selection ratios of 80% or less.

89. The analogous statistic used in epidemiology is called the relative risk. *See supra* note 38; Michael D. Green et al., Reference Guide on Epidemiology, § III.A, in this manual. Relative risks are usually quoted as decimals rather than percentages; for instance, a selection ratio of 75% corresponds to a relative risk of 0.75. A variation on this idea is the relative difference in the proportions, which expresses the proportion by which the probability of selection is reduced. Kairys et al., *supra* note 44, at 776, 789–90; *cf.* David C. Baldus & James W.L. Cole, Statistical Proof of Discrimination § 5.1, at 153 (1980 & Supp. 1987) (listing various ratios that can be used to measure disparities).

90. The Illinois Department of Employment Security tried to exploit this feature of the selection

mitted at identical rates. Combining two colleges with no association produces a university in which gender is associated strongly with admission. The explanation for this paradox: the business college, to which most of the women applied, admits relatively few applicants; the engineering college, to which most of the men applied, is easier to get into. This example illustrates a common issue: association can result from combining heterogeneous statistical material.[92]

C. Does a Graph Portray Data Fairly?

Graphs are useful for revealing key characteristics of a batch of numbers, trends over time, and the relationships among variables.[93]

1. How Are Trends Displayed?

Graphs that plot values over time are useful for seeing trends. However, the scales on the axes matter. In Figure 1, the federal debt appears to skyrocket during the Reagan and Bush administrations; in Figure 2, the federal debt appears to grow slowly.[94] The moral is simple: Pay attention to the markings on the axes to determine whether the scale is appropriate.

ratio in *Council 31, Am. Fed'n of State, County and Mun. Employees v. Ward*, 978 F.2d 373 (7th Cir. 1992). In January 1985, the department laid off 8.6% of the blacks on its staff in comparison with 3.0% of the whites. *Id.* at 375. Recognizing that these layoffs ran afoul of the 80% rule (since 3.0/8.6 = 35%, which is far less than 80%), the department instead presented the selection ratio for retention. *Id.* at 375–76. Since black employees were retained at 91.4/97.0 = 94% of the white rate, the retention rates showed no adverse impact under the 80% rule. *Id.* at 376. When a subsequent wave of layoffs was challenged as discriminatory, the department argued "that its retention rate analysis is the right approach to this case and . . . shows conclusively that the layoffs did not have a disparate impact." *Id.* at 379. The Seventh Circuit disagreed and, in reversing an order granting summary judgment to defendants on other grounds, left it to the district court on remand "to decide what method of proof is most appropriate." *Id.*

91. For women, the odds on rejection are 99 to 1; for men, 19 to 1. The ratio of these odds is 99/19. Likewise, the odds ratio for an admitted applicant being a man as opposed to a denied applicant being man is also 99/19.

92. Tables 3 and 4 are hypothetical, but closely patterned on a real example. *See* P.J. Bickel et al., *Sex Bias in Graduate Admissions: Data from Berkeley*, 187 Science 398 (1975). *See also* Freedman et al., *supra* note 16, at 17–20; Moore, *supra* note 16, at 246–47. The tables are an instance of "Simpson's Paradox." *See generally* Myra L. Samuels, *Simpson's Paradox and Related Phenomena*, 88 J. Am. Stat. Ass'n 81 (1993). Another perspective on Table 3 may be helpful. The college to which a student applies is a confounder. *See supra* § II.A.1. In the present context, confounders often are called "omitted variables." For opinions discussing the legal implications of omitted variables, see cases cited *supra* note 5 and *infra* note 230.

93. *See generally* William S. Cleveland, The Elements of Graphing Data (1985); David S. Moore & George P. McCabe, Introduction to the Practice of Statistics 3–20 (2d ed. 1993). Graphs showing relationships among variables are discussed *infra* § V.

94. *See* Howard Wainer, *Graphs in the Presidential Campaign*, Chance, Winter 1993, at 48, 50.

Figure 1. The federal debt skyrockets under Reagan–Bush.

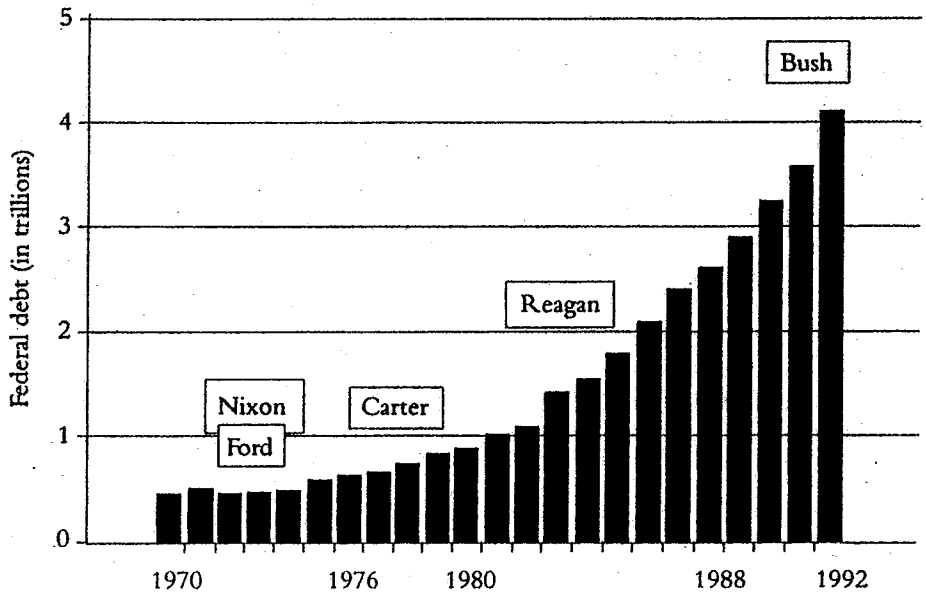
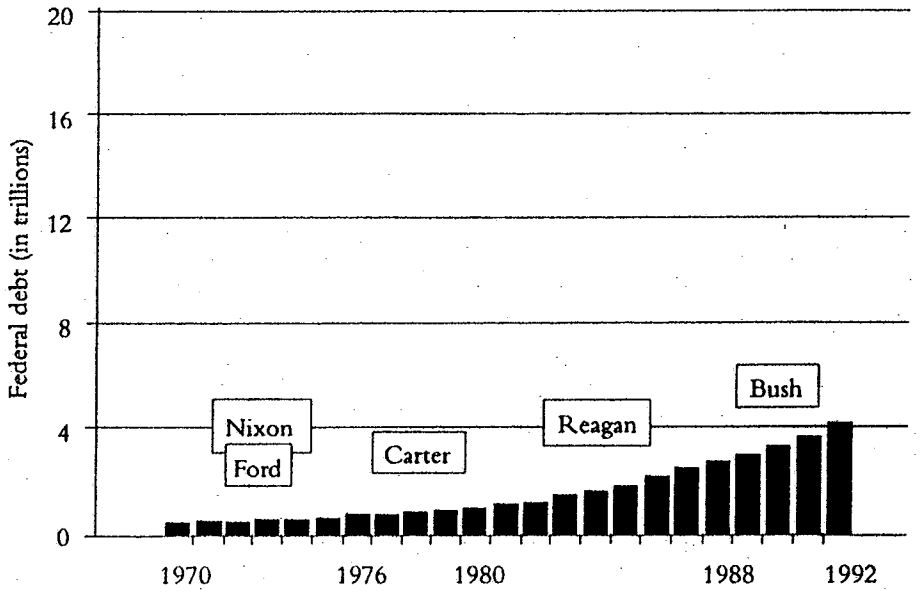


Figure 2. The federal debt grows steadily under Reagan–Bush.

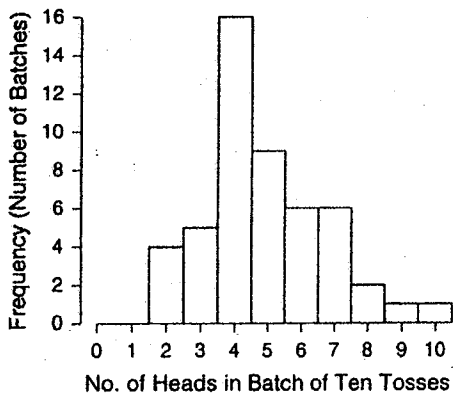


2. How Are Distributions Displayed?

A graph commonly used to display the distribution of data is the histogram.[95] One axis denotes the numbers, and the other indicates how often those fall within specified intervals (called "bins" or "class intervals"). For example, we flipped a quarter 10 times in a row and counted the number of heads in this "batch" of 10 tosses. With 50 batches, we obtained the following counts:[96]

7 7 5 6 8	4 2 3 6 5	4 3 4 7 4	6 8 4 7 4	7 4 5 4 3
4 4 2 5 3	5 4 2 4 4	5 7 2 3 5	4 6 4 9 10	5 5 6 6 4

Figure 3. Histogram showing how frequently various numbers of heads appeared in 50 batches of 10 tosses of a quarter.



The histogram is shown in Figure 3.[97] A histogram shows how the data are distributed over the range of possible values. The spread can be made to appear

95. For small batches of numbers, a "stem-and-leaf plot" may be more convenient. For instance, a stem-and-leaf plot for 11, 12, 23, 23, 23, 23, 33, 45, 69 is given below:

```
1 | 1 2
2 | 3 3 3 3
3 | 3
4 | 5
5 |
6 | 9
```

The numbers to the left of the line are the first digits; those to the right are the second digits. Thus, "2 | 3 3 3 3" stands for "23, 23, 23, 23."

96. The coin landed heads 7 times in the first 10 tosses; by coincidence, there were also 7 heads in the next 10 tosses; there were 5 heads in the third batch of 10 tosses; and so forth.

97. In Figure 3, the bin width is 1. There were no 0's or 1's in the data, so the bars over 0 and 1 disappear. There is a bin from 1.5 to 2.5; the four 2's in the data fall into this bin, so the bar over the

larger or smaller, however, by changing the scale of the horizontal axis. Likewise, the shape can be altered somewhat by changing the size of the bins.[98] It may be worth inquiring how the analyst chose the bin widths.

D. Is an Appropriate Measure Used for the Center of a Distribution?

Perhaps the most familiar descriptive statistic is the mean (or "arithmetic mean"). The mean can be found by adding up all the numbers and dividing by how many there are. By comparison, the median is defined so that half the numbers are bigger than the median, and half are smaller.[99] Yet a third statistic is the mode, which is the most common number in the data set. These statistics are different, although they are not always clearly distinguished.[100] The mean takes account of all the data—it involves the total of all the numbers; however, particularly with small data sets, a few unusually large or small observations may have too much influence on the mean. The median is resistant to such outliers.

To illustrate the distinction between the mean and the median, consider a report that the "average" award in malpractice cases skyrocketed from $220,000

interval from 1.5 to 2.5 has height four. There is another bin from 2.5 to 3.5, which catches five 3's; the height of the corresponding bar is five. And so forth.

All the bins in Figure 3 have the same width, so this histogram is just like a bar graph. However, data are often published in tables with unequal intervals. The resulting histograms will have unequal bin widths; bar heights should be calculated so that the areas (height × width) are proportional to the frequencies. In general, a histogram differs from a bar graph in that it represents frequencies by area, not height. *See* Freedman et al., *supra* note 16, at 31–41.

98. As the width of the bins decreases, the graph becomes more detailed. But the appearance becomes more ragged until finally the graph is effectively a plot of each datum. The optimal bin width "depends on the subject matter and the goal of the analysis." Cleveland, *supra* note 93, at 125.

99. Technically, at least half the numbers are at the median or larger; at least half are at the median or smaller. When the distribution is symmetric, the mean equals the median. The values diverge, however, when the distribution is asymmetric, or skewed. The distinction between the mean and the median is critical to the interpretation of the Railroad Revitalization and Regulatory Reform Act, 49 U.S.C. § 11503 (1988), which forbids the taxation of railroad property at a higher rate than other commercial and industrial property. To compare the rates, tax authorities often use the mean, whereas railroads prefer the median. The choice has important financial consequences, and much litigation has resulted. *See* David A. Freedman, *The Mean Versus the Median: A Case Study in 4-R Act Litigation*, 3 J. Bus. & Econ. Stat. 1 (1985).

100. In ordinary language, the arithmetic mean, the median, and the mode seem to be referred to interchangeably as "the average." In statistical parlance, the average is the arithmetic mean. The distinctions are brought out by the following question: How big an error would be made if every number in a batch were replaced by the "center" of the batch? The mode minimizes the number of errors; all errors count the same, no matter what their size. Similar distributions can have very different modes, and the mode is rarely used by statisticians. The median minimizes a different measure of error—the sum of all the differences between the center and the data points; signs are not taken into account when computing this sum, so positive and negative differences are treated the same way. The mean minimizes the sum of the squared differences.

in 1975 to more than $1 million in 1985.[101] The median award almost certainly was far less than $1 million,[102] and the apparently explosive growth may result from a few very large awards. Still, if the issue is whether insurers were experiencing more costs from jury verdicts, the mean is the more appropriate statistic: The total of the awards is directly related to the mean, not to the median.[103]

E. Is an Appropriate Measure of Variability Used?

The location of the center of a batch of numbers reveals nothing about the variations exhibited by these numbers.[104] Statistical measures of variability include the range, the interquartile range, and the standard deviation. The range is the difference between the largest number in the batch and the smallest. The range seems natural, and it indicates the maximum spread in the numbers, but it is generally the most unstable because it depends entirely on the most extreme values.[105] The interquartile range is the difference between the 25th and 75th percentiles.[106] The interquartile range contains 50% of the numbers and is resistant to changes in extreme values. The standard deviation is a sort of mean deviation from the mean.[107]

101. Kenneth Jost, *Still Warring Over Medical Malpractice: Time for Something Better*, A.B.A. J., May 1993, at 68, 70–71.

102. A study of cases in North Carolina reported an "average" (mean) award of about $368,000, and a median award of only $36,000. *Id.* at 71. In *TXO Production Corp. v. Alliance Resources Corp.*, 509 U.S. 443 (1993), briefs portraying the punitive damage system as out of control reported mean punitive awards, some ten times larger than the median awards described in briefs defending the current system of punitive damages. *See* Michael Rustad & Thomas Koenig, *The Supreme Court and Junk Social Science: Selective Distortion in Amicus Briefs*, 72 N.C. L. Rev. 91, 145–47 (1993). The mean differs so dramatically from the median because the mean takes into account (indeed, is heavily influenced by) the magnitudes of the few very large awards; the median screens these out. Of course, representative data on verdicts and awards are hard to find. For a study using a probability sample of cases, see Carol J. DeFrances et al., *Civil Jury Cases and Verdicts in Large Counties*, Bureau Just. Stats. Special Rep., July 1995, at 1.

103. To get the total award, just multiply the mean by the number of awards; by contrast, the total cannot be computed from the median. (The more pertinent figure for the insurance industry is not the total of jury awards, but actual claims experience including settlements; of course, even the risk of large punitive damage awards may have considerable impact.) These and related statistical issues are pursued further in, e.g., Theodore Eisenberg & Thomas A. Henderson, Jr., *Inside the Quiet Revolution in Products Liability*, 39 UCLA L. Rev. 731, 764–72 (1992); Scott Harrington & Robert E. Litan, *Causes of the Liability Insurance Crisis*, 239 Science 737, 740–41 (1988); Saks, *supra* note 81, at 1147, 1248–54.

104. The numbers 1, 2, 5, 8, 9 have 5 as their mean and median. So do the numbers 5, 5, 5, 5, 5. In the first batch, the numbers vary considerably about their mean; in the second, the numbers do not vary at all.

105. Typically, the range increases with the size of the sample, i.e., the number of units chosen for the sample.

106. By definition, 25% of the data fall below the 25th percentile, 90% fall below the 90th percentile, and so on. The median is the 50th percentile.

107. As discussed in the Appendix, when the distribution follows the normal curve, about 68% of the data will be within one standard deviation of the mean, and about 95% will be within two standard deviations of the mean. For other distributions, the proportions of the data within specified numbers of standard deviations will be different.

There are no hard and fast rules as to which statistic is the best. In general, the bigger these measures of spread are, the more the numbers are dispersed. Particularly in small data sets, the standard deviation can be influenced heavily by a few outlying values. To remove this influence, the mean and the standard deviation can be recomputed with the outliers discarded. Beyond this, any of the statistics can be supplemented with a figure that displays much of the data.[108]

IV. What Inferences Can Be Drawn from the Data?

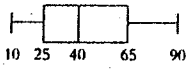
The inferences that may be drawn from a study depend on the quality of the data and the design of the study. As discussed in section II, the data might not address the issue of interest, might be systematically in error, or might be difficult to interpret due to confounding. We turn now to an additional concern—random error.[109] Are patterns in the data the result of chance? Would a pattern wash out if more data were collected?

The laws of probability are central to analyzing random error. By applying these laws, the statistician can assess the likely impact of chance error, using "standard errors," "confidence intervals," "significance probabilities," "hypothesis tests," or "posterior probability distributions." The following example illustrates the ideas. An employer plans to use a standardized examination to select trainees from a pool of 5,000 male and 5,000 female applicants. This total pool of 10,000 applicants is the statistical "population." Under Title VII of the Civil

Technically, the standard deviation is the square root of the variance; the variance is the mean square deviation from the mean. For instance, if the mean is 100, the datum 120 deviates from the mean by 20, and the square of 20 is $20^2 = 400$. If the variance (i.e., the mean of all the squared deviations) is 900, then the standard deviation is the square root of 900, that is, $\sqrt{900} = 30$. Among other things, taking the square root corrects for the fact that the variance is on a different scale than the measurements themselves. For example, if the measurements are of length in inches, the variance is in square inches; taking the square root changes back to inches.

To compare distributions on different scales, the coefficient of variation may be used: this statistic is the standard deviation, expressed as a percentage of the mean. For instance, consider the batch of numbers 1,4,4,7,9. The mean is $25/5 = 5$, the variance is $(16 + 1 + 1 + 4 + 16)/5 = 7.6$, and the standard deviation is $\sqrt{7.6} = 2.8$. The coefficient of variation is $2.8/5 = 56\%$.

108. For instance, the "five-number summary" lists the smallest value, the 25th percentile, the median, the 75th percentile, and the largest value. The five-number summary may be presented as a box plot. If the five numbers were 10, 25, 40, 65 and 90, the box plot would look like the following:



There are many variations on this idea in which the boundaries of the box, or the "whiskers" extending from it, represent slightly different points in the distribution of numbers.

109. Random error is also called sampling error, chance error, or statistical error. Econometricians use the parallel concept of random disturbance terms.

Rights Act, if the proposed examination excludes a disproportionate number of women, the employer needs to show that the exam is job related.[110]

To see whether there is disparate impact, the employer administers the exam to a sample of 50 men and 50 women drawn at random from the population of job applicants. In the sample, 29 of the men but only 19 of the women pass; the sample pass rates are therefore 29/50 = 58% and 19/50 = 38%. The employer announces that it will use the exam anyway, and several applicants bring an action under Title VII. Disparate impact seems clear. The difference in sample pass rates is 20 percentage points: 58% − 38% = 20 percentage points. The employer argues, however, that the disparity could just reflect random error. After all, only a small number of people took the test, and the sample could have included disproportionate numbers of high-scoring men and low-scoring women. Clearly, even if there were no overall difference in pass rates for male and female applicants, in some samples the men will outscore the women. More generally, a sample is unlikely to be a perfect microcosm of the population; statisticians call differences between the sample and the population, just due to the luck of the draw in choosing the sample, "random error" or "sampling error."

When assessing the impact of random error, a statistician might consider the following topics:

- *Estimation*. Plaintiffs use the difference of 20 percentage points between the sample men and women to estimate the disparity between all male and female applicants. How good is this estimate? Precision can be expressed using the "standard error" or a "confidence interval."
- *Statistical significance*. Suppose the defendant is right, and there is no disparate impact: in the population of all 5,000 male and 5,000 female applicants, pass rates are equal. How likely is it that a random sample of 50 men and 50 women will produce a disparity of 20 percentage points or more? This chance is known as a *p*-value. Statistical significance is determined by reference to the *p*-value, and "hypothesis testing" is the technique for computing *p*-values or determining statistical significance.[111]
- *Posterior probability*. Given the observed disparity of 20 percentage points in the sample, what is the probability that—in the population as a whole—men and women have equal pass rates? This question is of direct interest to the courts. For a subjectivist statistician, posterior probabilities may be com-

110. The seminal case is *Griggs v. Duke Power Co.*, 401 U.S. 424, 431 (1971). The requirements and procedures for the validation of tests can go beyond a simple showing of job relatedness. *See, e.g.*, Richard R. Reilly, *Validating Employee Selection Procedures*, *in* Statistical Methods in Discrimination Litigation, *supra* note 11, at 133; Michael Rothschild & Gregory J. Werden, *Title VII and the Use of Employment Tests: An Illustration of the Limits of the Judicial Process*, 11 J. Legal Stud. 261 (1982).

111. "Hypothesis testing" is also called "significance testing." For details on the example, see *infra* Appendix, especially note 245.

puted using "Bayes' rule." Within the framework of classical statistical theory, however, such a posterior probability has no meaning.[112]

- *Applicability of statistical models.* Statistical inference—whether done with confidence intervals or significance probabilities, by objective methods or subjective—depends on the validity of statistical models for the data. If the data are collected on the basis of a probability sample or a randomized experiment, there will be statistical models that fit the situation very well, and inferences based on these models will be quite secure. Otherwise, calculations are generally based on analogy: this group of people is like a random sample, that observational study is like a randomized experiment. The fit between the statistical model and the data may then require examination: how good is the analogy?

A. Estimation

1. What Estimator Should Be Used?

An estimator is a statistic computed from sample data and used to estimate a numerical characteristic of the population. For example, we used the difference in pass rates for a sample of men and women to estimate the corresponding disparity in the population of all applicants. In our sample, the pass rates were 58% and 38%; the difference in pass rates for the whole population was estimated as 20 percentage points: 58% − 38% = 20 percentage points. In more complex problems, statisticians may have to choose among several estimators. Generally, estimators that tend to make smaller errors are preferred. However, this idea can be made precise in more than one way,[113] leaving room for judgment in selecting an estimator.

2. What Is the Standard Error? The Confidence Interval?

An estimate based on a sample is likely to be off the mark, at least by a little, due to random error. The standard error gives the likely magnitude of this random error.[114] Whenever possible, an estimate should be accompanied by its standard

112. This classical framework is also called "objectivist" or "frequentist," by contrast with the "subjectivist" or "Bayesian" framework. In brief, objectivist statisticians view probabilities as objective properties of the system being studied. Subjectivists view probabilities as measuring subjective degrees of belief. Section IV.B.1 explains why posterior probabilities are excluded from the classical calculus, and section IV.C briefly discusses the subjectivist position. The procedure for computing posterior probabilities is presented *infra* Appendix. For more discussion, see David Freedman, *Some Issues in the Foundation of Statistics*, 1 Found. Sci. 19 (1995), *reprinted in* Topics in the Foundation of Statistics 19 (Bas C. van Fraasen ed., 1997).

113. Furthermore, reducing error in one context may increase error in other contexts; there may also be a trade-off between accuracy and simplicity.

114. "Standard errors" are also called "standard deviations," and courts seem to prefer the latter term, as do many authors.

error.[115] In our example, the standard error is about 10 percentage points: the estimate of 20 percentage points is likely to be off by something like 10 percentage points or so, in either direction.[116] Since the pass rates for all 5,000 men and 5,000 women are unknown, we cannot say exactly how far off the estimate is going to be, but 10 percentage points gauges the likely magnitude of the error.

Confidence intervals make the idea more precise. Statisticians who say that population differences fall within plus-or-minus 1 standard error of the sample differences will be correct about 68% of the time. To write this more compactly, we can abbreviate "standard error" as "SE." A 68% confidence interval is the range

estimate − 1 SE to estimate + 1 SE.

In our example, the 68% confidence interval goes from 10 to 30 percentage points. If a higher confidence level is wanted, the interval must be widened. The 95% confidence interval is about

estimate − 2 SE to estimate + 2 SE.

This runs from 0 to 40 percentage points.[117] Although 95% confidence intervals are used commonly, there is nothing special about 95%. For example, a 99.7% confidence interval is about

estimate − 3 SE to estimate + 3 SE.

This stretches from −10 to 50 percentage points.

The main point is that an estimate based on a sample will differ from the exact population value, due to random error; the standard error measures the likely size of the random error. If the standard error is small, the estimate probably is close to the truth. If the standard error is large, the estimate may be seriously wrong. Confidence intervals are a technical refinement, and

115. The standard error can also be used to measure reproducibility of estimates from one random sample to another. *See infra* note 237.

116. The standard error depends on the pass rates of men and women in the sample, and the size of the sample. With larger samples, chance error will be smaller, so the standard error goes down as sample size goes up. ("Sample size" is the number of subjects in the sample.) The Appendix gives the formula for computing the standard error of a difference in rates based on random samples. Generally, the formula for the standard error must take into account the method used to draw the sample and the nature of the estimator. Statistical expertise is needed to choose the right formula.

117. Confidence levels are usually read off the normal curve (*see infra* Appendix). Technically, the area under the normal curve between −2 and +2 is closer to 95.4% than 95.0%; thus, statisticians often use ±1.96 SEs for a 95% confidence interval. However, the normal curve only gives an approximation to the relevant chances, and the error in that approximation will often be larger than the difference between 95.4% and 95.0%. For simplicity, we use ±2 SEs for 95% confidence. Likewise, we use ±1 SE for 68% confidence, although the area under the curve between −1 and +1 is closer to 68.3%. The normal curve gives good approximations when the sample size is reasonably large; for small samples, other techniques should be used.

"confidence" is a term of art.[118] For a given confidence level, a narrower interval indicates a more precise estimate. For a given sample size, increased confidence can be attained only by widening the interval. A high confidence level alone means very little,[119] but a high confidence level for a small interval is impressive,[120] indicating that the random error in the sample estimate is low.

Standard errors and confidence intervals are derived using statistical models of the process that generated the data.[121] If the data come from a probability

118. In the standard frequentist theory of statistics, one cannot make probability statements about population characteristics. *See, e.g.*, Freedman et al., *supra* note 16, at 383–86; *infra* § IV.B.1. Consequently, it is imprecise to suggest that "[a] 95% confidence interval means that there is a 95% probability that the 'true' relative risk falls within the interval." DeLuca v. Merrell Dow Pharms., Inc., 791 F. Supp. 1042, 1046 (D.N.J. 1992), *aff'd*, 6 F.3d 778 (3d Cir. 1993). Because of the limited technical meaning of "confidence," it has been argued that the term is misleading and should be replaced by a more neutral one, such as "frequency coefficient," in courtroom presentations. David H. Kaye, *Is Proof of Statistical Significance Relevant?*, 61 Wash. L. Rev. 1333, 1354 (1986).

Another misconception is that the confidence level gives the chance that repeated estimates fall into the confidence interval. *E.g.*, Turpin v. Merrell Dow Pharms., Inc., 959 F.2d 1349, 1353 (6th Cir. 1992) ("a confidence interval of '95 percent between 0.8 and 3.10' . . . means that random repetition of the study should produce, 95 percent of the time, a relative risk somewhere between 0.8 and 3.10"); United States *ex rel.* Free v. Peters, 806 F. Supp. 705, 713 n.6 (N.D. Ill. 1992) ("A 99% confidence interval, for instance, is an indication that if we repeated our measurement 100 times under identical conditions, 99 times out of 100 the point estimate derived from the repeated experimentation will fall within the initial interval estimate . . . ."), *rev'd in part*, 12 F.3d 700 (7th Cir. 1993). However, the confidence level does not give the percentage of the time that repeated estimates fall in the interval; instead, it gives the percentage of the time that intervals from repeated samples cover the true value.

119. Statements about the confidence in a sample without any mention of the interval estimate are practically meaningless. In *Hilao v. Estate of Marcos*, 103 F.3d 767 (9th Cir. 1996), for instance, "an expert on statistics . . . testified that . . . a random sample of 137 claims would achieve 'a 95% statistical probability that the same percentage determined to be valid among the examined claims would be applicable to the totality of [9,541 facially valid] claims filed.'" *Id.* at 782. Unfortunately, there is no 95% "statistical probability" that a percentage computed from a sample will be "applicable" to a population. One can compute a confidence interval from a random sample and be 95% confident that the interval covers some parameter. That can be done for a sample of virtually any size, with larger samples giving smaller intervals. What is missing from the opinion is a discussion of the widths of the relevant intervals.

