

A Judge's Deskbook on the Basic Philosophies and Methods of Science

MODEL CURRICULUM



State Justice Institute

Additional funding provided by:

**The National Judicial College
The Federal Judicial Center
The Grant Sawyer Center for Justice Studies**

**at the University of Nevada, Reno
The Master of Judicial Studies Program
at the University of Nevada, Reno
The Permanency Planning for Children Department of the
National Council of Juvenile and Family Court Judges
U.S. Department of Justice, Office of Justice Programs,
Office of Juvenile Justice and Delinquency Prevention**

© March 1999

Acknowledgments ...

This document is the product of a national research project designed to investigate state trial court judges' understanding and knowledge of scientific methods and principles, opinions about the judge's role with respect to the admissibility of scientific evidence, knowledge about and attitudes toward the relatively new *Daubert* guidelines, and attitudes toward scientific evidence in general.

This curriculum, and the national research more generally, reflects a great deal of hard work and commitment from a variety of individuals. This curriculum was written by Shirley A. Dobbin, Ph.D., and Sophia I. Gatowski, Ph.D., Co-Project Directors. The authors and the Co-Principal Investigators on the research project, James T. Richardson, J.D., Ph.D., and Gerald P. Ginsburg, Ph.D., wish to thank a number of individuals for their contribution to this project.

Graduate research assistants Mara L. Merlino, Project Coordinator, and Veronica Dahir, Survey Manager, were integral to the success of the national research and to the completion of this curriculum; their hard work, long hours, and good humor were greatly appreciated. Thanks are also due to the staff of graduate and undergraduate students who served as interviewers and coders, especially Ryan Powell, Melissa Litchfield, Roselyn Polk, Cathy Felty, Jennifer Spencer, and Kimberly Surratt. We would also like to thank Denise Schaar-Buis, Project Administrator, Master of Judicial Studies, for her administrative support, Travis Linn, Dean, Reynolds School of Journalism, for his assistance in developing the website, and the members of the Advisory Board for their advice and guidance.

Thanks are especially owed to the more than 500 judges from around the country who served as pilot subjects, interviewees, focus group participants, and reviewers. This curriculum could not have been produced without their assistance -- especially the state trial court judges who participated in the national survey and spent, on average, an hour of their time on the telephone with survey staff.



Primary funding for the national research and the development of this curriculum was provided by the State Justice Institute.

This document was developed under grant number SJI-97-02B-D-030 from the State Justice Institute. The points of view expressed are those of the authors and do not necessarily represent the official position or policies of the State Justice Institute.

Additional funding was provided by:

- **The National Judicial College, Reno, Nevada**
- **The Federal Judicial Center, Washington, D.C.**
- **The Grant Sawyer Center for Justice Studies, University of Nevada, Reno**
- **The Master of Judicial Studies Program, University of Nevada, Reno**
- **The Permanency Planning for Children Department of the National Council for Juvenile and Family Court Judges, Reno, Nevada**

- **The U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention, Washington, D.C.**

This document was partially supported by Grant No. 96-CT-NX-001 from the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice. Points of view or opinions expressed are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. Reproduction of this publication for noncommercial educational and informational purposes is encouraged.

GRANT STAFF

Principal Investigators:

James T. Richardson, J.D., Ph.D.
Director, Master of Judicial Studies Program
Professor, Department of Sociology
University of Nevada, Reno/311
Reno, NV 89557

Gerald P. Ginsburg, Ph.D.
Professor Emeritus
Interdisciplinary Ph.D. Program in Social
Psychology/296
University of Nevada, Reno
Reno, NV 89557

Project Directors:

Shirley A. Dobbin, Ph.D.
Faculty Associate, Grant Sawyer Center
for Justice Studies
Research Specialist, National Council of
Juvenile and Family Court Judges
P.O. Box 8970
Reno, NV 89507

Sophia I. Gatowski, Ph.D.
Faculty Associate, Grant Sawyer Center
for Justice Studies
Research Specialist, National Council of
Juvenile and Family Court Judges
P.O. Box 8970
Reno, NV 89507

Graduate Research Assistants:

Mara L. Merlino
Project Coordinator/Research Assistant
Grant Sawyer Center for Justice Studies/313
University of Nevada, Reno
Reno, NV 89557

Veronica Dahir
Survey Manager/Research Assistant
Grant Sawyer Center for Justice Studies/313
University of Nevada, Reno
Reno, NV 89557

Advisory Board Members:

Richard A. Bjur, M.D., Ph.D.
Department of Pharmacology
University of Nevada, Reno
Reno, NV 89557

Krista R. Johns, J.D., Assistant Director
Permanency Planning for Children Dept.
National Council of Juvenile and Family
Court Judges
Reno, NV 89507

Joe Cecil, J.D., Ph.D., Director
Division of Research
Federal Judicial Center
One Columbus Circle, NE

Thomas Langhorne
Director of Educational Services
Supreme Court of Virginia

CHAPTER 1 Introduction and Overview

Our country's faith in science and its commitment to judicial resolution of disputes ensure that the judiciary will continue to be called upon to decide questions on the frontiers of science; these questions will arise in cases that raise profound social, economic, and public policy concerns. The time has come to give the judiciary the support it needs to perform this difficult task.¹

Trial judges at both state and federal levels are being called upon more frequently to make decisions involving proffered scientific evidence. All kinds of evidence from the physical, chemical, biological, and social sciences is being proffered in mass tort cases involving alleged damage to people or to the environment, as well as in cases as varied as murder trials and custody disputes. Psychological syndrome and 'profile' evidence is also proliferating rapidly, quite often in difficult cases involving allegations of child sexual abuse, spouse battering, and sexual assault. Indeed, the contemporary climate in the courtroom is one where science is proffered more frequently in more varied cases, requiring judges to become more discerning 'gatekeepers' as they dispense justice.

The scientific and technological communities of relevance to the court include:

- mathematical sciences (mathematics and statistics);
- computer science;
- physical sciences (e.g., chemistry and physics);
- life sciences (e.g., biology, medicine, agriculture);
- environmental sciences (e.g., atmospheric sciences, earth science, oceanography);
- social and behavioral sciences (e.g., psychology, economics, geography, political science, sociology, anthropology);
- forensic sciences (e.g., bite mark analysis)
- engineering (civil, chemical, electrical, industrial, mechanical).

Within the increasingly science-rich culture of the court, trial judges are being called upon to examine the methods, techniques, and underlying logic of science before making admissibility

rulings. In order to carry out their duties effectively, judges must be able to discern ‘good’ science from ‘bad’ science, which in turn means that judges must come to some understanding of the philosophy and sociology of science, and the proper ways of doing science. Judges must be able to assess the validity of purportedly scientific evidence proffered by attorneys and experts, whether the judge is handling a bench trial, deciding a motion *in limine* to suppress certain evidence, presiding over a pre-trial hearing concerning the evidence, or making a ruling on admissibility with the jury present. Whether or not judges possess the necessary level of scientific literacy required by such admissibility standards (and specifically the guidelines therein) is a matter of some controversy. Indeed, current practice and training of the judiciary may not sufficiently prepare them to perform the role of scientific evaluator.²

The courts’ ability to handle complex science-rich cases has recently been called into question, with widespread allegations that the judicial system is increasingly unable to manage and adjudicate science and technology issues. Critics have objected that judges cannot make appropriate decisions because they lack technical training, that jurors do not comprehend the complexity of the evidence they are supposed to analyze, and that the expert witnesses on whom the system relies are mercenaries whose biased testimony frequently produces erroneous and inconsistent determinations. If these claims go unanswered, or are not dealt with, confidence in the judiciary will be undermined as the public becomes convinced that the courts as now constituted are incapable of correctly resolving some of the most pressing legal issues of our day.³

The *Deskbook* that comprises one component of this curriculum is not intended to instruct judges about whether or not specific types of scientific evidence should be admissible. Rather, this *Deskbook* addresses critical issues with respect to the philosophy of science and the scientific method. It is hoped that this curriculum will aid in identifying issues and clarifying questions relevant to admissibility rulings, such as the relevance, reliability, and methodology of scientific evidence. Specific types of evidence are highlighted for illustrative purposes only. *The focus of this curriculum is the philosophy and methods of science, not the philosophy and methods of law.*

A number of very useful and comprehensive guides to scientific evidence for the judiciary already exist.⁴ However, this curriculum makes a unique contribution. It is unique because the content areas and issues highlighted were selected for inclusion in the curriculum on the basis of responses provided by 400⁵ state trial court judges around the country who participated in an extensive survey about scientific evidence. Information obtained from the survey about gaps in judges' knowledge and understanding of science and scientific method, as well as their experiences in dealing with various kinds of proffered scientific evidence, was used to shape the curriculum. This wealth of information about knowledge and practice was supplemented with an examination and analysis of case law to provide additional context. In essence, the curriculum presented in the following pages was developed, in large part, by judges, for judges.

Education of the judiciary becomes increasingly important as the system receives more and more cases that require judges to have some familiarity with scientific methodology and the factors scientists consider when they evaluate scientific work. Some comprehension of statistical and sampling concepts is crucial, given their importance in such disparate types of litigation as antitrust, discrimination, trademarks, and toxic torts. Education is therefore needed ... to provide basic information.⁶

Goals of the Curriculum

The goals of this curriculum are (1) to improve generally the knowledge base upon which judges draw to make their admissibility decisions and (2) to assist judges in articulating their decision-making process with respect to the admissibility of scientific evidence. Specifically, the goals of the curriculum are:

- ◆ To educate judges about the general principles of science and scientific methodology;
- ◆ To help judges become better informed decision-makers with respect to the admissibility of scientific evidence; and
- ◆ To provide judges with a general knowledge base so that they can be critical consumers of science and scientific evidence.

In order to better achieve these goals, a national survey of state trial court judges was conducted. The survey gathered information about judges' understanding and knowledge of scientific methods and principles, opinions about the judicial role with respect to the admissibility of scientific evidence, attitudes toward the *Daubert* guidelines, and attitudes toward scientific evidence more generally. This information provided a national snapshot of the state judiciary's experiences with, and understanding of, science and scientific evidence. It also served to highlight those areas in which judges are in need of the most guidance. This curriculum was designed specifically to address those areas — it is specifically tailored to meet the needs of the judiciary.

Organizational Overview

The structure and organization of this curriculum reflect the comments and recommendations offered by judges who participated in the national survey (at the end of the telephone survey judges were invited to make suggestions about the scope, nature, and organization of the curriculum), as well as comments provided by those judges who participated in focus groups and served as reviewers.

The vast majority of judges indicated that they wanted the curriculum to provide an overview of science, not an overview of law. Further, rather than an in-depth treatment of specific types of scientific evidence, judges indicated that they wanted an overview of the general methods and principles of science that can be applied across a variety of scientific disciplines. Thus, the focus of this curriculum is *science* — its philosophy and methods. Two specific types of scientific evidence, psychological and psychiatric evidence and DNA evidence, are included in this curriculum for illustrative purposes only. These specific types of evidence were chosen because, based upon the results of the national survey, judges find both types of evidence to be particularly problematic, and problematic for different reasons. Moreover, the selection of two such different forms of evidence serves to illustrate commonalities and differences between scientific disciplines.

This judicial curriculum is comprised of three parts: (1) the *Deskbook*; (2) an *Instructor's Manual*, to facilitate classroom teaching; and (3) a *Website*. The three components of this curriculum can be used either individually or in concert with each other. Each chapter in the *Deskbook* focuses on issues specific to the philosophy and conduct of science — e.g., the scientific method, including both quantitative and qualitative approaches, peer review and publication, and statistics. Each chapter presents an overview of the issues involved, raises issues to be addressed, and poses critical questions that can be adapted by the judiciary as decision-making aids. These critical questions are not presented as an exhaustive list that judges should pose when making admissibility decisions about scientific evidence (or any other decisions about the use of scientific evidence in the court). Rather, the questions are presented as indicators of the issues that scientific experts, and other purveyors of science, should address and comment on when proffering science for use in the court. These questions represent what judges should be listening for when scientific evidence is presented and what they should be asking about when the information is not forthcoming.

Each chapter also presents some questions for further reflection. The purpose of these questions is to encourage the reader to actively engage the material, to consider alternative points of view, and to actively reflect on his or her own decision-making process. At the end of each chapter a glossary of terms and a short list of suggested readings is provided, as well as a page for “judge’s notes.”

The *Instructor's Manual*, included as the final component of the *Deskbook*, contains chapter outlines, key concepts, suggested group exercises, and discussion questions. The *Instructor's Manual* has been written so that it can be used by instructors to facilitate classroom presentations of this curriculum or by individual judges using the *Deskbook* as an independent, self-directed learning tool.



A national survey of state trial court judges

A representative sample of state trial court judges was surveyed (N=400) to ascertain their general scientific literacy, understanding of admissibility criteria, and experience in dealing with various kinds of scientific evidence. Baseline data from this survey were crucial in the development of this curriculum (e.g., data guided selection of content areas and indicated appropriate areas for focused attention). The representative nature of the sample allowed for the development of a properly targeted and broadly applicable curriculum.

Research Objectives

- Objective 1: To assess trial judges' knowledge and understanding of scientific principles and practice and how those principles and practices relate to major decisions about the admissibility of scientific evidence.
- Objective 2: To examine case law, looking for patterns and trends, while highlighting problematic areas and applications for close attention, in order to better educate judges about especially difficult issues.
- Objective 3: To develop a curriculum tailored to the needs of state trial court judges, using results from objectives 1 and 2.

The Survey Sample

To be included in the sample, a judge had to be either (1) sitting on the bench of the state trial court of general jurisdiction, or (2) sitting on the bench of a court of special jurisdiction, the docket of which included cases likely to contain the types of evidence relevant to the study. The list of judges' names and information regarding the jurisdiction of their courts was obtained from the 1997-1998 edition of *The American Bench*. Judges in the sampling frame (i.e., *The American Bench*) were stratified first by federal circuit and then by state. These judges were then randomly, but proportionately, sampled. That is, judges were drawn to obtain a sample that was representative of both the geographical distribution of judges and the number of judges using our admission criteria in each state.

Securing Participation

Judges were initially sent a letter of introduction. The letter outlined the nature, purpose, and goal of the research, as well as the importance of the information to be obtained. Within a few weeks of mailing the letter of introduction, judges received follow-up telephone calls from a survey staff member. Following a script designed to encourage participation, the staff member discussed the research further with the judge. Judges were assured that their names would not be reported in connection with their responses. If the judge agreed to participate in the research, the staff member scheduled a date and time for the interview. The overall response rate for the survey was 67%.

Note that 67% is a very high response rate given the nature of the individuals to be surveyed (i.e., high status, remote professionals) and the amount of time necessary to complete the survey (i.e., telephone interviews averaged 60 minutes and included a mailed follow-up). This high response rate provides support for the claim that science and technology issues are becoming increasingly relevant to judges and that judges are becoming increasingly concerned with how best to manage and evaluate scientific evidence.

Endnotes:

1. William T. Golden, Co-Chair, and Joshua Lederberg, Co-Chair. (1993). Foreward, *Science and Technology in Judicial Decision Making: Creating Opportunities and Meeting Challenges*. Carnegie Commission on Science, Technology and Government.
2. *Supra*, note 1, Carnegie Commission; Jasanoff, S. (1993). "What Judges Should Know About the Sociology of Science." *Judicature*, Vol. 77(2), pg. 77-82; Miller, P.S., Rein, B.W., and Bailey, E.O. (1994). "Daubert and the Need for Scientific Literacy." *Judicature*, Vol. 77(5), pg. 254; Richardson, J.T., Ginsburg, G.P., Gatowski, S.I. and Dobbin, S.A. (1995). "The Problems of Applying *Daubert* to Psychological Syndrome Evidence." *Judicature*, Vol. 79(1), July-August, pg. 17-24.
3. *Supra*, note 1, Carnegie Commission, pg. 8.
4. See for example: Black, B., Klein, M.S., and Brunette, S.A. (1994). *Shepard's Expert and Scientific Evidence Quarterly - Evidence Series*; Federal Judicial Center (1994). *Reference Manual on Scientific Evidence*; Faigman, D.L., Kaye, D.H., Saks, M.J., and Sanders, J. (1997). *Modern Scientific Evidence: The Law and Science of Expert Testimony, Vol. I & Vol. II*. West; Clark, Boardman, Callaghan, and Parry, J.W. (1998). *National Benchbook on Psychiatric and Psychological Evidence and Testimony*. ABA Commission on Mental and Physical Disability Law and the State Justice Institute.
5. 400 judges served as interview subjects for the national survey of state trial court judges. Another 140 judges served as an educational comparison group, focus group participants, and reviewers. The responses of the 400 judges in the national survey are summarized in this curriculum.
6. *Supra* note 1, Carnegie Commission, pg. 44.

CHAPTER 2

of *A Judge's Deskbook on the Basic Philosophies and Methods of Science*,
by Shirley A. Dobbin, Ph.D, and Sophia I. Gatowski, Ph.D

The Judicial Role in Evidentiary Decision-Making

Until recently, the criteria for admissibility of scientific evidence in the U.S. federal system was guided largely by the *Frye*⁽¹⁾ decision, which emphasized the acceptance of the proffered theory or explanation by the relevant scientific community. *Frye* had given rise to the 'general acceptance' rule for admission of scientific evidence, which meant that expert evidence should not be admissible unless the methods and principles on which it was based had achieved widespread acceptance in the relevant discipline(s). *Frye*'s much criticized 'general acceptance' rule was overturned by the U.S. Supreme Court in June of 1993 in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*⁽²⁾ which outlined several relatively sophisticated guidelines for determining admissibility of scientific evidence.

Under *Daubert*, scientific evidence must be reviewed in terms of the following general guidelines: (1) whether the theory or technique can be, and has been, tested (i.e., a determination of its 'falsifiability'); (2) whether the evidence has been subjected to peer review and publication; (3) the 'known or potential error rate' associated with applications of a theory; and (4) the general acceptance of the theory or technique in question. It should be noted, however, that the approach to the admissibility of scientific evidence outlined in *Daubert* is intended to be a "flexible" one, and other cases listing criteria are mentioned favorably, especially *U.S. v. Downing* (1985).⁽³⁾ The application and utility of the *Daubert* guidelines, and the gatekeeping role of the judge, were recently upheld in the U.S. Supreme Court decision in *Kumho Tire Co. v. Carmichael* (1999).⁽⁴⁾

The Federal Rules of Evidence (FRE) themselves place limits on admissibility of purportedly scientific evidence. Under the Rules the trial judge must ensure, for instance, that any and all scientific testimony or evidence admitted is relevant and reliable, but not overly prejudicial.

FRE 401 and 402: All relevant evidence is admissible, except as otherwise provided ... Evidence which is not relevant is not admissible. Relevant evidence is defined as that which has any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.

FRE 403: Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.

FRE 702: If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.

While the primary shift in the U.S. standard of admissibility of scientific evidence has occurred at the federal court level (i.e., from *Frye's* 'general acceptance' to *Daubert's* more sophisticated guidelines), changes are also occurring at the level of individual states, which have their own autonomous court systems.

Although state courts are not required to adopt the FRE, and are therefore not required to consider the *Daubert* guidelines, states may choose to adopt the FRE and *Daubert* guidelines, to follow a 'general acceptance' or *Frye* rule, or to develop and implement some other admissibility rules. Whatever the specific admissibility rule, *all* trial court judges in the civil, criminal, juvenile and family court systems, and at both federal and state levels are being put in the position of deciding the relevancy, utility, and probative value of proffered scientific evidence -- whether using a *Frye* standard, a *Daubert* standard, or some other standard.

Before going any further, stop and reflect ...

How would you describe your decision-making role with respect to the admissibility of scientific evidence?

Has your role changed in recent years?

Do you think the judicial gatekeeping role, as articulated in *Daubert* and *Kumho*, is an appropriate one? Why or why not?

Table 1: Management Strategies for Handling Expert Scientific Evidence

	used exclusively in cases with difficult expert evidence	used in cases with various types of expert evidence	would not use at all
Define scope of expert testimony at a Rule 16 Conference (or state equivalent)	9%	39%	41%
Require or encourage early exchange of reports of prospective trial experts	9%	69%	9%
Require or encourage experts who will testify to specify to their areas of agreement and disagreement	5%	24%	65%
Hold a pre-trial hearing on the admissibility of expert testimony	12%	55%	24%
Ask the parties to provide special education or instruction to the court	11%	19%	65%
Engage in your own research in specific area of expert testimony	9%	31%	52%
Ask clarifying questions from the bench	6%	64%	19%
Designate critical scientific or technical issues for a separate trial	1%	5%	88%

Allow jurors to question experts directly or through the court	1%	9%	83%
Have experts for both sides testify sequentially on an issue before moving on to direct testimony on the next issue	2%	7%	86%
Limit the number of experts who testify on a particular issue	3%	39%	50%
Limit the amount of time devoted to expert testimony on a particular issue	3%	19%	72%
Permit expert testimony on videotape	3%	61%	25%
Appoint an independent expert of the court	5%	31%	58%

*Percentages do not necessarily add up to 100% because judges could endorse multiple strategies; "don't know" are not included in Table

Table 2: Problems with Expert Testimony

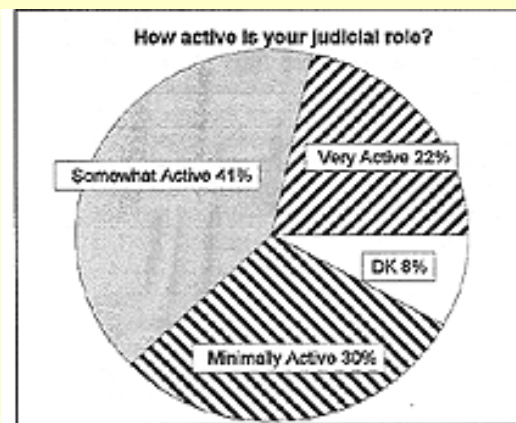
	1	2	3	4	5	Mean
	"very infreq"			"very freq"		
Delay in trial schedule caused by unavailability of experts	16%	27%	27%	20%	6%	2.96
Indigent party unable to retain expert to testify	29%	22%	18%	19%	8%	2.8
Failure of party(ies) to provide discoverable information concerning retained experts	25%	29%	26%	13%	3%	2.69
Excessive expense of party-hired experts	18%	15%	26%	26%	11%	3.2
Attorney(s) unable to adequately cross-examine expert(s)	21%	23%	33%	14%	6%	2.83
Experts abandon objectivity and become advocates for side that hired them	11%	15%	25%	30%	15%	3.48
Extensive disagreement among parties	6%	14%	32%	29%	15%	3.58
Expert testimony appears to be of questionable validity	20%	29%	34%	9%	5%	2.76
Expert testimony not comprehensible to trier of fact	26%	27%	28%	12%	5%	2.64
Expert poorly prepared to testify	32%	41%	18%	4%	2%	2.27
Disparity in level of competence of opposing experts	18%	33%	33%	13%	1%	2.65



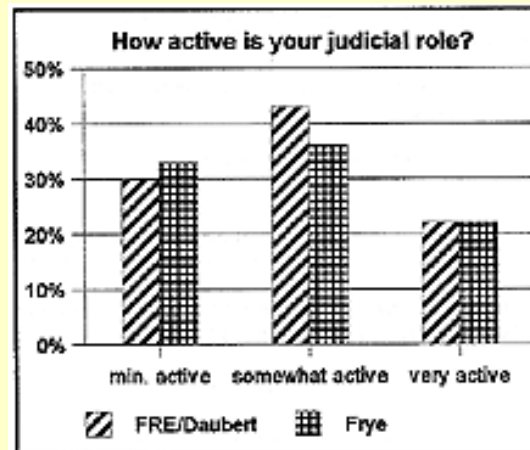
How do judges around the country view their decision-making role with respect to scientific evidence?

All of the judges interviewed (N=400) were asked how active a role they tend to take in determining the scientific merits of the proffered evidence. An overwhelming 91% of the judges believed that the role of gatekeeper was an appropriate one for a judge, irrespective of the admissibility standard followed in their state.

However, only 22% of all the judges described themselves as taking a "very active" role when deciding the admissibility of proffered scientific evidence. Most of the judges described themselves as taking either a "somewhat active" role (41%) or a "minimally active role" (30%). A few judges (8%) were unsure how to characterize their role in this way.



When comparing the responses of judges in states that follow the FRE or Daubert with those of judges in states that follow the Frye standard, there appears to be little difference in how they describe their role (68 judges were from states following some hybrid of Frye or some other standard).

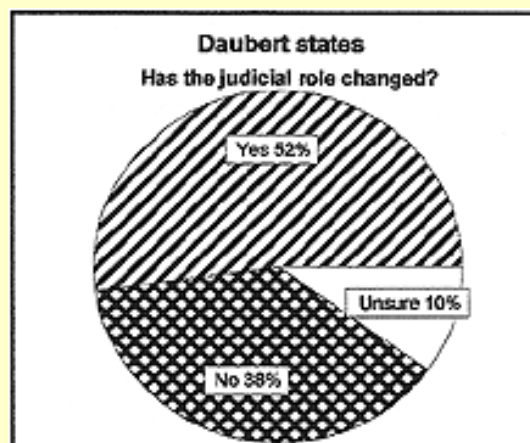


Looking just at the responses of judges from FRE/Daubert states (n=205). Those judges who indicated that their state followed the FRE or Daubert guidelines were asked if they thought the role of the judge with respect to admissibility decision-making had changed as a consequence of Daubert: 52% believed their role had changed; 38% believed their role had not changed; and 10% were unsure.

- 44% of the 54 judges who described themselves as taking a "minimally active role, believed that passage of the Daubert decision had changed their role, making it a more active one;
- 61% of the 78 judges who described themselves as taking a "somewhat active role in admissibility decision-making, believed that the passage of Daubert had increased their involvement in the admissibility decision-making process; and
- 59% of the 39 judges who described themselves as "very active," also believed that their role had become more active since the passage of Daubert.

Thus, for the majority of judges in FRE/Daubert states, the judicial role was seen to have become more active as a consequence of the passage of Daubert. Reasons given by the judges in states adopting Daubert for why they believed their judicial role had changed included:

"Daubert made it clear that the judge was to be more



of an active gatekeeper than before "

"[Under Daubert] the judge becomes more of a determiner of the reliability of the evidence."

"The role of the judge is greater because of the added procedures that may occur under Daubert, such as pre-trial admissibility hearings."

Judges who reported that Daubert had not changed the role of the judge, believed that under Daubert judges were still doing what they had always been doing. As one judge commented:

"Daubert put the judge back in the courtroom. Thing is, I've been there all along."

Endnotes:

1. *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923).
2. *Daubert v. Merrell Dow Pharmaceuticals, Inc.* 113 S. Ct. 2786 (1993).
3. *U.S. v. Downing* 753 F. 2d 1224 (3rd. Cir. 1985).
4. *Kumho Tire Co. v. Carmichael* (97-1709) 131 F. 3d, 1433. <http://supct.law.cornell.edu/supct/html/97-1709.ZS.html>
5. Data presented in these tables were collected in a replication and extension of a study conducted by the Federal Judicial Center with respect to the management of scientific evidence among federal judges.

SUGGESTED READINGS

Black, B., Ayala, F.J., and Saffran-Brinks, C. (1994). "Science and Law in the Wake of *Daubert*: A New Search for Scientific Knowledge." *Texas Law Review*, Vol. 72, pgs. 715-803.

Daubert v. Merrell Dow Pharmaceuticals 509 U.S. 579 (U.S. Supreme Court, June 29, 1993).
<http://supct.law.cornell.edu/supct/html/92-102.cpanel.html>

Faigman, D.L. (1995). "The Evidentiary Status of Social Science under *Daubert*: Is it 'Scientific,' 'Technical,' or 'Other' Knowledge." *Psychology, Public Policy, and Law*, Vol. 1(4), pgs. 960-979.

Faigman, D.L., Kaye, D.H., Saks, M.J., and Sanders, J. (1997). *Modern Scientific Evidence: The Law and Science of Expert Testimony, Vols. I and II*. Minnesota: West.

Faigman, D.L., Porter, E., and Saks, M. (1994). "Check Your Crystal Ball at the Courthouse Door Please: Exploring the Past, Understanding the Present, and Worrying About the Future of Scientific Evidence." *Cardozo Law Review*, Vol. 15, pgs. 1799-1835.

Federal Judicial Center. (1994). *Reference Manual on Scientific Evidence*. Washington: Government Printing Office.
http://www.fjc.gov/EVIDENCE/science/sc_ev_sec.html

Fenner, G.M. (1996). "The *Daubert* Handbook: The Case, Its Essential Dilemma, and Its Progeny." *Creighton Law Review*, Vol. 29, pg. 939.

General Electric Co. et al. v. Joiner 78 F. 3d 524 (11th Cir. 1996), 118 S. Ct. 512 (1997).
<http://supct.law.cornell.edu/supct/html/96-188.zx.html>

Gianelli, P.C., and Imwinkelried, E.J. (1993). *Scientific Evidence, Vols. I and II*. Michie.

Jasanoff, S. (1995). *Science at the Bar: Law, Science and Technology in America*. A Twentieth Century Fund Book.

Mealey's *Daubert* Report: Reviewing the Law and Admissibility of Expert Witness Testimony.

<http://www.mealeys.com/daubert.html>

Walker, L. and Monahan, J. (1996). "*Daubert* and the Reference Manual: An Essay on the Future of Science in Law." *Virginia Law Review*, Vol. 82, pg. 837.

[CHAPTER 1](#) --- [FRONT PAGE](#) --- [CHAPTER 3](#)

CHAPTER 3

An Introduction to the Philosophy of Science

Learning Objectives for Chapter 3

Upon completion of this chapter, the reader should be able to:

- ◆ Understand science as a particular way of thinking;
- ◆ Describe some of the basic differences between legal and scientific disciplines;
- ◆ Discuss the general tenets of fundamental perspectives in the philosophy and sociology of science;
- ◆ Understand why it is important that judges are informed about different philosophies of science; and
- ◆ Articulate a personal philosophy of science and how this personal philosophy might influence decision-making.

The goal of this chapter is not to provide an in-depth examination of the various philosophies of science, nor is it to create philosophers of science out of judges. Rather, the goal of this chapter is to present an overview of the general tenets of the philosophy and sociology of science. Moreover, judges are asked to reflect on how different philosophies of science, including their own, influence the presentation, examination, and admission of scientific evidence in court.

Daubert has been widely interpreted as an injunction that judges should “think like scientists” when assessing the validity of scientific evidence. On its face, this requirement seems both sensible and unproblematic. What could be more reasonable than to require that scientific evidence should conform to scientists’ own tests of validity? In practice, however, Daubert has given rise to three sets of difficulties that make the decision far from simple to implement: first, there is no single model of “good science” that can be mechanically applied to all offers of scientific evidence; second, the standards used by scientists may be impractical or impossible to achieve in legal proceedings; third, the law’s commitment to doing justice may conflict with full acceptance of scientists’ standards.¹

Before going any further, stop and reflect ...

- ◆ How do you define science?
- ◆ Do you think scientific knowledge can be distinguished from other forms of knowledge?
- ◆ If so, how do you distinguish science from other forms of knowledge?
- ◆ If not, why do you think such a distinction cannot be made?
- ◆ Do you think that the standards used by scientists can be translated appropriately into judicial reviews of scientific evidence?
- ◆ Do you think the law's commitment to justice conflicts with full acceptance of scientists' standards and practices?

Is it scientific or specialized knowledge and to what extent, if any, does that influence admissibility decision-making?

The *Daubert* decision did not set clear guidelines as to whether the four criteria articulated (falsifiability, error rate, peer review and publication, and general acceptance) should apply only to "scientific" knowledge or to "scientific, specialized, or technical knowledge." Nor did *Daubert* provide any guidance in how distinctions among different types of knowledge should be made.

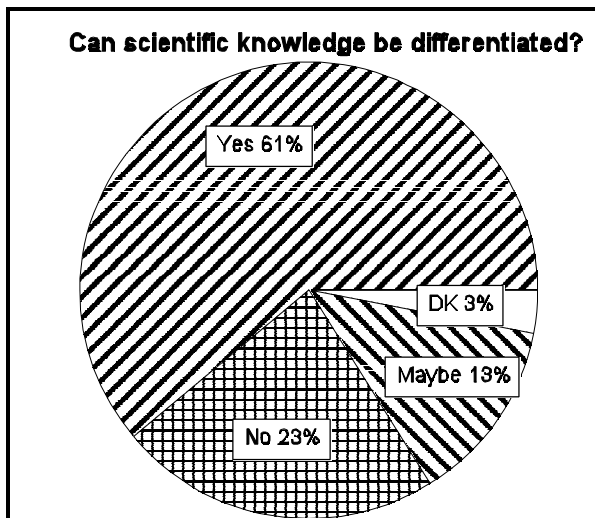
Whether or not the *Daubert* guidelines apply to all forms of specialized knowledge, or just "scientific" knowledge, was recently answered by the United States Supreme Court in *Kumho Tire Co. v. Carmichael* (1999). The United States Supreme Court, in reversing the Eleventh Circuit, held that the factors for a court to use in determining the reliability of a scientific theory or technique as set out in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, (1993), may apply to testimony of engineers and other experts who are not scientists. The Court noted that *Daubert* set forth a trial judge's "gatekeeping" obligation under FRE 702 to ensure that expert testimony is relevant and based on reliable scientific theories and that Rule 702 applies to all expert testimony because the language of 702 does not distinguish between 'scientific,' 'technical,' or 'other specialized' knowledge.



How do judges around the country distinguish scientific knowledge from technical or otherwise specialized knowledge?

More than half of the judges surveyed believed that “scientific knowledge” can be differentiated from “technical or otherwise specialized knowledge” (244 of 400 judges).

Those judges responding that “scientific” knowledge could be differentiated from other forms of knowledge were asked to discuss how they would make that distinction. The following general categories of responses were given as grounds for differentiating between “science” and other forms of knowledge (note that judges could give more than one response). It is interesting, that while more than half of the judges believed that “science” could be differentiated from other forms of knowledge, many of them had difficulty articulating how that distinction could or should be made.



- 59% the distinction would be made on a case-by-case basis, depending on the nature of the evidence, the purpose for which the evidence is being proffered, existing precedent, and so forth
- 21% technical knowledge is the application of facts or knowledge, while science is the generation of new knowledge
- 18% there is a distinction; but unsure on what grounds the distinction can or should be made
- 10% science involves the scientific method; experimentation
- 10% science is objective and less open to interpretation
- 7% science is theory-driven
- 5% technical knowledge refers to machines, software, equipment

One-third of all of the judges said that “scientific” knowledge could *not* be distinguished from other forms of knowledge for the purposes of deciding admissibility (132 of 400).



How do judges around the country distinguish scientific knowledge from technical or otherwise specialized knowledge? (Continued ...)