120. Conversely, a broad interval signals that random error is substantial. In *Cimino v. Raymark Industries, Inc.*, 751 F. Supp. 649 (E.D. Tex. 1990), the district court drew certain random samples from more than 6,000 pending asbestos cases, tried these cases, and used the results to estimate the total award to be given to all plaintiffs in the pending cases. The court then held a hearing to determine whether the samples were large enough to provide accurate estimates. The court's expert, an educational psychologist, testified that the estimates were accurate because the samples matched the population on such characteristics as race and the percentage of plaintiffs still alive. *Id.* at 664. However, the matches occurred only in the sense that population characteristics fell within very broad 99% confidence intervals computed from the samples. The court thought that matches within the 99% confidence intervals proved more than matches within 95% intervals. *Id.* Unfortunately, this is backwards. To be correct in a few instances with a 99% confidence interval is not very impressive—by definition, such intervals are broad enough to ensure coverage 99% of the time. *Cf.* Saks & Blanck, *supra* note 54.

121. Generally, statistical models enable the analyst to compute the chances of the various possible outcomes. For instance, the model may contain parameters, that is, numerical constants describing the population from which samples were drawn. *See infra* § V. That is the case for our example, where one

sample or a randomized controlled experiment,[122] the statistical model may be connected tightly to the actual data-collection process. In other situations, using the model may be tantamount to assuming that a sample of convenience is like a random sample, or that an observational study is like a randomized experiment.

Our example was based on a random sample, and that justified the statistical calculations.[123] In many contexts, the choice of an appropriate statistical model is not obvious.[124] When a model does not fit the data-collection process so well,

parameter is the pass rate of the 5,000 male applicants, and another parameter is the pass rate of the 5,000 female applicants. As explained in the Appendix, these parameters can be used to compute the chance of getting any particular sample difference. Using a model with known parameters to find the probability of an observed outcome (or one like it) is common in cases alleging discrimination in the selection of jurors. *E.g.*, Castaneda v. Partida, 430 U.S. 482, 496 (1977); Kaye, *supra* note 86, at 13; *cf.* Hazelwood Sch. Dist. v. United States, 433 U.S. 299, 311 n.17 (1977) (computing probabilities of selecting black teachers). But when the values of the parameters are not known, the statistician must work backwards, using the sample data to estimate the unknown population parameters. That is the kind of statistical inference described in this section.

122. *See supra* § II.A–B.

123. As discussed in the Appendix, large random samples give rise to certain normally distributed statistics. Partly because the Supreme Court used such a model in *Hazelwood* and *Castaneda*, courts and attorneys sometimes are skeptical of analyses that produce other types of random variables. *See, e.g.*, EEOC v. Western Elec. Co., 713 F.2d 1011 (4th Cir. 1983), *discussed in* David H. Kaye, *Ruminations on Jurimetrics: Hypergeometric Confusion in the Fourth Circuit*, 26 Jurimetrics J. 215 (1986). *But see* Branion v. Gramly, 855 F.2d 1256 (7th Cir. 1988) (questioning an apparently arbitrary assumption of normality), *discussed in* David H. Kaye, *Statistics for Lawyers and Law for Statistics*, 89 Mich. L. Rev. 1520 (1991) (defending the use of the normal approximation); Michael O. Finkelstein & Bruce Levin, *Reference Guide on Statistics: Non Lasciare Esperanza*, 36 Jurimetrics J. 201, 205 (1996) (review essay) ("The court was right to reject the normal distribution . . . ."). Whether a given variable is normally distributed is an empirical or statistical question, not a matter of law.

124. *See infra* § V. For examples of legal interest, see, e.g., Mary W. Gray, *Can Statistics Tell Us What We Do Not Want to Hear?: The Case of Complex Salary Structures*, 8 Stat. Sci. 144 (1993); Arthur P. Dempster, *Employment Discrimination and Statistical Science*, 3 Stat. Sci. 149 (1988). As one statistician describes the issue:

> [A] given data set can be viewed from more than one perspective, can be represented by a model in more than one way. Quite commonly, no unique model stands out as "true" or correct; justifying so strong a conclusion might require a depth of knowledge that is simply lacking. So it is not unusual for a given data set to be analyzed in several apparently reasonable ways. If conclusions are qualitatively concordant, that is regarded as grounds for placing additional trust in them. But more often, only a single model is applied, and the data are analyzed in accordance with it. . . .
>
> Desirable features in a model include (i) tractability, (ii) parsimony, and (iii) realism. That there is some tension among these is not surprising.
>
> *Tractability*. A model that is easy to understand and to explain is tractable in one sense. Computational tractability can also be an advantage, though with cheap computing available not too much weight can be given to it.
>
> *Parsimony*. Simplicity, like tractability, has a direct appeal, not wisely ignored—but not wisely over-valued either. If several models are plausible and more than one of them fits adequately with the data, then in choosing among them, one criterion is to prefer a model that is simpler than the other models.
>
> *Realism*. . . . First, does the model reflect well the actual [process that generated the data]? This question is really a host of questions, some about the distributions of the random errors, others about the mathematical relations among the [variables and] parameters. The second aspect of realism is sometimes called robustness.

estimates and standard errors will be less probative.[125]

Standard errors and confidence intervals generally ignore systematic errors such as selection bias or non-response bias; in other words, these biases are assumed to be negligible.[126] For example, one court—reviewing studies of whether a particular drug causes birth defects—observed that mothers of children with birth defects may be more likely to remember taking a drug during pregnancy than women with normal children.[127] This selective recall would bias comparisons between samples from the two groups of women. The standard error for the estimated difference in drug usage between the two groups ignores this bias; so does the confidence interval.[128] Likewise, the standard error does not address problems inherent in using convenience samples rather than random samples.[129]

B. *Significance Levels and Hypothesis Tests*

1. *What Is the p-value?*

In our example, 50 men and 50 women were drawn at random from 5,000 male and 5,000 female applicants. An exam was administered to this sample, and in the sample, the pass rates for the men and women were 58% and 38%, respectively. The sample difference in pass rates was 58% − 38% = 20 percentage points. The *p*-value answers the following question: If the pass rates among all 5,000 male applicants and 5,000 female applicants were identical, how probable would it be to find a discrepancy as big as or bigger than the 20 percentage point difference observed in our sample? The question is delicate, because the pass rates in the population are unknown—that is why a sample was taken in the first place.

> If the model is *false* in certain respects, how badly does that affect estimates, significance test results, etc., that are based on the flawed model?

Lincoln E. Moses, *The Reasoning of Statistical Inference*, *in* Perspectives on Contemporary Statistics, *supra* note 47, at 107, 117–18.

125. It still may be helpful to consider the standard error, perhaps as a minimal estimate for statistical uncertainty in the quantity being estimated.

126. For a discussion of such systematic errors, see *supra* § II.B.

127. Brock v. Merrell Dow Pharms., Inc., 874 F.2d 307, 311–12 (5th Cir.), *modified*, 884 F.2d 166 (5th Cir. 1989).

128. In *Brock*, the court stated that the confidence interval took account of bias (in the form of selective recall) as well as random error. 874 F.2d at 311–12. With respect, we disagree. Even if sampling error were nonexistent—which would be the case if one could interview every woman who had a child in the period that the drug was available—selective recall would produce a difference in the percentages of reported drug exposure between mothers of children with birth defects and those with normal children. In this hypothetical situation, the standard error would vanish. Therefore, the standard error could disclose nothing about the impact of selective recall. The same conclusion holds even in the presence of sampling error.

129. *See supra* § II.B.1.

The assertion that the pass rates in the population are the same is called the null hypothesis. The null hypothesis asserts that there is no difference between men and women in the whole population—differences in the sample are due to the luck of the draw. The p-value is the probability of getting data as extreme as, or more extreme than, the actual data, given that the null hypothesis is true:

$$p = \text{Probability}(\text{extreme data} \mid \text{null hypothesis in model})$$

In our example, $p = 5\%$. If the null hypothesis is true, there is only a 5% chance of getting a difference in the pass rates of 20 percentage points or more.[130] The p-value for the observed discrepancy is 5%, or .05.

In such cases, small p-values are evidence of disparate impact, while large p-values are evidence against disparate impact. Regrettably, multiple negatives are involved here. A statistical test is essentially an argument by contradiction. The "null hypothesis" asserts no difference in the population—that is, no disparate impact. Small p-values speak against the null hypothesis—there is disparate impact, because the observed difference is hard to explain by chance alone. Conversely, large p-values indicate that the data are compatible with the null hypothesis: the observed difference is easy to explain by chance. In this context, small p-values argue for the plaintiffs, while large p-values argue for the defense.[131]

Since p is calculated by assuming that the null hypothesis is correct (no real difference in pass rates), the p-value cannot give the chance that this hypothesis is true. The p-value merely gives the chance of getting evidence against the null hypothesis as strong or stronger than the evidence at hand—assuming the null hypothesis to be correct. No matter how many samples are obtained, the null hypothesis is either always right or always wrong. Chance affects the data, not the hypothesis. With the frequency interpretation of chance, there is no meaningful way to assign a numerical probability to the null hypothesis.[132]

130. *See infra* Appendix.

131. Of course, sample size must also be considered, among other factors. *See infra* § IV.C.

132. *See, e.g.*, The Evolving Role of Statistical Assessments as Evidence in the Courts, *supra* note 1, at 196–98; David H. Kaye, *Statistical Significance and the Burden of Persuasion*, Law & Contemp. Probs., Autumn 1983, at 13. Some opinions suggest a contrary view. *E.g.*, Vasquez v. Hillery, 474 U.S. 254, 259 n.3 (1986) ("the District Court . . . ultimately accepted . . . a probability of 2 in 1,000 that the phenomenon was attributable to chance"); EEOC v. Olson's Dairy Queens, Inc., 989 F.2d 165, 167 (5th Cir. 1993) ("Dr. Straszheim concluded that the likelihood that [the] observed hiring patterns resulted from truly race-neutral hiring practices was less than one chance in ten thousand"); Capaci v. Katz & Besthoff, Inc., 711 F.2d 647, 652 (5th Cir. 1983) ("the highest probability of unbiased hiring was 5.367×10^{-20}"). Such statements confuse the probability of the kind of outcome observed, which is computed under some model of chance, with the probability that chance is the explanation for the outcome.

In scientific notation, 10^{20} is 1 followed by 20 zeros, and 10^{-20} is the reciprocal of that number. The proverbial "one-in-a-million" is more dryly expressed as 1×10^{-6}.

Computing *p*-values requires statistical expertise. Many methods are available, but only some will fit the occasion. Sometimes standard errors will be part of the analysis, while other times they will not be. Sometimes a difference of 2 standard errors will imply a *p*-value of about .05, other times it will not. In general, the *p*-value depends on the model and its parameters, the size of the sample, and the sample statistics.[133]

Because the *p*-value is affected by sample size, it does not measure the extent or importance of a difference.[134] Suppose, for instance, that the 5,000 male and 5,000 female job applicants would differ in their pass rates, but only by a single percentage point. This difference might not be enough to make a case of disparate impact, but by including enough men and women in the sample, the data could be made to have an impressively small *p*-value. This *p*-value would confirm that the 5,000 men and 5,000 women have different pass rates, but it would not show the difference is substantial.[135] In short, the *p*-value does not measure the strength or importance of an association.

2. Is a Difference Statistically Significant?

Statistical significance is determined by comparing a *p*-value to a preestablished value, the significance level.[136] If an observed difference is in the middle of the distribution that would be expected under the null hypothesis, there is no surprise. The sample data are of the type that often would be seen when the null hypothesis is true: the difference is not significant, and the null hypothesis cannot be rejected. On the other hand, if the sample difference is far from the expected value—according to the null hypothesis—then the sample is unusual: the difference is "significant," and the null hypothesis is rejected. In our example, the 20 percentage point difference in pass rates for the men and women in the sample, whose *p*-value was about .05, might be considered significant at

133. In this context, a parameter is an unknown numerical constant that is part of the statistical model. *See supra* note 121.

134. Some opinions seem to equate small *p*-values with "gross" or "substantial" disparities. *E.g.*, Craik v. Minnesota St. Univ. Bd., 731 F.2d 465, 479 (8th Cir. 1984). Other courts have emphasized the need to decide whether the underlying sample statistics reveal that a disparity is large. *E.g.*, McCleskey v. Kemp, 753 F.2d 877, 892–94 (11th Cir. 1985), *aff'd*, 481 U.S. 279 (1987).

135. *Cf.* Frazier v. Garrison Indep. Sch. Dist., 980 F.2d 1514, 1526 (5th Cir. 1993) (rejecting claims of intentional discrimination in the use of a teacher competency examination that resulted in retention rates exceeding 95% for all groups).

136. Statisticians use the Greek letter alpha (α) to denote the significance level; α gives the chance of getting a "significant" result, assuming that the null hypothesis is true. Thus, α represents the chance of what is variously termed a "false rejection" of the null hypothesis or a "Type I error" (also called a "false positive" or a "false alarm"). For example, suppose $\alpha = 5\%$. If investigators do many studies, and the null hypothesis happens to be true in each case, then about 5% of the time they would obtain significant results—and falsely reject the null hypothesis.

the .05 level. If the threshold were set lower, say at .01, the result would not be significant.[137]

In practice, statistical analysts often use certain preset significance levels—typically .05 or .01.[138] The .05 level is the most common in social science, and an analyst who speaks of "significant" results without specifying the threshold probably is using this figure.[139] An unexplained reference to "highly significant" results probably means that p is less than .01.[140]

Since the term "significant" is merely a label for certain kinds of p-values, it is subject to the same limitations as are p-values themselves. Analysts may refer to a difference as "significant," meaning only that the p-value is below some threshold value. Significance depends not only on the magnitude of the effect, but also on the sample size (among other things). Thus, significant differences are evidence that something besides random error is at work, but they are not evidence that this "something" is legally or practically important. Statisticians distinguish between "statistical" and "practical" significance to make the point. When practical significance is lacking—when the size of a disparity or correlation is negligible—there is no reason to worry about statistical significance.[141]

As noted above, it is easy to mistake the p-value for the probability that there is no difference. Likewise, if results are significant at the .05 level, it is tempting to conclude that the null hypothesis has only a 5% chance of being correct.[142]

137. For another example of the relationship between a test statistic and significance, see *infra* § V.D.2.

138. The Supreme Court implicitly referred to this practice in *Castaneda v. Partida*, 430 U.S. 482, 496 n.17 (1977), and *Hazelwood School District v. United States*, 433 U.S. 299, 311 n.17 (1977). In these footnotes, the Court described the null hypothesis as "suspect to a social scientist" when a statistic from "large samples" falls more than "two or three standard deviations" from its expected value under the null hypothesis. Although the Court did not say so, these differences produce p-values of about .05 and .01 when the statistic is normally distributed. The Court's "standard deviation" is our "standard error."

139. Some have suggested that data not "significant" at the .05 level should be disregarded. *E.g.*, Paul Meier et al., *What Happened in* Hazelwood: *Statistics, Employment Discrimination, and the 80% Rule*, 1984 Am. B. Found. Res. J. 139, 152, *reprinted in* Statistics and the Law, *supra* note 1, at 1, 13. This view is challenged in, e.g., Kaye, *supra* note 118, at 1344 & n.56, 1345.

140. Merely labeling results as "significant" or "not significant" without providing the underlying information that goes into this conclusion is of limited value. *See, e.g.*, John C. Bailar III & Frederick Mosteller, *Guidelines for Statistical Reporting in Articles for Medical Journals: Amplifications and Explanations*, *in* Medical Uses of Statistics, *supra* note 28, at 313, 316.

141. *E.g.*, Waisome v. Port Auth., 948 F.2d 1370, 1376 (2d Cir. 1991) ("though the disparity was found to be statistically significant, it was of limited magnitude"); *cf.* Thornburg v. Gingles, 478 U.S. 30, 53–54 (1986) (repeating the district court's explanation of why "the correlation between the race of the voter and the voter's choice of certain candidates was [not only] statistically significant," but also "so marked as to be substantively significant, in the sense that the results of the individual election would have been different depending upon whether it had been held among only the white voters or only the black voters").

142. *E.g.*, *Waisome*, 948 F.2d at 1376 ("Social scientists consider a finding of two standard deviations significant, meaning there is about one chance in 20 that the explanation for a deviation could be random . . . ."); Rivera v. City of Wichita Falls, 665 F.2d 531, 545 n.22 (5th Cir. 1982) ("A variation

This temptation should be resisted. From the frequentist perspective, statistical hypotheses are either true or false; probabilities govern the samples, not the models and hypotheses. The significance level tells us what is likely to happen when the null hypothesis is correct; it cannot tell us the probability that the hypothesis is true. Significance comes no closer to expressing the probability that the null hypothesis is true than does the underlying *p*-value.[143]

C. Evaluating Hypothesis Tests

1. What Is the Power of the Test?

When a *p*-value is high, findings are not significant, and the null hypothesis is not rejected. This could happen for at least two reasons:

1. there is no difference in the population—the null hypothesis is true; or
2. there is some difference in the population—the null hypothesis is false—but, by chance, the data happened to be of the kind expected under the null hypothesis.

If the "power" of a statistical study is low, the second explanation may be plausible. Power is the chance that a statistical test will declare an effect when there is an effect to declare.[144] This chance depends on the size of the effect and

of two standard deviations would indicate that the probability of the observed outcome occurring purely by chance would be approximately five out of 100; that is, it could be said with a 95% certainty that the outcome was not merely a fluke."); Vuyanich v. Republic Nat'l Bank, 505 F. Supp. 224, 272 (N.D. Tex. 1980) ("[I]f a 5% level of significance is used, a sufficiently large *t*-statistic for the coefficient indicates that the chances are less than one in 20 that the true coefficient is actually zero."), *vacated*, 723 F.2d 1195 (5th Cir. 1984); Sheehan v. Daily Racing Form, Inc., 104 F.3d 940, 941 (7th Cir. 1997) ("An affidavit by a statistician . . . states that the probability that the retentions . . . are uncorrelated with age is less than 5 percent.").

143. For more discussion, see Kaye, *supra* note 118; *cf. infra* note 167.

144. More precisely, power is the probability of rejecting the null hypothesis when the alternative hypothesis is right. (On the meaning of "alternative hypothesis," see *infra* § IV.C.5.) Typically, this probability will depend on the values of unknown parameters, as well as the pre-set significance level α. Therefore, no single number gives the power of the test. One can specify particular values for the parameters and significance level and compute the power of the test accordingly. *See infra* Appendix for an example. Power may be denoted by the Greek letter beta (β).

Accepting the null hypothesis when the alternative is true is known as a "false acceptance" of the null hypothesis or a "Type II error" (also called a "false negative" or a "missed signal"). The chance of a false negative may be computed from the power, as $1 - \beta$. Frequentist hypothesis testing keeps the risk of a false positive to a specified level (such as $\alpha = .05$) and then tries to minimize the chance of a false negative ($1 - \beta$) for that value of α. Regrettably, the notation is in some degree of flux; many authors use β to denote the chance of a false negative; then, it is β that should be minimized.

Some commentators have claimed that the cutoff for significance should be chosen to equalize the chance of a false positive and a false negative, on the ground that this criterion corresponds to the "more-probable-than-not" burden of proof. Unfortunately, the argument is fallacious, because α and β do not give the probabilities of the null and alternative hypotheses; *see supra* § IV.B.2; *infra* note 167. *See* D.H. Kaye, *Hypothesis Testing in the Courtroom*, *in* Contributions to the Theory and Application of Statistics: A Volume in Honor of Herbert Solomon 331, 341–43 (Alan E. Gelfand ed., 1987); *supra* § IV.B.1; *infra* note 165.

the size of the sample. Discerning subtle differences in the population requires large samples; even so, small samples may detect truly substantial differences.[145]

When a study with low power fails to show a significant effect, the results are more fairly described as inconclusive than as negative: the proof is weak because power is low.[146] On the other hand, when studies have a good chance of detecting a meaningful association, failure to obtain significance can be persuasive evidence that there is no effect to be found.[147]

2. *One- or Two-tailed Tests?*

In many cases, a statistical test can be done either one-tailed or two-tailed. The second method will produce a *p*-value twice as big as the first method. Since

145. For simplicity, the numerical examples of statistical inference in this reference guide presuppose large samples. Some courts have expressed uneasiness about estimates or analyses based on small samples; indeed, a few courts have refused even to consider such studies or formal statistical procedures for handling small samples. *See, e.g.*, Bunch v. Bullard, 795 F.2d 384, 395 n.12 (5th Cir. 1986) (that 12 of 15 whites and only 3 of 13 blacks passed a police promotion test created a prima facie case of disparate impact; however, "[t]he district court did not perform, nor do we attempt, the application of probability theories to a sample size as small as this" because "[a]dvanced statistical analysis may be of little help in determining the significance of such disparities"); United States v. Lansdowne Swim Club, 713 F. Supp. 785, 809–10 (E.D. Pa. 1989) (collecting cases). Other courts have been more venturesome. *E.g.*, Bazemore v. Friday, 751 F.2d 662, 673 & n.9 (4th Cir. 1984) (court of appeals applied its own *t*-test rather than the normal curve to quartile rankings in an attempt to account for a sample size of nine), *rev'd on other grounds*, 478 U.S. 385 (1986).

Analyzing data from small samples may require more stringent assumptions, but there is no fundamental difference in the meaning of confidence intervals and *p*-values. If the assumptions underlying the statistical analysis are justified—and this can be more difficult to demonstrate with small samples—then confidence intervals and test statistics are no less trustworthy than those for large samples. Aside from the problem of choosing the correct analytical technique, the concern with small samples is not that they are beyond the ken of statistical theory, but that (1) the statistical tests involving small samples might lack power, and (2) the underlying assumptions may be hard to validate.

146. In our example, with $\alpha = .05$, power to detect a difference of 10 percentage points between the male and female job applicants is only about 1/6. *See infra* Appendix. Not seeing a "significant" difference therefore provides only weak proof that the difference between men and women is smaller than 10 percentage points. We prefer estimates accompanied by standard errors to tests because the former seem to make the state of the statistical evidence clearer: The estimated difference is 20 ± 10 percentage points, indicating that a difference of 10 percentage points is quite compatible with the data.

147. Some formal procedures are available to aggregate results across studies. *See In re* Paoli R.R. Yard PCB Litig., 916 F.2d 829 (3d Cir. 1990). In principle, the power of the collective results will be greater than the power of each study. *See, e.g.*, The Handbook of Research Synthesis 226–27 (Harris Cooper & Larry V. Hedges eds., 1993); Larry V. Hedges & Ingram Olkin, Statistical Methods for Meta-Analysis (1985); Jerome P. Kassirer, *Clinical Trials and Meta-Analysis: What Do They Do for Us?*, 327 New Eng. J. Med. 273, 274 (1992) ("[C]umulative meta-analysis represents one promising approach."); National Reseach Council, Combining Information: Statistical Issues and Opportunities for Research (1992); Symposium, *Meta-Analysis of Observational Studies*, 140 Am. J. Epidemiology 771 (1994). Unfortunately, the procedures have their own limitations. *E.g.*, Diana B. Petitti, Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis Methods for Quantitative Synthesis in Medicine (2d ed. 2000); Michael Oakes, Statistical Inference: A Commentary for the Social and Behavioural Sciences 157 (1986) ("a retrograde development"); John C. Bailar III, *The Promise and Problems of Meta-Analysis*, 337 New Eng. J. Med. 559 (1997) (editorial); Charles Mann, *Meta-Analysis in the Breech*, 249 Science 476 (1990).

small p-values are evidence against the null hypothesis, a one tailed test seems to produce stronger evidence than a two-tailed test. However, this difference is largely illusory.[148]

Some courts have expressed a preference for two-tailed tests,[149] but a rigid rule is not required if p-values and significance levels are used as clues rather than as mechanical rules for statistical proof. One-tailed tests make it easier to reach a threshold like .05, but if .05 is not used as a magic line, then the choice between one tail and two is less important—as long as the choice and its effect on the p-value are made explicit.[150]

3. How Many Tests Have Been Performed?

Repeated testing complicates the interpretation of significance levels. If enough comparisons are made, random error almost guarantees that some will yield "significant" findings, even when there is no real effect. Consider the problem of deciding whether a coin is biased. The probability that a fair coin will produce ten heads when tossed ten times is $(1/2)^{10} = 1/1{,}024$. Observing ten heads in the first ten tosses, therefore, would be strong evidence that the coin is biased. Nevertheless, if a fair coin is tossed a few thousand times, it is likely that at least one string of ten consecutive heads will appear. The test—looking for a run of ten heads—can be repeated far too often.

148. In our pass rate example, the p-value of the test is approximated by a certain area under the normal curve. The one-tailed procedure uses the "tail area" under the curve to the right of 2, giving $p = .025$ (approximately). The two-tailed procedure contemplates the area to the left of -2, as well as the area to the right of 2. Now there are two tails, and $p = .05$. *See infra* Appendix (figure 13); Freedman et al., *supra* note 16, at 549–52.

According to formal statistical theory, the choice between one tail or two can sometimes be made by considering the exact form of the "alternative hypothesis." *See infra* § IV.C.5. In our example, the null hypothesis is that pass rates are equal for men and women in the whole population of applicants. The alternative hypothesis may exclude a priori the possibility that women have a higher pass rate, and hold that more men will pass than women. This asymmetric alternative suggests a one-tailed test. On the other hand, the alternative hypothesis may simply be that pass rates for men and women in the whole population are unequal. This symmetric alternative admits the possibility that women may score higher than men, and points to a two-tailed test. *See, e.g.*, Freedman et al., *supra* note 16, at 551. Some experts think that the choice between one-tailed and two-tailed tests can often be made by considering the exact form of the null and alternative hypothesis.

149. *See, e.g.*, Baldus & Cole, *supra* note 89, § 9.1, at 308 n.35a; The Evolving Role of Statistical Assessments as Evidence in the Courts, *supra* note 1, at 38–40 (citing EEOC v. Federal Reserve Bank, 698 F.2d 633 (4th Cir. 1983), *rev'd on other grounds sub nom.* Cooper v. Federal Reserve Bank, 467 U.S. 867 (1984)); Kaye, *supra* note 118, at 1358 n.113; David H. Kaye, *The Numbers Game: Statistical Inference in Discrimination Cases*, 80 Mich. L. Rev. 833 (1982) (citing Hazelwood Sch. Dist. v. United States, 433 U.S. 299 (1977)). Arguments for one-tailed tests are discussed in Finkelstein & Levin, *supra* note 1, at 125–26; Richard Goldstein, *Two Types of Statistical Errors in Employment Discrimination Cases*, 26 Jurimetrics J. 32 (1985); Kaye, *supra* at 841.

150. One-tailed tests at the .05 level are viewed as weak evidence—no weaker standard is commonly used in the technical literature.

Such artifacts are commonplace. Since research that fails to uncover significance is not usually published, reviews of the literature may produce an unduly large number of studies finding statistical significance.[151] Even a single researcher may search for so many different relationships that a few will achieve statistical significance by mere happenstance. Almost any large data set—even pages from a table of random digits—will contain some unusual pattern that can be uncovered by a diligent search. Having detected the pattern, the analyst can perform a statistical test for it, blandly ignoring the search effort. Statistical significance is bound to follow. Ten heads in the first ten tosses means one thing; a run of ten heads somewhere along the way in a few thousand tosses of a coin means quite another.

There are statistical methods for coping with multiple looks at the data, which permit the calculation of meaningful *p*-values in certain cases.[152] However, no general solution is available, and the existing methods would be of little help in the typical case where analysts have tested and rejected a variety of regression models before arriving at the one considered the most satisfactory. In these situations, courts should not be overly impressed with claims that estimates are significant. Instead, they should be asking how analysts developed their models.[153]

4. Tests or Interval Estimates?

Statistical significance depends on the *p*-value, and *p*-values depend on sample size. Therefore, a "significant" effect could be small. Conversely, an effect that is "not significant" could be large.[154] By inquiring into the magnitude of an effect, courts can avoid being misled by *p*-values. To focus attention where it belongs—on the actual size of an effect and the reliability of the statistical analysis—interval estimates may be valuable.[155] Seeing a plausible range of values for the quantity of interest helps describe the statistical uncertainty in the estimate.

In our example, the 95% confidence interval for the difference in the pass rates of men and women ranged from 0 to 40 percentage points. Our best

151. *E.g.*, Stuart J. Pocock et al., *Statistical Problems in the Reporting of Clinical Trials: A Survey of Three Medical Journals*, 317 New Eng. J. Med. 426 (1987).

152. *See, e.g.*, Rupert G. Miller, Jr., Simultaneous Statistical Inference (2d ed. 1981).