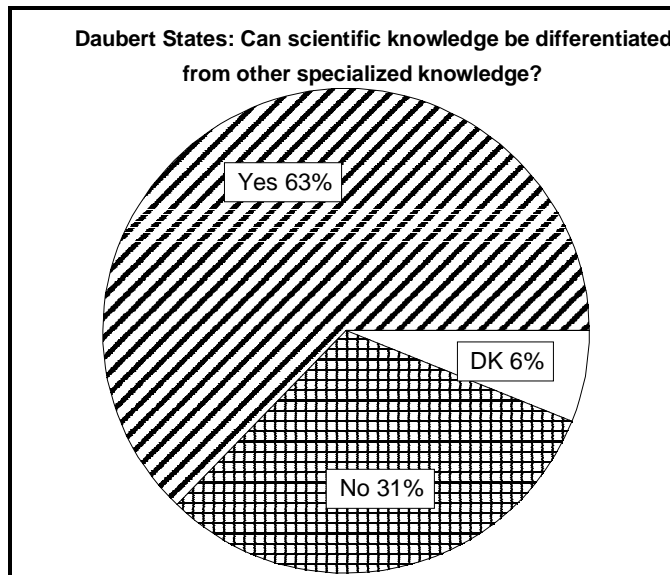
Those judges who believed that the distinction between “scientific knowledge” and “technical or otherwise specialized knowledge” *could not* be made, gave the following reasons:

- 50% there is no substantive difference between the two forms of knowledge
- 27% admissibility guidelines can be applied to both forms of knowledge, therefore the distinction is irrelevant for the purposes of deciding admissibility
- 17% deciding whether the evidence is “scientific” is an issue for the jury, not the judge
- 16% each form of knowledge has elements of the other: technology is applied science, science is inclusive of technology

Looking just at the responses of judges in *Daubert* states, almost 3/4 of the judges believed that “scientific knowledge” could be or might be differentiated from “technical or otherwise specialized knowledge” (63% or 130 of 205 judges).

Of those 130 judges, 31% said they would make the distinction on a case-by-case basis. 26% of those judges defined science as the generation of new knowledge and technical or specialized knowledge as the application of scientific knowledge.

31% of the judges in *Daubert* states believed that “science” could *not* be distinguished from other forms of knowledge (64 of 205 judges). 39% of these 64 judges indicated that the issue of deciding whether or not particular evidence was scientific was irrelevant, stating either that the *Daubert* guidelines are equally applicable to both scientific knowledge and technical or specialized knowledge or that the focus should be on the jury’s level of understanding and not on the classification of the evidence. 30% of judges indicated that there is no substantive difference between scientific knowledge and technical or specialized knowledge.



➡ *The Interface Between Science and Law*

Both science and law are embedded in a common cultural context and both share norms of rationality and causality. Yet, when one discipline (e.g., the law) seeks to understand another (e.g., science) a dilemma may result because each approaches uncertainty in a different way. When asked "How do you know which decision is the right one?," each relies on different methods, each uses a different set of concepts and theories to describe the same phenomenon, and each reflects different underlying assumptions and values. That is, each discipline incorporates and reflects different ways of meaning and different ways of knowing -- each discipline views the event through a different "lens." Thus, the perspective from which uncertainty is approached and the perspective from which questions are asked, whether legal or scientific, influences both the answer to the question and the means by which the answer is discovered.

- Science is creative and forward-focused. Law looks to the past for justifications and methods (i.e., *stare decisis*).
- Science relies on the scientific method to find "truth." The scientific method stresses objectivity and tries to eliminate bias and error. Law relies on the adversarial process to find "truth." Bias in the selection and presentation of evidence is assumed and accepted as part of the adversarial process.
- Scientists view their research as a body of working assumptions, of contingent and sometimes competing claims. Even when core insights are validated over time, the details of these hypotheses are subject to revision and refinement as a result of open criticism within the scientific communities. The scientific process is ongoing and scientists can suspend judgment until a later point in time. In the legal system, however, all of the players are forced to make decisions at a particular moment in time; the decision cannot be suspended to some time in an unspecified future.

- Scientific knowledge is based on empirical findings and is probabilistic. Law tries to give the appearance of certainty. During a trial, “burden of proof” and “reasonable doubt” are argued. But, once a decision is made, it is assumed to be the “truth” and, for the most part, irrevocable.

Many [science and technology] issues are legally relevant because the applicable substantive law raises the question of how to proceed in a world of imperfect knowledge. Lawyers and scientists approach uncertainty in ways that are characteristic of the goals their respective disciplines are seeking to achieve. In the courts, scientific knowledge must inform the choice, but abdication to the scientist is incompatible with the judge's responsibility to decide the law.²

It is important that judges have a basic understanding of the philosophies of science and an understanding of how differences in philosophy within scientific disciplines influence areas selected for study, the process of discovery, the nature of the questions asked and answers found, and the manner in which interpretations are drawn. It is also important to realize that differences in philosophy of science, when placed within the legal context, influence what evidence and which expert witnesses are selected and presented, and what details are attended to in direct and cross examination.

It is also important to recognize that everyone, including judges, have a personal philosophy of science and that this personal philosophy, whether we are conscious of it or not, influences how we judge, critique, and endorse various scientific claims. Because of the gatekeeping role judges play in admissibility decision-making, it is important that judges recognize that their personal philosophy of science might influence decision-making about what constitutes legitimate and admissible scientific evidence. The first steps to reducing the influence of personal philosophies on legal decision-making are to recognize, articulate, and question the philosophy that one holds.

THE PHILOSOPHY OF SCIENCE

The philosophy of science encompasses questions about science in general (e.g., Are some theoretical entities real?), about particular groups of sciences (e.g., Can social objects be studied in the same way as natural ones?), and about individual sciences (e.g., What are the implications of relativity theory for our concepts of space and time?). The philosophy of science arose as a discipline separate from the more general theory of knowledge in the mid-nineteenth century -- at about the time that distinct sciences bearing names such as 'physics,' 'chemistry,' and 'biology' were becoming professionalized.

Why is it important for judges to understand the various perspectives of the philosophy of science?

An understanding of the philosophy of science:

- challenges the notion that objective scientific 'truth' is out there to be discovered;
- provides a more complete picture of the processes of science and scientific discovery;
- provides a more principled method for evaluating the validity, reliability, and applicability of scientific evidence;
- provides context for the analysis of the judicial role in admissibility decision-making; and
- makes explicit the implicit assumptions underlying various scientific methods and theories.

Before going any further, stop and reflect ...

- ◆ What are the distinctive characteristics of the quest for knowledge that we call science?
- ◆ What distinguishes sciences like astrophysics and geology from pseudo-sciences like astrology and phrenology, and from non-sciences like literary criticism and religion?
- ◆ How should we classify fields like sociology, economics, engineering, or psychology?

Questions about the nature of science are not merely academic -- where the line is drawn between science and non-science can determine what subjects are taught in schools, what forms of treatment are covered by medical insurance, and what evidence is admissible in court.

➡ *Science as a Way of Thinking*

Scientists seek answers to their own questions. Their work is built on highly refined skills in asking and answering questions -- knowing how to ask questions is as important as knowing how to go about answering them. The essence of science is the process of carefully composing questions and then systematically seeking their answers to gain a better understanding of a problem or puzzle. Science involves a process of inquiry, a particular way of thinking.

The essence of modern science is the way of thinking, the discipline in asking and answering questions. It is intellectual and logical thought and the demands for confirming evidence that characterize science, *not* the technologies.

The essence of science is its systematic, disciplined way of thinking aimed at gaining knowledge about a problem or puzzle. Science places heavy demands on the adequacy of its information and on the processes applied to that information.

Three traits generally distinguish scientific knowledge from other forms of knowledge:

- science seeks the *systematic organization of information* about the world and, in so doing, discovers new relationships among natural phenomena;
- science endeavors to explain *why* phenomena occur and *how* they are related; and
- scientific explanations must be formulated in a way that makes them subject to *empirical testing*.

Science does not advance simply by accumulating observations and experimental results and then extracting a theory or hypothesis from them. Rather, science progresses through an ongoing and cyclical process of discovery.

Scientific progress begins with the invention of a possible world, or a fragment thereof, which [is] ... compared by experimentation with the real world. And it [is] this constant dialogue between imagination and experimentation that allow[s] one to form an increasingly fine-grained conception of what is called reality.³

➡ ***The Problem of Induction***

Induction is the process whereby scientists infer, on the basis of a number of observations or experiments, that some theory is true. Thus, inductively derived theory begins with a solid base of empirical observations and gradually builds up to more abstract levels of theoretical explanation. What all inductive inferences have in common, is that they start with particular premises about a *finite* number of past observations, and then, based upon these past observations, infer general conclusions about the future.

<p>Induction: reasoning from the particular to the general; an inductive theory begins with specific observations and infers general conclusions</p>

Critics of induction argue that inferences about the future, based upon observations of the past, do not guarantee that the inference will be supported. That is, it is unclear how a finite amount of information about what has happened in the past can guarantee that a natural pattern will occur, or continue to occur, in the future. What rules out the possibility that the course of nature may change and that the patterns observed so far turn out to be a poor guide to the future? If the only evidence available is simply that some particular observation has occurred frequently in the past, or that some particular natural law has worked so far, then how can one be sure that past events, or particular laws, will not be disproved by future occurrences? The problem of induction calls the authority of laws such as these into question. The views of Sir Karl Popper provide one influential response to the problem of induction.

➡ ***Popper, Deduction, and Falsification***

Sir Karl Raimund Popper (1902-1994) was one of the most well known philosophers of science. Educated in Vienna in mathematics, physics, and philosophy, Popper's ideas have

relevance and impact in science, business, politics, art, music, and the law. Indeed, Popper's philosophy of science has become increasingly influential within the law with the passage of *Daubert* and its reliance on falsification.

In Popper's view, science is *not* characterized by induction. Popper denied that scientists start with observations and then infer a general theory. Rather, Popper argued that scientists first put forward a general theory and then make specific predictions based upon that theory. These predictions are then compared with observations to see whether the theory is supported. If the theory is not supported, then the theory is experimentally falsified, and a new, alternative theory must be found. If, on the other hand, the tests support the theory, then scientists will continue to uphold it -- not as proven truth but as theory that has the potential to be falsified; a theory that has "not yet been disconfirmed."

Thus, Popper was arguing that science rests on **deduction** rather than induction. That is, the general theory guides the researcher in making and testing specific predictions about future events. These deductions (i.e., predictions) are then empirically tested through experimental research and support or lack of support for the theory is obtained. If we look at science in this way, argued Popper, then we see that it does not need induction. According to Popper, the inferences which matter to science are *refutations*, which take some failed prediction as the premise and conclude that the theory behind the prediction is false. In his first book, *The Logic of Scientific Discovery* (1934), Popper proposed the '**falsification**' principle. Popper proposed that a theory is a scientific theory if and only if there are possible observations that would refute it. Thus, a

Deduction: reasoning from the general to the particular; a deductive theory begins with a construct or theory, makes specific predictions about the constructs, and then empirically tests the predictions

Falsification: the refutation, or potential refutation, of a theory; a theory or technical procedure is falsifiable if it can be subjected to testing which could prove it to be incorrect or false

It is important to distinguish the concept of falsifiability from similar terms with very different meanings, such as:

"*falsify*" (to alter something with the intent to defraud)

"*falsity*" (the quality of being false; making a false assertion)

theory is only a scientific theory if there is the potential for falsification. In order for a theory to have the potential for falsification, the theory itself must be empirically testable.

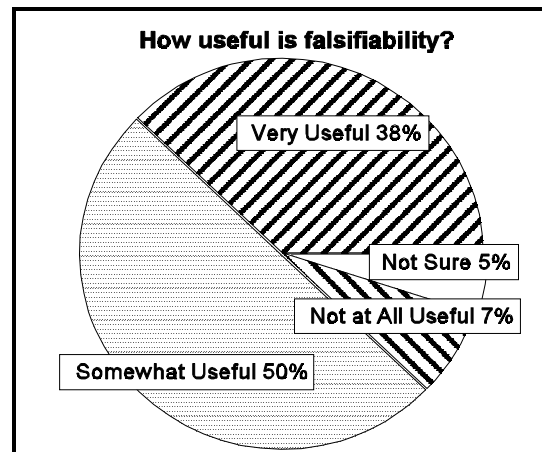
Popper used the principle of falsifiability to distinguish genuine science from traditional belief systems like astrology, as well as from Marxism, psychoanalysis, and various other modern disciplines he called 'pseudo-sciences.' According to Popper, the central claims of these theories are unfalsifiable; that is, the central claims have not been, nor can they be, empirically tested. Psychoanalytic theories, for example, predict that adult neuroses are due to childhood traumas. Psychoanalysts argue that the observation of a neurotic adult with clearly identified childhood traumas supports psychoanalytic theory. However, when a neurotic adult who has no identifiable childhood traumas is observed, psychoanalysts argue that his trauma is actually unconscious, private, and not available to the conscious mind (therefore the person is unaware of the trauma having occurred). For Popper, this is the antithesis of scientific reasoning. When both the presence and absence of an event or observation (in this case childhood trauma) are said to support a prediction, there is no way to empirically test the underlying claim or inference – the claim is unfalsifiable.



To what extent do judges around the country find the concept of falsifiability a useful criterion for critically evaluating scientific evidence?

All judges in the survey sample (N=400), even those not in FRE/*Daubert* states, were asked how useful they thought the concept of falsifiability is for admissibility decision-making.

The vast majority of judges, regardless of the admissibility standard followed in their state, believed falsifiability to be a useful guideline for determining the admissibility of scientific evidence (38% indicating that it is “very useful” and 50% indicating that it is “somewhat useful”).

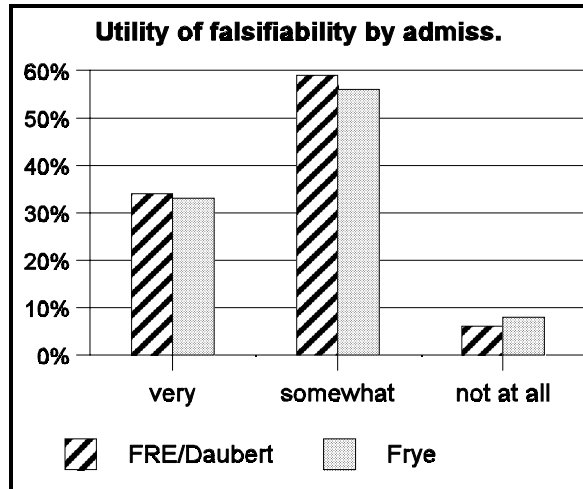




To what extent do judges around the country find the concept of falsifiability a useful criterion for critically evaluating scientific evidence? (Continued ...)

When responses are separated into those provided by judges in *FRE/Daubert* states and those in *Frye* states, very similar responses are found with respect to the utility of the guideline. It is interesting to note the similarity in how judges following different admissibility standards rated the utility of falsifiability. As shown in the table, the ratings between the two groups are almost identical.

As a follow-up to the question asking about the utility of the guideline, judges were asked how they would use the guideline of “falsifiability” when scientific evidence is proffered in court. The question was designed in such a way as to allow the researchers to infer how well the judge actually understood the specific concept.



For coding purposes, responses were coded as “judge understands concept,” “judge’s understanding of concept questionable,” and “judge clearly does not understand concept.” Although a definition of the concept was not provided as an explicit component of the question, if the judge asked for a definition the interviewer provided a standardized, pre-tested definition. The interviewer also noted that the definition had been asked for and given. If a definition of the concept was not asked for by the respondent, it was not given.

With respect to falsifiability, in order for a response to be coded as “judge understands concept,” the judge had to explicitly include reference to testability, test and disproof, or prove wrong a theory or hypothesis, proof/disproof, or validity. If the judge did not explicitly refer to any of these concepts, but appeared to ‘talk around the issue,’ or just refer to a test of a theory or hypothesis, then the judge’s understanding was coded as questionable. That is, the researchers were unable to confidently infer that the judge truly understood the scientific meaning of the concept. The researchers realize that this criteria sets a high threshold of understanding and that some of the judges whose responses were coded as questionable may, in actuality, understand the concept. However, the threshold was set deliberately high so that researchers could be sure with a high level of confidence that the judge truly understood the meaning of the concept.



To what extent do judges around the country understand the scientific meaning of falsifiability?

Sample comment coded as “judge understands concept”

“I would want to know to what extent the theory has been properly and sufficiently tested and whether or not there has been research that has attempted to prove the theory to be wrong”

Sample comment coded as “judge’s understanding of concept questionable”

“I would want to know if the theory has been tested”

Sample comment coded as “judge clearly does not understand concept”

“If there is white-out on the page then the document has been falsified”

As discussed, the vast majority of judges surveyed in both FRE/*Daubert* and *Frye* states, believe falsifiability to be a useful guideline for determining the admissibility of proffered scientific evidence. However, the results of the survey indicate that judges do not fully understand the scientific meaning of falsifiability and that, as a result, they are unsure how to properly utilize the concept as an admissibility guideline. Indeed, when asked a question about how they would apply the guideline of falsifiability, the majority of judges expressed some hesitancy or uncertainty. From the answers that were provided, the researchers could only infer a true understanding of the concept in 5% of the responses. Moreover, for judges in *Daubert* states, a true understanding of the concept could only be inferred in 3% of the responses.

Before going any further, stop and reflect ...