153. *See, e.g.*, On Model Uncertainty and Its Statistical Implications: Lecture Notes in Econometric and Mathematical Systems (Theo K. Dijkstra ed., 1988); Frank T. Denton, *Data Mining As an Industry*, 67 Rev. Econ. & Stat. 124 (1985). Intuition may suggest that the more variables included in the model, the better. However, this idea often seems to be wrong. Complex models may reflect only accidental features of the data. Standard statistical tests offer little protection against this possibility when the analyst has tried a variety of models before settling on the final specification.

154. *See supra* § IV.B.1.

155. An interval estimate may be composed of a point estimate—like the sample mean used to estimate the population mean—together with its standard error; or the point estimate and standard error can be combined in a confidence interval.

estimate is that the pass rate for men is 20 percentage points higher than for women; and the difference may plausibly be as little as 0 or as much as 40 percentage points. The *p*-value does not yield this information. The confidence interval contains the information provided by a significance test—and more.[156] For instance, significance at the .05 level can be read off the 95% confidence interval.[157] In our example, zero is at the extreme edge of the 95% confidence interval, so we have "significant" evidence that the true difference in pass rates between male and female applicants is not zero. But there are values very close to zero inside the interval.

On the other hand, suppose a significance test fails to reject the null hypothesis. The confidence interval may prevent the mistake of thinking there is positive proof for the null hypothesis. To illustrate, let us change our example slightly: say that 29 men and 20 women passed the test. The 95% confidence interval goes from -2 to 38 percentage points. Because a difference of zero falls within the 95% confidence interval, the null hypothesis—that the true difference is zero—cannot be rejected at the .05 level. But the interval extends to 38 percentage points, indicating that the population difference could be substantial. Lack of significance does not exclude this possibility.[158]

5. What Are the Rival Hypotheses?

The *p*-value of a statistical test is computed on the basis of a model for the data—the null hypothesis. Usually, the test is made in order to argue for the alternative hypothesis—another model. However, on closer examination, both models may prove to be unreasonable.[159] A small *p*-value means something is going on, besides random error; the alternative hypothesis should be viewed as one possible explanation—out of many—for the data.[160]

156. Accordingly, it has been argued that courts should demand confidence intervals (whenever they can be computed) to the exclusion of explicit significance tests and *p*-values. Kaye, *supra* note 118, at 1349 n.78; *cf.* Bailar & Mosteller, *supra* note 140, at 317.

157. Instead of referring to significance at the .05 level, some writers refer to "the 95 percent confidence level that is often used by scientists to reject the possibility that chance alone accounted for observed differences." Carnegie Comm'n on Science, Tech. & Gov't, Science and Technology in Judicial Decision Making: Creating Opportunities and Meeting Challenges 28 (1993).

158. We have used two-sided intervals, corresponding to two-tailed tests. One-sided intervals, corresponding to one-tailed tests, also are available.

159. Often, the null and alternative hypotheses are statements about possible ranges of values for parameters in a common statistical model. *See, e.g.*, *supra* note 148. Computations of standard errors, *p*-values, and power all take place within the confines of this basic model. The statistical analysis looks at the relative plausibility for competing values of the parameters, but makes no global assessment of the reasonableness of the basic model.

160. *See, e.g.*, Paul Meier & Sandy Zabell, *Benjamin Peirce and the Howland Will*, 75 J. Am. Stat. Ass'n 497 (1980) (competing explanations in a forgery case). Outside the legal realm there are many intriguing examples of the tendency to think that a small *p*-value is definitive proof of an alternative hypothesis, even though there are other plausible explanations for the data. *See, e.g.*, Freedman et al., *supra* note 16, at 562–63; C.E.M. Hansel, ESP: A Scientific Evaluation (1966).

In *Mapes Casino, Inc. v. Maryland Casualty Co.*,[161] for example, the court recognized the importance of explanations that the proponent of the statistical evidence had failed to consider. In this action to collect on an insurance policy, Mapes Casino sought to quantify the amount of its loss due to employee defalcation. The casino argued that certain employees were using an intermediary to cash in chips at other casinos. It established that over an 18-month period, the win percentage at its craps tables was 6%, compared to an expected value of 20%. The court recognized that the statistics were probative of the fact that *something* was wrong at the craps tables—the discrepancy was too big to explain as the mere product of random chance. But it was not convinced by plaintiff's alternative hypothesis. The court pointed to other possible explanations (Runyonesque activities like "skimming," "scamming," and "crossroading") that might have accounted for the discrepancy without implicating the suspect employees.[162] In short, rejection of the null hypothesis does not leave the proffered alternative hypothesis as the only viable explanation for the data.[163]

In many studies, the validity of the model is secured by the procedures used to collect the data. There are formulas for standard errors and confidence intervals that hold when random samples are used. *See supra* §§ II.B, IV.A.2. There are statistical tests for comparing two random samples, or evaluating the results of a randomized experiment. *See supra* §§ II.A, IV.B.2. In such examples, the statistical procedures flow from the sampling method and the design of the study. On the other hand, if samples of convenience are used, or subjects are not randomized, the validity of the statistical procedures can be contested. *See* Freedman et al., *supra* note 16, at 387–88, 424, 557–65.

161. 290 F. Supp. 186 (D. Nev. 1968).

162. *Id.* at 193. "Skimming" consists of "taking off the top before counting the drop," "scamming" is "cheating by collusion between dealer and player," and "crossroading" involves "professional cheaters among the players." *Id.* In plainer language, the court seems to have ruled that the casino itself might be cheating, or there could have been cheaters other than the particular employees identified in the case. At the least, plaintiff's statistical evidence did not rule out such possibilities.

163. *Compare* EEOC v. Sears, Roebuck & Co., 839 F.2d 302, 312 & n.9, 313 (7th Cir. 1988) (EEOC's regression studies showing significant differences did not establish liability because surveys and testimony supported the rival hypothesis that women generally had less interest in commission sales positions), *with* EEOC v. General Tel. Co., 885 F.2d 575 (9th Cir. 1989) (unsubstantiated rival hypothesis of "lack of interest" in "non-traditional" jobs insufficient to rebut prima facie case of gender discrimination); *cf. supra* § II.A (problem of confounding); *infra* note 230 (effect of omitting important variables from a regression model).

D. Posterior Probabilities

Standard errors, p-values, and significance tests are common techniques for assessing random error. These procedures rely on the sample data, and are justified in terms of the "operating characteristics" of the statistical procedures.[164] However, this frequentist approach does not permit the statistician to compute the probability that a particular hypothesis is correct, given the data.[165] For instance, a frequentist may postulate that a coin is fair: it has a 50-50 chance of landing heads, and successive tosses are independent; this is viewed as an empirical statement—potentially falsifiable—about the coin. On this basis, it is easy to calculate the chance that the coin will turn up heads in the next ten tosses:[166] the answer is 1/1,024. Therefore, observing ten heads in a row brings into serious question the initial hypothesis of fairness. Rejecting the hypothesis of fairness when there are ten heads in ten tosses gives the wrong result—when the coin is fair—only one time in 1,024. That is an example of an operating characteristic of a statistical procedure.

But what of the converse probability: if a coin lands heads ten times in a row, what is the chance that it is fair?[167] To compute such converse probabilities, it is necessary to postulate initial probabilities that the coin is fair, as well as probabilities of unfairness to various degrees.[168] And that is beyond the scope of frequentist statistics.[169]

164. "Operating characteristics" are the expected value and standard error of estimators, probabilities of error for statistical tests, and related quantities.

165. *See supra* § IV.B.1; *infra* Appendix. Consequently, quantities such as p-values or confidence levels cannot be compared directly to numbers like .95 or .50 that might be thought to quantify the burden of persuasion in criminal or civil cases. *See* Kaye, *supra* note 144; D.H. Kaye, *Apples and Oranges: Confidence Coefficients and the Burden of Persuasion*, 73 Cornell L. Rev. 54 (1987).

166. Stated slightly more formally, if the coin is fair and each outcome is independent (the hypothesis), then the probability of observing ten heads (the data) is $\Pr(\text{data} \mid H_0) = (1/2)^{10} = 1/1{,}024$, where H_0 stands for the hypothesis that the coin is fair.

167. We call this a "converse probability" because it is of the form $\Pr(H_0 \mid \text{data})$ rather than $\Pr(\text{data} \mid H_0)$; an equivalent phrase, "inverse probability," also is used. The tendency to think of $\Pr(\text{data} \mid H_0)$ as if it were the converse probability $\Pr(H_0 \mid \text{data})$ is the "transposition fallacy." For instance, most United States senators are men, but very few men are senators. Consequently, there is a high probability that an individual who is a senator is a man, but the probability that an individual who is a man is a senator is practically zero. For examples of the transposition fallacy in court opinions, see cases cited *supra* note 142. *See also* Committee on DNA Forensic Science: An Update, *supra* note 60, at 133 (describing the fallacy in cases involving DNA identification evidence as the "prosecutor's fallacy"). The frequentist p-value, $\Pr(\text{data} \mid H_0)$, is generally not a good approximation to the Bayesian $\Pr(H_0 \mid \text{data})$; the latter includes considerations of power and base rates.

168. *See infra* Appendix.

169. In some situations, the probability of an event on which a case depends can be computed with objective methods. However, these events are measurable outcomes (like the number of heads in a series of tosses of a coin) rather than hypotheses about the process that generated the data (like the claim that the coin is fair). For example, in *United States v. Shonubi*, 895 F. Supp. 460 (E.D.N.Y. 1995), *rev'd*,

In the Bayesian or subjectivist approach, probabilities represent subjective degrees of belief rather than objective facts. The observer's confidence in the hypothesis that a coin is fair, for example, is expressed as a number between zero and one;[170] likewise, the observer must quantify beliefs about the chance that the coin is unfair to various degrees—all in advance of seeing the data.[171] These subjective probabilities, like the probabilities governing the tosses of the coin, are set up to obey the axioms of probability theory. The probabilities for the various hypotheses about the coin, specified before data collection, are called prior probabilities.

These prior probabilities can then be updated, using "Bayes' rule," given data on how the coin actually falls.[172] In short, Bayesian statisticians can compute posterior probabilities for various hypotheses about the coin, given the data.[173] Although such posterior probabilities can pertain directly to hypotheses of legal interest, they are necessarily subjective, for they reflect not just the data but also

103 F.3d 1085 (2d Cir. 1997), a government expert estimated for sentencing purposes the total quantity of heroin that a Nigerian defendant living in New Jersey had smuggled (by swallowing heroin-filled balloons) in the course of eight trips to and from Nigeria. He applied a method known as "resampling" or "bootstrapping." Specifically, he drew 100,000 independent simple random samples of size seven from a population of weights distributed as in customs data on 117 other balloon swallowers caught in the same airport during the same time period; he discovered that for 99% of these samples, the total weight was at least 2090.2 grams. 895 F. Supp. at 504. Thus, the researcher reported that "there is a 99% chance that Shonubi carried at least 2090.2 grams of heroin on the seven [prior] trips . . . ." *Id.* However, the Second Circuit reversed this finding for want of "specific evidence of what Shonubi had done." 103 F.3d at 1090. Although the logical basis for this "specific evidence" requirement is unclear, a difficulty with the expert's analysis is apparent. Statistical inference generally involves an extrapolation from the units sampled to the population of all units. Thus, the sample needs to be representative. In *Shonubi*, the government used a sample of weights, one for each courier on the trip at which that courier was caught. It sought to extrapolate from these data to many trips taken by a single courier—trips on which that other courier was not caught.

170. Here "confidence" has the meaning ordinarily ascribed to it rather than the technical interpretation applicable to a frequentist "confidence interval." Consequently, it can be related to the burden of persuasion. *See* Kaye, *supra* note 165.

171. For instance, let p be the unknown probability that coin lands heads: What is the chance that p exceeds .6? The Bayesian statistician must be prepared to answer all such questions. Bayesian procedures are sometimes defended on the ground that the beliefs of any rational observer must conform to the Bayesian rules. However, the definition of "rational" is purely formal. *See* Peter C. Fishburn, *The Axioms of Subjective Probability*, 1 Stat. Sci. 335 (1986); David Kaye, *The Laws of Probability and the Law of the Land*, 47 U. Chi. L. Rev. 34 (1979).

172. *See infra* Appendix.

173. *See generally* George E.P. Box & George C. Tiao, Bayesian Inference in Statistical Analysis (Wiley Classics Library ed., John Wiley & Sons, Inc. 1992) (1973). For applications to legal issues, see, e.g., Aitken et al., *supra* note 45, at 337–48; David H. Kaye, *DNA Evidence: Probability, Population Genetics, and the Courts*, 7 Harv. J.L. & Tech. 101 (1993).

the subjective prior probabilities—that is, the degrees of belief about the various hypotheses concerning the coin specified prior to obtaining the data.[174]

Such analyses have rarely been used in court,[175] and the question of their forensic value has been aired primarily in the academic literature.[176] Some statisticians favor Bayesian methods,[177] and some legal commentators have proposed their use in certain kinds of cases in certain circumstances.[178]

V. Correlation and Regression

Regression models are often used to infer causation from association; for example, such models are frequently introduced to prove disparate treatment in discrimination cases, or to estimate damages in antitrust actions. Section V.D explains the ideas and some of the pitfalls. Sections V.A–C cover some preliminary material, showing how scatter diagrams, correlation coefficients, and regression lines can be used to summarize relationships between variables.

174. In this framework, the question arises of whose beliefs to use—the statistician's or the factfinder's. *See, e.g.*, Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 Harv. L. Rev. 489 (1970) (proposing that experts give posterior probabilities for a wide range of prior probabilities, to allow jurors to use their own prior probabilities or just to judge the impact of the data on possible values of the prior probabilities). *But see* Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 Harv. L. Rev. 1329 (1971) (arguing that efforts to describe the impact of evidence on a juror's subjective probabilities would unduly impress jurors and undermine the presumption of innocence and other legal values).

175. The exception is paternity litigation; when genetic tests are indicative of paternity, testimony as to a posterior "probability of paternity" is common. *See, e.g.*, 1 Modern Scientific Evidence: The Law and Science of Expert Testimony, *supra* note 3, § 19–2.5.

176. *See, e.g.*, Probability and Inference in the Law of Evidence: The Uses and Limits of Bayesianism (Peter Tillers & Eric D. Green eds., 1988); Symposium, *Decision and Inference in Litigation*, 13 Cardozo L. Rev. 253 (1991). The Bayesian framework probably has received more acceptance in explicating legal concepts such as the relevance of evidence, the nature of prejudicial evidence, probative value, and burdens of persuasion. *See, e.g.*, Richard D. Friedman, *Assessing Evidence*, 94 Mich. L. Rev. 1810 (1996) (book review); Richard O. Lempert, *Modeling Relevance*, 75 Mich. L. Rev. 1021 (1977); D.H. Kaye, *Clarifying the Burden of Persuasion: What Bayesian Decision Rules Do and Do Not Do*, 3 Int'l J. Evidence & Proof 1 (1999).

177. *E.g.*, Donald A. Berry, *Inferences Using DNA Profiling in Forensic Identification and Paternity Cases*, 6 Stat. Sci. 175, 180 (1991); Stephen E. Fienberg & Mark J. Schervish, *The Relevance of Bayesian Inference for the Presentation of Statistical Evidence and for Legal Decisionmaking*, 66 B.U. L. Rev. 771 (1986). Nevertheless, many statisticians question the general applicability of Bayesian techniques: The results of the analysis may be substantially influenced by the prior probabilities, which in turn may be quite arbitrary. *See, e.g.*, Freedman, *supra* note 112.

178. *E.g.*, Joseph C. Bright, Jr. et al., *Statistical Sampling in Tax Audits*, 13 L. & Soc. Inquiry 305 (1988); Ira Mark Ellman & David Kaye, *Probabilities and Proof: Can HLA and Blood Group Testing Prove Paternity?*, 54 N.Y.U. L. Rev. 1131 (1979); Finkelstein & Fairley, *supra* note 174; Kaye, *supra* note 173.

A. Scatter Diagrams

The relationship between two variables can be graphed in a scatter diagram.[179] Data on income and education for a sample of 350 men, ages 25 to 29, residing in Texas[180] provide an example. Each person in the sample corresponds to one dot in the diagram. As indicated in Figure 4, the horizontal axis shows the person's education, and the vertical axis shows his income. Person A completed 8 years of schooling (grade school) and had an income of $19,000. Person B completed 16 years of schooling (college) and had an income of $38,000.

Figure 4. Plotting a scatter diagram. The horizontal axis shows educational level and the vertical axis shows income.

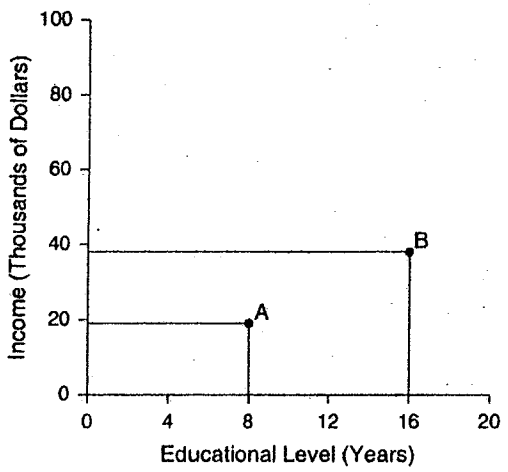
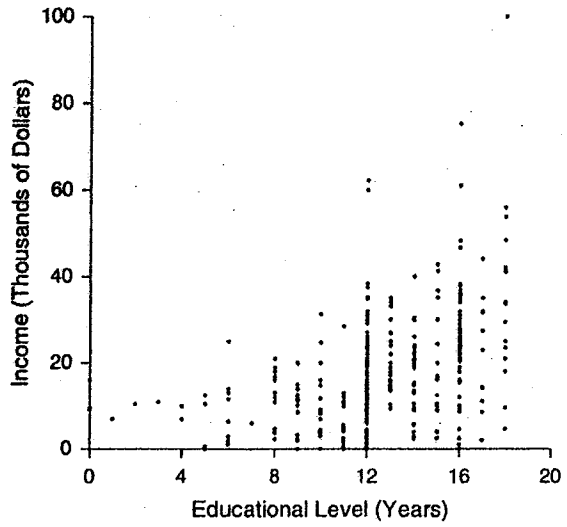


Figure 5 is the scatter diagram for the Texas data. The diagram confirms an obvious point. There is a "positive association" between income and education: in general, persons with a higher educational level have higher incomes. However, there are many exceptions to this rule, and the association is not as strong as one might expect.

179. These diagrams are also referred to as scatterplots or scattergrams.

180. These data are from a public-use data tape, Bureau of the Census, U.S. Dep't of Commerce, for the March 1988 Current Population Survey. Income and education (years of schooling completed) are self-reported. Income is truncated at $100,000 and education at 18 years.

Figure 5. Scatter diagram for income and education: men age 25 to 29 in Texas.[181]



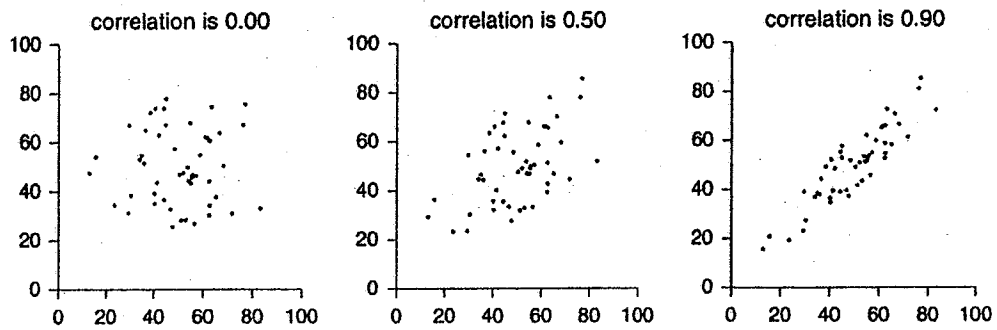
B. Correlation Coefficients

Two variables are positively correlated when their values tend to go up or down together.[182] Income and education in Figure 5 provides an example. The correlation coefficient (usually denoted by the letter r) is a single number that reflects the strength of an association. Figure 6 shows the values of r for three scatter diagrams.

181. Education may be compulsory, but the Current Population Survey generally finds a small percentage of respondents who report very little schooling. Such respondents will be found at the lower left corner of the scatter diagram.

182. Many statistics and displays are available to investigate association. The most common are the correlation coefficient and the scatter diagram.

Figure 6. The correlation coefficient measures the strength of linear association.



A correlation coefficient of 0 indicates no linear association between the variables, while a coefficient of +1 indicates a perfect linear relationship: all the dots in the scatter diagram fall on a straight line that slopes up. The maximum value for r is +1. Sometimes, there is a negative association between two variables: large values of one tend to go with small values of the other. The age of a car and its fuel economy in miles per gallon provide an example. Negative association is indicated by negative values for r. The extreme case is an r of −1, indicating that all the points in the scatter diagram lie on a straight line which slopes down.

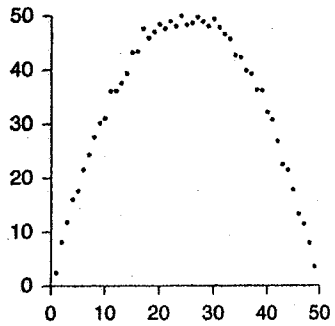
Moderate associations are the general rule in the social sciences; correlations larger than, say, 0.7 are quite unusual in many fields. For example, the correlation between college grades and first-year law school grades is under 0.3 at most law schools, while the correlation between LSAT scores and first-year law grades is generally about 0.4.[183] The correlation between heights of fraternal twins is about 0.5, while the correlation between heights of identical twins is about 0.95. In Figure 5, the correlation between income and education was 0.43. The correlation coefficient cannot capture all the underlying information. Several issues may arise in this regard, and we consider them in turn.

183. Linda F. Wightman, Predictive Validity of the LSAT: A National Summary of the 1990–1992 Correlation Studies 10 (1993); *cf.* Linda F. Wightman & David G. Muller, An Analysis of Differential Validity and Differential Prediction for Black, Mexican-American, Hispanic, and White Law School Students 11–13 (1990). A combination of LSAT and undergraduate grade point average has a higher correlation with first-year law school grades than either item alone. The multiple correlation coefficient is typically about 0.5. Wightman, *supra*, at 10.

1. Is the Association Linear?

The correlation coefficient is designed to measure linear association. Figure 7 shows a strong nonlinear pattern with a correlation close to zero. When the scatter diagram reveals a strong nonlinear pattern, the correlation coefficient may not be a useful summary statistic.

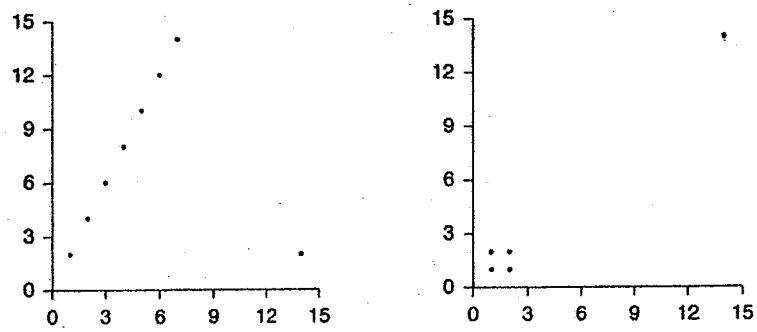
Figure 7. The correlation coefficient only measures linear association. The scatter diagram shows a strong nonlinear association with a correlation coefficient close to zero.



2. Do Outliers Influence the Correlation Coefficient?

The correlation coefficient can be distorted by outliers—a few points that are far removed from the bulk of the data. The left hand panel in Figure 8 shows that one outlier (lower right hand corner) can reduce a perfect correlation to nearly nothing. Conversely, the right hand panel shows that one outlier (upper right hand corner) can raise a correlation of zero to nearly one.

Figure 8. The correlation coefficient can be distorted by outliers. The left hand panel shows an outlier (in the lower right hand corner) that destroys a nearly perfect correlation. The right hand panel shows an outlier (in the upper right hand corner) that changes the correlation from zero to nearly one.



3. Does a Confounding Variable Influence the Coefficient?

The correlation coefficient measures the association between two variables. Investigators—and the courts—are usually more interested in causation. Association is not necessarily the same as causation. As noted in section II.A, the association between two variables may be driven largely by a "third variable" that has been omitted from the analysis. For an easy example, among school children, there is an association between shoe size and vocabulary. However, learning more words does not cause feet to get bigger, and swollen feet do not make children more articulate. In this case, the third variable is easy to spot—age. In more realistic examples, the driving variable may be harder to identify.

Technically, third variables are called confounders or confounding variables.[184] The basic methods of dealing with confounding variables involve controlled experiments[185] or the application, typically though a technique called "multiple regression,"[186] of "statistical controls."[187] In many examples, association really does reflect causation, but a large correlation coefficient is not enough to warrant such a conclusion. A large value of r only means that the dependent variable

184. *See supra* § II.A.1.

185. *See supra* § II.A.2.

186. Multiple regression analysis is discussed *infra* § V.D and again in Daniel L. Rubinfeld, Reference Guide on Multiple Regression, § II, in this manual.

187. For the reasons stated *supra* § II.A, efforts to control confounding in observational studies are generally less convincing than randomized controlled experiments.

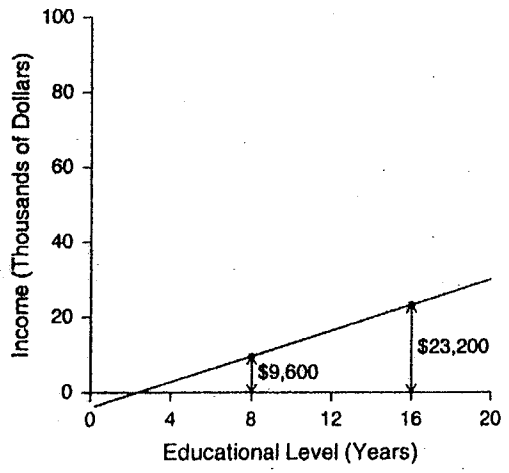
marches in step with the independent one—for any number of possible reasons, ranging from causation to confounding.[188]

C. Regression Lines

The regression line can be used to describe a linear trend in the data. The regression line for income on education in the Texas sample is shown in Figure 9. The height of the line estimates the average income for a given educational level. For example, the average income for people with eight years of education is estimated at $9,600, indicated by the height of the line at eight years; the average income for people with sixteen years of education is estimated at about $23,200.

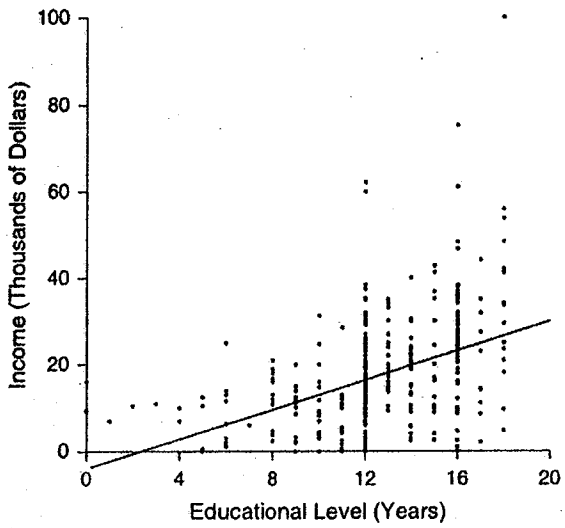
Figure 10 repeats the scatter diagram for income and education (see Figure 5); the regression line is plotted too. In a general way, the line shows the average trend of income as education increases. Thus, the regression line indicates the extent to which a change in one variable (income) is associated with a change in another variable (education).

Figure 9. The regression line for income on education, and its estimates.



188. The square of the correlation coefficient, r^2, is sometimes called the proportion of variance "explained." However, "explained" is meant in a purely technical sense, and large values of r^2 need not point to a causal explanation.

Figure 10. Scatter diagram for income and education, with the regression line indicating the trend.



1. *What Are the Slope and Intercept?*

The regression line can be described in terms of its slope and intercept.[189] In Figure 10, the slope is $1,700 per year. On average, each additional year of education is associated with an additional $1,700 of income. The intercept is –$4,000. This is an estimate of the average income for persons with zero years of education. The estimate is not a good one, for such persons are far from the center of the diagram. In general, estimates based on the regression line become less trustworthy as we move away from the bulk of the data.