- ◆ How is a ‘Popperian’ view of science reflected in current evidentiary law?
- ◆ Does strict adherence to a ‘Popperian’ view of science pose particular challenges for certain kinds of scientific evidence (e.g., psychological evidence)?
- ◆ Many modern philosophers of science argue that a reliance on the falsification principle does not accurately reflect a modern view of science. Do you think a ‘Popperian’ definition of science is a modern one? Why or why not?
- ◆ What form might potential challenges to Popper and the concept of falsification take?

➡ ***Kuhn, Paradigms, and Revolutions***

Thomas Samuel Kuhn (1922-1996) was an American philosopher and historian of science. Educated at Harvard as a physicist, Kuhn turned to the history of science to write *The Structure of Scientific Revolutions* in 1962. In this work he contrasted the historical development of science with the traditional view of science as a purely rational, objective, enterprise.

Scientific practice, Kuhn argued, is governed by a **paradigm**, a world-view sanctioned by the scientific community. A paradigm refers to a collection of procedures or ideas that instruct scientists, implicitly as well as explicitly, what to believe and how to work. Most scientists, Kuhn argued, never question the paradigm. They solve puzzles and problems whose solutions reinforce and extend the scope of the paradigm rather than challenge it. This is what Kuhn called normal science. However, some puzzles may prove so difficult or intractable that they become anomalies. If anomalies build up, argued Kuhn, a science may enter a crisis state, where the credibility of the existing paradigm becomes increasingly poor. A candidate for a new paradigm may emerge to challenge the existing one, often created and supported by younger scientists who have much to gain and little to lose in terms of their reputations.

Paradigm: a world-view that reflects assumptions about the social world and how science should be conducted, as well as what constitutes legitimate problems, methods, solutions, and standards of 'proof'

The most controversial aspect of Kuhn's theory is his claim that a revolution takes place when paradigms are switched and that science is a progression of revolutions. Because a paradigm is a world-view, it is difficult for adherents to step outside of the paradigm and evaluate the pros and cons of an alternative paradigm; scientists tend to be closely wedded to a particular scientific world-view and to the theoretical and methodological assumptions that underlie it. However, according to Kuhn, once a scientist adopts a new paradigm, she will not be able to revert back to her original paradigm -- the original paradigm is forever changed. Thus,

according to Kuhn's view, science progresses through "paradigm shifts." It is important to recognize, however, that there is no guarantee that science is actually progressing toward anything -- least of all toward "the truth."⁴

Kuhn's view diverges significantly from Popper's. Kuhn argued that falsification is not possible because it implies the existence of absolute standards of evidence, which transcend any individual paradigm. A new paradigm may solve puzzles better than the old one did, and it may yield more practical applications, but, as stated by Kuhn, "you cannot simply describe the other science as false."⁵ According to Kuhn, it is the incompleteness and imperfection of an existing data-theory fit that defines the puzzles that characterize normal science. If, as Popper suggested, failure to fit were grounds for theory rejection, then, according to Kuhn, all theories would be rejected at all times.

Kuhn's theory remains controversial, but it is important not only because it was the first significant attempt to reach a descriptive rather than a normative or rule-governed account of science, but also because of the interest it has gathered beyond philosophers of science.

Central Tenets of Kuhn's Philosophy of Science

- Scientific practice is governed by a "paradigm," a world-view that is sanctioned by the scientific community.
- Science progresses through a series of "paradigm shifts."
- Falsification is not possible because it implies the existence of absolute standards of evidence, which transcend any individual paradigm.

It is important to note that falsification can refer to both the falsification of a paradigm or to the falsification of a particular theory. For the practical purposes of judicial decision-making however, the guideline of falsifiability refers to particular theories. The question becomes one of "Is the theory testable?" Thus, for the practical purposes of evaluating scientific theories, the concept of falsifiability is a useful tool, although not a definitive criterion.

Before going any further, stop and reflect ...

- ◆ Does Kuhn, in your opinion, pose a significant challenge to Popper?
- ◆ To what extent, if any, does a 'Kuhnian' philosophy of science challenge current evidentiary law?

➔ ***Paradigms, Negotiations, and Boundary-Drawing***

The sociology of science explores the social character of science, with special reference to the social production of scientific knowledge. Sociologists of science are interested in the social context within which science is practiced. Questions of interest for sociologists of science might be, for example:

- ◆ Does science progress in a rational way, or is it subject to historical and cultural factors, like other systems of belief?
- ◆ Is the only difference between science and other disciplines the authority our society invests in science?
- ◆ To what extent do the socially instilled biases of scientists influence the science they produce?
- ◆ What role does politics play in science?

Critics of the judicial system's handling of scientific claims often have an idealized view of science that many scientists themselves reject. Although these critics accept the indeterminacy of legal concepts, they speak of scientific "facts" as though they were objectively true, and they berate the legal system for failing to adduce conclusive evidence, even though scientists themselves concede that scientific hypotheses remain open to challenge until the incentives for attacking them disappear.⁶

In present day society, the term 'science' has great potency. Not only is 'science' more or less equivalent to 'valid knowledge,' but it also merges with 'technology' -- the useful application of knowledge. Consequently, those people known as 'scientists' are widely regarded as the purveyors of a superior kind of knowledge which represents the real world with a degree of

precision and reliability that makes possible extensive control over its natural processes. Sociologists of science view scientific knowledge as deriving, not from the impartial application of clear technical criteria, but from socially contingent formulations that have been deemed adequate by specific groups in particular cultural and social situations.

A major theory within the sociology of science is the theory of **social constructivism**. According to the social constructivist perspective, the ‘facts’ that scientists present to the world are not merely raw observations. Rather, scientific ‘facts’ are socially constructed through the institutions and social processes of science. Observations achieve the status of ‘facts’ only if they are produced in accordance with prior agreements about the

Social Constructivism: the philosophical position that truth is contingent and conditional and there are multiple perspectives and multiple realities

Researchers who hold a constructivist perspective argue that people in different geographic, cultural, or social locations may construe knowledge, truth, and relevance in different ways, and that each of these different ways of knowing may be legitimate and worthy.

correctness of theories, experimental methods, instrumentation, validation procedures, and review processes. These agreements, in turn, are socially derived through continual negotiation and renegotiation among relevant bodies of scientists.

Before going any further, stop and reflect ...

- ◆ To what extent do you agree with the claim that the process of science is shaped by political and social forces? Can you think of some specific examples?
- ◆ To what extent, if any, do theories derived from the social constructivist perspective challenge more traditional philosophies of science, such as the philosophy of Karl Popper?
- ◆ To what extent, if any, does the social constructivist perspective challenge current evidentiary law?
- ◆ Is there a place for this perspective in evidentiary law?

This brief overview of three philosophical approaches to science in no way represents a comprehensive or theoretically sophisticated discussion of the various philosophies of science and their points of convergence and debate. It is not the intent of this overview to do so. Rather, the goal is to challenge the notion that “science” is a given and that it exists independent of personal, social, and political influences. It is important to recognize that the scientific process itself is multi-leveled and multi-faceted.

➡ *A Brief Look at Other Influential Philosophers of Science*

◆ **Paul Feyerabend**

In his most influential book, *Against Method* (1975), Feyerabend argued that philosophy cannot provide a methodology or rationale for science, since there is no rationale to explain. Feyerabend sought to show that there is no logic to science; scientists create and adhere to scientific theories for what are ultimately subjective and even irrational reasons. According to Feyerabend, scientists can and must do whatever is necessary to advance their view.

◆ **Larry Laudan**

Laudan sees science operating within a conceptual framework that he calls a "research tradition." The research tradition consists of a number of specific theories, along with a set of metaphysical and conceptual assumptions that are shared by those scientists who adhere to the tradition. A major function of the research tradition is to provide a set of methodological and philosophical guidelines for the further development of the tradition. Following both Kuhn and Popper, Laudan argues that the objective of science is to solve problems -- that is, to provide "acceptable answers to interesting questions."⁷ However, for Laudan, the "truth" of a theory is irrelevant as an appraisal criterion. Rather, the key question is whether the theory offers an explanation for problems that arise when we encounter something in the natural or social environment which clashes with our preconceived notions or is otherwise in need of explanation.

Before going any further, stop and reflect ...

In her paper “Judging Science: Issues, Assumptions, Models (1997),” Professor Sheila Jasanoff presents five possible models judges may follow in coming to terms with the separate values and goals of science and litigation:

The Inquisitor

- experts are appointed by and are answerable to the judge, who also questions witnesses and conducts formal fact-finding; rests upon the assumption that neutral or unbiased experts exist and that they can be identified by impartial judges

The Gatekeeper

- role envisioned by *Daubert*; rests upon the assumption that science operates according to objective standards that can be clearly understood and applied by judges; standards may vary from case to case, but judges are seen as capable of identifying science that is so substandard as to merit exclusion

The Referee

- judges are likely to view the parties’ scientific claims as driven by interests and contaminated by bias; instead of screening the evidence according to some “objective” criteria of scientific validity, the refereeing judge may attempt to use perceived weaknesses in the parties’ scientific arguments to steer the litigants toward settlement

The Mediator

- the mediating judge may shape the discovery process and other pretrial proceedings so as to promote a sharpening of the scientific issues and, where possible, a negotiated resolution of significant scientific disputes

The Judge

- judge rejects the mythical notions of “pure science” and “junk science” and understands and recognizes how bias creeps into scientific inquiry and the differences between legitimately different viewpoints and truly marginal forms of inquiry; the judge holds the conviction that courts are not a forum for resolving scientific disputes definitively, but rather for doing justice on a case-by-case basis with the aid of all available scientific knowledge that meets threshold tests of relevance and reliability

-
- ◆ What do you think of the judicial roles articulated above?
 - ◆ In what way do you think a philosophy of science might influence decision-making about the admissibility of different kinds of scientific evidence? Consider how each of these conceptions of the judicial role rests on different assumptions or philosophies of science.
 - ◆ What might some of the advantages and disadvantages of each of these roles be?
 - ◆ Does one of these conceptions of the judicial role more accurately represent how you perceive your role with respect to scientific evidence? Which one and why?
 - ◆ Given this overview of some of the fundamental distinctions in philosophies of science, has your view of science and your conception of your judicial role changed in any way? How and why?

Endnotes:

1. Jasanoff, S. (1997). *Judging Science: Issues, Assumptions, Models*.
2. *Science and Technology in Judicial Decision Making: Creating Opportunities and Meeting Challenges*. Carnegie Commission on Science, Technology and Government (1993), pg. 24.
3. Molecular biologist and Nobel Laureate, Francois Jacob.
4. Kuhn, T.S. (1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press., pg. 170.
5. *Ibid.*
6. Jasanoff, S. (1993). "What Judges Should Know About the Sociology of Science." *Judicature*, Vol. 77(2), pgs. 77-82.
7. Laudan, L. (1977). *Progress and its Problems*. Berkeley, C.A.: University of California Press. pg. 13.

GLOSSARY

deduction	involves reasoning from the general to the particular; a deductive theory begins with a construct (theory), makes specific predictions about the construct, and then empirically tests the predictions
falsification	based upon the philosophy of Sir Karl Popper, falsification is said to be the cornerstone of science -- that is, a theory is only scientific to the extent that there is a potential for falsification; the goal of falsification is to refute (prove incorrect) a theory based upon observations gained through the scientific method; a theory or technique is falsifiable if it has been, or has the potential to be, tested
induction	induction involves reasoning from the particular to the general; an inductive theory begins with specific observations and infers general conclusions
paradigm	based upon the philosophy of Thomas Kuhn, a paradigm is a world-view that reflects certain assumptions, both implicit and explicit, about the social world, what constitutes a scientific problem or issue, what the appropriate methods of investigation are, and what constitutes appropriate solutions and standards of proof

social constructivism philosophical position that truth is contingent and conditional and that there are multiple perspectives and multiple realities; the belief that people in different geographic, cultural, and social locations construe knowledge, truth, and relevance in different ways, and that each of these ways of knowing is legitimate and worthy of consideration

SUGGESTED READINGS

Chalmers, A.F. (1982). *What is This Thing Called 'Science'? An Assessment of the Nature and Status of Science and its Methods*. University of Queensland Press.

Feyerabend, P. (1978). *Against Method*. London: Verso.

Feyerabend, P. (1981). *The Problems of Empiricism*. Cambridge: Cambridge University Press.

Jasanoff, S. (1993). "What Judges Should Know About the Sociology of Science." *Judicature*, Vol. 77(2), pgs. 77-82.

Jasanoff, S. (1995). *Science at the Bar: Law, Science, and Technology in America*. Cambridge: Harvard University Press.

Knorr-Cetina, K.D. (1981). *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. New York: Pergamon.

Kuhn, T.S. (1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lakatos, I. (1974). "Falsification and the Methodology of Scientific Research Programs." In *Criticism and the Growth of Knowledge*, Imre Lakatos and Alan Musgrave (Eds.), Cambridge: Cambridge University Press, pgs. 91-195.

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge: Harvard University Press.

Laudan, L. (1977). *Progress and its Problems*. Berkeley, C.A.: University of California Press.

Popper, K.S. (1959). *The Logic of Scientific Discovery*. New York: Basic Books.

Judge's Notes:

CHAPTER 4

Quantitative and Qualitative Research

Learning Objectives for Chapter 4

Upon completion of this chapter, the reader should be able to:

- ◆ Understand the differences between quantitative and qualitative research, including:
 - ◆ the differing assumptions underlying the two approaches;
 - ◆ the methods typical of each approach; and
- ◆ Understand and discuss how these two approaches to research differentially influence the scientific questions asked, the methodologies employed, and the conclusions drawn, and why this is important to consider.

It is important to recognize that systematic observation and testing can be accomplished using a wide variety of methods. Many people think of scientific inquiry strictly in terms of laboratory experimentation. However, it is neither possible nor desirable to study all phenomena of interest under controlled laboratory conditions.

The design of any study begins with the selection of a topic and a research methodology. These initial decisions reflect assumptions about the social world, how science should be conducted, and what constitutes legitimate problems, solutions, and criteria of "proof." Different approaches to research encompass both theory and method. Two general approaches are widely recognized: quantitative research and qualitative research.

Quantitative research is an inquiry into an identified problem, based on testing a theory, measured with numbers, and analyzed using statistical techniques. The goal of quantitative methods is to determine whether the predictive generalizations of a theory hold true.

By contrast, a study based upon a **qualitative** process of inquiry has the goal of understanding a social or human problem from multiple perspectives. Qualitative research is conducted in a natural setting and involves a process of building a complex and holistic picture of the phenomenon of interest.

The selection of which research approach is appropriate in a given study should be based upon the problem of interest, resources available, the skills and training of the researcher, and the audience for the research. Although some research may incorporate both quantitative and qualitative methodologies, in their 'pure' form there are significant differences in the assumptions underlying these approaches, as well as in the data collection and analysis procedures used.

Why is the distinction between quantitative and qualitative research important?

It is important to be able to identify and understand the research approach underlying any given study because the selection of a research approach influences the questions asked, the methods chosen, the statistical analyses used, the inferences made, and the ultimate goal of the research.

When critically reviewing scientific research, the questions asked, and the answers given, will differ depending upon whether the research is quantitative or qualitative.

 ***Quantitative Methods***

Assumptions Underlying Quantitative Methods

- reality is objective, "out there," and independent of the researcher -- therefore reality is something that can be studied objectively;
- the researcher should remain distant and independent of what is being researched;
- the values of the researcher do not interfere with, or become part of, the research -- research is value-free;
- research is based primarily on deductive forms of logic and theories and hypotheses are tested in a cause-effect order; and
- the goal is to develop generalizations that contribute to theory that enable the researcher to predict, explain, and understand some phenomenon.

Three general types of quantitative methods:

1. **Experiments** → True experiments are characterized by random assignment of subjects to experimental conditions and the use of experimental controls.
2. **Quasi-Experiments** → Quasi-experimental studies share almost all the features of experimental designs *except* that they involve non-randomized assignment of subjects to experimental conditions.
3. **Surveys** → Surveys include cross-sectional and longitudinal studies using questionnaires or interviews for data collection with the intent of estimating the characteristics of a large population of interest based on a smaller sample from that population.

➔ *Qualitative Methods*

Assumptions Underlying Qualitative Methods

- multiple realities exist in any given situation -- the researcher's, those of the individuals being investigated, and the reader or audience interpreting the results; these multiple perspectives, or voices, of informants (i.e., subjects) are included in the study;
- the researcher interacts with those he studies and actively works to minimize the distance between the researcher and those being researched;
- the researcher explicitly recognizes and acknowledges the value-laden nature of the research;
- research is context-bound;
- research is based on inductive forms of logic; categories of interest emerge from informants (subjects), rather than being identified *a priori* by the researcher;
- the goal is to uncover and discover patterns or theories that help explain a phenomenon of interest; and
- determinations of accuracy involve verifying the information with informants or "triangulating" among different sources of information (e.g., collecting information from different sources).

Three general types of qualitative methods:

1. **Case Studies** → In a case study the researcher explores a single entity or phenomenon ('the case') bounded by time and activity (e.g., a program, event, institution, or social group) and collects detailed information through a variety of data

collection procedures over a sustained period of time. The case study is a descriptive record of an individual's experiences and/or behaviors kept by an outside observer.

2. **Ethnographic Studies** → In ethnographic research the researcher studies an intact cultural group in a natural setting over a specific period of time. A cultural group can be any group of individuals who share a common social experience, location, or other social characteristic of interest -- this could range from an ethnographic study of rape victims in crisis shelters, to children in foster care, to a study of a cultural group in Africa.
3. **Phenomenological Studies** → In a phenomenological study, human experiences are examined through the detailed description of the people being studied -- the goal is to understand the 'lived experience' of the individuals being studied. This approach involves researching a small group of people intensively over a long period of time.

???

Questions to consider when evaluating scientific evidence ...

- ◆ Was a quantitative or qualitative research approach adopted?
- ◆ Was the research approach appropriate given the problem investigated and the goals of the research?
- ◆ Was the process of investigation consistent with the underlying assumptions of the research used?
- ◆ Were appropriate types of conclusions drawn given the research approach used?

Before going any further, stop and reflect ...

- ◆ What are some examples of quantitative research that might be proffered as evidence in court?
- ◆ What are some examples of qualitative research that might be proffered as evidence in court?
- ◆ To what extent, if any, would your critical review of these different types of evidence differ?

CRITICAL QUESTIONS REVIEWED

- ◆ Was the quantitative or qualitative approach adopted?
- ◆ Was the research approach selected appropriate given the problem investigated and the goals of the research?
- ◆ Was the process of investigation consistent with underlying assumptions of the research approach used?
- ◆ Were appropriate types of conclusions drawn given the research approached used?

GLOSSARY

case studies	the researcher explores a single entity or phenomenon ("the case") bounded by time and activity (e.g., a program, event, process, institution, or social group) and collects detailed information through a variety of data collection procedures over a sustained period of time
ethnographic studies	the researcher studies an intact cultural group in a natural setting over a specific period of time; a cultural group can be any group of individuals who share a common social experience, location, or other social characteristic of interest
experimental studies	characterized by random assignment of subjects to experimental conditions and the use of experimental controls
phenomenological studies	human experiences are examined through the detailed description of the people being studied -- the goal is to understand the "lived experience" of the individuals being studied; involves studying a small group of people intensively over a long period of time
qualitative research	a process of inquiry with the goal of understanding a social or human problem from multiple perspectives; conducted in a natural setting with a goal of building a complex and holistic picture of the phenomenon of interest
quantitative research	an inquiry into an identified problem, based on testing a theory composed of variables, measured with numbers, and analyzed using statistical techniques; the goal is to determine whether the predictive generalizations of a theory hold true
quasi-experimental studies	share almost all the features of experimental designs <i>except</i> that they involve non-randomized assignment of subjects to experimental conditions
random assignment	all subjects have an equal chance of being assigned to a given experimental condition; a procedure used to ensure that experimental conditions do not differ significantly from each other
survey	questionnaires or interviews for data collection with the intent of generalizing from a sample population to a larger population of interest

SUGGESTED READINGS

Creswell, J.W. (1994). *Research Design: Qualitative and Quantitative Approaches*. Sage Publications: Thousand Oaks, CA.

Mason, J. (1996). *Qualitative Research*. Thousand Oaks: Sage Publications.

Judge's Notes:

In designing and carrying out research, the researcher moves through all phases of the research process beginning with ideas that are then refined and developed into one or more specific questions. The researcher then designs the procedures to be used to answer the questions and proceeds with the observations. The design phase is crucial in scientific research. The researcher must carefully and systematically plan each step, asking and answering a variety of conceptual and procedural questions along the way. The process begins with an idea that is refined into a statement of the problem.

Recall that science is generally characterized as:

- the systematic organization of information about the world with the goal of discovering new relationships among natural phenomena;
- endeavoring to explain why phenomena occur and how they are related; and
- the formulation of explanations in such a way that they can be empirically tested.

I. Statement of the Problem

What a researcher identifies as a problem or an issue worthy of study is often influenced by attitudes about what constitutes legitimate science and what constitutes a legitimate problem to study, disciplinary training in particular theories and methodologies, the process of peer review, and the degree of institutional and financial support for different types of research (e.g., universities, research institutions, or funding agencies).

A researcher's initial idea may be somewhat vague but it does identify the variables to be studied. The researcher's initial ideas are transformed into a **statement of the problem** by building a prediction into the question. In experimentation, the problem statement focuses on a causal prediction (Does variable A cause a specific change in variable B?). The nature of the expected

Recall that philosophers of science from a social constructivist perspective argue that scientific discovery is embedded in a social and political context.

Statement of the Problem

(i) the identification of at least two variables

(ii) a statement about an expected relationship between the identified variables

(iii) an indication of the nature or direction (e.g., increase or decrease) of the causal effect

effect is also clearly stated (i.e., whether the predicted effect of variable A is to increase or decrease the level or occurrence of variable B).

Articulating the statement of the problem is an important early step in designing experimental research. In articulating the statement of the problem, the researcher should have clearly identified and defined the major variables to be studied and clearly specified the nature of the predicted effect of one variable on another variable.

???

Questions to consider when evaluating scientific evidence ...

- ◆ Do you have a clear understanding of what the research was designed to study?
- ◆ Do you understand the nature of the predicted relationship? That is, did the researcher clearly articulate the statement of the problem?

II. Hypothesis Development and Hypothesis-Testing

Most scientific research studies are attempts to test an **hypothesis** formulated by the researcher. An hypothesis is a type of idea; it states that two or more variables are expected to be related to one another. The research hypothesis has its beginning in initial ideas which are often vague and overly general. The researcher must carefully refine these initial ideas into a statement of the

Hypothesis: a claim or prediction that two or more variables are expected to be related to one another

problem drawing on observations of the phenomenon, as well as a thorough review of previous research.

A well constructed hypothesis should be internally consistent and logical. An hypothesis that is clearly illogical or self-contradictory on its face should be rejected. An hypothesis should also be examined to ensure that it really provides insight and understanding into why observed phenomena occur; an hypothesis that requires constant modification to explain away contradictory results should be treated with skepticism. And a valid hypothesis in experimental research should also be able to survive experimental or observational tests that will show it to be false if it is wrong.

The researcher designs a study or experiment to test an hypothesis. **Hypothesis-testing** is a critical part of the experimental research process. The researcher makes a specific prediction concerning the outcome of the experiment. If the prediction is confirmed by the results of the experiment, the hypothesis is supported. If the prediction is not confirmed, the researcher will either reject the hypothesis or conduct further research using different methods to test the hypothesis.

Hypothesis-testing: the process of systematically testing an hypothesis

Developing the **research hypothesis** is a major task for the researcher. It is the research hypothesis that is tested through the processes of systematically making, measuring, analyzing, and interpreting empirical observations under controlled conditions.

Research Hypothesis: (i) Identifies and operationalizes the independent and dependent variables; (ii) states the relationship between the independent and dependent variables; and (iii) allows for the possibility of empirically testing the relationship

The research hypothesis is a complex statement that actually incorporates two hypotheses:

- i. The Null Hypothesis; and

ii. The Experimental (or Causal or Alternative) Hypothesis.

i. *The Null Hypothesis (H_0)*

The null hypothesis is what its name suggests; null means ‘none.’ The null hypothesis states that there is no difference between the two conditions beyond chance differences (e.g., Variable A has no effect on Variable B). If a statistically significant difference is found, the null hypothesis is rejected. If the difference is found to be within the limits of chance, it is concluded that there is insufficient evidence to reject the null hypothesis.

ii. *The Experimental (or Causal or Alternative) Hypothesis (H_1)*

The experimental hypothesis states that a particular variable has a predicted effect on another variable. The nature of this effect or relationship can be stated in two ways: (1) The manipulation of the first variable causes an increase in the level of the second variable (an increase in variable A causes an increase in variable B); or (2) the manipulation of the first variable causes a decrease in the second variable (an increase in variable A causes a decrease in variable B).

The statement of the problem is converted into a research hypothesis when the theoretical concepts in the problem statement are described in terms of their procedures for measurement or manipulation. This process is called operationalization, or creating an **operational definition** of the concepts. By combining the statement of the problem and operational definitions within experimental research, the researcher makes a prediction about the effects of the specific, operationally defined independent variable on the specific, operationally defined dependent variable.

<p>Operational Definition: a description of an independent or dependent variable, stated in terms of how the variable is to be measured or manipulated</p>

The **independent variable** (IV) is the presumed *cause* of some outcome under study; changes in an independent variable are hypothesized to have an effect on the outcome or behavior of interest. The independent variable is the variable that is experimentally manipulated by the researcher.

Independent Variable: the presumed *cause* of some outcome under study; the experimentally manipulated variable; changes in an independent variable are hypothesized to have an effect on the outcome or behavior of interest

The **dependent variable** (DV) is a presumed *effect*. The dependent variable is predicted to change as a result of the manipulation of the independent variable. The value of the dependent variable (e.g., score) is dependent on the value of the independent variable.

Dependent Variable: a measure of presumed *effect* in a study; predicted to change as a result of the manipulation of the independent variable; the value of the dependent variable (e.g., score) is dependent on the value of the independent variable

It is very important that the independent variable and the dependent variable are clearly defined and clearly articulated. The operational definition of both the independent and dependent variable should accurately define what is meant by a particular variable. If the independent and dependent variables are not clearly defined and articulated -- that is, they are not clearly operationalized -- then it is difficult to determine with any degree of certainty whether or not the researcher has actually studied what she intended to study and whether the results of the experiment are valid and reliable.



Questions to consider when evaluating scientific evidence ...

- ◆ How were independent and dependent variables operationalized? That is, do you clearly understand what each variable means (as indicated by operational definitions)?
- ◆ Did the operational definitions adequately capture the full conceptual meaning of the variables?
- ◆ Do you have a clear understanding of how the researcher intended to measure changes in the variables?
- ◆ Did the researcher actually measure changes in the variables in the way originally intended?

III. Experimental Research Design

The components of the experimental process discussed so far underscore that careful and systematic planning and development are critical for well conducted scientific research. The research design is the detailed customized process for the systematic testing of the research hypothesis. The research design should have been clearly described in detail – it serves as a road map for other researchers to follow when replicating the research and it provides a basis for a critical review of the research methodology. Note that the experimental design refers to both the activity involved in the detailed planning of the experiment and to the detailed plan itself.

The Concept of Falsifiability Revisited: Testing the Null Hypothesis

Popper argued that the objective of testing is the refutation of the hypothesis. When a theory's predictions are falsified, the theory will be rejected. Those theories that survive falsification are said to be corroborated and tentatively accepted, but not positively confirmed. Popper's notion of falsification is based upon the assumption that even after the testing and corroboration of predictions, an hypothesis has only achieved the status of 'not yet disconfirmed.'

In the process of hypothesis-testing, a finding that the difference between two experimental conditions is within the realm of chance results in a failure to reject the null hypothesis. Following the logic of Popper's concept of falsification, the null hypothesis has not yet been disconfirmed -- but the potential for rejecting the null hypothesis exists. That is, it is still possible for the null hypothesis to be rejected given future tests of that hypothesis. According to Popper, a theory that does not have the potential to be falsified in this way is not scientific.

It is important to note, however, that while it may be the case that the null hypothesis cannot be absolutely accepted, in many practical contexts decisions have to be made and actions have to be taken as though the null hypothesis were true. This is especially true in applied research where decisions have to be made based upon imperfect knowledge.

There is a great deal of variability in how research is designed and conducted. Each type of design carries with it strengths and weaknesses and different designs are more appropriate for answering certain kinds of research questions than others. The task of the researcher is to develop a research design that properly and appropriately tests the research hypothesis.

◆ Experimental Variance

Variance is a necessary part of experimentation – without variation there would be no differences to test. When an experiment is conducted, the researcher predicts variation and hopes to determine that the variation between two or more research groups is due to experimental manipulation of the independent variable. However, as much as variation between experimental groups is the goal, the researcher (and those evaluating the research) must be cautious about unwanted or extraneous variation. Unwanted variation

can occur in any study and can threaten the validity of the study by allowing for alternative explanations of results. This reduces confidence in drawing causal inference, in generalizing beyond the sample, and in interpreting results. Two primary forms of variance are:

- i. Systematic Between-Groups Variance; and
- ii. Non-Systematic Within-groups Variance.

i. Systematic Between-Groups Variance

For purposes of illustration, let us assume that a study has three levels of the independent variable (e.g., three different dosages of a given drug) and three groups, each of which gets one of the dosages. The researcher predicts that the dependent measure will differ across each of the three groups depending on the level of the independent variable (drug dosage) for that group. If there is not a significantly high between-groups variance – that is, the groups are essentially the same on the dependent measure – then the independent variable had no effect. Thus, a significantly high between-groups variance is needed to support the research hypothesis that the independent variable influenced the dependent variable as predicted.

However, even if the between-group variance is high, the researcher must be careful in drawing conclusions about a causal relationship. The significant difference may be due to either the systematic effects of the independent variable as predicted by the research hypothesis (**experimental variance**) or it may be due to systematic effects of uncontrolled extraneous variables (**extraneous variance**), or a combination of the two. Thus, between-groups

Systematic Between-Groups Variance: systematic differences between research groups on measures of the dependent variable as a function of different levels of the independent variable; includes influence of both experimental variance and extraneous variance

Experimental Variance: differences between research groups are due to systematic effects of the independent variable on the dependent variable as predicted by the research hypothesis

Extraneous Variance: differences between research groups are due to uncontrolled or extraneous variables rather than to the systematic effects of the independent variable on the dependent variable

variance is a function of both experimental effects and confounding effects. High experimental variance is important for the experiment. High extraneous variance is a serious problem and it makes conclusions about causality difficult, if not impossible, to draw. Thus, it is critically important that researchers seek to maximize the experimental variance and control for, or minimize, the extraneous variance.

ii Non-Systematic Within-Groups Variance

The term **error variance** is often used to denote the non-systematic within-groups variability. Error variance is due to random factors that affect only some subjects within a group. For example, some subjects may score lower than other subjects in the group because they are tired or anxious, or because of individual differences in personality, motivation, or interest. Error variance is also increased by experimenter errors or equipment variations that cause measurement errors for some subjects in a group but not for others. Error variance is the variation among individuals within a group that is due to chance factors. Because no two individuals are alike and no procedures are perfect, there will always be some degree of error variance. Error variance is random and therefore has random effects. If random errors cause some subjects to score lower than others, it is reasonable to assume that random error will cause other subjects to score higher-- that is, the effects of within-groups random error tend to cancel each other out.

Non-Systematic Within-in Group Variance: differences between subjects in the same group that are due to random influences; variability within a group

Error Variance: random error that results from differences between subjects, experimenter errors, and/or equipment variations

To show a causal effect of the independent variable on the dependent variable, the experimental variance must be high and not be masked by too much extraneous or error variance. The greater the extraneous and/or error variance, the more difficult it becomes to show the effects of systematic, experimental variance. In experimentation, each study is designed to maximize the experimental

variance, control extraneous variance, and minimize error variance.

- ◆ Researchers must be aware of, explicitly acknowledge, and act to minimize threats to validity.
- ◆ Researchers must ensure that subjects in different research groups are as similar as possible.
- ◆ Researchers must ensure that subjects in experimental and control groups are treated in the same way with the exception of the experimental manipulation.

When critically evaluating research, it is important to ask some fundamental questions with respect to the research design:

- ◆ Did the researcher use an experimental or quasi-experimental design?
- ◆ Did the researcher appropriately control for extraneous, or confounding, variables that might influence the nature and strength of the relationship between the variables?
- ◆ Did the researcher appropriately acknowledge and diminish threats to validity?

The answers to these questions place boundaries and limits on the type of interpretations that can be made about the research results obtained and the level of confidence with which one can make causal inferences about the relationship between the independent and dependent variables.

- ◆ *Did the researcher use an experimental or quasi-experimental design?*

The primary difference between experimental and quasi-experimental research design is the degree of control the researcher exercises over the research situation and whether or not the researcher is able to randomly assign individuals (i.e., subjects) to different research groups.

Pure **experimental designs** are characterized by the ability to **randomly assign** subjects to different experimental conditions. Moreover, in pure experiments researchers have a great deal of control over the research groups and the research environment. In experimental research the researcher is generally able to manipulate the research situation or condition, to make causal predictions about the outcome, and to observe the resulting outcome. It is because the researcher has the ability to randomly assign subjects and manipulate the research situation, that he can draw causal inferences about the effect of one variable (i.e., the independent variable) on another variable (i.e., the dependent variable).

In pure experiments the control of variance is maximized through the use of random assignment of subjects and control groups. A **control group** serves as a basis of comparison for some other, experimental group. The ideal control group is similar to the experimental group on all variables except the variable of interest (the independent variable). This is achieved through random assignment of subjects. There are a large number of experimental research designs that vary in complexity, however an in-depth discussion of experimental research design is beyond the scope and purpose of this *Deskbook*.

Both pure experimental and quasi-experimental designs incorporate experimental manipulations, outcome measures, and comparison groups. Both experimental and quasi-experimental designs are concerned with testing causal relationships.

Random assignment to experimental groups is the distinguishing characteristic between pure experiments and quasi-experiments. True experiments utilize random assignment, while quasi-experiments do not.
Control Group: a group of subjects used in experimental research that serves as a basis of comparison for experimental groups; the ideal control group is similar to the experimental group(s) on all variables except the independent variable

Two Examples of an Experimental Design

Randomized, Posttest-Only, Control-Group Design

In a randomized, posttest-only, control-group design subjects are selected from a specified population of interest according to some sampling procedure (discussed further in Chapter 6), and randomly assigned to either an experimental (treatment) group or the control (non-treatment) group; the experimental manipulation is carried out and the specified dependent variable is measured; differences in the posttest scores of the two groups are then measured.

Randomized, Pretest-Posttest, Control-Group Design

In a randomized, pretest-posttest, control-group design subjects are selected from a specified population of interest according to some sampling procedure, and randomly assigned to either the experimental (treatment) group or the control (non-treatment) group; subjects in both groups are measured on the dependent variable (pretest); the experimental manipulation of the independent variable is carried out and the specified dependent variable is measured again (posttest); the differences between the pretest and posttest measure of each group are then compared.