The slope has the same limitations as the correlation coefficient in measuring the degree of association:[190] (1) It only measures linear relationships; (2) it may

189. The regression line, like any straight line, has an equation of the form $y = mx + b$. Here, m is the slope, that is, the change in y per unit change in x. The slope is the same anywhere along the line. Mathematically, that is what distinguishes straight lines from curves. The intercept b is the value of y when x is zero. The slope of a line is akin to the grade of a road; the intercept gives the starting elevation. In Figure 9, the regression line estimates an average income of $23,200 for people with 16 years of education. This may be computed from the slope and intercept as follows:

($1,700 per year) × 16 years – $4,000 = $27,200 – $4,000 = $23,200

190. In fact, the correlation coefficient is the slope of the regression line if the variables are "standardized," that is, measured in terms of standard deviations away from the mean.

be influenced by outliers; and (3) it does not control for the effect of other variables. With respect to (1), the slope of $1,700 per year presents each additional year of education as having the same value, but some years of schooling surely are worth more and others less. With respect to (3), the association between education and income graphed in Figure 10 is partly causal, but there are other factors to consider, including the family backgrounds of the people in the sample. For instance, people with college degrees probably come from richer and better educated families than those who drop out after grade school. College graduates have other advantages besides the extra education. Factors like these must have some effect on income. That is why statisticians use the qualified language of "on average" and "associated with."[191]

2. *What Is the Unit of Analysis?*

If association between the characteristics of individuals is of interest, these characteristics should be measured on individuals. Sometimes the individual data are not available, but rates or averages are; correlations computed from rates or averages are termed "ecological." However, ecological correlations generally overstate the strength of an association. An example makes the point. The average income and average education can be determined for the men living in each state. The correlation coefficient for these 50 pairs of averages turns out to be 0.66. However, states do not go to school and do not earn incomes. People do. The correlation for income and education for all men in the United States is only about 0.44.[192] The correlation for state averages overstates the correlation for individuals—a common tendency for such ecological correlations.[193]

Ecological correlations are often used in cases claiming a dilution in the voting strength of a racial minority. In this type of voting rights case plaintiffs must prove three things: (1) the minority group constitutes a majority in at least one district of a proposed plan; (2) the minority group is politically cohesive, that is, votes fairly solidly for its preferred candidate; and (3) the majority group votes sufficiently as a bloc to defeat the minority-preferred candidate.[194] The first test is called compactness. The second and third tests deal with racially polarized voting.

191. Many investigators would use multiple regression to isolate the effects of one variable on another—for instance, the independent effect of education on income. Such efforts may run into problems. *See generally supra* § II.A, *infra* § V.D.

192. Correlations are computed from a public-use data tape, Bureau of the Census, Dep't of Commerce, for the March 1993 Current Population Survey.

193. The ecological correlation uses only the average figures, but within each state there is a lot of spread about the average. The ecological correlation overlooks this individual variation.

194. *See* Thornburg v. Gingles, 478 U.S. 30, 50–51 (1986) ("First, the minority group must be able to demonstrate that it is sufficiently large and geographically compact to constitute a majority in a single-member district. . . . Second, the minority group must be able to show that it is politically

Of course, the secrecy of the ballot box means that racially polarized voting cannot be directly observed.[195] Instead, plaintiffs in these voting rights cases rely on scatter diagrams and regression lines to estimate voting behavior by racial or ethnic groups. The unit of analysis is typically the precinct; hence, the technique is called "ecological regression." For each precinct, public records may suffice to determine the percentage of registrants in each racial or ethnic group, as well as the percentage of the total vote for each candidate—by voters from all demographic groups combined. The statistical issue, then, is to estimate how each demographic subgroup voted.

Figure 11 provides an example. Each point in the scatter diagram shows data for a precinct in the 1982 Democratic primary election for auditor in Lee County, South Carolina. The horizontal axis shows the percentage of registrants who are white. The vertical axis shows the "turnout rate" for the white candidate.[196] The regression line is plotted too. In this sort of diagram, the slope is often interpreted as the difference between the white turnout rate and the black turnout rate for the white candidate; the intercept would be interpreted as the black turnout rate for the white candidate.[197] However, the validity of such estimates is contested in statistical literature.[198]

cohesive. . . . Third, the minority must be able to demonstrate that the white majority votes sufficiently as a bloc to enable it . . . usually to defeat the minority's preferred candidate."). In subsequent cases, the Court has emphasized that these factors are not sufficient to make out a violation of section 2 of the Voting Rights Act. *E.g.*, Johnson v. De Grandy, 512 U.S. 997, 1011 (1994) ("*Gingles* . . . clearly declined to hold [these factors] sufficient in combination, either in the sense that a court's examination of relevant circumstances was complete once the three factors were found to exist, or in the sense that the three in combination necessarily and in all circumstances demonstrated dilution.").

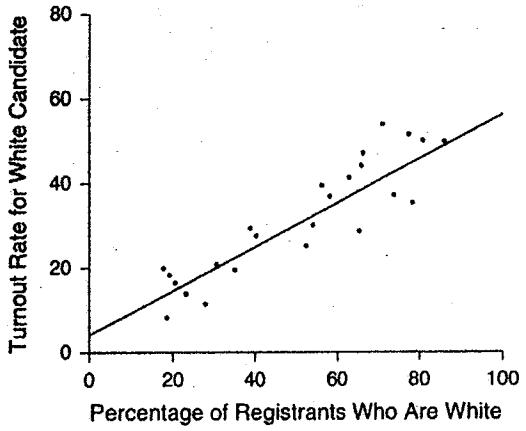
195. Some information could be obtained from exit polls. *E.g.*, Aldasoro v. Kennerson, 922 F. Supp. 339, 344 (S.D. Cal. 1995).

196. By definition, the turnout rate equals the number of votes for the candidate, divided by the number of registrants; the rate is computed separately for each precinct.

197. Figure 11 contemplates only one white candidate; more complicated techniques could be used if there were several candidates of each race. The intercept of the line is 4% and the slope is .52. Plaintiffs would conclude that only 4% of the black registrants voted for the white candidate, while 4% + 52% = 56% of the white registrants voted for the white candidate, which demonstrates polarization.

198. For further discussion of the problem of ecological regression in this context, see Stephen P. Klein & David A. Freedman, *Ecological Regression in Voting Rights Cases*, Chance, Summer 1993, at 38; Bernard Grofman & Chandler Davidson, Controversies in Minority Voting: The Voting Rights Act in Perspective (1992). The use of ecological regression increased considerably after the Supreme Court noted in *Thornburg v. Gingles*, 478 U.S. 30, 53 n.20 (1986), that "[t]he District Court found both methods [extreme case analysis and bivariate ecological regression analysis] standard in the literature for the analysis of racially polarized voting." *See, e.g.*, Teague v. Attala County, 92 F.3d 283, 285 (5th Cir. 1996) (one of "two standard methods for analyzing electoral data"); Houston v. Lafayette County, 56 F.3d 606, 612 (5th Cir. 1995) (holding that district court erred in ignoring ecological regression results). Nevertheless, courts have cautioned against "overreliance on bivariate ecological regression" in light of the inherent limitations of the technique (Lewis v. Alamance County, 99 F.3d 600, 604 n.3 (4th Cir. 1996)), and some courts have found ecological regressions unconvincing. *E.g.*, Aldasoro v. Kennerson,

Figure 11. Turnout rate for the white candidate plotted against the percentage of registrants who are white. Precinct-level data, 1982 Democratic Primary for Auditor, Lee County, South Carolina.[199]



D. *Statistical Models*

Statistical models are widely used in the social sciences and in litigation.[200] For example, the census suffers an undercount, more severe in certain places than others; if some statistical models are to be believed, the undercount can be corrected—moving seats in Congress and millions of dollars a year in entitlement funds.[201] Other models purport to lift the veil of secrecy from the ballot

922 F. Supp. 339 (S.D. Cal. 1995); Romero v. City of Pomona, 665 F. Supp. 853, 860 (C.D. Cal. 1987), *aff'd*, 883 F.2d 1418 (9th Cir. 1989); *cf.* Johnson v. Miller, 864 F. Supp. 1354, 1390 (S.D. Ga. 1994) ("mind-numbing and contradictory statistical data," including bivariate ecological regression, established "that some degree of vote polarization exists, but not in alarming quantities. Exact levels are unknowable."), *aff'd*, 515 U.S. 900 (1995).

Redistricting plans based predominantly on racial considerations are unconstitutional unless narrowly tailored to meet a compelling state interest. Shaw v. Reno, 509 U.S. 630 (1993). Whether compliance with the Voting Rights Act can be considered a compelling interest is an open question, but efforts to sustain racially motived redistricting on this basis have not fared well before the Supreme Court. *See* Abrams v. Johnson, 521 U.S. 74 (1997); Shaw v. Hunt, 517 U.S. 899 (1996); Bush v. Vera, 517 U.S. 952 (1996).

199. Data from James W. Loewen & Bernard Grofman, *Recent Developments in Methods Used in Vote Dilution Litigation*, 21 Urb. Law. 589, 591 tbl.1 (1989).

200. The frequency with which regression models are used is no guarantee that they are the best choice for a particular problem. *See, e.g.*, David W. Peterson, *Reference Guide on Multiple Regression*, 36 Jurimetrics J. 213, 214–15 (1996) (review essay). On the factors that might justify the choice of a particular model, see Moses, *supra* note 124.

201. *See supra* note 43.

box, enabling the experts to determine how racial or ethnic groups have voted—a crucial step in litigation to enforce minority voting rights.[202] This section discusses the statistical logic of regression models.[203]

A regression model attempts to combine the values of certain variables (the independent variables) in order to get expected values for another variable (the dependent variable). The model can be expressed in the form of a regression equation. A simple regression equation has only one independent variable; a multiple regression equation has several independent variables. Coefficients in the equation will often be interpreted as showing the effects of changing the corresponding variables. Sometimes, this interpretation can be justified. For instance, Hooke's law describes how a spring stretches in response to the load hung from it: strain is proportional to stress.[204] There will be a number of observations on a spring. For each observation, the physicist hangs a weight on the spring, and measures its length. A statistician could apply a regression model to these data: for quite a large range of weights,[205]

$$\text{length} = a + b \times \text{weight} + \varepsilon. \qquad (1)$$

The error term, denoted by the Greek letter epsilon (ε), is needed because measured length will not be exactly equal to $a + b \times$ weight. If nothing else, measurement error must be reckoned with. We model ε as a draw made at random with replacement from a box of tickets. Each ticket shows a potential error, which will be realized if that ticket is drawn. The average of all the potential errors in the box is assumed to be zero. In more standard statistical terminology, the εs for different observations are assumed to be "independent and identically distributed, with mean zero."[206]

In equation (1), a and b are parameters, unknown constants of nature that characterize the spring: a is the length of the spring under no load, and b is elasticity, the increase in length per unit increase in weight.[207] These parameters

202. *See supra* § V.C.2.

203. For a more detailed treatment, see Daniel L. Rubinfeld, Reference Guide on Multiple Regression at app., in this manual.

204. This law is named after Robert Hooke (England, 1653–1703).

205. The dependent or response variable in equation (1) is the length of the spring, on the left hand side of the equation. There is one independent or explanatory variable on the right hand side—weight. Since there is only one explanatory variable, equation (1) is a simple regression equation.

Hooke's law is only an approximation, although it is a very good one. With large enough weights, a quadratic term will be needed in equation (1). Moreover, beyond some point, the spring exceeds its elastic limit and snaps.

206. For some purposes, it is also necessary to assume that the errors follow the normal distribution.

207. *Cf. supra* note 121 (defining the term "parameter").

are not observable,[208] but they can be estimated by "the method of least squares."[209] In statistical notation, estimates are often denoted by hats; thus, $\hat{a}$ is the estimate for a, and $\hat{b}$ is the estimate for b.[210] Basically, the values of $\hat{a}$ and $\hat{b}$ are chosen to minimize the sum of the squared "prediction errors."[211] These errors are also called "residuals": they measure the difference between the actual length and the predicted length, the latter being $\hat{a} + \hat{b} \times$ weight.[212]

$$\text{residual} = \text{actual length} - \hat{a} - \hat{b} \times \text{weight} \qquad (2)$$

Of course, no one really imagines there to be a box of tickets hidden in the spring. However, the variability of physical measurements (under many but by no means all circumstances) does seem to be remarkably like the variability in draws from a box.[213] In short, the statistical model corresponds rather closely to the empirical phenomenon.

1. A Social Science Example

We turn now to social science applications of the kind that might be seen in litigation. A case study would take us too far afield, but a stylized example of regression analysis used to demonstrate sex discrimination in salaries may give the idea.[214] We use a regression model to predict salaries (dollars per year) of employees in a firm using three explanatory variables: education (years of schooling completed), experience (years with the firm), and a dummy variable for

208. It might seem that a is observable; after all, one can measure the length of the spring with no load. However, the measurement is subject to error, so one observes not a but $a + \varepsilon$. *See* equation (1). The parameters a and b can be estimated, even estimated very well, but they cannot be observed directly.

209. The method was developed by Adrien-Marie Legendre (France, 1752–1833) and Carl Friedrich Gauss (Germany, 1777–1855) to fit astronomical orbits.

210. Another convention is use Greek letters for the parameters and English letters for the estimates.

211. Given trial values for a and b, one computes residuals as in equation (2), and then the sum of the squares of these residuals. The "least squares" estimates $\hat{a}$ and $\hat{b}$ are the values of a and b that minimize this sum of squares. These least squares values can be computed from the data by a mathematical formula. They are the intercept and slope of the regression line. *See supra* § V.C.1; Freedman et al., *supra* note 16, at 208–10.

212. The residual is observable, but because the estimates $\hat{a}$ and $\hat{b}$ are only approximations to the parameters a and b, the residual is only an approximation to the error term in equation (1). The term "predicted value" is used in a specialized sense, because the actual values are available too; statisticians often refer to "fitted value" rather than "predicted value," to avoid possible misinterpretations.

213. This is Gauss's model for measurement error. *See* Freedman et al., *supra* note 16, at 450–52.

214. For a more extended treatment of the concepts, see Daniel L. Rubinfeld, Reference Guide on Multiple Regression, at app., in this manual.

gender, taking the value 1 for men and 0 for women.[215] The equation is[216]

$$\text{salary} = a + b \times \text{education} + c \times \text{experience} + d \times \text{gender} + \varepsilon \qquad (3)$$

Equation (3) is a statistical model for the data, with unknown parameters a, b, c, and d; here, a is the intercept and the others are regression coefficients; ε is an unobservable error term. This is a formal analog of Hooke's law, shown as equation (1); the same assumptions are made about the errors. In other words, an employee's salary is determined as if by computing

$$a + b \times \text{education} + c \times \text{experience} + d \times \text{gender} \qquad (4)$$

then adding an error drawn at random from a box of tickets. The expression (4) is the expected value for salary given the explanatory variables (education, experience, gender); the error term in equation (3) represents deviations from the expected.

The parameters in equation (3) are estimated from the data using least squares. If the estimated coefficient for the dummy variable turns out to be positive and statistically significant (by a t-test[217]), that would be taken as evidence of disparate impact: men earn more than women, even after adjusting for differences in background factors that might affect productivity. Education and experience are entered into equation (3) as statistical controls, precisely in order to claim that adjustment has been made for differences in backgrounds.

Suppose the estimated equation turns out as follows:

$$\text{predicted salary} = \$7{,}100 + \$1{,}300 \times \text{education} + \$2{,}200 \times \text{experience} + \$700 \times \text{gender} \qquad (5)$$

That is, $\hat{a} = \$7{,}100$, $\hat{b} = \$1{,}300$, and so forth. According to equation (5), every extra year of education is worth on average $1,300; similarly, every extra year of experience is worth on average $2,200; and, most important, the company gives men a salary premium of $700 over women with the same education and expe-

215. A dummy variable takes only two values (e.g., 0 and 1) and serves to identify two mutually exclusive and exhaustive categories.

216. In equation (3), the variable on the left hand side, salary, is the response variable. On the right hand side are the explanatory variables—education, experience, and the dummy variable for gender. Because there are several explanatory variables, this is a multiple regression equation rather than a simple regression equation; *cf. supra* note 205.

Equations like (3) are suggested, somewhat loosely, by "human capital theory." However, there remains considerable uncertainty about which variables to put into the equation, what functional form to assume, and how error terms are supposed to behave. Adding more variables is no panacea. *See* Peterson, *supra* note 200, at 214–15.

217. *See infra* § V.D.2.

rience, on average. For example, a male employee with 12 years of education (high school) and 10 years of experience would have a predicted salary of

$$\$7{,}100 + \$1{,}300 \times 12 + \$2{,}200 \times 10 + \$700 \times 1$$
$$= \$7{,}100 + \$15{,}600 + \$22{,}000 + \$700 = \$45{,}400 \qquad (6)$$

A similarly situated female employee has a predicted salary of only

$$\$7{,}100 + \$1{,}300 \times 12 + \$2{,}200 \times 10 + \$700 \times 0$$
$$= \$7{,}100 + \$15{,}600 + \$22{,}000 + \$0 = \$44{,}700 \qquad (7)$$

Notice the impact of the dummy variable: $700 is added to equation (6), but not to equation (7).

A major step in proving discrimination is establishing that the estimated coefficient of the dummy variable—$700 in our numerical illustration—is statistically significant. This depends on the statistical assumptions built into the model. For instance, each extra year of education is assumed to be worth the same (on average) across all levels of experience, both for men and women. Similarly, each extra year of experience is worth the same across all levels of education, both for men and women. Furthermore, the premium paid to men does not depend systematically on education or experience. Ability, quality of education, or quality of experience are assumed not to make any systematic difference to the predictions of the model.[218]

The assumptions about the error term—that the errors are independent and identically distributed from person to person in the data set—turn out to be critical for computing *p*-values and demonstrating statistical significance. Regression modeling that does not produce statistically significant coefficients is unlikely to establish discrimination, and statistical significance cannot be established unless stylized assumptions are made about unobservable error terms.[219]

The typical regression model is based on a host of such assumptions; without them, inferences cannot be drawn from the model. With Hooke's law—equation (1)—the model rests on assumptions that are relatively easy to validate experimentally. For the salary discrimination model—equation (3)—validation seems more difficult.[220] Court or counsel may well inquire: What are the assumptions behind the model, and why do they apply to the case at bar? In this regard, it is important to distinguish between situations where (1) the nature of the relationship between the variables is known and regression is being used to make quantitative estimates, and (2) where the nature of the relationship is largely unknown and regression is being used to determine the nature of the relation-

218. Technically, these omitted variables are assumed to be uncorrelated with the error term in the equation.

219. *See supra* note 124.

220. Some of the material in this section is taken from Freedman, *supra* note 112, at 29–35.

ship—or indeed whether any relationship exists at all. The statistical basis for regression theory was developed to handle situations of the first type, with Hooke's law being an example. The basis for the second type of application is analogical, and the tightness of the analogy is a critical issue.

2. *Standard Errors, t-statistics, and Statistical Significance*

Statistical proof of discrimination depends on the significance of $\hat{d}$ (the estimated coefficient for gender); significance is determined by the t-test, using the standard error of $\hat{d}$. The standard error of $\hat{d}$ measures the likely difference between $\hat{d}$ and d, the difference being due to the action of the error term in equation (3). The t-statistic is $\hat{d}$ divided by its standard error. For example, in equation (5), $\hat{d}$ = \$700. If the standard error of $\hat{d}$ is \$325, then t = \$700/\$325 = 2.15. This is significant, that is, hard to explain as the mere product of random chance. Under the null hypothesis that $d = 0$, there is only about a 5% chance that the absolute value of t (denoted $|t|$) is greater than 2. A value of t greater than 2 would therefore demonstrate statistical significance.[221] On the other hand, if the standard error is \$1,400, then t = \$700/\$1,400 = 0.5, and the discrepancy could easily result from chance. Of course, the parameter d is only a construct in a model. If the model is wrong, the standard error, t-statistic, and significance level are rather difficult to interpret.

Even if the model is granted, there is a further issue: the 5% is a probability for the data given the model, namely, $P(|t| > 2 \mid d = 0)$. However, the 5% is often misinterpreted as $P(d = 0 \mid \text{data})$. This misinterpretation is commonplace in the social science literature, and it appears in some opinions describing expert testimony.[222] For an objectivist statistician, $P(d = 0 \mid \text{data})$ makes no sense: parameters do not exhibit chance variation. For a subjectivist statistician, $P(d = 0 \mid \text{data})$ makes good sense, but its computation via the t-test could be seriously in error, because the prior probability that $d = 0$ has not been taken into account.[223]

3. *Summary*

The main ideas of regression modeling can be captured in a hypothetical exchange between a plaintiff seeking to prove salary discrimination and a company denying that allegation. Such a dialog might proceed as follows:

1. Plaintiff argues that the defendant company pays male employees more than females, which establishes prima facie case of discrimination.[224]

221. The cutoff at 2 applies to large samples. Small samples require higher thresholds.

222. *See supra* § IV.B and notes 142, 167.

223. For an objectivist, the vertical bar "|" in $P(|t| > 2 \mid d = 0)$ means "computed on the assumption that." For a subjectivist, the bar would signify a conditional probability. *See supra* § IV.B.1, C; *infra* Appendix.

224. The conditions under which a simple disparity between two groups amounts to a prima facie case that shifts the burden of proof to the defendant in Title VII and other discrimination cases have yet

2. The company responds that the men are paid more because they are better educated and have more experience.
3. Plaintiff tries to refute the company's theory by fitting a regression equation like equation (5). Even after adjusting for differences in education and experience, men earn $700 a year more than women, on average. This remaining difference in pay shows discrimination.
4. The company argues that a small difference like $700 could be the result of chance, not discrimination.
5. Plaintiff replies that the coefficient of "gender" in equation (5) is statistically significant, so chance is not a good explanation for the data.

Statistical significance is determined by reference to the observed significance level, which is usually abbreviated to p.[225] The p-value depends not only on the $700 difference in salary levels, but also on the sample size, among other things.[226] The bigger the sample, other things being equal, the smaller is p—and the tighter is plaintiff's argument that the disparity cannot be explained by chance. Often, a cutoff at 5% is used; if p is less than 5%, the difference is "statistically significant."[227]

In some cases, the p-value has been interpreted as the probability that defendants are innocent of discrimination. However, such an interpretation is wrong: p merely represents the probability of getting a large test statistic, given that the model is correct and the true coefficient of "gender" is zero.[228] Therefore, even if the model is undisputed, a p-value less than 50% does not necessarily demonstrate a "preponderance of the evidence" against the null hypothesis. Indeed, a p-value less than 5% or 1% might not meet the preponderance standard.

In employment discrimination cases, and other contexts too, a wide variety of models are used. This is perhaps not surprising, for specific equations are not dictated by the science. Thus, in a strongly contested case, our dialog would be likely to continue with an exchange about which model is better. Although

to be articulated clearly and comprehensively. *Compare* EEOC v. Olson's Dairy Queens, Inc., 989 F.2d 165, 168 (5th Cir. 1993) (reversing district court for failing to find a prima facie case from the EEOC's statistics on the proportion of African-Americans in defendant's workforce as compared to the proportion of food preparation and service workers in the Houston Standard Metropolitan Statistical Area), *with* Wilkins v. University of Houston, 654 F.2d 388 (5th Cir. 1981) (holding that the district court correctly found that plaintiffs' proof of simple disparities in faculty salaries of men and women did not constitute a prima facie case), *vacated and remanded on other grounds*, 459 U.S. 809 (1982), *aff'd on remand*, 695 F.2d 134 (5th Cir. 1983). *See generally*, D.H. Kaye, *Statistical Evidence: How to Avoid the "Diderot Effect" of Getting Stumped*, Inside Litig., Apr. 1988, at 21. Richard Lempert, *Befuddled Judges: Statistical Evidence in Title VII Cases*, *in* Controversies in Civil Rights (Bernard Grofman ed., forthcoming 2000).

225. *See supra* § IV.B.1.

226. The p-value depends on the estimated value of the coefficient and its standard error. These quantities can be computed from (1) the sample size, (2) the means and SDs of the variables, and (3) the correlations between pairs of variables. The computation is rather intricate.

227. *See supra* § IV.B.2.

228. *See supra* §§ IV.B, V.D.2.

statistical assumptions[229] are challenged in court from time to time, arguments more commonly revolve around the choice of variables. One model may be questioned because it omits variables that should be included—for instance, skill levels or prior evaluations;[230] another model may be challenged because it includes "tainted" variables reflecting past discriminatory behavior by the firm.[231] Frequently, each side will have its own equations and its own team of experts; the court then must decide which model—if either—fits the occasion.[232]

229. *See generally supra* § V.D.1 (discussion following equation (7)); Finkelstein & Levin, *supra* note 1, at 397–403; Daniel L. Rubinfeld, Reference Guide on Multiple Regression, in this manual. One example of a statistical assumption is the independence from subject to subject of the error term in equation (3); another example is that the errors have mean zero and constant variance.

230. *E.g.*, Smith v. Virginia Commonwealth Univ., 84 F.3d 672 (4th Cir. 1996) (dispute over omitted variables precludes summary judgment). *Compare* Bazemore v. Friday, 478 U.S. 385 (1986), *on remand*, 848 F.2d 476 (4th Cir. 1988) *and* Sobel v. Yeshiva Univ., 839 F.2d 18, 34 (2d Cir. 1988) (failure to include variables for scholarly productivity did not vitiate plaintiffs' regression study of salary differences because "Yeshiva's experts . . . [offered] no reason, in evidence or analysis, for concluding that they correlated with sex"), *with* Penk v. Oregon State Bd. of Higher Educ., 816 F.2d 458, 465 (9th Cir. 1987) ("Missing parts of the plaintiffs' interpretation of the board's decision-making equation included such highly determinative quality and productivity factors as teaching quality, community and institutional service, and quality of research and scholarship . . . that . . . must have had a significant influence on salary and advancement decisions.") *and* Chang v. University of R.I., 606 F. Supp. 1161, 1207 (D.R.I. 1985) (plaintiff's regression not entitled to substantial weight because the analyst "excluded salient variables even though he knew of their importance").

The same issue arises, of course, with simpler statistical models, such as those used to assess the difference between two proportions. *See, e.g.*, Sheehan v. Daily Racing Form, Inc., 104 F.3d 940, 942 (7th Cir. 1997) ("Completely ignored was the more than remote possibility that age was correlated with a legitimate job-related qualification, such as familiarity with computers. Everyone knows that younger people are on average more comfortable with computers than older people are, just as older people are on average more comfortable with manual-shift cars than younger people are.").

231. Michael O. Finkelstein, *The Judicial Reception of Multiple Regression Studies in Race and Sex Discrimination Cases*, 80 Colum. L. Rev. 737 (1980).

232. *E.g.*, *Chang*, 606 F. Supp. at 1207 ("it is plain to the court that [defendant's] model comprises a better, more useful, more reliable tool than [plaintiff's] counterpart"); Presseisen v. Swarthmore College, 442 F. Supp. 593, 619 (E.D. Pa. 1977) ("[E]ach side has done a superior job in challenging the other's regression analysis, but only a mediocre job in supporting their own . . . and the Court is . . . left with nothing."), *aff'd*, 582 F.2d 1275 (3d Cir. 1978).

Appendix

A. Probability and Statistical Inference

The mathematical theory of probability consists of theorems derived from axioms and definitions. The mathematical reasoning is not controversial, but there is some disagreement as to how the theory should be applied; that is, statisticians may differ on the proper interpretation of probabilities in specific applications. There are two main interpretations. For a subjectivist statistician, probabilities represent degrees of belief, on a scale between 0 and 1. An impossible event has probability 0, an event that is sure to happen has probability 1. For an objectivist statistician, probabilities are not beliefs; rather, they are inherent properties of an experiment. If the experiment can be repeated, then in the long run, the relative frequency of an event tends to its probability. For instance, if a fair coin is tossed, the probability of heads is 1/2; if the experiment is repeated, the coin will land heads about one-half the time. If a fair die is rolled, the probability of getting an ace (one spot) is 1/6; if the die is rolled many times, an ace will turn up about one-sixth of the time.[233] (Objectivist statisticians are also called frequentists, while subjectivists are Bayesians, after the Reverend Thomas Bayes, England, c.1701–1761.)

Statisticians also use conditional probability, that is, the probability of one event given that another has occurred. For instance, suppose a coin is tossed twice. One event is that the coin will land HH. Another event is that at least one H will be seen. Before the coin is tossed, there are four possible, equally likely, outcomes: HH, HT, TH, TT. So the probability of HH is 1/4. However, if we know that at least one head has been obtained, then we can rule out two tails TT. In other words, given that at least one H has been obtained, the conditional probability of TT is 0, and the first three outcomes have conditional probability 1/3 each. In particular, the conditional probability of HH is 1/3. This is usually written as $P(\text{HH} \mid \text{at least one H}) = 1/3$. More generally, the probability of any event B is denoted as $P(B)$; the conditional probability of B given A is written as $P(B \mid A)$.