The highest degree of control is obtained with experiments that allow causal conclusions to be drawn with the highest degree of confidence. However, there are times when the standards of a ‘true’ or ‘pure’ experiment cannot be met, but the researcher still wants to answer a causal question. In these situations, a **quasi-experimental design** (“quasi” means “similar to”) can be used.

Quasi-experimental designs have the same general form as experimental designs including a causal hypothesis and some type of manipulation to compare two or more conditions or groups. However, in quasi-experimental designs researchers have less control over the research environment and do not randomly assign individuals to different research groups. That is, quasi-experimental research does not use random assignment to create equivalent comparison groups from which experimental cause is inferred. Instead, comparisons are made between non-equivalent groups that differ from each other in ways other than the

Quasi-Experimental Design: same general form as experimental designs including a causal hypothesis and some type of manipulation to compare two or more conditions or groups, but researchers have less control over the research environment and do not randomly assign individuals to different research groups (non-equivalent comparison groups)

presence or absence of some experimental variable whose effects are being tested. Quasi-experimentation requires separating the effects of an experimental manipulation from the effects due to the original non-comparability of the research groups. In order to separate these effects, the researcher must explicate specific threats to valid causal inferences and find some way to overcome, or at least minimize, these threats. The random assignment of subjects to experimental groups (or conditions) in pure experiments prevents most threats to validity, but in quasi-experiments the threats must be explicitly identified and handled. As with experimental designs, there are a variety of quasi-experimental research designs.

Nonequivalent Control Group Design

The best way to test causal hypotheses with confidence is to compare groups that are created by the researcher through random assignment. This makes it likely that the groups are equivalent at the beginning of the study. However, there are circumstances in which subjects cannot be randomly assigned and therefore, experimental conditions may not be equivalent at the beginning of the study.

Typically, for nonequivalent groups a pretest and posttest measure is taken for each group on the dependent variable and the difference between these two measures is taken as a measure of how each group changed as a result of the manipulation. Because the groups may differ on characteristics other than the variables of interest, it is critically important that the researcher identify, measure, and attempt to rule out extraneous, or confounding, variables.

When evaluating scientific research, the distinction between experimental and quasi-experimental methods, and the implications of those distinctions, is an important one to recognize. For example, whether the research is experimental or quasi-experimental will influence whether random assignment of subjects and control groups are used; the extent to which confounding variables need to be explicated and controlled for; the extent to which a researcher can make a causal inference; and the degree to which one can have confidence in that inference.

- ◆ ***Did the researcher appropriately control for extraneous, or confounding variables that might influence the nature and strength of the relationship between the variables?***

When critically evaluating the results of any experimental or quasi-experimental research, it is important to consider whether or not the results might be due to, or at least affected by, the influence of some extraneous or confounding variable. In experimental research cause-and-effect conclusions are justified only when, all other things being equal, the manipulation of the independent variable leads to a resultant change in the dependent variable. The fact that other things must be kept equal, means that the experimental and control groups must be as identical as possible. Other possibly influencing variables must be held constant so that the only thing that really varies among the groups is the independent variable.

When other factors are inadvertently allowed to vary, these factors *confound* the results. **Confounding variables** can take a variety of forms. For example, confounds might include

Confounding Variable: an extraneous variable that interferes with the researcher's ability to draw a causal connection between the independent variable and the dependent variable; the influence of the independent variable (Variable A) on the dependent variable (Variable B) cannot be disentangled from the possible influence of a confounding variable (Variable C).

- Do changes in Variable A cause changes in Variable B, or only in the presence of Variable C?
- Are changes in Variable B the consequence of Variable C rather than Variable A?
- Each confounding variable is a threat to the validity of the experiment.

differences in the characteristics of the subjects (e.g., in a study of the effects of a particular kind of medication, the fact that all of the subjects in group 1 are over the age of 50 and all the subjects in group 2 are under the age of 50 would be a confound -- the potential relationship between taking the medication and some observed change might be due more to differences in age between the two groups than to dosage), characteristics of the research environment (e.g., differences in the results of an experiment conducted in two labs might be due more to the cleanliness of the labs), or characteristics of the instruments used (e.g., perhaps one instrument is better calibrated than another).

Confounding variables should have been anticipated by the researcher and their potential influence should have been reduced through experimental controls. Controlling or eliminating confounding variables is crucial. Each confounding variable is a threat to the validity of the experiment. This is especially true for quasi-experimental research designs.

◆ *Did the researcher appropriately acknowledge, control for, and diminish threats to validity?*

A major concern in research is the **validity** of the procedures and conclusions.

- A valid measure measures what it is supposed to measure
- A valid research design tests what it is supposed to test

Regardless of whether it has its basis in the physical, biological, social, or behavioral sciences, research influences our personal choices, the quality of our lives, our culture, and our environment. For example, psychological testing is prevalent in modern American society; it is estimated that hundreds of millions of achievement and intelligence tests are administered each year. Advances in the biological sciences, such as DNA testing, have impacted case law and the administration of justice; people who were

convicted for crimes they did not commit are being freed from prison on the strength of DNA evidence, while others who might otherwise have gone unpunished for their crimes are being rightfully prosecuted. Children are being removed from their families on the basis of predictions of future likelihood of parental abuse or on assessments of poor parental competency. Examining the validity of research enables us to say with some confidence that we have achieved a degree of understanding with respect to some phenomenon; that we are measuring what we intend to measure; that the results achieved can safely be said to be specific to the conditions we created; and that the evidence we have gathered truly supports the conclusions we have drawn.

There are four main types of validity that are important to consider when evaluating the conclusions and interpretations made about a scientific study.

Types of Validity

- **Statistical Conclusion Validity**
- **Construct Validity**
- **External Validity**
- **Internal Validity**

◆ Statistical Conclusion Validity

When statistical procedures are used to test the null hypothesis a statement is being made about the statistical validity of the results. A threat to statistical conclusion validity occurs when there are concerns about the adequacy and appropriateness of the conclusion to reject or fail to reject the null hypothesis. There are three primary threats to statistical conclusion validity:

- the possibility that the measures (e.g., calibrated instruments in a laboratory, surveys) used to assess the dependent

variable are unreliable -- that is, the measures cannot be depended upon to measure true changes;

- the possibility that experimental treatments are not consistently implemented across subjects, time, or experimenters; and
- the possibility that the researcher has violated the assumptions that underlie specific statistical tests.

◆ **Construct Validity**

Hypotheses are bound to theoretical ideas or constructs. Construct validity refers to how well the study's results support the theory or constructs behind the research and whether the theory supported by the findings provides the best available theoretical explanation of the results. There must be congruence between the conceptual definition and the operational definition. The level of congruence between the conceptual and operational definition is closely related to the level of confidence the researcher can have in the construct validity of the theory under study. In order to help reduce threats to construct validity the researcher should have clearly stated definitions and carefully built the hypotheses on solid, well-validated constructs. The theoretical bases for the research should be clear and well supported, with rival theories carefully ruled out.

◆ **External Validity**

In the strictest sense, the results of an experiment are limited to those subjects and conditions used in the particular experiment. However, researchers typically want to be able to generalize the results beyond the specific conditions and subjects, and to be able to apply the findings to other similar subjects and conditions. External validity refers to the degree to which researchers can generalize the results of the research to other subjects, conditions, times, and places.

Sampling and External Validity

External validity, or the extent to which the conclusions reached in an experiment generalize both to and across groups, settings, or times, is an important goal in the research enterprise. To achieve external validity, the research must accurately identify the target population -- that is, the population to which the results will be generalized.

A **sample** consists of members of the population who have been selected for observation in an empirical study. One of the assumptions upon which the use of a sample rests is that the sample is representative of the larger population of interest. If the sample is not representative, then sampling bias exists and generalizations made on the basis of the results obtained from the sample are likely to be inaccurate and lack external validity. (Sampling will be discussed further in Chapter 6).

◆ Internal Validity

Internal validity is of great concern to the researcher because it involves the very essence of experimentation -- the demonstration of causality. In an experiment, internal validity addresses the question: "Was the independent variable, and not some confounding (extraneous) variable, responsible for the observed changes in the dependent variable?" Internal validity refers to the approximate validity with which the researcher can infer that a relationship between two variables is causal or that the absence of a relationship implies the absence of cause. Threats to internal validity are especially salient for quasi-experimental designs.

Internal validity:

- is there a causal relationship between variable A and variable B?
- once it has been established that a causal relationship exists, the direction of causality must be measured
 - Does variable A cause variable B?
 - Does variable B cause variable A?
 - Is the relationship between variable A and variable B contingent on the existence of variable C?

Internal validity refers to the causal relationship between the independent variable and the dependent variable. Internal validity is concerned with whether the change in the dependent variable is caused by a change in the independent variable, or by an extraneous variable or variables or artifacts of the research methodology.

It is possible that there exists more than one threat to internal validity in any given situation. The net biasing effect of internal validity threats depends on the number of threats, whether the existing threats to internal validity are similar or different in the direction of bias, and on the magnitude of any bias they cause independently. The more numerous and powerful the validity threats and the more similar the direction of their effects, the more likely it is that a false causal inference will be made.

In a classic text, Cook and Campbell¹ summarized the major types of confounding variables that can affect the results of experimental, and especially quasi-experimental, research designs and thus lead to erroneous causal inferences. The primary threats to internal validity are briefly presented below. A detailed discussion of the various threats to internal validity is beyond the scope and purpose of this *Deskbook*.

Assessing Threats to Internal Validity

Maturation

- Are the observed results due to maturational changes in the subjects or to the sheer passage of time? For example, is it possible that the observed effect might be due to the subjects growing older, wiser, stronger, or more experienced rather than to the experimental manipulation?

History

- Did some event, unrelated to the research, occur between the pretest and posttest that could have influenced the outcome and account for the observed results?

Assessing Threats to Internal Validity (Continued)

Testing

- Are changes in subject scores a function of practice effects? That is, is it possible that subject familiarity with a test enhanced performance when re-taking the same test at a later point in time?

Instrumentation

- Did the measuring instrument change over time?
- Did human observers become more proficient in administering tests and observing change and could this account for overall observed changes?

Regression to the Mean

- Were subjects selected because they scored extreme scores on the pretest measure? For example, is it possible that changes in subject scores between the pretest and post-test occurred because subjects were originally selected on the basis of their extreme pretest scores (either very high or very low) and because scores have a natural tendency to be less extreme on the posttest regardless of the experimental manipulation?

Selection

- Are comparison or control groups equivalent on all important characteristics? Might observed experimental effects be due to the fact that different people were selected, or selected themselves, into the research groups?

Attrition

- Are observed differences between research groups due to the fact that subjects with certain characteristics dropped out of the study while other subjects did not?

Diffusion of Treatment

- Are subjects in different experimental conditions able to communicate and exchange information? That is, did earlier subjects have the opportunity to “give away” the procedures and goals of the research to those subjects scheduled later in the study?

Sequencing Effects

- Did every subject proceed through each experimental condition in the same order? That is, could their experiences in earlier conditions influence their experiences in later conditions?

There are also threats to internal validity that are due to subject and experimenter factors. The expectations and biases of both the

researcher and the subjects can systematically affect the results of the experiment in subtle ways, thus reducing the study's validity. Factors such as motivation, knowledge, expectations, and information or misinformation about the study can be powerful influences on behavior.

Threats to Internal Validity from Subject and Experimenter Effects

SUBJECT EFFECTS

Demand Characteristics: cues implicit in the research setting or procedures, or unintentionally communicated by the researcher, that provide information to the subjects on how they should behave and/or information about the purpose and goals of the research

Placebo Effects: merely being in the experiment produces the predicted effect, regardless of whether or not they were actually part of the experimental group

EXPERIMENTER EFFECTS

Experimenter Expectancies: the expectation of the scientist that the research may turn out in a particular way or have particular findings may influence data selection, the way in which the research is designed, the statistical procedures used, and the interpretation and results. The scientist may unintentionally design and conduct the research so that it supports the expected results

Recall that the major objective of an experiment is to demonstrate with confidence that the manipulated independent variable is the major *cause* of the observed changes in the dependent variable. When causality is not clear because some variable other than the independent variable may have caused the effect, then the internal validity of the research is threatened. While it may be that the independent variable did have some causal influence on the dependent variable, potential threats to internal validity that have not been explicitly identified and handled by the researcher undermine the ability to infer a direct causal relationship between the independent and dependent variable. Thus, threats to internal validity reduce the confidence one can have in the causal relationship between the independent and dependent variable.

Experimental control procedures are needed to counteract threats to validity so that researchers can have confidence in their conclusions. Many control procedures are available to meet the variety of threats to validity. But not every threat to validity is likely to occur in every experiment; thus, not every control measure is needed in every experiment. Although some control procedures are of general value and therefore applicable to nearly all scientific studies, many of the available controls must be carefully selected to meet the particular threats to validity present in the study. However, it is important to realize that controls are necessary in all kinds of research.

The Concept of Experimental Control

The term 'control' is used in several different ways in research design:

- the ability to control the situation in which an experiment is being conducted so as to keep out extraneous forces;
- the ability to determine which subjects/conditions receive a particular experimental manipulation at a particular time; and
- the attempt to control for the knowledge, experiences, attitudes, and expectations of respondents.

However it is used, the major function of experimental controls is to rule out (or severely diminish) threats to a valid causal inference.

Experimental Controls to Reduce Threats to Validity (Continued)

2. Control over subject and experimenter effects (continued)

- the use of deception in which the researcher misinforms the subjects about the purpose of the experiment, the hypotheses being tested, etc. (note: ethical standards require that subjects are "debriefed" at the conclusion of the experiment and told what the real purpose of the experiment was and why deception was used)

3. Control achieved through the selection and assignment of subjects

- subjects must be carefully selected; the target population must be clearly defined and sampling procedures clearly specified

4. Control achieved through specific experimental design

- random assignment of subjects to different experimental conditions

Experimental Controls to Reduce Threats to Validity

Four general types of experimental controls to reduce threats to validity:

1. General control procedures

- preparation of the research setting to eliminate as many extraneous variables as possible (e.g., interfering visual stimuli, other people, etc.)
- careful selection of reliable and valid measures
- explicit specification of operational terms, conditions, procedures, instruments, etc.

2. Control over subject and experimenter effects

- single blind procedures: the subjects do not know whether they have been assigned to the experimental group or the control group
- double blind procedures: the researcher does not know which condition subjects have been assigned to and subjects do not know which condition they are in (e.g., experimental group or control group)
- automation and/or standardization of instructions to subjects and procedures used
- use of objective measures that are empirically observable and precisely defined (i.e., measures that do not require any subjective interpretation on the part of the researcher)
- use of multiple observers and a comparison of the data for agreement (inter-rater reliability)

Threats to Validity in Review

Statistical Conclusion Validity

- a threat to statistical conclusion validity occurs when there are concerns about the adequacy and appropriateness of the conclusion to reject or fail to reject the null hypothesis

Construct Validity

- a threat to construct validity exists when a construct or theory other than those that underlie the experiment allows for an alternative explanation for the results; it is the theoretical construct that is confounded

External Validity

- a threat to external validity occurs whenever the researcher cannot confidently generalize results from the study's particular subjects, times, and settings to other conditions

Internal Validity

- a threat to internal validity exists when some uncontrolled for factor other than the independent variable may be responsible for the results; it is the independent variable that is confounded



Questions to consider when evaluating scientific evidence ...

- ◆ Did the researcher clearly outline the research design? Do you have a good general overview of the major research steps involved?
- ◆ Did the researcher provide adequate justifications for why decisions were made, especially if alternative methods were also appropriate?
- ◆ Did the researcher use an experimental or quasi-experimental design?
- ◆ If an experimental design, was random assignment used? Was an appropriate control group used?
- ◆ If a quasi-experimental design, were appropriate steps taken to avoid, minimize or eliminate potential threats to validity? Was the appropriate level of care taken in expressing the strength or nature of the causal relationship?
- ◆ Did the researcher appropriately control for extraneous, or confounding, variables that might influence the nature and strength of the relationship between the variables?
- ◆ Did the researcher appropriately acknowledge and diminish threats to validity?

Before going any further, stop and reflect ...

- ◆ What underlying assumptions about science do validity concerns reflect?
- ◆ Why is validity, especially internal validity, such an important concept in the scientific method?
- ◆ How might validity threats influence the weight given to evidence about the causal connection between silicone breast implants and auto-immune disease, between some environmental agent and cancer, or between child abuse and repressed memories?

IV. Measurement

Measurement in experimental and quasi-experimental research can take a variety of forms. However, what is most important is to assess the reliability of the measures used. **Reliability** is concerned with whether repeated efforts to measure the same phenomenon come up with the same results. Much of the work on reliability has been done in conjunction with testing.

Test-retest procedures examine the consistency of answers given by the same people to the same test items, or a parallel set of items, at two different test administrations. Test-retest reliability is reported as a correlation (a value of 1.0 indicates a test with perfect reliability; a value of 0.0 indicates the test has no reliability). When a measure consists of a number of items, **internal consistency** checks look at people's responses to different subsets of items in the instrument.

Reliability is not an inherent quality of the measure but a quality of the measure used in a particular

Reliability: the extent to which multiple measures of a phenomenon produce the same results; the extent to which a measure is free from random error

Test-retest: a measure of stability which examines reliability over time; the same test is administered at two different points in time (to the same group of people and a correlation between the two tests is determined)

Internal Consistency Check: responses to different subsets of questions are compared for consistency; reliability check

context. Even when an existing measure is reported to have high reliability, it may not be very reliable when used by different raters with different subjects.