Two events A and B are independent if the conditional probability of B given that A occurs is equal to the conditional probability of B given that A does not occur. Statisticians often use "$\sim A$" to denote the event that A does not occur, so A and B are independent if $P(B \mid A) = P(B \mid \sim A)$. If A and B are inde-

233. Probabilities may be estimated from relative frequencies, but probability itself is a subtler idea. For instance, suppose a computer prints out a sequence of ten letters H and T (for heads and tails), which alternate between the two possibilities H and T as follows: H T H T H T H T H T. The relative frequency of heads is 5/10 or 50%, but it is not at all obvious that the chance of an H at the next position is 50%.

pendent, then the probability that both occur is equal to the product of the probabilities:

$$P(A \text{ and } B) = P(A) \times P(B) \tag{1}$$

This is the multiplication rule (or product rule) for independent events. If events are dependent, then conditional probabilities must be used:

$$P(A \text{ and } B) = P(A) \times P(B \mid A) \tag{2}$$

This is the multiplication rule for dependent events.

Assessing probabilities, conditional probabilities, and independence is not entirely straightforward. Inquiry into the basis for expert judgment may be useful, and casual assumptions about independence should be questioned.[234]

Bayesian statisticians assign probabilities to hypotheses as well as to events; indeed, for them, the distinction between hypotheses and events may not be a sharp one. If H_0 and H_1 are two hypotheses[235] which govern the probability of an event A, a Bayesian statistician might use the multiplication rule (2) to find that

$$P(A \text{ and } H_0) = P(A \mid H_0)\, P(H_0) \tag{3a}$$

and

$$P(A \text{ and } H_1) = P(A \mid H_1)\, P(H_1) \tag{3b}$$

Reasoning further that $P(A) = P(A \text{ and } H_0) + P(A \text{ and } H_1)$, the statistician would conclude that

$$P(H_0 \mid A) = \frac{P(A \mid H_0)P(H_0)}{P(A \mid H_0)P(H_0) + P(A \mid H_1)P(H_1)} \tag{4}$$

This is a special case of Bayes' rule, which yields the conditional probability of hypothesis H_0 given that event A has occurred. For example, H_0 might be the hypothesis that blood found at the scene of a crime came from a person unrelated to the defendant; H_1 might deny H_0 and assert that the blood came from the defendant; and A could be the event that blood from both the crime scene and the defendant is type A. Then $P(H_0)$ is the prior probability of H_0, based on subjective judgment, while $P(H_0 \mid A)$ is the posterior probability—the prior probability updated using the data. Here, we have observed a match in type A blood,

234. For problematic assumptions of independence in litigation, see, e.g., *Branion v. Gramly*, 855 F.2d 1256 (7th Cir. 1988); *People v. Collins*, 438 P.2d 33 (Cal. 1968); D.H. Kaye, *The Admissibility of "Probability Evidence" in Criminal Trials* (pts. 1 & 2), 26 Jurimetrics J. 343 (1986), 27 Jurimetrics J. 160 (1987).

235. H_0 is read "H-sub-zero," while H_1 is "H-sub-one."

which occurs in about 42% of the population, so $P(A \mid H_0) = 0.42$.[236] Because the defendant has type A blood, the match probability given that the blood came from him is $P(A \mid H_1) = 1$. If the prior probabilities were, say, $P(H_0) = P(H_1) = 0.5$, then according to (4), the posterior probability would be

$$P(H_0 \mid A) = \frac{0.42 \times 0.5}{0.42 \times 0.5 + 1 \times 0.5} = 0.30 \qquad (5)$$

Conversely, the posterior probability that the blood is from the defendant would be

$$P(H_1 \mid A) = 1 - P(H_0 \mid A) = 0.70 \qquad (6)$$

Thus, the data make it more probable that the blood is the defendant's: the probability rises from the prior value of $P(H_1) = 0.50$ to the posterior value of $P(H_1 \mid A) = 0.70$.

A frequentist statistician would be hesitant to quantify the probability of hypotheses like H_0 and H_1. Such a statistician would merely report that if H_0 is true, then the probability of type A blood is 42%, whereas if H_1 is true, the probability is 100%.

More generally, H_0 could refer to parameters in a statistical model. For example, H_0 might specify equal selection rates for a population of male and female applicants; H_1 might deny H_0 and assert that the selection rates are not equal; and A could be the event that a test statistic exceeds 2 in absolute value. In such situations, the frequentist statistician would compute $P(A \mid H_0)$ and reject H_0 if this probability fell below a figure such as 0.05.

B. *Technical Details on the Standard Error, the Normal Curve, and Significance Levels*

This section of the Appendix describes several calculations for the pass rate example of section IV. In that example, the population consisted of all 5,000 men and 5,000 women in the applicant pool. Suppose by way of illustration that the pass rates for these men and women were 60% and 35%, respectively; so the "population difference" is 60% − 35% = 25 percentage points. We chose 50 men at random from the population, and 50 women. In our sample, the pass rate for the men was 58% and the pass rate for the women was 38%, so the sample difference was 58% − 38% = 20 percentage points. Another sample might have pass rates of 62% and 36%, for a sample difference of 62% − 36% = 26 percentage points. And so forth.

236. Not all statisticians would accept the identification of a population frequency with $P(A \mid H_0)$; indeed, H_0 has been translated into a hypothesis that the true donor has been randomly selected from the population, which is a major step needing justification.

In principle, we can consider the set of all possible samples from the population, and make a list of the corresponding differences. This is a long list. Indeed, the number of distinct samples of 50 men and 50 women that can be formed is immense—nearly 5×10^{240}, or 5 followed by 240 zeros. Our sample difference was chosen at random from this list. Statistical theory enables us to make some precise statements about the list, and hence about the chances in the sampling procedure.

- The average of the list—that is, the average of the differences over the 5×10^{240} possible samples—equals the difference between the pass rates of all 5,000 men and 5,000 women. In more technical language, the expected value of the sample difference equals the population difference. Even more tersely, the sample difference is an unbiased estimate of the population difference.
- The standard deviation (SD) of the list—that is, the standard deviation of the differences over the 5×10^{240} possible samples—is equal to[237]

$$\sqrt{\frac{5{,}000 - 50}{5{,}000 - 1}} \times \sqrt{\frac{P_{men}\,(1 - P_{men})}{50} + \frac{P_{women}\,(1 - P_{women})}{50}} \qquad (7)$$

In expression (7), P_{men} stands for the proportion of the 5,000 male applicants who would pass the exam, and P_{women} stands for the corresponding proportion of women. With the 60% and 35% figures we have postulated, the standard deviation of the sample differences would be 9.6 percentage points:

$$\sqrt{\frac{5{,}000 - 50}{5{,}000 - 1}} \times \sqrt{\frac{.60\,(1 - .60)}{50} + \frac{.35\,(1 - .35)}{50}} = .096 \qquad (8)$$

Figure 12 shows the histogram for the sample differences.[238] The graph is drawn so the area between two values gives the relative frequency of sample

237. *See, e.g.*, Freedman et al., *supra* note 16, at 414, 503–04; Moore & McCabe, *supra* note 93, at 590–91. The standard error for the sample difference equals the standard deviation of the list of all possible sample differences, making the connection between standard error and standard deviation. If we drew two samples at random, the difference between them would be on the order of $\sqrt{2} \approx 1.4$ times this standard deviation. The standard error can therefore be used to measure reproducibility of sample data. On the standard deviation, see *supra* § III.E; Freedman et al., *supra* note 16, at 67–72.

238. The "probability histogram" in Figure 12 shows the "distribution" of the sample differences, indicating the relative likelihoods of the various ranges of possible values; likelihood is represented by

differences falling in that range, among all 5×10^{240} possible samples. For instance, take the range from 20 to 30 percentage points. About half the area under the histogram falls into this range. Therefore, given our assumptions, there is about a 50% chance that for a sample of 50 men and 50 women chosen at random, the difference between the pass rates for the sample men and women will be in the range from 20 to 30 percentage points. The "central limit theorem" establishes that the histogram for the sample differences follows the normal curve, at least to a good approximation. Figure 12 shows this curve for comparison.[239] The main point is that chances for the sample difference can be approximated by areas under the normal curve.

Generally, we do not know the pass rates P_{men} and P_{women} in the population. We chose 60% and 35% just by way of illustration. Statisticians would use the pass rates in the sample—58% and 38%—to estimate the pass rates in the population. Substituting the sample pass rates into expression (7) yields

$$\sqrt{\frac{5{,}000 - 50}{5{,}000 - 1}} \times \sqrt{\frac{.58\,(1 - .58)}{50} + \frac{.38\,(1 - .38)}{50}} = .097 \qquad (9)$$

That is about 10 percentage points—the standard error reported in section IV.A.2.[240]

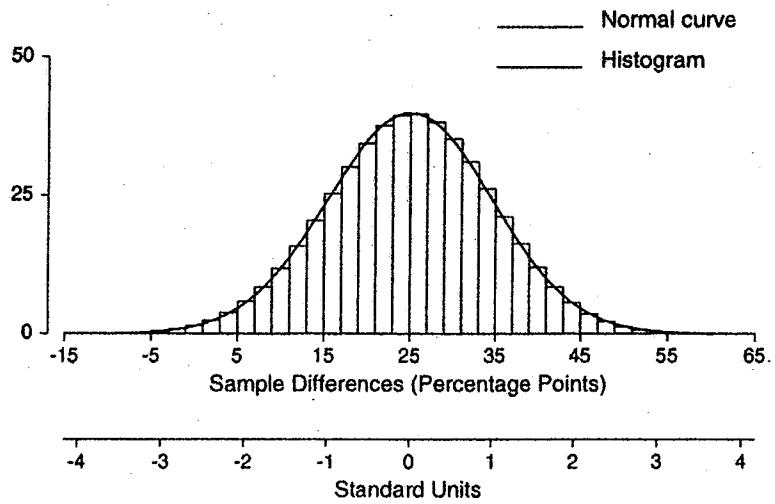
area. The lower horizontal scale shows "standard units," that is, deviations from the expected value relative to the standard error. In our example, the expected value is 25 percentage points and the standard error is 9.6 percentage points. Thus, 35 percentage points would be expressed as (35 – 25)/9.6 = 1.04 standard units. The vertical scale in the figure shows probability per standard unit. Probability is measured on a percentage scale, with 100% representing certainty; the maximum shown on the vertical scale in the figure is 50, i.e., 50% per standard unit. *See* Freedman et al., *supra* note 16, at 80, 315.

239. The normal curve is the famous bell-shaped curve of statistics, whose equation is

$$y = \frac{100\%}{\sqrt{2\pi}}\, e^{-x^2/2}$$

240. There is little difference between (8) and (9)—the standard error does not depend very strongly on the pass rates.

Figure 12. The distribution of the sample difference in pass rates when $P_{men} = 60\%$ and $P_{women} = 35\%$



To sum up, the histogram for the sample differences follows the normal curve, centered at the population difference. The spread is given by the standard error. That is why confidence levels can be based on the standard error, with confidence levels read off the normal curve: 68% of the area under the curve is between −1 and 1, while 95% is between −2 and 2, and 99.7% is between −3 and 3, approximately.

We turn to *p*-values.[241] Consider the null hypothesis that the men and women in the population have the same overall pass rates. In that case, the sample differences are centered at zero, because $P_{men} - P_{women} = 0$. Since the overall pass rate in the sample is 48%, we use this value to estimate both P_{men} and P_{women} in expression (7):

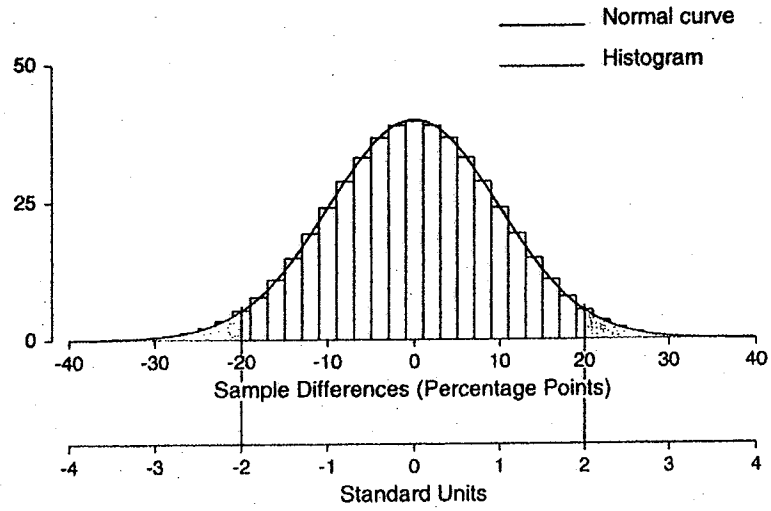
$$\sqrt{\frac{5{,}000 - 50}{5{,}000 - 1}} \times \sqrt{\frac{.48\,(1 - .48)}{50} + \frac{.48\,(1 - .48)}{50}} = .099 \qquad (10)$$

Again, the standard error (SE) is about 10 percentage points. The observed difference of 20 percentage points is 20/10 = 2.0 SEs. As shown in Figure 13, differences of that magnitude or larger have about a 5% chance of occurring:

241. *See supra* § IV.B.1.

About 5% of the area under the normal curve lies beyond ±2. (In Figure 13, this tail area is shaded.) The *p*-value is about 5%.[242]

Figure 13. *p*-value for observed difference of 20 percentage points, computed using the null hypothesis. The chance of getting a sample difference of 20 points in magnitude (or more) is about equal to the area under the normal curve beyond ±2. That shaded area is about 5%.



Finally, we calculate power.[243] We are making a two-tailed test at the .05 level. Instead of the null hypothesis, we assume an alternative: In the applicant pool, 55% of the men would pass, and 45% of the women. So there is a difference of 10 percentage points between the pass rates. The distribution of sample differences would now be centered at 10 percentage points (see Figure 14). Again, the sample differences follow the normal curve. The true SE is about 10

242. Technically, the *p*-value is the chance of getting data as extreme as, or more extreme than, the data at hand. *See supra* § IV.B.1. That is the chance of getting a difference of 20 percentage points or more on the right, together with the chance of getting −20 or less on the left. This chance equals the area under the histogram to the right of 19, together with the area to the left of −19. (The rectangle whose area represents the chance of getting a difference of 20 is included, and likewise for the rectangle above −20.) The area under the histogram may in turn be approximated by the area under the normal curve beyond ± 1.9, which is 5.7%. *See, e.g.*, Freedman et al., *supra* note 16, at 318. Keeping track of the edges of the rectangles is called the "continuity correction." *Id.* The histogram is computed assuming pass rates of 48% for the men and the women. Other values could be dealt with in a similar way. *See infra* note 245.

243. *See supra* note 144.

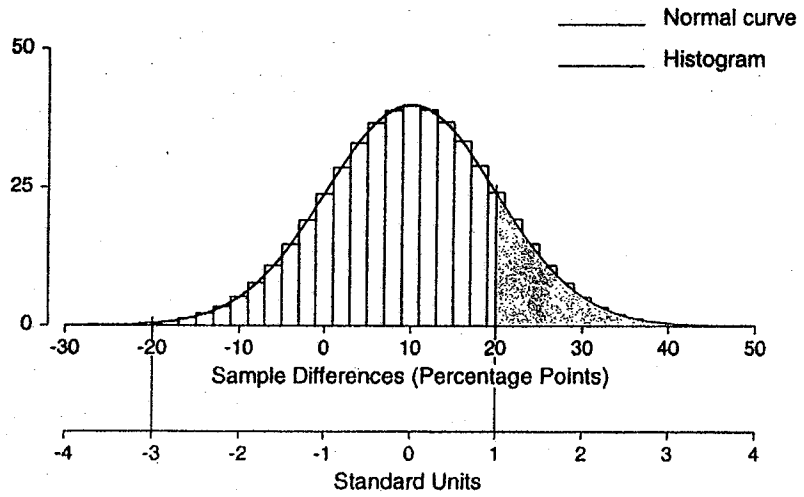
percentage points by equation (1), and the SE estimated from the sample will be about the same. On that basis, only sample differences larger than 20 percentage points or smaller than −20 points will be declared significant.[244] About 1/6 of the area under the normal curve in Figure 14 lies in this region.[245] Therefore, the power of the test against the specified alternative is only about 1/6. In the figure, it is the shaded area that corresponds to power.

Figures 12, 13, and 14 have the same shape: the central limit theorem is at work. However, the histograms are centered differently, because the values of P_{men} and P_{women} are different in all three figures. Figure 12 is centered at 25 percentage points, reflecting our illustrative values of 60% and 35% for the pass rates. Figure 13 is centered at zero, because it is drawn according to the requirements of the null hypothesis. Figure 14 is centered at 10, because the alternative hypothesis is used to determine the center, rather than the null hypothesis.

244. The null hypothesis asserts a difference of zero. In Figure 13, 20 percentage points is 2 SEs to the right of the value expected under the null hypothesis; likewise, −20 is 2 SEs to the left. However, Figure 14 takes the alternative hypothesis to be true; on that basis, the expected value is 10 instead of zero, so 20 is 1 SE to the right of the expected value, while −20 is 3 SEs to the left.

245. Let t = sample difference/SE, where the SE is estimated from the data, as in expression (10). One formal version of our test rejects the null hypothesis if $|t| \geq 2$. To find the power, we replace the estimated SE by the true SE, computed as in expression (7); and we replace the probability histogram by the normal curve. These approximations are quite good. The size can be approximated in a similar way, given a common value for the two population pass rates. Of course, more exact calculations are possible. *See supra* note 242.

Figure 14. Power when $P_{men} = 55\%$ and $P_{women} = 45\%$. The chance of getting a significant difference (at the 5% level, two-tailed) is about equal to the area under the normal curve, to the right of +1 or to the left of −3. That shaded area is about 1/6. Power is about 1/6, or 17%.



Glossary of Terms

The following terms and definitions are adapted from a variety of sources, including Michael O. Finkelstein & Bruce Levin, Statistics for Lawyers (1990), and David A. Freedman et al., Statistics (3d ed. 1998).

adjust for. See control for.

alpha (α). A symbol often used to denote the probability of a Type I error. See Type I error; size. Compare beta.

alternative hypothesis. A statistical hypothesis that is contrasted with the null hypothesis in a significance test. See statistical hypothesis; significance test.

area sample. An area sample is a probability sample in which the sampling frame is a list of geographical areas. That is, the researchers make a list of areas, choose some at random, and interview people in the selected areas. This is a cost-effective way to draw a sample of people. See probability sample; sampling frame.

arithmetic mean. See mean.

average. See mean.

Bayes' rule. An investigator may start with a subjective probability (the "prior") that expresses degrees of belief about a parameter or a hypothesis. Data are collected according to some statistical model, at least in the investigator's opinion. Bayes' rule gives a procedure for combining the prior with the data to compute the "posterior" probability, which expresses the investigator's belief about the parameter or hypothesis given the data. See Appendix.

beta (β). A symbol sometimes used to denote power, and sometimes to denote the probability of a Type II error. See Type II error; power. Compare alpha.

bias. A systematic tendency for an estimate to be too high or too low. An estimate is "unbiased" if the bias is zero. (Does not mean prejudice, partiality, or discriminatory intent.) See non-sampling error. Compare sampling error.

bin. A class interval in a histogram. See class interval; histogram.

binary variable. A variable that has only two possible values (e.g., gender). Also called a "dummy variable."

binomial distribution. A distribution for the number of occurrences in repeated, independent "trials" where the probabilities are fixed. For example, the number of heads in 100 tosses of a coin follows a binomial distribution. When the probability is not too close to zero or one and the number of trials is large, the binomial distribution has about the same shape as the normal distribution. See normal distribution; Poisson distribution.

blind. See double-blind experiment.

bootstrap. Also called resampling; Monte Carlo method. A procedure for estimating sampling error by constructing a simulated population on the basis of the sample, then repeatedly drawing samples from this simulated population.

categorical data; categorical variable. See qualitative variable. Compare quantitative variable.

central limit theorem. Shows that under suitable conditions, the probability histogram for a sum (or average or rate) will follow the normal curve.

chance error. See random error; sampling error.

chi-squared (χ^2). The chi-squared statistic measures the distance between the data and expected values computed from a statistical model. If χ^2 is too large to explain by chance, the data contradict the model. The definition of "large" depends on the context. See statistical hypothesis; significance test.

class interval. Also, bin. The base of a rectangle in a histogram; the area of the rectangle shows the percentage of observations in the class interval. See histogram.

cluster sample. A type of random sample. For example, one might take households at random, then interview all people in the selected households. This is a cluster sample of people: a cluster consists of all the people in a selected household. Generally, clustering reduces the cost of interviewing. See multistage cluster sample.

coefficient of determination. A statistic (more commonly known as R^2) that describes how well a regression equation fits the data. See R-squared.

coefficient of variation. A statistic that measures spread relative to the mean: SD/mean, or SE/expected value. See expected value; mean; standard deviation; standard error.

collinearity. See multicollinearity.

conditional probability. The probability that one event will occur given that another has occurred.

confidence coefficient. See confidence interval.

confidence interval. An estimate, expressed as a range, for a quantity in a population. If an estimate from a large sample is unbiased, a 95% "confidence interval" is the range from about two standard errors below to two standard errors above the estimate. Intervals obtained this way cover the true value about 95% of the time, and 95% is the "confidence level" or the "confidence coefficient." See unbiased estimator; standard error. Compare bias.

confidence level. See confidence interval.

confounding. See confounding variable; observational study.

confounding variable; confounder. A variable that is correlated with the independent variables and the dependent variable. An association between the dependent and independent variables in an observational study may not be causal, but may instead be due to confounding. See controlled experiment; observational study.

consistency; consistent. See consistent estimator.

consistent estimator. An estimator that tends to become more and more accurate as the sample size grows. Inconsistent estimators, which do not become more accurate as the sample gets large, are seldom used by statisticians.

content validity. The extent to which a skills test is appropriate to its intended purpose, as evidenced by a set of questions that adequately reflect the domain being tested.

continuous variable. A variable that has arbitrarily fine gradations, such as a person's height. Compare discrete variable.

control for. Statisticians may "control for" the effects of confounding variables in nonexperimental data by making comparisons for smaller and more homogeneous groups of subjects, or by entering the confounders as explanatory variables in a regression model. To "adjust for" is perhaps a better phrase in the regression context, because in an observational study the confounding factors are not under experimental control; statistical adjustments are an imperfect substitute. See regression model.

control group. See controlled experiment.

controlled experiment. An experiment where the investigators determine which subjects are put into the "treatment group" and which are put into the "control group." Subjects in the treatment group are exposed by the investigators to some influence—the "treatment"; those in the control group are not so exposed. For instance, in an experiment to evaluate a new drug, subjects in the treatment group are given the drug, subjects in the control group are given some other therapy; the outcomes in the two groups are compared to see whether the new drug works.

"Randomization"—that is, randomly assigning subjects to each group—is usually the best way to assure that any observed difference between the two groups comes from the treatment rather than pre-existing differences. Of course, in many situations, a randomized controlled experiment is impractical, and investigators must then rely on observational studies. Compare observational study.

convenience sample. A non-random sample of units, also called a "grab sample." Such samples are easy to take, but may suffer from serious bias. Mall samples are convenience samples.

correlation coefficient. A number between −1 and 1 that indicates the extent of the linear association between two variables. Often, the correlation coefficient is abbreviated as "*r*."

covariance. A quantity that describes the statistical interrelationship of two variables. Compare correlation coefficient; standard error; variance.

covariate. A variable that is related to other variables of primary interest in a study; a measured confounder; a statistical control in a regression equation.

criterion. The variable against which an examination or other selection procedure is validated. See predictive validity.

data. Observations or measurements, usually of units in a sample taken from a larger population.

dependent variable. See independent variable.

descriptive statistics. Like the mean or standard deviation, used to summarize data.

differential validity. Differences in the correlation between skills test scores and outcome measures across different subgroups of test-takers.

discrete variable. A variable that has only a finite number of possible values, such as the number of automobiles owned by a household. Compare continuous variable.

distribution. See frequency distribution; probability distribution; sampling distribution.

disturbance term. A synonym for error term.

double-blind experiment. An experiment with human subjects in which neither the diagnosticians nor the subjects know who is in the treatment group or the control group. This is accomplished by giving a placebo treatment to patients in the control group. In a *single-blind experiment*, the patients do not know whether they are in treatment or control; however, the diagnosticians have this information.

dummy variable. Generally, a dummy variable takes only the values 0 or 1, and distinguishes one group of interest from another. See binary variable; regression model.

econometrics. Statistical study of economic issues.

epidemiology. Statistical study of disease or injury in human populations.

error term. The part of a statistical model that describes random error, i.e., the impact of chance factors unrelated to variables in the model. In econometric models, the error term is called a "disturbance term."

estimator. A sample statistic used to estimate the value of a population parameter. For instance, the sample mean commonly is used to estimate the popu-

lation mean. The term "estimator" connotes a statistical procedure, while an "estimate" connotes a particular numerical result.

expected value. See random variable.

experiment. See controlled experiment; randomized controlled experiment. Compare observational study.

explanatory variable. See independent variable, regression model.

factors. See independent variable.

Fisher's exact test. When comparing two sample proportions, for instance the proportions of whites and blacks getting a promotion, an investigator may wish to test the null hypothesis that promotion does not depend on race. Fisher's exact test is one way to arrive at a *p*-value. The calculation is based on the hypergeometric distribution. For more details, see Michael O. Finkelstein & Bruce Levin, Statistics for Lawyers 156–59 (1990). See hypergeometric distribution; *p*-value; significance test; statistical hypothesis.

fitted value. See residual.

fixed significance level. Also alpha; size. A pre-set level, such as 0.05 or 0.01; if the *p*-value of a test falls below this level, the result is deemed "statistically significant." See significance test. Compare observed significance level; *p*-value.

frequency distribution. Shows how often specified values occur in a data set.

Gaussian distribution. A synonym for the normal distribution. See normal distribution.

general linear model. Expresses the dependent variable as a linear combination of the independent variables plus an error term whose components may be dependent and have differing variances. See error term; linear combination; variance. Compare regression model.

grab sample. See convenience sample.

heteroscedastic. See scatter diagram.

histogram. A plot showing how observed values fall within specified intervals, called "bins" or "class intervals." Generally, matters are arranged so the area under the histogram, but over a class interval, gives the frequency or relative frequency of data in that interval. With a probability histogram, the area gives the chance of observing a value that falls in the corresponding interval.

homoscedastic. See scatter diagram.

hypergeometric distribution. Suppose a sample is drawn at random without replacement, from a finite population. How many times will items of a certain type come into the sample? The hypergeometric distribution gives the probabilities. For more details, see 1 William Feller, An Introduction to Prob-

ability Theory and its Applications 41–42 (2d ed. 1957). Compare Fisher's exact test.

hypothesis. See alternative hypothesis; null hypothesis; one-sided hypothesis; significance test; statistical hypothesis; two-sided hypothesis.

hypothesis test. See significance test.

independence. Events are independent when the probability of one is unaffected by the occurrence or non-occurrence of the other. Compare conditional probability.

independent variable. Independent variables (also called explanatory variables or factors) are used in a regression model to predict the dependent variable. For instance, the unemployment rate has been used as the independent variable in a model for predicting the crime rate; the unemployment rate is the independent variable in this model, and the crime rate is the dependent variable. See regression model. Compare dependent variable.

indicator variable. See dummy variable.

interquartile range. Difference between 25th and 75th percentile. See percentile.

interval estimate. A "confidence interval," or an estimate coupled with a standard error. See confidence interval; standard error. Compare point estimate.

least squares. See least squares estimator; regression model.

least squares estimator. An estimator that is computed by minimizing the sum of the squared residuals. See residual.

level. The level of a significance test is denoted alpha (α). See alpha; fixed significance level; observed significance level; p-value; significance test.

linear combination. To obtain a linear combination of two variables, multiply the first variable by some constant, multiply the second variable by another constant, and add the two products. For instance, $2u + 3v$ is a linear combination of u and v.

loss function. Statisticians may evaluate estimators according to a mathematical formula involving the errors, i.e., differences between actual values and estimated values. The "loss" may be the total of the squared errors, or the total of the absolute errors, etc. Loss functions seldom quantify real losses, but may be useful summary statistics and may prompt the construction of useful statistical procedures. Compare risk.

lurking variable. See confounding variable.

mean. Also, the average; the expected value of a random variable. The mean is one way to find the center of a batch of numbers: add up the numbers, and

divide by how many there are. Weights may be employed, as in "weighted mean" or "weighted average." See random variable. Compare median; mode.

median. The median is another way to find the center of a batch of numbers. The median is the 50th percentile. Half the numbers are larger, and half are smaller. (To be very precise: at least half the numbers are greater than or equal to the median; at least half the numbers are less than or equal to the median; for small data sets, the median may not be uniquely defined.) Compare mean; mode; percentile.

meta-analysis. Attempts to combine information from all studies on a certain topic. For example, in the epidemiologic context, a meta-analysis may attempt to provide a summary odds ratio and confidence interval for the effect of a certain exposure on a certain disease.

mode. The most commonly observed value. Compare mean; median.

model. See probability model; regression model; statistical model.

multicollinearity. Also, collinearity. The existence of correlations among the "independent variables" in a regression model. See independent variable; regression model.

multiple comparison. Making several statistical tests on the same data set. Multiple comparisons complicate the interpretation of a p-value. For example, if 20 divisions of a company are examined, and one division is found to have a disparity "significant" at the 0.05 level, the result is not surprising; indeed, it should be expected under the null hypothesis. Compare p-value; significance test; statistical hypothesis.

multiple correlation coefficient. A number that indicates the extent to which one variable can be predicted as a linear combination of other variables. Its magnitude is the square root of R^2. See linear combination; R-squared; regression model. Compare correlation coefficient.

multiple regression. A regression equation that includes two or more independent variables. See regression model. Compare simple regression.

multivariate methods. Methods for fitting models with multiple variables, especially, multiple response variables; occasionally, multiple explanatory variables. See regression model.

multi-stage cluster sample. A probability sample drawn in stages, usually after stratification; the last stage will involve drawing a cluster. See cluster sample; probability sample; stratified random sample.

natural experiment. An observational study in which treatment and control groups have been formed by some natural development; however, the assignment of subjects to groups is judged akin to randomization. See observational study. Compare controlled experiment.

nonresponse bias. Systematic error created by differences between respondents and nonrespondents. If the nonresponse rate is high, this bias may be severe.

non-sampling error. A catch-all term for sources of error in a survey, other than sampling error. Non-sampling errors cause bias. One example is selection bias: the sample is drawn in a way that tends to exclude certain subgroups in the population. A second example is non-response bias: people who do not respond to a survey are usually different from respondents. A final example: response bias arises, for instance, if the interviewer uses a loaded question.

normal distribution. Also, Gaussian distribution. The density for this distribution is the famous "bell-shaped" curve. Statistical terminology notwithstanding, there need be nothing wrong with a distribution that differs from the normal.

null hypothesis. For example, a hypothesis that there is no difference between two groups from which samples are drawn. See significance test; statistical hypothesis. Compare alternative hypothesis.

observational study. A study in which subjects select themselves into groups; investigators then compare the outcomes for the different groups. For example, studies of smoking are generally observational. Subjects decide whether or not to smoke; the investigators compare the death rate for smokers to the death rate for non-smokers. In an observational study, the groups may differ in important ways that the investigators do not notice; controlled experiments minimize this problem. The critical distinction is that in a controlled experiment, the investigators intervene to manipulate the circumstances of the subjects; in an observational study, the investigators are passive observers. (Of course, running a good observational study is hard work, and may be quite useful.) Compare confounding variable; controlled experiment.

observed significance level. A synonym for p-value. See significance test. Compare fixed significance level.

odds. The probability that an event will occur divided by the probability that it will not. For example, if the chance of rain tomorrow is 2/3, then the odds on rain are $(2/3)/(1/3) = 2/1$, or 2 to 1; the odds against rain are 1 to 2.

odds ratio. A measure of association, often used in epidemiology. For instance, if 10% of all people exposed to a chemical develop a disease, compared to 5% of people who are not exposed, then the odds of the disease in the exposed group are $10/90 = 1/9$, compared to $5/95 = 1/19$ in the unexposed group. The odds ratio is $19/9 = 2.1$. An odds ratio of 1 indicates no association. Compare relative risk.

one-sided hypothesis. Excludes the possibility that a parameter could be, e.g., less than the value asserted in the null hypothesis. A one-sided hypothesis leads to a one-tailed test. See significance test; statistical hypothesis; compare two-sided hypothesis.

one-tailed test. See significance test.

outlier. An observation that is far removed from the bulk of the data. Outliers may indicate faulty measurements and they may exert undue influence on summary statistics, such as the mean or the correlation coefficient.

***p*-value.** The output of a statistical test. The probability of getting, just by chance, a test statistic as large as or larger than the observed value. Large *p*-values are consistent with the null hypothesis; small *p*-values undermine this hypothesis. However, *p* itself does not give the probability that the null hypothesis is true. If *p* is smaller than 5%, the result is said to be "statistically significant." If *p* is smaller than 1%, the result is "highly significant." The *p*-value is also called "the observed significance level." See significance test; statistical hypothesis.

parameter. A numerical characteristic of a population or a model. See probability model.

percentile. To get the percentiles of a data set, array the data from the smallest value to the largest. Take the 90th percentile by way of example: 90% of the values fall below the 90th percentile, and 10% are above. (To be very precise: at least 90% of the data are at the 90th percentile or below; at least 10% of the data are at the 90th percentile or above.) The 50th percentile is the median: 50% of the values fall below the median, and 50% are above. When the LSAT first was scored on a 10–50 scale in 1982, a score of 32 placed a test taker at the 50th percentile; a score of 40 was at the 90th percentile (approximately). Compare mean; median; quartile.

placebo. See double-blind experiment.

point estimate. An estimate of the value of a quantity expressed as a single number. See estimator. Compare confidence interval; interval estimate.

Poisson distribution. The Poisson distribution is a limiting case of the binomial distribution, when the number of trials is large and the common probability is small. The "parameter" of the approximating Poisson distribution is the number of "trials" times the common probability, which is the "expected" number of events. When this number is large, the Poisson distribution may be approximated by a normal distribution.

population. Also, universe. All the units of interest to the researcher. Compare sample; sampling frame.

posterior probability. See Bayes' rule.

power. The probability that a statistical test will reject the null hypothesis. To compute power, one has to fix the size of the test and specify parameter values outside the range given in the null hypothesis. A powerful test has a good chance of detecting an effect, when there is an effect to be detected. See beta; significance test. Compare alpha; size; *p*-value.

practical significance. Substantive importance. Statistical significance does not necessarily establish practical significance. With large samples, small differences can be statistically significant. See significance test.

predicted value. See residual.

predictive validity. A skills test has predictive validity to the extent that test scores are well correlated with later performance, or more generally with outcomes that the test is intended to predict.

prior probability. See Bayes' rule.

probability. Chance, on a scale from 0 to 1. Impossibility is represented by 0, certainty by 1. Equivalently, chances may be quoted in percent; 100% corresponds to 1, while 5% corresponds to .05, and so forth.

probability density. Describes the probability distribution of a random variable. The chance that the random variable falls in an interval equals the area below the density and above the interval. (However, not all random variables have densities.) See probability distribution; random variable.

probability distribution. Gives probabilities for possible values or ranges of values of a random variable. Often, the distribution is described in terms of a density. See probability density.

probability histogram. See histogram.

probability model. Relates probabilities of outcomes to parameters; also, statistical model. The latter connotes unknown parameters.

probability sample. A sample drawn from a sampling frame by some objective chance mechanism; each unit has a known probability of being sampled. Such samples minimize selection bias, but can be expensive to draw.

psychometrics. The study of psychological measurement and testing.

qualitative variable; quantitative variable. A "qualitative" or "categorical" variable describes qualitative features of subjects in a study (e.g., marital status—never-married, married, widowed, divorced, separated). A "quantitative" variable describes numerical features of the subjects (e.g., height, weight, income). This is not a hard-and-fast distinction, because qualitative features may be given numerical codes, as in a "dummy variable." Quantitative variables may be classified as "discrete" or "continuous." Concepts like the mean and the standard deviation apply only to quantitative variables. Compare continuous variable; discrete variable; dummy variable. See variable.

quartile. The 25th or 75th percentile. See percentile. Compare median.

***R*-squared (R^2).** Measures how well a regression equation fits the data. R^2 varies between zero (no fit) and one (perfect fit). R^2 does not measure the validity of underlying assumptions. See regression model. Compare multiple correlation coefficient; standard error of regression.

random error. Sources of error that are haphazard in their effect. These are reflected in the "error term" of a statistical model. Some authors refer to "random error" as "chance error" or "sampling error." See regression model.

random variable. A variable whose possible values occur according to some probability mechanism. For example, if a pair of dice are thrown, the total number of spots is a random variable. The chance of two spots is 1/36, the chance of three spots is 2/36, and so forth; the most likely number is 7, with chance 6/36.

The "expected value" of a random variable is the weighted average of the possible values; the weights are the probabilities. In our example, the expected value is

$$\frac{1}{36} \times 2 + \frac{2}{36} \times 3 + \frac{3}{36} \times 4 + \frac{4}{36} \times 5 + \frac{5}{36} \times 6 + \frac{6}{36} \times 7$$
$$+ \frac{5}{36} \times 8 + \frac{4}{36} \times 9 + \frac{3}{36} \times 10 + \frac{2}{36} \times 11 + \frac{1}{36} \times 12 = 7$$

In many problems, the weighted average is computed with respect to the density; then sums must be replaced by integrals. The expected value need not be a possible value for the random variable.

Generally, a random variable will be somewhere around its expected value, but will be off (in either direction) by something like a standard error (SE) or so. If the random variable has a more or less normal distribution, there is about a 68% chance for it to fall in the range "expected value − SE" to "expected value + SE." See normal curve; standard error.

randomization. See controlled experiment; randomized controlled experiment.

randomized controlled experiment. A controlled experiment in which subjects are placed into the treatment and control groups at random—as if by lot, that is, by randomization. See controlled experiment. Compare observational study.

range. The difference between the biggest and the smallest values in a batch of numbers.

regression coefficient. A constant in a regression equation. See regression model.

regression diagnostics. Procedures intended to check whether the assumptions of a regression model are appropriate.

regression equation. See regression model.

regression line. The graph of a (simple) regression equation.

regression model. A "regression model" attempts to combine the values of certain variables (the "independent" or "explanatory" variables) in order to get expected values for another variable (the "dependent" variable). Sometimes, "regression model" refers to a probability model for the data; if no qualifications are made, the model will generally be linear, and errors will be assumed independent across observations, with common variance; the coefficients in the linear combination are called "regression coefficients"; these are parameters. At times, "regression model" refers to an equation (the "regression equation") estimated from data, typically by least squares.

For example, in a regression study of salary differences between men and women in a firm, the analyst may include a "dummy variable" for gender, as well as "statistical controls" like education and experience to adjust for productivity differences between men and women. The dummy variable would be defined as 1 for the men, 0 for the women. Salary would be the dependent variable; education, experience, and the dummy would be the independent variables. See least squares; multiple regression; random error; variance. Compare general linear model.

relative risk. A measure of association used in epidemiology. For instance, if 10% of all people exposed to a chemical develop a disease, compared to 5% of people who are not exposed, then the disease occurs twice as frequently among the exposed people: the relative risk is 10%/5% = 2. A relative risk of 1 indicates no association. For more details, see Abraham M. Lilienfeld & David E. Lilienfeld, Foundations of Epidemiology 209 (2d ed. 1980). Compare odds ratio.

reliability. The extent to which a measuring instrument gives the same results on repeated measurement of the same thing. Compare validity.

resampling. See bootstrap.

residual. The difference between an actual and a "predicted" value. The predicted value comes typically from a regression equation, and is also called the "fitted value." See regression model; independent variable.

response variable. See independent variable.

risk. Expected loss. "Expected" means on average, over the various data sets that could be generated by the statistical model under examination. Usually, risk cannot be computed exactly but has to be estimated, because the parameters in the statistical model are unknown and must be estimated. See loss function; random variable.

robust. A statistic or procedure that does not change much when data or assumptions are modified slightly.

sample. A set of units collected for study. Compare population.

sample size. The number of units in a sample.

sampling distribution. The distribution of the values of a statistic, over all possible samples from a population. For example, suppose a random sample is drawn. Some values of the sample mean are more likely, others are less likely. The "sampling distribution" specifies the chance that the sample mean will fall in one interval rather than another.

sampling error. A sample is part of a population. When a sample is used to estimate a numerical characteristic of the population, the estimate is likely to differ from the population value because the sample is not a perfect microcosm of the whole. If the estimate is unbiased, the difference between the estimate and the exact value is "sampling error." More generally,

$$\text{estimate} = \text{true value} + \text{bias} + \text{sampling error}.$$

Sampling error is also called "chance error" or "random error." See standard error. Compare bias; non-sampling error.

sampling frame. A list of units designed to represent the entire population as completely as possible. The sample is drawn from the frame.

scatter diagram. Also, scatterplot; scatter diagram. A graph showing the relationship between two variables in a study. Each dot represents one subject. One variable is plotted along the horizontal axis, the other variable is plotted along the vertical axis. A scatter diagram is "homoscedastic" when the spread is more or less the same inside any vertical strip. If the spread changes from one strip to another, the diagram is "heteroscedastic."

selection bias. Systematic error due to non-random selection of subjects for study.

sensitivity. In clinical medicine, the probability that a test for a disease will give a positive result given that the patient has the disease. Sensitivity is analogous to the power of a statistical test. Compare specificity.

sensitivity analysis. Analyzing data in different ways to see how results depend on methods or assumptions.

significance level. See fixed significance level; p-value.

significance test. Also, statistical test; hypothesis test; test of significance. A significance test involves formulating a statistical hypothesis and a test statistic, computing a p-value, and comparing p to some pre-established value ("alpha") to decide if the test statistic is "significant." The idea is to see whether the data conform to the predictions of the null hypothesis. Generally, a large

test statistic goes with a small p-value; and small p-values would undermine the null hypothesis.

For instance, suppose that a random sample of male and female employees were given a skills test and the mean scores of the men and women were different—in the sample. To judge whether the difference is due to sampling error, a statistician might consider the implications of competing hypotheses about the difference in the population. The "null hypothesis" would say that on average, in the population, men and women have the same scores: the difference observed in the data is then just due to sampling error. A "one-sided alternative hypothesis" would be that on average, in the population, men score higher than women. The "one-tailed" test would reject the null hypothesis if the sample men score substantially higher than the women—so much so that the difference is hard to explain on the basis of sampling error.

In contrast, the null hypothesis could be tested against the "two-sided alternative" that on average, in the population, men score differently than women—higher or lower. The corresponding "two-tailed" test would reject the null hypothesis if the sample men score substantially higher or substantially lower than the women.

The one-tailed and two-tailed tests would both be based on the same data, and use the same t-statistic. However, if the men in the sample score higher than the women, the one-tailed test would give a p-value only half as large as the two-tailed test, that is, the one-tailed test would appear to give stronger evidence against the null hypothesis. See p-value; statistical hypothesis; t-statistic.

significant. See p-value; practical significance; significance test.

simple random sample. A random sample in which each unit in the sampling frame has the same chance of being sampled. One takes a unit at random (as if by lottery), sets it aside, takes another at random from what is left, and so forth.

simple regression. A regression equation that includes only one independent variable. Compare multiple regression.

size. A synonym for alpha (α).

specificity. In clinical medicine, the probability that a test for a disease will give a negative result given that the patient does not have the disease. Specificity is analogous to $1 - \alpha$, where α is the significance level of a statistical test. Compare sensitivity.

spurious correlation. When two variables are correlated, one is not necessarily the cause of the other. The vocabulary and shoe size of children in elementary school, for instance, are correlated—but learning more words will not make the feet grow. Such non-causal correlations are said to be "spuri-

ous." (Originally, the term seems to have been applied to the correlation between two rates with the same denominator: even if the numerators are unrelated, the common denominator will create some association.) Compare confounding variable.

standard deviation (SD). The SD indicates how far a typical element deviates from the average. For instance, in round numbers, the average height of women age 18 and over in the United States is 5 feet 4 inches. However, few women are exactly average; most will deviate from average, at least by a little. The SD is sort of an average deviation from average. For the height distribution, the SD is 3 inches. The height of a typical woman is around 5 feet 4 inches, but is off that average value by something like 3 inches.

For distributions that follow the normal curve, about 68% of the elements are in the range "mean – SD" to "mean + SD." Thus, about 68% of women have heights in the range 5 feet 1 inch to 5 feet 7 inches. Deviations from the average that exceed three or four SDs are extremely unusual. Many authors use "standard deviation" to also mean standard error. See standard error.

standard error (SE). Indicates the likely size of the sampling error in an estimate. Many authors use the term "standard deviation" instead of standard error. Compare expected value; standard deviation.

standard error of regression. Indicates how actual values differ (in some average sense) from the fitted values in a regression model. See regression model; residual. Compare R-squared.

standardization. See standardized variable.

standardized variable. Transformed to have mean zero and variance one. This involves two steps: (1) subtract the mean, (2) divide by the standard deviation.

statistic. A number that summarizes data. A "statistic" refers to a sample; a "parameter" or a "true value" refers to a population or a probability model.

statistical controls. Procedures that try to filter out the effects of confounding variables on non-experimental data, for instance, by "adjusting" through statistical procedures (like multiple regression). Variables in a multiple regression equation. See multiple regression; confounding variable; observational study. Compare controlled experiment.

statistical hypothesis. Data may be governed by a probability model; "parameters" are numerical characteristics describing features of the model. Generally, a "statistical hypothesis" is a statement about the parameters in a probability model. The "null hypothesis" may assert that certain parameters have specified values or fall in specified ranges; the alternative hypothesis would specify other values or ranges. The null hypothesis is "tested" against the data

with a "test statistic"; the null hypothesis may be "rejected" if there is a "statistically significant" difference between the data and the predictions of the null hypothesis.

Typically, the investigator seeks to demonstrate the alternative hypothesis; the null hypothesis would explain the findings as a result of mere chance, and the investigator uses a significance test to rule out this explanation. See significance test.

statistical model. See probability model.

statistical test. See significance test.

statistical significance. See p-value.

stratified random sample. A type of probability sample. One divides the population up into relatively homogeneous groups called "strata," and draws a random sample separately from each stratum.

systematic sampling. The elements of the population are numbered consecutively as 1, 2, 3 . . . . Then, every kth element is chosen. If $k = 10$, for instance, the sample would consist of items 1, 11, 21 . . . . Sometimes the starting point is chosen at random from 1 to k.

t-statistic. A test statistic, used to make the "t-test." The t-statistic indicates how far away an estimate is from its expected value, relative to the standard error. The expected value is computed using the null hypothesis that is being tested. Some authors refer to the t-statistic, others to the "z-statistic," especially when the sample is large. In such cases, a t-statistic larger than 2 or 3 in absolute value makes the null hypothesis rather unlikely—the estimate is too many standard errors away from its expected value. See statistical hypothesis; significance test; t-test.

t-test. A statistical test based on the t-statistic. Large t-statistics are beyond the usual range of sampling error. For example, if t is bigger than 2, or smaller than -2, then the estimate is "statistically significant" at the 5% level: such values of t are hard to explain on the basis of sampling error. The scale for t-statistics is tied to areas under the normal curve. For instance, a t-statistic of 1.5 is not very striking, because 13% = 13/100 of the area under the normal curve is outside the range from -1.5 to 1.5. On the other hand, $t = 3$ is remarkable: only 3/1,000 of the area lies outside the range from -3 to 3. This discussion is predicated on having a reasonably large sample; in that context, many authors refer to the "z-test" rather than the t-test.

For small samples drawn at random from a population known to be normal, the t-statistic follows "Student's t-distribution" (when the null hypothesis holds) rather than the normal curve; larger values of t are required to achieve "significance." A t-test is not appropriate for small samples drawn

from a population that is not normal. See *p*-value; significance test; statistical hypothesis.

test statistic. A statistic used to judge whether data conform to the null hypothesis. The parameters of a probability model determine expected values for the data; differences between expected values and observed values are measured by a "test statistic." Such test statistics include the chi-squared statistic (χ^2) and the *t*-statistic. Generally, small values of the test statistic are consistent with the null hypothesis; large values lead to rejection. See *p*-value; statistical hypothesis; *t*-statistic.

time series. A series of data collected over time, for instance, the Gross National Product of the United States from 1940 to 1990.

treatment group. See controlled experiment.

two-sided hypothesis. An alternative hypothesis asserting that the values of a parameter are different from—either greater than or less than—the value asserted in the null hypothesis. A two-sided alternative hypothesis suggests a two-tailed test. See statistical hypothesis; significance test. Compare one-sided hypothesis.

two-tailed test. See significance test.

Type I error. A statistical test makes a "Type I error" when (1) the null hypothesis is true and (2) the test rejects the null hypothesis, i.e., there is a false positive. For instance, a study of two groups may show some difference between samples from each group, even when there is no difference in the population. When a statistical test deems the difference to be "significant" in this situation, it makes a Type I error. See significance test; statistical hypothesis. Compare alpha; Type II error.

Type II error. A statistical test makes a "Type II error" when (1) the null hypothesis is false and (2) the test fails to reject the null hypothesis, i.e., there is a false negative. For instance, there may not be a "significant" difference between samples from two groups when, in fact, the groups are different. See significance test; statistical hypothesis. Compare beta; Type I error.

unbiased estimator. An estimator that is correct on average, over the possible data sets. The estimates have no systematic tendency to be high or low. Compare bias.

uniform distribution. For example, a whole number picked at random from 1 to 100 has the uniform distribution: all values are equally likely. Similarly, a uniform distribution is obtained by picking a real number at random between 0.75 and 3.25: the chance of landing in an interval is proportional to the length of the interval.

validity. The extent to which an instrument measures what it is supposed to, rather than something else. The validity of a standardized test is often indicated (in part) by the correlation coefficient between the test scores and some outcome measure.

variable. A property of units in a study, which varies from one unit to another. For example, in a study of households, household income; in a study of people, employment status (employed, unemployed, not in labor force).

variance. The square of the standard deviation. Compare standard error; covariance.

***z*-statistic.** See *t*-statistic.

***z*-test.** See *t*-test.

References on Statistics

General Surveys

David Freedman et al., Statistics (3d ed. 1998).

Darrell Huff, How to Lie with Statistics (1954).

Gregory A. Kimble, How to Use (and Misuse) Statistics (1978).

David S. Moore, Statistics: Concepts and Controversies (3d ed. 1991).

David S. Moore & George P. McCabe, Introduction to the Practice of Statistics (2d ed. 1993).

Michael Oakes, Statistical Inference: A Commentary for the Social and Behavioral Sciences (1986).

Perspectives on Contemporary Statistics (David G. Hoaglin & David S. Moore eds., 1992).

Statistics: A Guide to the Unknown (Judith M. Tanur et al. eds., 2d ed. 1978).

Hans Zeisel, Say It with Figures (6th ed. 1985).

Reference Works for Lawyers and Judges

David C. Baldus & James W.L. Cole, Statistical Proof of Discrimination (1980 & Supp. 1987).

David W. Barnes & John M. Conley, Statistical Evidence in Litigation: Methodology, Procedure, and Practice (1986).

James Brooks, A Lawyer's Guide to Probability and Statistics (1990).

Michael O. Finkelstein & Bruce Levin, Statistics for Lawyers (1990).

1 & 2 Modern Scientific Evidence: The Law and Science of Expert Testimony (David L. Faigman et al. eds., 1997)

Ramona Paetzold & Steven L. Willborn, The Statistics of Discrimination: Using Statistical Evidence in Discrimination Cases (1994)

Panel on Statistical Assessments as Evidence in the Courts, National Research Council, The Evolving Role of Statistical Assessments as Evidence in the Courts (Stephen E. Fienberg ed., 1989).

Statistical Methods in Discrimination Litigation (David H. Kaye & Mikel Aickin eds., 1986).

Hans Zeisel & David Kaye, Prove It with Figures: Empirical Methods in Law and Litigation (1997)

General Reference

International Encyclopedia of Statistics (William H. Kruskal & Judith M. Tanur eds., 1978).

Lindquist & Vennum PLLP

DEPOSITION EXHIBIT
FORD
PO 4 9-15-06

Final Report: Sportsman's Warehouse Recognition Assessment

Prepared by

WINKELMAN CONSULTING

PO BOX 11375 • FARGO, ND 58106-1375

701-237-2283 • Office/Voice Mail

701-237-6877 • Fax Machine

701-799-0877 • Cell

MRWinkelman@cableone.net

Final Report Delivered September 2006

TABLE OF CONTENTS

EXECUTIVE SUMMARY **1-1**

FINDINGS SUMMARY & GENERAL CONCLUSIONS **2-1**

PURPOSE #1: ASSESS THE IMPLIED MEANING OF THE WORDS "SPORTSMAN'S WAREHOUSE". 2-1

FIGURE 1A. WHEN YOU HEAR THE WORDS "SPORTSMAN'S WAREHOUSE", DO YOU THINK THESE WORDS ARE MOST LIKELY 2-1

FIGURE 1B. WHEN YOU HEAR THE WORDS "SPORTSMAN'S WAREHOUSE", DO YOU THINK THESE WORDS ARE MOST LIKELY 2-2

FIGURE 2. TOP-OF-MIND AND UNAIDED AWARENESS OF SPORTSMAN'S WAREHOUSE. 2-3

FIGURE 3. WORDS USED MOST FREQUENTLY TO DESCRIBE A LARGE STORE THAT SELLS MAINLY HUNTING, FISHING, CAMPING, AND OUTDOOR GEAR? 2-4

PURPOSE #2: ASSESS THE IMPLIED MEANING OF THE "SPORTSMAN'S WAREHOUSE" SAMPLE LOGO 2-8

FIGURE 4. AFTER SEEING THIS SAMPLE LOGO, DO YOU THINK IT IS MOST LIKELY... 2-9

METHODOLOGY **3-1**

PURPOSE 3-1

SAMPLING FRAME & SAMPLE SIZE 3-1

COLLECTION TECHNIQUE & TIMING 3-2

MARGIN OF ERROR 3-3

THE QUESTIONNAIRE **4-1**

CONSUMER SURVEY 4-2

CONTINGENCY TABLES **5-1**

Section

1

EXECUTIVE SUMMARY

A "summary" report that included both consumers and vendors was provided previously. This "final" report was prepared to both present consumer results only and enhance the detail of the previous report.

- **Assess the implied meaning of the words "Sportsman's Warehouse".**

  - Nearly one of every two respondents felt the words "Sportsman's Warehouse" are at least the name of a specific store or chain that sells mainly hunting, fishing, camping, or outdoor gear. In contrast, only about one of every three respondents felt the words "Sportsman's Warehouse" are most likely general words that can be used to describe any large store that sells mainly hunting, fishing, camping, or outdoor gear.

  - Nearly one of every twenty respondents named Sportsman's Warehouse first (top-of-mind awareness), while more than one of every ten named Sportsman's Warehouse (unaided awareness, which includes top-of-mind awareness). These types of awareness are primarily measures of the longevity and marketing strength of an organization. In short, the more established an organization is in a market and/or the more an organization advertises, the higher its top-of-mind and/or unaided awareness will be.

  - The results indicate that a minuscule proportion of respondents used the words "sportsman(s)" or "sportsman(s) warehouse" as a generic term for a large store that sells mainly hunting, fishing, camping, and outdoor gear.

- **Assess the implied meaning of the "Sportsman's Warehouse" sample logo.**

  - Of those who viewed the sample logos, nearly eight of every ten felt the Sportsman's Warehouse sample logo is at least a logo that could be used to identify one specific store or chain of stores. In comparison, more than eight of every ten felt the Bass Pro Shop sample logo is at least a logo that could be used to identify one specific store or chain of stores.

Section 2

FINDINGS SUMMARY & GENERAL CONCLUSIONS

When reviewing this report, the findings and conclusions will be more thoroughly understood if several other sections of the report are also reviewed. First, the questionnaires in Section 4 provide the actual phrasing for each question. A solid understanding of the context in which each question was asked will enable you to more accurately interpret the findings. Second, the contingency tables in Section 5 provide more detailed results.

Purpose #1: Assess the implied meaning of the words "Sportsman's Warehouse".

- **Conclusion #1:** Nearly one of every two respondents felt the words "Sportsman's Warehouse" are at least the name of a specific store or chain that sells mainly hunting, fishing, camping, or outdoor gear.

  - Respondents without top-of-mind or unaided awareness of Sportsman's Warehouse were asked if, when they hear the words "Sportsman's Warehouse", they think these words are most likely the name of one specific store or chain of stores, words that describe one specific store or chain of stores, or general words that can be used to describe any large store that sells mainly hunting, fishing, camping, or outdoor gear. Those who felt these words are the name of or describe one specific store or chain of stores, were also asked to name the location of a Sportsman's Warehouse.