Reliability & Validity in Measurement

Observed Score

- The observed score is just that, the score you would observe in a research setting. The observed score is comprised of a true score and an error score.

$$\text{True Score} + \text{Error Score} = \text{Observed Score}$$

True Score

- The true score is a theoretical concept. Why is it theoretical? Because there is no way to really know what the true score is (unless you're God). The true score reflects the true value of a variable.

Error Score

- The error score is the reason why the observed is different from the true score. The error score is further broken down into method error and random error.

$$\text{Method Error} + \text{Random Error} = \text{Error Score}$$

Method Error

- refers to anything that causes a difference between the observed score and true score due to the testing situation.
 - for example, any type of disruption (loud music, talking, traffic) that occurs while students are taking a test may cause the students to become distracted and may affect their scores on the test.

Random Error

- is caused by any factors related to the characteristic of the person taking the test that may randomly affect measurement.
 - an example of trait error at work is when individuals are tired, hungry, or unmotivated. These characteristics can affect their performance on a test, making the scores seem worse than they would be if the individuals were alert, well-fed, or motivated.

The Relationship Between Reliability and Validity

- a measurement can be reliable, but not valid -- a measure can consistently obtain the same score, but the score itself may not be measuring what it was intended to measure



Questions to consider when evaluating scientific evidence ...

- ◆ Were the measures used reliable?
- ◆ How was the reliability of the measures determined?

V. Data Collection

There is a great deal of diversity in how data can be collected. But, regardless of data collection procedures used, the procedures should have been well articulated as part of the research design and justifications should have been given for the selection of one collection procedure over another potentially appropriate procedure. As discussed earlier, it is important that potential confounds are identified and minimized.

VI. Data Analysis

Chapter 9 presents a brief overview of basic statistical procedures and principles. As with the methodologies used, the data analysis plan should have been carefully laid out in the design of the research. Different statistical tests rest on different assumptions about the underlying population of study, the nature of the inference being made, and the appropriate boundaries within which statistical significance can be determined. Inappropriate selection of statistical tests, violations of underlying statistical assumptions, and incorrect inferences of causality and significance, represent significant threats to statistical conclusion validity and to the confidence with which results can be accepted.

VII. Generalization

As briefly discussed in the section on external validity, the results of an experiment are technically limited to those subjects and conditions used in the particular experiment. However, it is typically the case that researchers want to be able to generalize the results beyond the specific conditions and subjects, and to be able to apply the findings to other similar subjects and conditions. The extent to which research findings are generalizable to another group of subjects, to other situations, and other times, depends a great deal on the amount of care taken in designing the research and minimizing threats to validity (especially threats to external validity). Although it is often the case that the results of well designed research can be generalized, there must be some common relationship or characteristic across individuals, situations, or time.

In most cases, determining the generalizability of research findings to other people, settings, and time involves a replication of the research on another group of subjects, or in a different location, or under different constraints. As discussed at the beginning of this chapter, the research process is cyclical -- findings of one study provide questions for the next study.

CRITICAL QUESTIONS REVIEWED

- ◆ Do you have a clear understanding of what the research was designed to study? Do you understand the nature of the predicted relationship? That is, did the researcher clearly articulate the statement of the problem?
- ◆ How were independent and dependent variables operationalized? That is, do you clearly understand what each variable means (as indicated by operational definition)?
- ◆ Do the operational definitions adequately capture the conceptual meaning of the variables?
- ◆ Do you have a clear understanding of how the researcher measured changes in the variables?
- ◆ Did the researcher clearly outline the research design? Do you have a good general overview of the major research steps involved?
- ◆ Did the researcher provide adequate justifications for why decisions were made, especially if alternative methods were also appropriate?
- ◆ Did the researcher use an experimental or quasi-experimental design?
- ◆ If an experimental design, was random assignment used? Was an appropriate control group used?
- ◆ If a quasi-experimental design, were appropriate steps taken to avoid, minimize or eliminate potential threats to validity? Was the appropriate level of care taken in expressing the strength or nature of the causal relationship?
- ◆ Did the researcher appropriately control for extraneous, or confounding, variables that might influence the nature and strength of the relationship between the variables?
- ◆ Did the researcher appropriately acknowledge and diminish threats to validity?
- ◆ How was the reliability of the measures determined?

Endnotes:

1. Cook, T. D. & Campbell, D.T. (1979). *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Houghton Mifflin Company: Boston.
-

GLOSSARY

attrition	a threat to internal validity when subjects drop out of a study differentially; if the number and type of people dropping out is not distributed across groups, there may be a biasing effect
cause	three elements to establishing cause: covariation between variables; temporal ordering; and no (or severely minimized) confounds
confounding variable	an extraneous variable, unrelated to the experimental relationship of interest, that interferes with the researcher's ability to draw a causal connection between the independent variable and the dependent variable; the influence of the independent variable (Variable A) on the dependent variable (Variable B) cannot be disentangled from the possible influence of a confounding variable (Variable C)
construct validity	refers to how well the study's results support the theory or constructs behind the research and whether the theory supported by the findings provides the best available theoretical explanation of the results
control (experimental)	the major function of experimental controls is to rule out (or severely diminish) threats to a valid causal inference; control is used in three general ways: the ability to control the situation in which an experiment is being conducted so as to keep out extraneous forces; the ability to determine which subjects/conditions receive a particular experimental manipulation at a particular time; and the attempt to control for the knowledge, experiences, attitudes, and expectations of subjects
control group	a group of subjects used in experimental research that serves as a basis of comparison for other (experimental) groups; the ideal control group is similar to the experimental group on all variables except the independent variable
correlation	two events are correlated when the presence of a high value of one variable is regularly associated with a high or low value of another
covariation	two events vary together; a change in one variable is associated with, but not necessarily caused by, another variable

deception	the researcher misinforms the subjects about the purpose of the experiment, the hypothesis being tested, etc.; a technique to control for subject and experimenter effects that may threaten the validity of the experiment
demand characteristics	cues implicit in the research setting or procedures, or unintentionally communicated by the researcher, that provide information to the subjects on how they should behave and/or information about the purpose and goals of the research
dependent variable	a measure of presumed <i>effect</i> in a study; the DV is predicted to change as a result of the manipulation of the independent variable
diffusion of treatment	a threat to internal validity when subjects in different experimental conditions are in close proximity and are able to communicate with each other; earlier subjects may "give away" the procedures used to those scheduled later
double blind procedure	the researcher does not know which condition subjects have been assigned to and subjects do not know which condition they have been assigned to; a technique to reduce threats to validity
error variance	random error that results from differences between subjects, experimenter errors, and/or equipment variations
experimental hypothesis	states the effect the independent variable is predicted to have on the dependent variable
experimental studies	characterized by the ability of the researcher to: manipulate the situation or condition; make predictions about the outcome; and observe the resulting outcome; because of the use of manipulation and experimental controls, it is possible to make causal inferences about the effect of the manipulation on the outcome
experimenter effects	the possibility that the expectations, motivations, and biases of the researcher may systematically affect the results of the experiment in subtle ways, thus reducing the study's validity
experimenter expectancies	the expectation of the scientist that the research may turn out in a particular way or have particular findings may influence data selection, the way in which the research is designed, the statistical procedures used and the interpretation and results
experimental variance	differences between research groups are due to systematic effects of the independent variable on the dependent variable as predicted by the research hypothesis
extraneous variance	differences between research groups are due to uncontrolled or extraneous variables rather than to the systematic effects of the independent variable on the dependent variable

external validity	the degree to which researchers can generalize the results of the research to other subjects, conditions, times, and places
falsifiability	a theory is only scientific to the extent that there is a potential for falsification; the goal of falsification is to refute (prove incorrect) a theory based upon observations gained through the scientific method
history effect	a threat to internal validity when an observed effect might be due to an event which takes place between the pretest and the posttest, when this event is not the event of interest, or when the outcome or the effects of the experiment might be due to different life experiences of the subjects
hypothesis	a type of idea; it states that two or more variables are expected to be related to one another
hypothesis-testing	the process of systematically testing an hypothesis
independent variable	the presumed <i>cause</i> of some outcome under study; the IV is the experimentally manipulated variable; changes in an independent variable are hypothesized to have an effect on the outcome or behavior of interest
instrumentation	a threat when an effect might be due to a change in the measuring instrument between pretest and posttest or to the researchers becoming more proficient over time in administering tests or in making observations
internal consistency	responses to different subsets of questions are compared for consistency; reliability check
internal validity	refers to the approximate validity with which the researcher can infer that a relationship between two variables is causal or that the absence of a relationship implies the absence of cause
maturation effect	a threat to internal validity when an observed effect might be due to the respondent's growing older, wiser, stronger, and more experienced and not due to the experimental manipulation
negative correlation	an increase in one variable is associated with a decrease in another variable
non-experimental studies	descriptive rather than predictive; they can demonstrate that a relationship exists between antecedent relationships and outcomes, but in most cases they cannot establish a causal connection
non-systematic within-in group variance	differences between subjects in the same group that are due to random influences; variability within a group
null hypothesis	the null hypothesis states that there is no difference between the two conditions beyond chance differences; if a statistically significant difference is found, the null hypothesis is rejected, if the difference is found to be within the limits of chance, it is

concluded that there is insufficient evidence to reject the null hypothesis

operational definition	a description of an independent or dependent variable, stated in terms of how the variable is to be measured or manipulated
placebo effect	merely being in the experiment produces the predicted effect, regardless of whether or not they were actually part of the experimental group
positive correlation	an increase in one variable is associated with an increase in another variable
quasi-experimental studies	incorporates experimental manipulations, outcome measures and comparison groups, but does <i>not</i> involve the random assignment of subjects to different experimental conditions
random assignment	all subjects have an equal chance of being assigned to a given experimental condition; a procedure used to ensure that experimental conditions do not differ significantly from each other
regression toward the mean	a threat to internal validity when the researcher selects subjects because their scores on a measure are extreme because on the second testing, scores have a tendency to be less extreme regardless of the experimental manipulation
reliability	the extent to which multiple measures of a phenomenon produce the same results; the extent to which a measure is free from random error
research hypothesis	a precise and formal statement of the research question; identifies and operationalizes the independent and dependent variables; states a relationship clearly between the independent and dependent variables; and clearly allows for the possibility of empirically testing the relationship
sample	consists of members of the population who have been selected for observation in an empirical study; a sample should be representative of the larger population of interest; if the sample is not representative, then sampling bias exists and generalizations made on the basis of the results obtained from the sample are likely to be inaccurate and lack external validity
selection	a threat to internal validity when care is not taken by the researcher to insure that two or more groups being compared are equivalent before the manipulation begins
sequencing effects	a threat when subjects are exposed to more than one experimental condition, their experiences in earlier conditions may influence their experiences in later conditions; if the order of presentation of conditions for all subjects is condition A, followed by condition B, followed by condition C, then systematic confounding effects can occur
single blind procedure	the subjects do not know whether they have been assigned to the experimental or control conditions; a technique to reduce

	threats to validity
statement of the problem	includes: identification of at least two variables; a statement about an expected relationship between identified variables; and an indication of the nature of the causal effect
subject effects	the possibility that the expectations, motivations, and biases of the subjects may systematically affect the results of the experiment in subtle ways, thus reducing the study's validity
statistical conclusion validity	are the results due to some systematic factor or are they due merely to chance variations (e.g., are measures reliable? have statistical assumptions been violated?)
systematic between-groups variance	systematic differences between research groups on measures of the dependent variable as a function of different levels of the independent variable; includes influence of both experimental variance and extraneous variance
temporal ordering	in order for a variable to cause a change in the other, the cause must precede the effect; a necessary component
test-retest	procedures used to examine consistency in answers given by the same people to the same test items or parallel set of items, or at two different test administrations; reported as a correlation; reliability check
testing effect	a threat to internal validity when an effect might be due to the number of times particular responses are measured or to the fact that the questionnaire itself alerts subjects to the subject matter or the goals of the research
validity	concerned with whether the experiment/instruments/measures are really testing what they are supposed to test

SUGGESTED READINGS:

Topic-related resources prepared for the judiciary and legal community:

Black, B., Hoffman, J.M, Dunbar, J.F., Hogan, C.A., & Lavender, G.W. (1997). "The Law of Expert Testimony - A Post-*Daubert* Analysis." In Black, B. & Lee, P.W. (Eds.) *Expert Evidence: A Practitioner's Guide to Law, Science, and FJC Manual*, West Publishing,

pgs. 34-46.

For more technical treatments of the topic:

Cook, T. D. & Campbell, D.T. (1979). *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Houghton Mifflin Company: Boston.

Manly, B.F.J. (1992). *The Design and Analysis of Research Studies*. Cambridge: Cambridge University Press.

Montgomery, D.C. (1991). *Design and Analysis of Experiments, 3rd Edition*. New York: John Wiley

Judge's Notes:

CHAPTER 6

An Introduction to the Scientific Methods of Survey Research

Learning Objectives for Chapter 6

Upon completion of this chapter, the reader should be able to:

- ◆ Articulate the three components of total survey design;
- ◆ Understand different sampling procedures used in survey research and how to evaluate them;
- ◆ Understand the importance of considering non-response rates;
- ◆ Understand the importance of good question construction and the implications of poor question construction;
- ◆ Recognize potential biases in the manner in which questions are designed and survey instruments are constructed;
- ◆ Understand the importance of pre-testing survey instruments; and
- ◆ Critically evaluate survey methodology and results.

A survey is a set of one or more questions asked of respondents (i.e., subjects). A survey can ask people about their attitudes, beliefs, plans, health, work, income, life satisfactions and concerns, consumer preferences, political views, and so on. Virtually any human issue can be surveyed.

Some Examples of What Surveys Can Measure

- ◆ Attitudes and Preferences
- ◆ Beliefs
- ◆ Past Experiences
- ◆ Levels of Knowledge
- ◆ Census Information

➡ *Surveys in Court*

Courts have been slow to admit survey evidence. In the 1930s, courts viewed such evidence as hearsay and hence ruled it inadmissible, particularly since those surveyed could not be cross-examined in court. By the 1950s, some courts held that surveys were not hearsay since they were not being used to prove the truth of what respondents said. Other courts were accepting of surveys as evidence of "present state of mind, attitude, or belief," a

recognized exception to the hearsay rule. Today, in the case of surveys, the question for the court has become: "Was the survey conducted in accordance with generally accepted survey principles, and were the results used in a statistically correct way?"¹

Some Examples of How Survey Evidence Might be Used in Court

- ◆ Obscenity/Contemporary Community Standards
- ◆ Trademark Infringement
- ◆ Deceptive Advertising
- ◆ Employment Discrimination
- ◆ Assessment of Damages
- ◆ Mass Tort Aggregation
- ◆ Change of Venue
- ◆ Expansion of Voir Dire

This chapter provides an overview of survey methodology. It became clear from the results of the national survey that judges had little knowledge about how best to evaluate survey methodology and survey results. Indeed, many judges expressed distrust of survey results and dissatisfaction with their ability to adequately address methodological issues pertaining to survey research. While it is true that the results of poorly designed and poorly conducted surveys should be viewed with skepticism, the same can be said for the results of poorly designed experimental or quasi-experimental research (e.g., if significant threats to validity have not been appropriately addressed). Like other research methodologies, survey research relies upon a range of methodologies and procedures that have been agreed upon by those who do the research to be characteristic of well designed and properly conducted surveys.

It is important to recognize, however, that survey research does present some unique challenges and that it is uniquely prone to the influence of bias. Indeed, how specific questions are written, the order in which questions are presented, and the way in which questions are

asked, and of whom they are asked, all influence the validity and reliability of the survey results. Thus, the claim that survey questions can be designed to get whatever the researcher wants, is, to some extent, true. However, such surveys are not representative of well designed surveys. Well designed survey research provides a powerful tool for gathering valid, reliable, and useful information. Throughout this chapter, the methods and techniques of properly designed survey research will be presented. In addition, attention will be drawn to those areas which are potentially problematic and require particular focus.



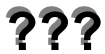
To what extent do judges around the country define surveys as “scientific”?

Of those judges surveyed who believed that “scientific” knowledge can be distinguished from other forms of technical or specialized knowledge (n=244), only 11% believed that surveys constitute “science.” The majority of judges (80%) believed that surveys constitute a form of technical knowledge. A number of the judges (19%) expressed the opinion that the results obtained from surveys are too subjective and sufficiently lacking in validity and reliability to be considered scientific. Some judges recognized the importance of sampling procedures, question construction and the like, but expressed uncertainty about how best to evaluate the survey instrument. 18% of the judges believed that survey results reflected nothing more than questions designed to get what the researcher wanted and statistics manipulated to provide a particular result.

Before going any further, stop and reflect ...

- ◆ How might the results of a survey be used in a specific case? Can you think of some examples of cases in which survey results might be proffered?
- ◆ Would you classify survey research as “scientific” or as technical knowledge? Why?
- ◆ How is survey research like other scientific methods? How is it unlike other scientific methodologies?
- ◆ Given what you know about the scientific method, what information would you expect to hear from an expert presenting results obtained from a survey?

Like all measurement techniques in all scientific disciplines, survey measurement is not error-free. The procedures used to conduct a survey have a major effect on the likelihood that the resulting data will describe accurately what is intended to be described. The content and execution of a survey must be scrutinized to determine if the survey provides relevant, reliable, and valid data on the issue before the court. The purpose and goals of the survey should be clearly stated at the outset, and they should be relevant to the issue at hand. That is, a statement of the problem to be studied should be clearly articulated.