    Figure 1a shows that 47.2% of all respondents felt the words "Sportsman's Warehouse" are at least the name of one specific store or chain of stores. In contrast, slightly more than one of every three respondents felt the words "Sportsman's Warehouse" are most likely general words that can be used to describe any large store that sells mainly hunting, fishing, camping, or outdoor gear.

Figure 1a. When you hear the words "Sportsman's Warehouse", do you think these words are most likely...
(Includes all consumers)

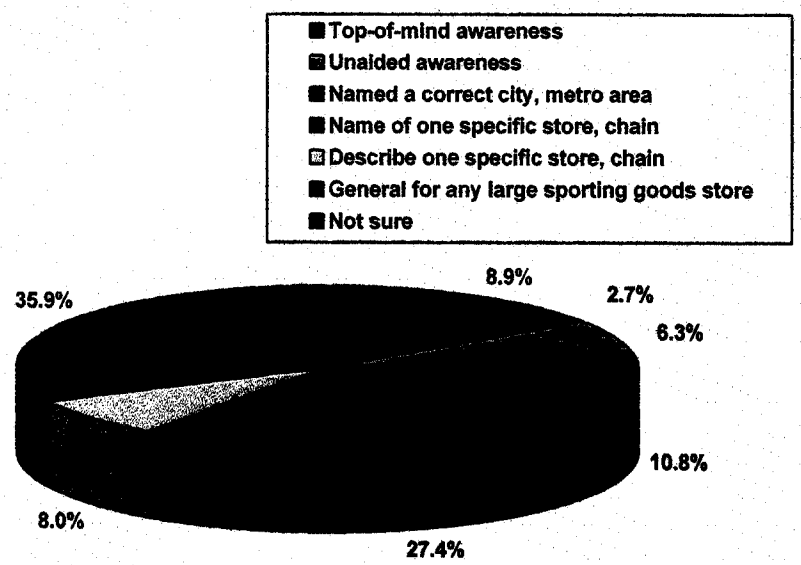
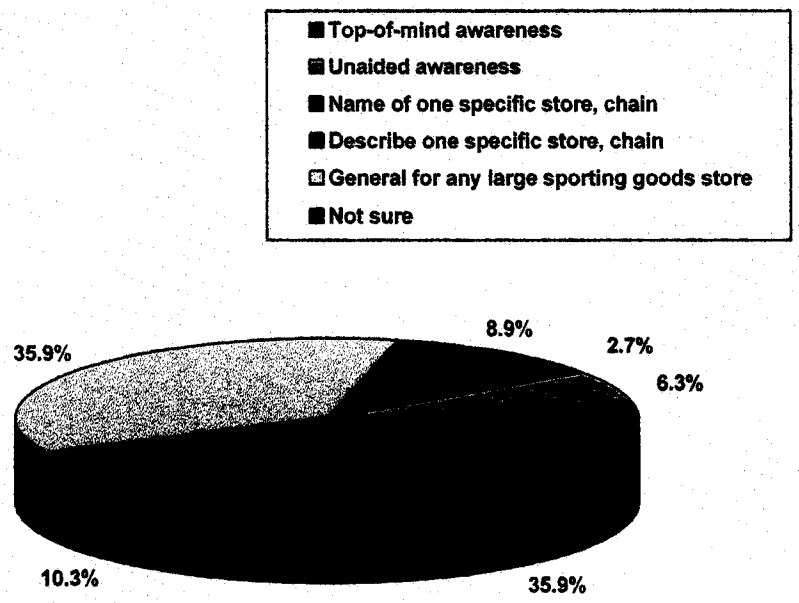


Figure 1b. When you hear the words "Sportsman's Warehouse", do you think these words are most likely...
(Includes all consumers)



C Please note that the results in Figure 1a are based on the combined results for questions A3a to A5. Therefore, the figure indicates each respondent's highest level of awareness for Sportsman's Warehouse or recognition for the words Sportsman's Warehouse for these questions.

Figure 1b is based upon the combined results of questions A3a to A4. Again, the figure indicates each respondent's highest level of awareness for Sportsman's Warehouse or recognition for the words Sportsman's Warehouse for these questions. When looking at these results, roughly 45% of the respondents felt the words "Sportsman's Warehouse" are at least the name of a specific store or chain that sells mainly hunting, fishing, camping, or outdoor gear.

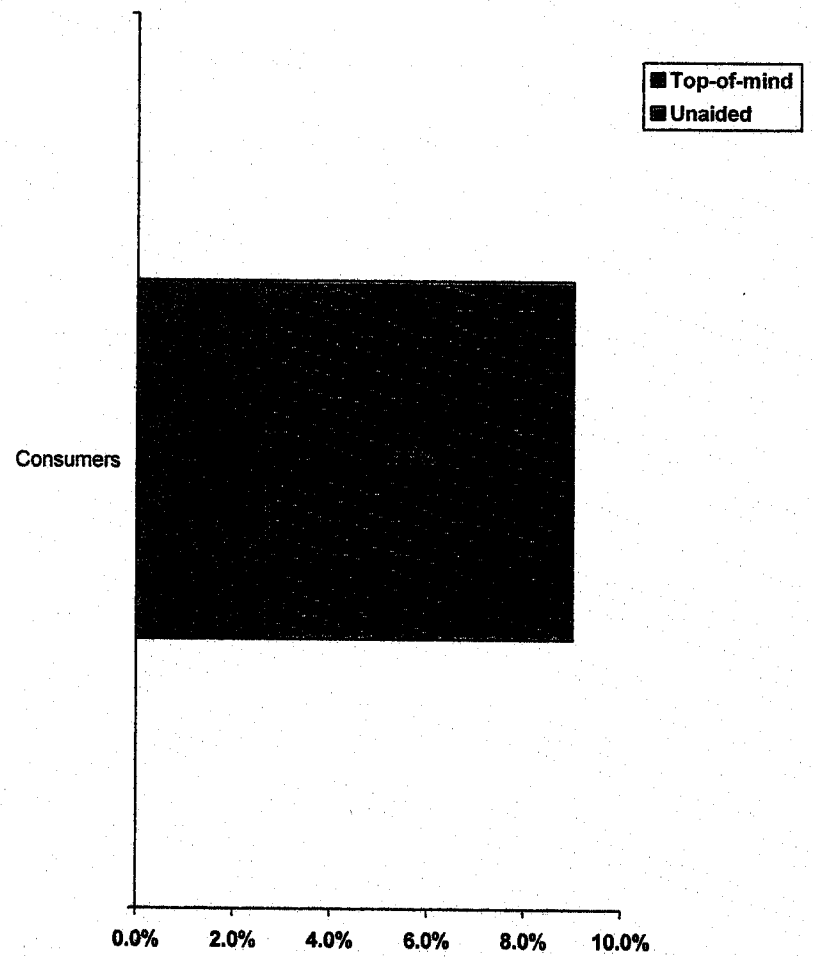
- **Conclusion #2:** Nearly one of every twenty respondents named Sportsman's Warehouse first (top-of-mind awareness), while more than one of every ten named Sportsman's Warehouse (unaided awareness, which includes top-of-mind awareness). These types of awareness are primarily measures of the longevity and marketing strength of an organization. In short, the more established an organization is in a market and/or the more an organization advertises, the higher its top-of-mind and/or unaided awareness will be.

  - All respondents were asked to name the one store that comes to mind first *and* other stores that come to mind when they think about large stores that sell mainly hunting, fishing, camping, or outdoor gear. As can be seen again in **Figure 2**, 3.7% named Sportsman's Warehouse first (top-of-mind awareness), while another 7.0% (10.7% overall) named Sportsman's Warehouse (unaided awareness, which includes top-of-mind awareness).

    For this question, respondents were allowed to give up to seven responses. Therefore, the results will exceed 100% when the percentages are added together.

  - Please note that the results in Figure 2 are based on the combined results for questions A3a to A3g. Therefore, they indicate each respondent's highest level of awareness for Sportsman's Warehouse.

Figure 2. Top-of-mind and unaided awareness of Sportsman's Warehouse.
(Includes all respondents)



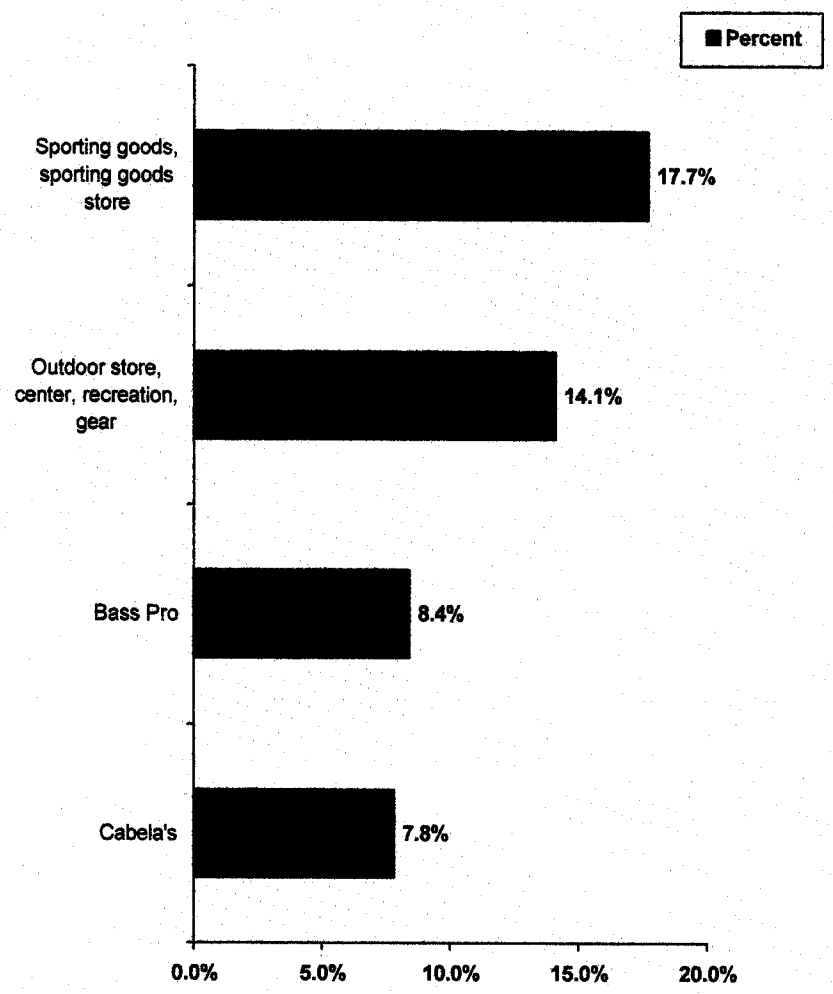
- **Conclusion #3:** The results indicate that a minuscule proportion of respondents used the words "sportsman(s)" or "sportsman(s) warehouse" as a generic term for a large store that sells mainly hunting, fishing, camping, and outdoor gear.

  - Before being asked any questions related to the words "Sportsman's Warehouse", all respondents were asked what word or words they would use to describe a large store that sells mainly hunting, fishing, camping, and outdoor gear. **Figure 3** shows that the words used most frequently for a large store that sells mainly hunting, fishing, camping, and outdoor gear were the words sporting goods or outdoor, followed by Bass Pro and Cabela's.

    For this open-ended question, respondents were allowed to give up to three responses. Therefore, the results will exceed 100% when the percentages are added together.

    - The contingency tables in Section Five show that less than 2% of the respondents used the words "sportsman(s) warehouse" to describe a large store that sells mainly hunting, fishing, camping, and outdoor gear.

Figure 3. Words used most frequently to describe a large store that sells mainly hunting, fishing, camping, and outdoor gear?
(Includes all consumers)



To better understand some of the findings summarized above, some additional explanation may be helpful.

The purpose of question A2 was to prove or disprove whether the words "sportsman's warehouse" are a generic term that a large number of outdoors people "use to describe a large store that sells mainly hunting, fishing, camping, and outdoor gear." Therefore, the lack of "probe phrases" is not "very unusual" and is appropriate in this case since:

The question does not ask the respondent for only one word or phrase,

A fair number of respondents did provide more than one word or phrase,

The question accurately obtained the dominant word or words each respondent uses to describe this type of business, and

The purpose of this question was not to measure all of the words or phrases each respondent uses to describe this type of business.

If the purpose of question A2 had been to collect as many thoughts or opinions as we can, then "probe phrases" would be appropriate. For example, the purpose of questions A3a and A3b-g was to measure top-of-mind and unaided awareness of large stores that sell mainly hunting, fishing, camping, or outdoor gear by soliciting up to seven store names from respondents. Based on this purpose and the fact that question A3a asked respondents "...When you think of names of large stores that sell mainly hunting, fishing, camping, or outdoor gear, what is the **one** store name that comes to mind first?" the probe question (A3b-g) of "What other store names come to mind when you think about large stores that sell mainly hunting, fishing, camping, or outdoor gear?" is appropriate.

When looking at the coding related to questions A3a and A3b-g, 'Sportsman's Warehouse' responses were assigned to the "Sportsman's Warehouse" brand name. In this situation, coding that "Sportsman's Warehouse" as a brand name is a logical and appropriate grouping since, to the best of my knowledge, Bass Pro Shops has not used the "Bass Pro Shops Sportsman's Warehouse" name or logo in the eight of the nine states in which the consumer surveys were conducted -- Colorado, Iowa, Mississippi, Nevada, Oklahoma, Pennsylvania, South Carolina, Tennessee (signage removed in 2005), and Texas.

As can be seen in Section 4, question A4 asks "When you hear the words "Sportsman's Warehouse", do you think these words are most likely (1) The name of one specific store or chain of stores, (2) Words that describe one specific store or chain of stores, or (3) General words that can be used to describe any large store that sells mainly hunting, fishing, camping, or outdoor gear?" There are many reasons why we used three response choices without reading the "not sure" response for this question.

First, the purpose of this question was to obtain information that determines whether or not consumers recognize the words/name Sportsman's Warehouse as a single or sole source for hunting, fishing, camping, or outdoor gear. To assess perceived single or sole source, we used three response choices that are similar to those we have used in several other surveys and cover the entire array of unique, meaningful response choice options: (1) name of a single source, (2) describes a single source, or (3) does not describe a single source.

Second, it is my experience that, in many cases, a single closed-ended question with an odd number of response choices is more effective in obtaining accurate data than other response options. For example:

The five-point scale of very likely, likely, somewhat likely, not very likely, or not at all likely has repeatedly proven to more accurately estimate actual consumer behavior than the "traditional" four-point scale of very likely, somewhat likely, somewhat unlikely, or very unlikely. Therefore, having three favorable and two unfavorable responses enhances the quality of the data. Similarly, increasing to four response choices or decreasing to two response choices for question A4 would have hampered our ability to collect accurate data.

In this case, we could have also used an approach that utilizes multiple two response questions such as:

(1) Do you think the words Sportsman's Warehouse are the name of one specific store or chain or (2) Don't you think the words Sportsman's Warehouse are the name of one specific store or chain?

(1) Do you think the words Sportsman's Warehouse describe one specific store or chain of stores or (2) Don't you think the words Sportsman's Warehouse describe one specific store or chain of stores?

However, it is my experience that, because this technique tends to inflate the "do" responses, it would not have provided an accurate snapshot of respondents' true options related to perceiving the words Sportsman's Warehouse as a single or sole source.

In conclusion, despite "theoretical" risks that some associate with a closed-ended question using an odd number of response choices, I deemed the format used for this study to be superior to other options in producing the most accurate data.

The "not sure" response choice was not read for A4. While some theorists claim that the proportion of "not sure" responses may increase slightly (20% to 25%), it does not significantly affect the relative distribution of respondents selecting a listed choice. Assuming that this is true, since only 8.9% of the respondents gave a "not sure" response to A4, the proportion of "not sure" responses would have likely increased only 20% to 25% –increasing the "not sure" responses to a maximum of 11.1% with minimal impact on the other response choices.

	Actual	20% change in "not sure"	25% change in "not sure"
Name of store	35.9%	35.1%	34.9%
Describes store	10.3%	10.1%	10.0%
General words	35.9%	35.1%	34.9%
Not sure	8.9%	10.7%	11.1%
Total	91.0%	91.0%	91.0%
Increase in "Not sure"	n.a.	1.8%	2.2%

For question A5 (To the best of your knowledge, in what city or town is a "Sportsman's Warehouse" located in your state), some metro areas named include both a "Sportsman's Warehouse" and "Bass Pro" store. The cities named by respondents were grouped based on typical consumer habits/behavior and/or other facts.

First, when people from my community (Fargo, ND-Moorhead, MN) go shopping in the Twin Cities (Minneapolis-St. Paul, MN area), for example, they typically describe the location of where they shopped as "Minneapolis" even though they actually shopped at stores/malls in Edina, Bloomington, or some other city in the metro area. Therefore, responses were grouped into appropriate "metro" areas as described on MapQuest.com.

Second, assigning "common" metro areas to "Sportsman's Warehouse" is logical and appropriate since, to the best of my knowledge, Bass Pro Shops has not used the "Bass Pro Shops Sportsman's Warehouse" name or logo in the states in which the consumer surveys were conducted.

Purpose #2: Assess the implied meaning of the "Sportsman's Warehouse" sample logo.

- **Conclusion #4:** Of those who viewed the sample logos, nearly eight of every ten felt the Sportsman's Warehouse sample logo is at least a logo that could be used to identify one specific store or chain of stores. In comparison, more than eight of every ten felt the Bass Pro Shop sample logo is at least a logo that could be used to identify one specific store or chain of stores.

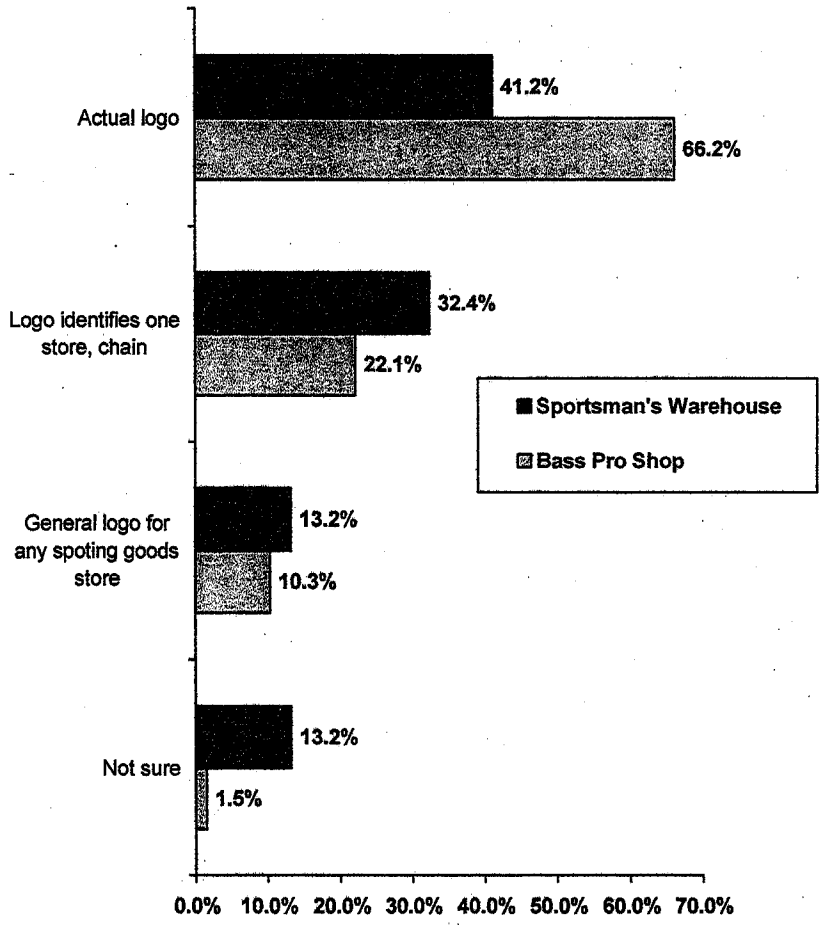
  - Those who were willing and able to access the Internet were asked to view sample logos for Sportsman's Warehouse and Bass Pro Shop. After reviewing each logo they were asked if they think that sample logo is most likely an actual logo that is currently used by one specific store or chain of stores, a logo that could be used to identify one specific store or chain of stores, or a general logo that could be used by any sporting goods store to let consumers know their stores sell mainly hunting, fishing, camping, or outdoor gear.

Sportsman's Warehouse Sample Logo

Bass Pro Shop Sample Logo

C **Figure 4** shows that 76.0% of these respondents felt the Sportsman's Warehouse sample logo is at least a logo that could be used to identify one specific store or chain of stores, while 84.0% felt the Bass Pro Shop sample logo is at least a logo that could be used to identify one specific store or chain of stores. Given the margin of error associated with these questions (roughly ±9% to ±10%), one cannot say that the recognition of one sample logo is significantly higher (or lower) than the other.

Figure 4. After seeing this sample logo, do you think it is most likely...
(Includes only phone/web consumers)



Again, to better understand some of the findings summarized in above, some additional explanation may be helpful.

For questions B3b (Sportsman's Warehouse) and B3b (Bass Pro Shop Sportsman's Warehouse), respondents were asked if, "After seeing this sample logo, do you think it is most likely (1) An actual logo that is currently used by one specific store or chain of stores, (2) A logo that could be used to identify one specific store or chain of stores, or (3) A general logo that could be used by any sporting goods store to let consumers know their stores sell mainly hunting, fishing, camping, or outdoor gear?" As was noted previously for question A4, a three response question was used since:

The purpose of these questions was to obtain information that determines whether or not consumers recognize or associate each sample logo with a single or sole source for hunting, fishing, camping, or outdoor gear.

The three response choices are similar to those we have used in several other surveys and cover the entire array of unique, meaningful response choice options.

This format was deemed to be superior to other two and four response choice options in producing the most accurate data.

Based on theoretical projections, having a passive "not sure" response choice would not significantly affect the relative distribution of respondents selecting a listed choice.

Sportsman's Warehouse	Actual	20% change in "not sure"	25% change in "not sure"
Logo for one store	41.2%	39.9%	39.6%
Could be logo for one store	32.4%	31.4%	31.2%
General logo	13.2%	12.8%	12.7%
Not sure	13.2%	15.8%	16.5%
Total	100.0%	100.0%	100.0%
Increase in "Not sure"	n.a.	2.6%	3.3%

Bass Pro Shop Sportsman's Warehouse	Actual	20% change in "not sure"	25% change in "not sure"
Logo for one store	66.2%	66.0%	65.9%
Could be logo for one store	22.1%	22.0%	22.0%
General logo	10.3%	10.3%	10.3%
Not sure	1.5%	1.8%	1.9%
Total (rounding error)	100.1%	100.1%	100.1%
Increase in "Not sure"	n.a.	0.3%	0.4%

Typically, in some trademark surveys, additional questions are required to adjust for demand effects and/or alternative explanations for the results. In this case, however, the following were not deemed to be appropriate:

> The purpose of these questions was to obtain information that determines whether or not consumers recognize or associate each sample logo with a single or sole source for hunting, fishing, camping, or outdoor gear. Since the words sportsman's warehouse appeared in both sample logos, any demand effect should have an equal impact on both logos and, therefore, would not justify the added cost needed to conduct a separate logos study.

> In my opinion, a control group and/or question(s) were not needed for the following reasons:

> > Previous experience on similar studies that use control questions has shown that 1% to 8% of respondents generally "guess" when responding to a two response question. However, the proportion of people who "guess" is typically smaller when more response choices are provided.

Use of control groups requires one to ask the "test" question(s) of one-half of the respondents and the "control" questions of the other half, thus reducing the sample size, increasing the margin of error or allowable fluctuation, and decreasing the statistical accuracy for the "key" questions. Therefore, using a control group would leave us with the following options:

> Keep the sample size, budget, and time frame the same and increase the maximum margin of error for "test" question A4 from ±4.5% to ±6.4% and the maximum margin of error for "test" questions B3b and B4b from ±11.9% to ±16.8%, or

> Double the sample size, budget, and time frame to keep the maximum margin of error for question A4 at ±4.5% and the maximum margin of error for questions B3b and B4b at ±11.9%.

Section

3

METHODOLOGY

Purpose

The purpose of this research study was to obtain information that will be used to determine consumer's recognition of the words/name Sportsman's Warehouse and the Sportsman's Warehouse logo. To do this, the study focused on the following objectives:

Assess the implied meaning of the words "Sportsman's Warehouse".

Assess the implied meaning of the "Sportsman's Warehouse" sample logo.

Sampling Frame & Sample Size

The primary population for this study consisted of telephone households who have expressed an interest in fishing, hunting, and/or camping and live in a state in which, based on their website information, there is both a Bass Pro Shop and a Sportsman's Warehouse. Telephone numbers for households who have expressed an interest in fishing, hunting, and/or camping were obtained from Survey Sampling International (SSI), a list company that specializes in generating samples for survey research. A stratified (by state), random-systematic (the standard "replicate" method utilized by SSI for a vast majority of the samples they generate) sampling technique was employed in this study to select households. Only persons who had purchased fishing, hunting, camping, and/or other outdoor gear in the past twelve months were allowed to participate in the survey. To obtain the highest response rate possible, a relatively small number of telephone numbers/replicates were introduced into the active call list, telephone numbers/replicates were added to the active call list only when they were needed to replace unusable numbers, and, once a telephone number was added to the active call list, it was called each day that calls were dialed from that date until the completion date.

Due to (1) the relatively low proportion of the general population who hunt/fish/camp and (2) the time/budget constraints for this study, we determined that the use of a "pre-qualified" sample would be most appropriate for this study. SSI uses national mail surveys, national email surveys, and other national sources to identify those with interests in a variety of areas.

This sample cannot be deemed to be a "convenience sample." In short, a convenience sample is a sample that merely includes respondents or participants who happen to be at a certain place at a certain time. The respondents or participants included in a convenience sample are <u>not</u> representative of the larger population and, therefore, the results of a convenience sample <u>cannot</u> be projected to the larger population with any known level of statistical certainty. A good example of a convenience sample would be mall intercept interviews. Mall intercept interviews simply represent the skewed segment of shoppers who happen to frequent that mall and were available be at the times the interviews were being completed and, therefore, <u>cannot</u> be used to estimate the proportion of the larger population who have a certain opinion, knowledge, or anticipated behavior.

From this sample, 474 interviews were completed with consumers. The next chart summarizes the call disposition for the calls made to complete these interviews.

		Consumer	
Total Numbers Available		6,440	
Unusable Numbers		4,902	
	Wrong Number		129
	Disconnected		733
	Language Barrier		51
	Not Qualified		1,647
	Quota Full		979
	Fax/Modem		223
	Duplicate Number		4
	Do Not Call		137
	Privacy Manager		607
	Hard Refusal		392
Remaining Useable Numbers		1,538	
	Busy		65
	No Answer/Machine		999
	Completed Interviews		474
Response Rate			31%
Estimated Eligible Numbers from			
Remaining Useable Numbers			1,055
Estimated Response Rate			45%

The secondary population for this study consisted of Sportsman's Warehouse vendors who sell fishing, hunting, and/or camping to retailers. For each vendor, the contact persons name and telephone number were obtained from Sportsman's Warehouse. No sampling technique was employed for this segment since we attempted to contact all 52 vendors provided by Sportsman's Warehouse.

From this sample, 10 interviews were completed with consumers. Since the sole purpose of surveying this segment was to this provide supplemental marketing information for Sportsman's Warehouse, the results for this segment are not included in this report.

Collection Technique & Timing

All data was collected through the use of telephone interviews and a web site. For consumers, telephone interviews were conducted from May 31 to June 12, 2006, while the telephone interviews for vendors were only completed from June 12-13, 2006. The data collection was completed in compliance with specifications established by Winkelman Consulting. Interviewing was supervised and performed by trained personnel from KLJ Solutions.

Data for the web portion of the survey was collected from May 31 to June 21, 2006.

Margin of Error

The 474 completed consumer questionnaires provide a 95% confidence level with an overall minimum and maximum margin of error of ±2.7% and ±4.5%, respectively, in estimating the proportion of the population who possess a certain characteristic or opinion. In other words, if 100 samples (all having a total of 474 completed consumer questionnaires) were drawn from this population, approximately 95 of the samples would have proportions within ±2.7% and ±4.5% of the proportions of the entire population for the characteristic or opinion being measured.

The margin of error explained previously only applies to responses of the *entire* sample. As shown in the next chart, the margin of error will be larger when looking at the responses of smaller segments.

			Margin of Error for results at or about...				
Populations	**Completions**	**Households**	10%/90%	20%/80%	30%/70%	40%/60%	50%/50%
All Consumers	474	100,000	2.7%	3.6%	4.1%	4.4%	4.5%
Web Survey Consumers	68	14,346	7.1%	9.5%	10.9%	11.6%	11.9%

** The maximum margin of error is shown in the "50%/50%" column and the minimum margin of error is shown in the "10%/90%" column.*

Section

4

THE QUESTIONNAIRE

The questionnaire in this report was used for interviews conducted with consumers for this study.

Consumer Survey

Case #: _____ Employee #: _____ Date: ___/___/___ Start Time: ____ : ____ am/pm Resp. #: _____

Phase One

Intro. **Hi, my name is *(Your First Name)* from Winkelman Consulting in Fargo, North Dakota and I'm not selling anything – I'm simply conducting a short survey about stores that sell hunting, fishing, camping, and outdoor gear in your state. The survey will only take about 1 to 3 minutes, there are no right or wrong answers, and your answers will be kept confidential. This survey is for people who have bought hunting, fishing, camping, or outdoor gear in the past 12 months. Using a scale of within the past 30 days, 1 to 6 months, 7 to 12 months, more than 12 months, or never; how recently have you bought any…**

	30 days	1 to 6 months	7 to 12 months	More than 12 months	Never	Not sure
I1a. **Fishing gear?**	1	2	3	4	5	91
I1b. **Hunting gear?**	1	2	3	4	5	91
I1c. **Camping gear?**	1	2	3	4	5	91
I1d. **Other outdoor gear?**	1	2	3	4	5	91

DC1. *[IF the respondent has bought any of the above in the past 12 months (I1a =< 3 OR I1b =< 3 OR I1c =< 3 OR I1d =< 3), then SKIP to DC3 (DC1-2 = 999).]* **Is there another person in your household who has bought hunting, fishing, camping, or outdoor gear in the past 12 months?**

1..... Yes

Discontinue Close (DC1) ← ← 2..... No, Not Sure, No Response

DC2. **May I please talk to that person?**

Start over at "Intro" ← ← ← ← ← ← ← ← 1..... Yes

Record first name & call back date/time ← ← 2..... Yes, but currently not available

Discontinue Close (DC2) ← ← ← ← ← ← ← 3..... No, Refused *[ONLY use this response choice if they "refuse" to allow you to speak to the other person.] [If they say "no" because the other person is not there right now, then select response choice "2" above.]*

DC3. **Do you or any members of your household work for a store that sells mainly hunting, fishing, camping, or outdoor gear?**

Discontinue Close (DC3) ← ← 1..... Yes

2..... No, Not Sure, No Response

A1. **First, do you have a computer that you can use to access the Internet?** *[The quota will be set after 1-2 days of calling have been completed.]*

1..... Yes

2..... No → → → *If quota is full, then Discontinue Close (DC4)*

91..... Not sure/No response → → → *If quota is full, then Discontinue Close (DC4)*

Copyright © 2006 Winkelman Consulting

A2. **I'd like you tell me what words or words you use to describe a large store that sells mainly hunting, fishing, camping, and outdoor gear. For example, people may use the words "Mexican restaurant" to describe a business that serves meals that include mainly Mexican foods. Also, people may use the word "Thermos" to describe a travel vacuum bottle or container a person can carry with them and use to keep cold beverages cold or warm beverages warm. So, what word or words would you use to describe a large store that sells mainly hunting, fishing, camping, and outdoor gear?** *[Correctly type in the exact word or words stated by the respondent! Do not paraphrase, abbreviate, or group comments!]*

a.
b.
c.

A3a. **Now, I'd like you to think about the names of large stores that sell mainly hunting, fishing, camping, or outdoor gear. When you think of names of large stores that sell mainly hunting, fishing, camping, or outdoor gear, what is the one store name that comes to mind first?** *[If the respondent states a name that is not exactly as written in the response choice, then select "other" and type in the name exactly as stated.]*

1 ..... Bass Pro or Bass Pro Shop
2 ..... Bass Pro Outdoor World or Bass Pro Shop Outdoor World
3 ..... Bass Pro Sportsman's Center or Bass Pro Shop Sportsman's Center
4 ..... Bass Pro Sportsman's Warehouse or Bass Pro Shop Sportsman's Warehouse
5 ..... Cabela's
6 ..... Dick's Sporting Goods
7 ..... Gander Mountain
8 ..... Scheel's or Scheel's Allsports
9 ..... Sportsman's Warehouse
10..... WalMart
90..... Other *[Specify: _____]*
SKIP to A4 ← ← ← 91..... Not sure, No response

A3b-g. *[IF A3a = 91, then SKIP to A4 (A3b = 92, A3c = 93, A3d = 94, A3e = 95, A3f = 96, A3g = 97).]* **What other store names come to mind when you think about large stores that sell mainly hunting, fishing, camping, or outdoor gear?** *[Probe for up to 6 responses] [If the respondent states a name that is not exactly as written in the response choice, then select other and type in the name exactly as stated.]*

	B	C	D	E	F	G
1. Bass Pro or Bass Pro Shop	1	1	1	1	1	1
2. Bass Pro Outdoor World or Bass Pro Shop Outdoor World	2	2	2	2	2	2
3. Bass Pro Sportsman's Center or Bass Pro Shop Sportsman's Center	3	3	3	3	3	3
4. Bass Pro Sportsman's Warehouse or Bass Pro Shop Sportsman's Warehouse	4	4	4	4	4	4
5. Cabela's	5	5	5	5	5	5
6. Dick's Sporting Goods	6	6	6	6	6	6
7. Gander Mountain	7	7	7	7	7	7
8. Scheel's or Scheel's Allsports	8	8	8	8	8	8
9. Sportsman's Warehouse	9	9	9	9	9	9
10. WalMart	10	10	10	10	10	10
90. Other *[Specify: _____]*	90	90	90	90	90	90
91. Not sure, No response	92	93	94	95	96	97

IF A3b = 92, then SKIP to A4 (A3c = 93, A3d = 94, A3e = 95, A3f = 96, A3g = 97).
IF A3c = 93, then SKIP to A4 (A3d = 94, A3e = 95, A3f = 96, A3g = 97).
IF A3d = 94, then SKIP to A4 (A3e = 95, A3f = 96, A3g = 97).
IF A3e = 95, then SKIP to A4 (A3f = 96, A3g = 97).
IF A3f = 96, then SKIP to A4 (A3g = 97).

A4. *[IF "Sportsman's Warehouse" was named in A3b-g (A3b = 9 OR A3c = 8 OR A3d = 9 OR A3e = 9 OR A3f = 9 OR A3g = 9), then SKIP to B1 (A4-A5cstate = 992).] [IF "Sportsman's Warehouse" was named in A3a (A3a = 9), then SKIP to B1 (A4-A5cstate = 991).]* **When you hear the words "Sportsman's Warehouse", do you think these words are most likely...**

3 ..... **The name of one specific store or chain of stores,**
4 ..... **Words that describe one specific store or chain of stores, or**
SKIP to B1 ← ← ← 5 ..... **General words that can be used to describe any large store that sells mainly hunting, fishing, camping, or outdoor gear?**
SKIP to B1 ← ← ← 91..... *[Do Not Read]* Not sure, No response
991..... *[Do NOT include in response choices seen by interviewers]* Top-of-mind awareness of Sportsman's Warehouse
992..... *[Do NOT include in response choices seen by interviewers]* Unaided awareness of Sportsman's Warehouse

A5. *[IF A4 >= 5, then SKIP to B1 (A5acity-A5cstate = 993).]* **To the best of your knowledge, in what city or town is a "Sportsman's Warehouse" located in your state?**

a. City:	a. State:
b. City:	b. State:
c. City:	c. State:

Phase Two

B1. *[IF they do not have Internet access (A1 >= 2), then SKIP to D1 (B1-C2d = 999).]* **For this interview, I will also want you to look at some sample logos, and then answer a few questions about the sample logos. To do this, I will need you to go to an Internet Web site that has the sample logos. Can you talk to me on this telephone and access the Internet at the same time?**

1..... Yes
Get callback information ← ← 2..... No, but can be online and talk on a different number
Name & alternate number and then call back immediately!

SKIP to C1a ← ← ← ← ← ← 3..... No, have no other phone number
SKIP to C1a ← ← ← ← ← ← 91..... Not sure/No response

Can Access Internet While on the Telephone

B2a. *[IF B1 >= 3, then SKIP to C1a (B2a-B4b = 998).]* **Is your computer currently online?**
SKIP to B2c ← ← ← 1 ..... Yes
2 ..... No
91..... Not sure/No response

B2b. **Can you please go to your computer and let me know when you are online.** *[Once they say they are online, continue]*

B2c. **First, you will want to open Internet Explorer or the program you use for going to the Internet. The first screen that pops up should be your web browser. In the address box at the top of your web browser, type 66.223.28.200/20 and then either hit the enter key or click on the "go" button. When you see the web page that says, "Winkelman Consulting - Welcome to our sample logo web site!", let me know.** *[Read as 66 period 223 period 28 period 200 slash 20 – the slash to be used is the slash (/) on the same key as the "question mark".] [Once they say they see the Web page, continue] [If the respondent has "pop up blocking" on their web browser, the logo will not pop up right away. To see the logos, they will need to click on the "pop up bar", then select "Allow pop ups from this site". Then, the logo should appear.]*

[If respondents has problem getting online or to the web site, ask them to see if they can figure out what is wrong and then schedule a call back date/time. Continue on from this point when you call back.]

Copyright © 2006 Winkelman Consulting

B3a. **Next, I'm going to give you the user name and password you need to go to the next page. Before I do that, however, let me tell you what will happen once you enter this information. As soon as you click on the submit button, a sample logo will appear for about 10 seconds, then disappear. Once the sample logo disappears, let me know. First, click on the "user name" box, then type in this 5-digit "user name" number:** *(Username from list)*. **Now, click on the "password box", then type in this 5-digit "password" number:** *(Password from list)*. **Last, get ready to look at the sample logo and then click on the "continue" button.** *[Once they say the picture disappeared, continue]*

[Once the ID & password have been entered and the "continue button" clicked, the respondent will not be able to access the web site again using the same number. If the respondent accidentally closes the website or experiences a technical problem before the interview is done, get their first name, schedule a callback date/time, ask the supervisor for an extra "user name" and "password", and then call them back.] [Programming note: we MUST retain the original "user name" and "password". Copy to original "user name" to "oldid" and copy the original "password" to "oldpassword", then copy the "extra" "user name" and "password" into the fields the needed for the script above.]

B3b. **After seeing this sample logo, do you think it is most likely...**

1 ..... **An actual logo that is currently used by one specific store or chain of stores,**
2 ..... **A logo that could be used to identify one specific store or chain of stores, or**
3 ..... **A general logo that could be used by any sporting goods store to let consumers know their stores sell mainly hunting, fishing, camping, or outdoor gear?**
91..... *[Do Not Read]* Not sure, No response

B4a. **Now, I'm going to give you the password you need to see the next page. This time, once you enter the password, another sample logo will appear for about 10 seconds, then disappear. When the sample logo disappears, let me know. Click on the password box. Then type in the number "1". Then click on the "continue" button.** *[Once they say the picture disappeared, continue]*

B4b. **After seeing this sample logo, do you think it is most likely...**

1 ..... **An actual logo that is currently used by one specific store or chain of stores,**
2 ..... **A logo that could be used to identify one specific store or chain of stores, or**
3 ..... **A general logo that could be used by any sporting goods store to let consumers know their stores sell mainly hunting, fishing, camping, and outdoor gear?**
91..... *[Do Not Read]* Not sure, No response

Cannot Access Internet While on the Telephone

C1a. *[IF B1 <= 2, then SKIP to D1 (C1a-C2d = 996).]* **Could you please get a pen and paper so you can write down the web site address and the ID and password needed to access the web site.** *[When they have the writing materials, then continue.]* **The web site address is 66.223.28.200/40, your 5-digit ID number is** *(Username from list)* **and your 5-digit password number is** *(Password from list)*. **As soon as we are done with this interview, you can go to the web site, view the sample logos, and answer two quick questions. If your web browser is set to block "pop ups", the logos will not pop up right away. To see the logos, they will need to click on the "pop up bar" near the top of your web browser page, then select "Allow pop ups from this site". Then, the first logo should appear.** *[Read as 66 period 223 period 28 period 200 slash 40 – the slash to be used is the slash on the same key as the "question mark".] [Once they say they have written everything down, then continue]*

C1b. **Do you have an email address that I can send a reminder email to that provides the web site address and the password you will need to access the sample logos?**

1 ..... Yes
SKIP to D1 ← ← ← 2 ..... No
SKIP to D1 ← ← ← 91..... Not sure, refused

C2a. *[IF C1b >= 2, then SKIP to D1 (C2a-C2c = 995).]* **What is your email address?**

C2b. **Let me read your email address back to you so I can make sure I typed it correctly?** [Interviewers: *When on the phone, letters like "b", "d", "p", "t", etc. can count alike. When confirming the address, say the letters ("b" as in "boy", "d" as in "dog", "p" as in "pencil", "t" as in truck, etc. (see list)) rather than words.] [Correct email address in C2a as needed, then retype correct email address into C2b.]*

C2c. *[IF C2b = C2a, then SKIP to C2d (C2c = Missing).]* **I'm sorry, I must have typed your email address incorrectly. Can you give it to me again?** *[Return to C2a.]*

C2d. **In the next day or so, you should receive your reminder email with the web site address and the ID and password needed to access the website. As soon as you get the email, please go to the website, view the sample logos, and answer the questions if you haven't already done so before you get the email.**

Demographics

D1. **For classification and statistical analysis purposes, may I have your age please?** *[Do Not Read categories UNLESS reluctant or refuse, then ask, "**Please stop me when I read the age group which includes you...**"]*

1 ..... **24 or younger,**
2 ..... **25 to 34,**
3 ..... **35 to 44,**
4 ..... **45 to 54,**
5 ..... **55 to 64, or**
6 ..... **65 or older?**
91 ..... *[Do Not Read]* Choose not to answer

D2. **How long have you lived in your state – would you say...**

1 ..... **5 years or less,**
2 ..... **6 to 10 years,**
3 ..... **11 to 20 years,**
4 ..... **21 to 40 years, or**
5 ..... **More than 40 years?**
6 ..... *[Do Not Read]* Not sure, No response

D3. *[Do Not Read]* Enter respondent's gender.

1 ..... Male
2 ..... Female

VERIFICATION CLOSE: **That's all the questions I have for you. Lastly, let me verify that I dialed ____-_____. Again, my name is *(Your First Name)*, and on occasion a small percentage of people are called back just to verify that this interview actually took place. May I please have your first name, and first name only, so my supervisor will know who to ask for in case this interview is verified. Thank you for your time and have a good (evening/day).**

Resp. First Name: _____ Phone: (_____) _____-_____ End Time: _____

D4. *[Do Not Read]* Enter zip code from list. ___ ___ ___ ___ ___

D5. *[Do Not Read]* Enter state code from list. ___ ___

DISCONTINUE CLOSE: **That's all the questions I have for you. Thank you for your time and have a good (evening/day).**

Section

5

CONTINGENCY TABLES

Contingency tables are commonly referred to as "cross-tabs". They present the findings in an easy-to-understand, table form and provide the categorical data that is used most frequently in marketing. We strongly recommend that you review these tables and use them to facilitate any major decisions you make.

The contingency tables on the following pages show the proportion of all respondents who gave various responses to each question, as well as the proportion of specific sample segments that provided a particular response.

Please note the tables are separated into sections. The tables in each section have the same "banners" or sample segments across the top. Within each section, the tables are in order by question number, which appears on the top, left-hand side of each table.

NOTE: The size and page numbering of the tables in the electronic version of the report will not be identical to that in the printed report.

		Total	Survey Segment: Consumers
A2a-c. What word or words would you use to describe a large store that sells mainly hunting, fishing, camping, and outdoor gear?	Academy	2.5%	2.5%
	Ace Hardware	.4%	.4%
	All sports, all around, all in one, all purpose, one stop	1.9%	1.9%
	American store	.2%	.2%
	Anything store	.2%	.2%
	Authentic	.2%	.2%
	Awesome, amazing	.8%	.8%
	Bait shop	.2%	.2%
	Bass Outdoors	.2%	.2%
	Bass Outlets	.4%	.4%
	Bass Pro, Bass Pro Shop	8.4%	8.4%
	Big, big box, big book	.8%	.8%
	Buckhorn	.2%	.2%
	Burger King, have it your way	.2%	.2%
	Cabela's	7.8%	7.8%
	Camping, camping store, camping suplies	2.1%	2.1%
	Camping World	.4%	.4%
	Catalogue store	.2%	.2%
	Clothing World	.4%	.4%
	Complete, comprehensive	.6%	.6%
	Convenient, handy	.6%	.6%
	Corporate store, chain	.4%	.4%
	Department store	.4%	.4%
	Depot	.2%	.2%
	Dick's	.6%	.6%
	Discount store	.4%	.4%
	Eastern Mountain Sports	.2%	.2%
	Enjoyment, entertaining, exciting, experience, extravagant	1.1%	1.1%
	Equipment	.2%	.2%
	Everything store	1.3%	1.3%
	Expensive	.6%	.6%
	Fabulous, fantastic, fun, favorite	1.9%	1.9%

	Fishing, fishing store, fishing gear	3.2%	3.2%
	Fishing & hunting, store, gear	.8%	.8%
	Full gear	.2%	.2%
	Gaming	.4%	.4%
	Gander Mountain	1.3%	1.3%
	Gart/Garth Brothers	.2%	.2%
	Giant, huge	.8%	.8%
	Good, very good, great	1.3%	1.3%
	Grandpa's	.2%	.2%
	Gun store, gun shop	.4%	.4%
	Hardware store	.4%	.4%
	Heaven, paradise, dream world	1.5%	1.5%
	Hunter's haven	.2%	.2%
	Hunting, hunting store, hunting gear	3.6%	3.6%
	Hunting dog training, gear	.2%	.2%
	Hunting season	.2%	.2%
	Ice chest	.2%	.2%
	Jordan Creek	.2%	.2%
	Interesting	1.3%	1.3%
	K-Mart	.4%	.4%
	Large store, selection	1.1%	1.1%
	Man store, manly, rugged folks	.6%	.6%
	Mega store	.2%	.2%
	Midway	.2%	.2%
	Mossy Oak	.2%	.2%
	Mountains	.2%	.2%
	Multi-purpose	.4%	.4%
	Name brands	.2%	.2%
	Non-Wal-Mart	.2%	.2%
	Out World	.2%	.2%
	Outdoor store, center, recreation, gear	14.1%	14.1%
	Outdoor sports store, center	1.9%	1.9%
	Outdoor World	1.5%	1.5%
	Outfitter	1.1%	1.1%
	Outlet	.2%	.2%
	Overwhelming	.2%	.2%
	Pro Bass, Pro Bass Shop	.8%	.8%
	PX	.2%	.2%
	Recreation store	.6%	.6%
	REI	.8%	.8%
	Retail store	.2%	.2%
	Right place	.2%	.2%

	RV, travel store	.4%	.4%
	Scheel's	.2%	.2%
	Specialty store, specialized	.4%	.4%
	Sporting goods, sporting goods store	17.7%	17.7%
	Sporting superstore	.2%	.2%
	Sports store, center, shop, gear	3.2%	3.2%
	Sports Academy	.4%	.4%
	Sports warehouse	.4%	.4%
	Sports World	.2%	.2%
	Sportsman(s) Warehouse	.6%	.6%
	Sportsman(s)	.8%	.8%
	Sportsman(s) outfitter	.2%	.2%
	Sportsman(s) store	.4%	.4%
	Super center	.2%	.2%
	Swapmeet	.2%	.2%
	Tamitta	.2%	.2%
	Terrific, wonderful	.4%	.4%
	"The" Store	.2%	.2%
	Toy store	.4%	.4%
	Variety, variety store, well-stocked	.8%	.8%
	Wal-Mart	2.7%	2.7%
	Warehouse	1.1%	1.1%
	Misc. others	.8%	.8%
	Not sure, no response	9.1%	9.1%
	No second response	83.8%	83.8%
	No third response	95.6%	95.6%
Total Responses		474	474

Includes all respondents

		Total	Survey Segment. Consumers
A3a. When you think of names of large stores that sell mainly hunting, fishing, camping, or outdoor gear, what is the one store name that comes to mind first?	Bass Pro or Bass Pro Shop	26.6%	26.6%
	Bass Pro Outdoor World or Bass Pro Shop Outdoor World	.6%	.6%
	Bass Pro Sportsman's Warehouse or Bass Pro Shop Sportsman's	.2%	.2%
	Cabela's	30.0%	30.0%
	Dick's Sporting Goods	1.9%	1.9%
	Gander Mountain	3.2%	3.2%
	Scheel's or Scheel's Allsports	1.3%	1.3%
	Sportsman's Warehouse	2.7%	2.7%
	WalMart, Sam's	8.0%	8.0%
	Academy	4.6%	4.6%
	Ace Hardware	.2%	.2%
	Backwoods	.4%	.4%
	Baron's	.2%	.2%
	Bass	.2%	.2%
	Big 5, Big Five	.8%	.8%
	Big Book	.2%	.2%
	Big Buck(s)	1.5%	1.5%
	Boot Box	.2%	.2%
	Camping World	.2%	.2%
	Coleman	.2%	.2%
	Dunham's	.2%	.2%
	Eddie Bauer	.4%	.4%
	Farm Fleet	.2%	.2%
	Fin & Feather	.2%	.2%
	French Creek Outfitters	.4%	.4%
	Gart/Garth Brothers	.6%	.6%
	Gene Taylor's	.4%	.4%
	L L Bean	.8%	.8%
	Orvis	.2%	.2%
	Pro Bass, Pro Bass Shop	.2%	.2%
	REI	1.5%	1.5%
	Sears	1.1%	1.1%
	Sports Academy	.6%	.6%
	Sports Authority	.6%	.6%
	Misc. others	4.0%	4.0%
	Not sure, No response	5.1%	5.1%
Total Responses		474	474

Includes all respondents

		Total	Survey Segment. Consumers
A3a-g. Unaided awareness of stores that sell mainly fishing, hunting, camping and outdoor gear.	Bass Pro or Bass Pro Shop	49.2%	49.2%
	Bass Pro Outdoor World or Bass Pro Shop Outdoor World	1.1%	1.1%
	Bass Pro Sportsman's Warehouse or Bass Pro Shop Sportsman's	.4%	.4%
	Cabela's	47.5%	47.5%
	Dick's Sporting Goods	8.2%	8.2%
	Gander Mountain	14.6%	14.6%
	Scheel's or Scheel's Allsports	4.4%	4.4%
	Sportsman's Warehouse	9.1%	9.1%
	WalMart, Sam's	28.7%	28.7%
	Academy	11.6%	11.6%
	Ace Hardware	.6%	.6%
	Backwoods	.4%	.4%
	Baron's	.2%	.2%
	Bass	.2%	.2%
	Big 5, Big Five	2.7%	2.7%
	Big Book	.2%	.2%
	Big Buck(s)	1.9%	1.9%
	Boot Box	.2%	.2%
	Camping World	1.7%	1.7%
	Campmor	.4%	.4%
	Coleman	.4%	.4%
	Dunham's	.6%	.6%
	Eddie Bauer	.8%	.8%
	Farm Fleet	.4%	.4%
	Fin & Feather	.8%	.8%
	French Creek Outfitters	.6%	.6%
	Gart/Garth Brothers	3.2%	3.2%
	Gene Taylor's	.4%	.4%
	Great Outdoors	.8%	.8%
	K-Mart	3.4%	3.4%
	L L Bean	2.7%	2.7%
	Lowe's	.6%	.6%
	Orvis	.8%	.8%
	Oshman's	.6%	.6%
	Outdoor America	.6%	.6%
	Pro Bass, Pro Bass Shop	.4%	.4%
	Outdoor World	.2%	.2%
	REI	4.0%	4.0%
	Sears	1.9%	1.9%
	Sports Academy	.8%	.8%
	Sports Authority	2.3%	2.3%
	Surplus City	.6%	.6%
	Target	1.3%	1.3%
	Misc. others	27.8%	27.8%
	Not sure, No response	5.1%	5.1%
	No second response	21.3%	21.3%
	No third response	57.4%	57.4%
	No fourth response	83.1%	83.1%
	No fifth response	95.1%	95.1%
	No sixth response	98.3%	98.3%
	No seventh response	99.8%	99.8%
Total Responses		474	474

Includes all respondents

		Total	Survey Segment. Consumers
A3-A4. When you hear the words Sportsman's Warehouse, do you think these words are most likely...	Top-of-mind awareness of Sportsman's Warehouse	2.7%	2.7%
	Unaided awareness of Sportsman's Warehouse	6.3%	6.3%
	The name of one specific store or chain of stores	35.9%	35.9%
	Words that describe one specific store or chain of stores	10.3%	10.3%
	General words that describe any large sporting goods store	35.9%	35.9%
	Not sure, No response	8.9%	8.9%
Total Responses		474	474

Includes all respondents

		Total	Survey Segment. Consumers
A3-A5. Overall awareness and knowledge of Sportsman's Warehouse.	Top-of-mind awareness	2.7%	2.7%
	Unaided awareness	6.3%	6.3%
	Named a SW city/area	10.8%	10.8%
	SW is name of one specific store or chain	27.4%	27.4%
	SW are words that describe one specific store or chain	8.0%	8.0%
	SW are words that describe any large sporting goods store	35.9%	35.9%
	Not sure, no response	8.9%	8.9%
Total Responses		474	474

Includes all respondents

		Total	Survey Segment. Consumers
A1. Do you have a computer that you can use to access the Internet?	Yes	69.2%	69.2%
	No	30.6%	30.6%
	Not sure, No response	.2%	.2%
Total Responses		474	474

Includes all respondents

		Total	Survey Segment. Consumers
B1. Can you talk to me on this telephone and access the Internet at the same time?	Yes	11.6%	11.6%
	No	87.8%	87.8%
	Not sure, No response	.6%	.6%
Total Responses		328	328

Includes all respondents who have Internet access

		Total	Survey Segment. Consumers
B3b. (Sportsman's Warehouse sample logo) After seeing this sample logo, do you think it is most likely...	An actual logo that is currently used by one specific store	41.2%	41.2%
	A logo that could be used to identify one specific store or	32.4%	32.4%
	A general logo that could be used by any sporting goods stor	13.2%	13.2%
	Not sure, No response	13.2%	13.2%
Total Responses		68	68

Includes all respondents who responded to the web portion of the survey

		Total	Survey Segment. Consumers
B4b. (Bass Pro Shop sample logo) After seeing this sample logo, do you think it is most likely...	An actual logo that is currently used by one specific store	66.2%	66.2%
	A logo that could be used to identify one specific store or	22.1%	22.1%
	A general logo that could be used by any sporting goods stor	10.3%	10.3%
	Not sure, No response	1.5%	1.5%
Total Responses		68	68

Includes all respondents who responded to the web portion of the survey

		Total	Survey Segment. Consumers
D1. Respondent's age.	24 or younger	3.8%	3.8%
	25 to 34	14.1%	14.1%
	35 to 44	19.8%	19.8%
	45 to 54	22.6%	22.6%
	55 to 64	22.8%	22.8%
	65 or older	14.6%	14.6%
	Choose not to answer	2.3%	2.3%
Total Responses		474	474

Includes all respondents

		Total	Survey Segment. Consumers
D2. How long have you (lived in your state/sold gear to retailers)?	5 years or less,	4.4%	4.4%
	6 to 10 years,	7.4%	7.4%
	11 to 20 years,	10.5%	10.5%
	21 to 40 years, or	33.3%	33.3%
	More than 40 years?	42.2%	42.2%
	Not sure, No response	2.1%	2.1%
Total Responses		474	474

Includes all respondents

		Total	Survey Segment. Consumers
D3. Respondent's gender.	Male	63.5%	63.5%
	Female	36.5%	36.5%
Total Responses		474	474

Includes all respondents

		Total	Survey Segment. Consumers
D5. State of residence.	Colorado	11.6%	11.6%
	Iowa	12.9%	12.9%
	Mississippi	13.3%	13.3%
	Nevada	9.1%	9.1%
	Oklahoma	13.7%	13.7%
	Pennsylvania	11.2%	11.2%
	South Carolina	9.1%	9.1%
	Tennessee	9.9%	9.9%
	Texas	9.3%	9.3%
Total Responses		474	474

Includes all respondents