Questions to consider when evaluating survey evidence ...

- ◆ Was the survey designed to address the relevant question of interest?
- ◆ Was the design of the survey, its administration, and the interpretation of the results appropriately controlled to ensure objectivity?
- ◆ Were the experts who designed, conducted, and/or analyzed the survey appropriately skilled and experienced?
- ◆ Were the experts who presented the results of the surveys conducted by others appropriately qualified by skill and experience?

Cross-Sectional vs. Longitudinal Survey Designs

Cross-Sectional Design

- ◆ A cross-sectional design involves administering the survey to a group of people at a given point in time, yielding data on the measured characteristics as they exist at the time of the survey. The information can be completely descriptive or it can involve testing relationships among different characteristics of the sample.

Longitudinal Design

- ◆ The longitudinal or panel design involves administering the survey to the same group of people at different points in time. Longitudinal surveys make it possible to assess changes over time within individuals. It is often difficult, however, to obtain subjects who are willing to be surveyed several times and often large numbers of people drop out of the study before it is completed.

Different Methods of Data Collection: In-Person, Telephone, and Mail Surveys

In-Person Interviews

Advantages: probably the most effective way of enlisting cooperation; rapport and confidence-building are facilitated; longer interviews are easier to complete; interviewer can address respondent concerns directly; interviewer controls the flow, pace, and order of the interview

Disadvantages: can be very costly; requires highly trained interviewers who are located near interview respondents; data collection typically takes longer; some respondents (e.g., those in areas with high crime rates, or rural areas) may be more easily accessed by other data collection modes

Telephone Interviews

Advantages: lower cost than in-person interviews; helpful if population to be sampled is large and/or geographically diverse; better access to certain populations; shorter data collection period; interviewer staffing and supervision easier (e.g., interviewers do not have to be located near to sample); random digit dialing (RDD) techniques can be used (RDD provides coverage of households with both listed and unlisted numbers by generating numbers at random from the frame of all possible telephone numbers -- e.g., numbers with a particular prefix); better response rate than mail surveys

Disadvantages: omits respondents without telephones; non-response is higher than with in-person interviewing; limits use of visual aids and interviewer observations; may be less appropriate for sensitive or personal questions

Mail Surveys

Advantages: relatively low cost; requires minimal staff and facilities; provide access to widely dispersed samples or to samples that might otherwise be inaccessible; respondents can take time to think about answers or look up records

Disadvantages: difficult to enlist cooperation; requires up-to-date and complete addresses; often have low response rates; some potential respondents may not have the necessary reading and writing skills to complete the survey without assistance

To determine the relevance, reliability, and validity of a survey, the **total survey design**² must always be reviewed. In evaluating the total survey design, three different methodological components of the survey must be evaluated critically:

- I. Sampling Procedures
- II. Question Construction

III. Interviewing Procedures

These three components taken together constitute what is called the total survey design. It is not uncommon for researchers to fail to design high-quality procedures in all three of these areas. Researchers often pay attention to only one or two of the primary design features. Current best practice in survey research, however, requires an examination of *all three* design areas. As with any kind of scientific evidence, it is important to critically evaluate survey methodology.

Sources of Error in Survey Research

- ◆ Sampling Procedures Used
- ◆ Question Construction
- ◆ Interviewers
- ◆ Coding of Responses
- ◆ Data Entry
- ◆ Interpretation

Methodological Issues to Consider

- ◆ Sampling Procedures
- ◆ Question Design
- ◆ Demand Characteristics
- ◆ Interviewer Biases
- ◆ Interpretation of Results
- ◆ Internal Validity
- ◆ External Validity

I. *Sampling Procedures*

◆ **Defining the Population**

One of the first steps in designing a survey, or in deciding whether an existing survey is relevant, is to identify the **target population**. The target population consists of all the **population elements** (e.g., objects, individuals, or other social units) whose attitudes, perceptions, behaviors, or knowledge the survey results are intended to represent.

Frequently, however, the target population includes members who are inaccessible or who cannot be identified in advance. As a consequence, some decisions must be made when identifying the **sample population** who will actually be sampled and from whom data will

Target Population: the larger population that consists of all the elements (e.g., objects, individuals, or other social units) whose attitudes, perceptions, behaviors, or knowledge the survey results are intended to represent; the larger population of interest

Population Element: a single member of the population

Sample Population: the smaller population of elements from whom information will actually be gathered; information gathered from the sample population will be used to infer information about the target population

Inferences will be drawn about the target population based upon the information from the sample population.

actually be gathered.

Inferences will be drawn about the target population of interest based upon the information obtained from the sample population. It is therefore critically important that the sample population be properly selected and that it is an accurate reflection of the target population. Moreover, when evaluating the relevance of survey results, it is important to consider whether the target population used for the survey is relevant to the specific issue in question.

A survey report must include a clear definition and description of both the target population and the sample population, as well as a discussion of the differences between the two and an evaluation of how those differences might influence the results of the survey and their interpretation. A survey report must also include a detailed discussion of how the sample population was selected. The way to evaluate a sample population is not by the characteristics of the sample, but by the *process* by which the sample was selected.

???

Questions to consider when evaluating survey evidence ...

- ◆ Was the target population identified appropriately and defined properly?
- ◆ Was the defined target population relevant to the issue in question?
- ◆ Did the population members constitute individuals whose attitudes and/or behaviors are relevant to the dispute?
- ◆ Did the sample population adequately reflect the target population — the individuals whose attitudes and behaviors are relevant to the issue in question?

◆ **The Sampling Frame**

The sampling frame is the set of objects, events, or people that has a chance to be selected, given the sampling approach that is chosen. Any sample

Sampling Frame: the set of objects, events, or people that has a chance to be selected; the sampling frame includes an identification of the sources (e.g., telephone book) from which elements will be drawn, a specification of the probability for selection, and a detailed discussion of the sampling process

selection procedure will give some elements (e.g., individuals) a chance to be included in the sample. The first step in evaluating the quality of a sample is to define the sampling frame. Most sampling schemes fall into three general classes:

1. sampling from a more or less complete list of individuals in the population to be studied;
2. sampling from a set of people who go somewhere or do something that enables them to be sampled (e.g., patients receiving a particular type of medical treatment); and
3. sampling in two or more stages, with the first stage involving sampling something other than the individuals to be selected, but within which the individuals are contained (e.g., a city block or metropolitan area). In two or more steps, these primary units are sampled, and eventually a list of individuals (or other identified sampling units) is created from which a final sample selection is made.

There are three characteristics of a sampling frame that should be evaluated:

- i. Comprehensiveness;
- ii. Probability of Selection; and
- iii. Efficiency.

i. Comprehensiveness

A sample can only be representative of the sampling frame -- that is, the population that actually had a chance to be selected. Most sampling approaches leave out at least a few people from the population the researcher wants to study. Although some sample lists (e.g., registered voters, telephone directories, people with driver's licenses, homeowners) cover large segments of some populations, they also omit major segments with distinctive characteristics (e.g., a telephone directory excludes those individuals with unlisted telephone numbers; a list of homeowners excludes people who rent). A key part of any sampling scheme is determining the percentage of the study population that has a chance of being selected and the extent to which those that are excluded are distinctive.

ii. *Probability of Selection*

It is essential that the researcher, or those evaluating the research, be able to specify the probability of selection for each individual selected (e.g., individuals who appear on a list more than once have a higher chance of selection than those individuals who appear on a list only once).

iii. *Efficiency*

In some cases, sampling frames include units that are not among those that the researcher wants to sample and should be deleted. The researcher must be able to identify how the appropriate persons in the sample were included and how the inappropriate persons were excluded.

???

Questions to consider when evaluating survey evidence ...

- ◆ Was the sampling frame appropriately specified — sources of information, sampling procedures to be used, probability of an element being selected?
- ◆ Was the sampling frame appropriately comprehensive — did it include all necessary population elements?

◆ **Probability Sampling**

When a sample is chosen for a study, the primary objective is to draw a sample that approximates the target population as closely as possible. This is accomplished by selecting the members of the population for inclusion in the sample so that every member of the population has a known and specified probability of being included in, or excluded from, the sample as every other member. This is known as **probability sampling**. 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 Sampling: every member of the population has a known and specified probability of being included in the sample

A probability sample is drawn from a population at random. That is, no systematic bias is permitted to creep into the selection process so that more people of any one kind get included in the sample than ought to given their numbers in the population and the goals of the survey. Although a variety of probability sampling techniques exist, and they range in level of complexity, a review of the primary types of sampling techniques provides a basic understanding of the sampling process.

- *Simple Random Sampling*

Simple random sampling approximates drawing a sample out of a hat -- members of a population are selected one at a time, independent of each other and without replacement. Once a unit is selected, it has no further chance to be selected (e.g., once a name is drawn out of a hat, it is not replaced and therefore cannot be re-selected).

For example, assume the researcher has a list of 8,500 individuals and each individual appears on the list only once. The goal of the survey is to select a simple random sample of 100 individuals. The researcher would number each individual from 1 to 8,500 and then using a computer, a table of random numbers, or through some other means of generating random numbers, the researcher would produce 100 different numbers occurring between 1 and 8,500. The individuals corresponding to the 100 numbers chosen would constitute a simple random sample of that population of 8,500.

- *Systematic Sampling*

When drawing a systematic sample from a list, the researcher first determines the number of entries on the list and the number of elements from the list that are to be selected. Dividing the latter by the former will produce a sampling fraction.

For example, assume there is a list of 8,500 people and the researcher requires a sample of 100 -- 1/85 of the list (100/8,500 or 1 out of every 85 individuals) is to be included in the sample. In order to select a systematic sample, the starting point is designated by choosing

a random number from 1 to 85 and then from that number taking every 85th person on the list (e.g., 24 is randomly selected as the starting number, then person 109 is selected, followed by person 194, 279, etc.). It is important to examine the population list to ensure that the list is not ordered in some systematic way or according to some re-occurring pattern that will affect the sample (e.g., if a list is always ordered as male/female then it may result in an over-representation of one gender).

- *Stratified Sampling*

When a simple random sample is drawn, each new selection is independent and unaffected by any selections that come before it. As a result of this process, any of the characteristics of the sample may, by chance, differ somewhat from the population from which it is drawn. Often, little is known about the characteristics of individual population members before data collection. It is not uncommon, however, that at least a few characteristics of the population can be identified at the time of sampling. When this is the case, there is a possibility of structuring the sampling process to reduce the normal sampling variation, thereby producing a sample that is likely to be more efficient in its ability to reflect the total population.

- *Differential Probabilities of Selection*

Sometimes stratification is used as a first step to vary the probability of selection across various population subgroups. For example, a target population may be subdivided on the basis of geographic location, membership in some group, ethnicity, and so forth. It is important to note that each element in the target population can only belong to one **stratum**. That is, each stratum must be mutually exclusive.

<p>Stratum: a population characteristic (e.g., gender, ethnicity, geographic location) the basis on which the population can be divided; each population element can only belong to one stratum, and each stratum is mutually exclusive</p>
--

The probability of selecting any given element within a stratum may differ across strata. For example, stratum 1 is comprised of all African American individuals living in City X at a given

point in time (100,000 African American individuals) and stratum 2 is comprised of all Hispanic American individuals living in the same city at the same time (10,000 Hispanic American individuals). There are more African American individuals living in the city than Hispanic American individuals, therefore the probability of a specific African American individual being selected in stratum 1 (1/100,000) is less than the probability of a specific Hispanic American individual being selected (1/10,000). If the researcher wants a sample population that contains an equal number of African American and Hispanic American individuals, he must either over-sample from stratum 2 (Hispanic Americans) (i.e., increase the probability of a specific individual being selected) or under-sample from stratum 1 (African Americans) (i.e., decrease the probability of a specific individual being selected).

???

Questions to consider when evaluating survey evidence ...

- ◆ Was the sampling frame clearly defined?
- ◆ Was the sampling frame comprehensive?
- ◆ Were probabilities for selection of elements known?
- ◆ Was the sample drawn using specifiable probabilities of selection?
- ◆ Was the sample selection free of bias, or were some categories of people likely to be omitted or over- or under-represented?

◆ **The Importance of Considering Non-Responses**

The accuracy of any particular inference from the sample to the target population depends on who provides an answer to a particular question. In every survey, there are some people who agree to be respondents and answer every question, others agree to be respondents but do not answer every question, and there are still others who refuse to be respondents. There are three categories of those selected to be in a sample who do *not* actually provide data:

1. those whom the data collection procedures do not reach, thereby not giving them a chance to answer the questions (e.g., wrong telephone numbers, people who screen calls with answering machines);
2. those asked to provide data but who refuse to; and
3. those asked to provide data but who are unable to perform the task required of them (e.g., people who do not speak the researcher's language, people whose reading and writing skills preclude filling out self-administered questionnaires).

The effect of non-responses on survey estimates depends on the percentage not responding and the extent to which those not responding are biased -- that is, the extent to which non-respondents differ systematically from those in the sample who did respond.

Examples of Biases in Non-Responses

Mail Surveys

- people who have a particular interest in the subject matter or the research itself are more likely to return mail questionnaires than those who are less interested;
- better-educated or more literate people usually return mail questionnaires more quickly than those with less education

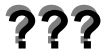
Telephone Surveys

- certain types of people will tend not to be home during certain times of the day (e.g., if telephone surveys are conducted between 9 a.m. and 5 p.m. Mondays to Fridays the sample will yield a high proportion of homemakers, retired people, and the unemployed).

Although there is no agreed-upon standard for a minimum acceptable response rate, it is generally the case that a survey with a higher response rate will produce a better and less biased sample than one that has a higher level of non-response.

Before going any further, stop and reflect ...

- ◆ What might some consequences of non-response be for the validity and reliability of survey results?



Questions to consider when evaluating survey evidence ...

- ◆ Was the most appropriate mode of data collection (e.g., in-person, telephone, or mail) selected?
- ◆ Were the appropriate steps taken to reduce non-responses?
- ◆ Were the disadvantages of the selected collection mode acknowledged?

II. Question Construction

Before going any further, stop and reflect ...

- ◆ A concern expressed by many judges in the national survey was that survey questions could be designed to elicit whatever information the researcher wanted. Think about how you would ask questions on a survey. How might you shape a question to elicit a certain response?
- ◆ Recognizing that the way in which questions are designed may influence the answer given, can you think of some ways in which researchers can reduce the potential bias of question design?

◆ Designing Good Questions

Good questions are *reliable* (providing consistent measures in comparable situations) and *valid* (answers correspond to what they are intended to measure). Designing good, reliable and valid questions presents a number of challenges to the survey researcher. For example, the researcher has to ensure that the questions are written in such a way that the respondent will understand what the question means and that the respondent's understanding of the question matches what the researcher intended. The researcher also has to construct questions in such a way as to minimize biases that may be inherent in the way in which the question is worded, multiple choice answers are provided, rating scales (e.g., on a scale of 0 to 5) are developed, questions are ordered, and so forth.³

Characteristics of Good Questions

- ◆ Questions should be clearly written, complete, and concise
- ◆ Question meaning should be clear and consistent for all respondents
- ◆ Questions should not be double-barreled (i.e., a single question that actually contains two questions)
- ◆ Questions should not include double negatives
- ◆ Questions and response alternatives should not be "loaded" (e.g., response alternatives should present both sides of the issue; not use overly emotion-laden terminology)
- ◆ Questions should have mutually exclusive response categories

◆ The Use of Filter Questions

Filter questions are used to screen out respondents who do not have an opinion on, or any knowledge of, the issue being addressed. For example, 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. There are three approaches researchers generally use to deal with a 'don't know' possibility.

Filter Question: a question used to screen out respondents who do not have an opinion on, or any knowledge of, the issue being addressed

1. Simply ask the question directly and rely on the respondent to volunteer a 'don't know' response. Faced with a direct question, however, respondents may be unwilling to admit a lack of knowledge and instead guess at the answer.
2. The survey can include a **quasi-filter question** to reduce guessing by providing a 'don't know' response alternative. By signaling to the respondent that it is acceptable not to know the answer, the filter reduces the demand for an answer and, as a result, the inclination to hazard a guess is reduced. Respondents are more likely to endorse a 'don't know' if it is mentioned explicitly by the interviewer than if it is merely accepted when the respondent spontaneously offers it as a response.

Quasi-filter Question: providing the respondent with a "don't know" or "no opinion" response option; reduces the demand for an answer and reduces guessing on part of respondent

3. The survey can include **full-filter questions**, that is, questions that lay the ground work for the substantive question by first asking the respondent if he has an opinion about the issue. The interviewer then asks the substantive question only of those respondents who indicate that they have an opinion on the issue.

Full-filter Question: ask the respondent if she has an opinion on the issue and then only ask the follow-up question of those that have an opinion

The choice among these three options and the way they are used can affect the rate of ‘don't know’ and ‘no opinion’ responses that a given question will evoke. For example, respondents are more likely to say they have no opinion when a full-filter question is used than if a quasi-filter question is used. It is important to recognize that the use of full-filter questions may produce an under-reporting of opinions. For example, full-filter questions may discourage respondents who actually have opinions from offering them by conveying the implicit suggestion that the respondent can avoid difficult or time-consuming follow-up questions by saying that he has no opinion.

In sum, a survey that uses full-filter questions tends to provide a conservative estimate of the number of respondents holding an opinion, while a survey that uses neither full-filter nor quasi-filter questions tends to over-estimate the number of respondents with opinions because some respondents offering opinions are guessing.

◆ **Open-Ended vs. Close-Ended Questions**

When constructing a questionnaire, the researcher must decide what form, or forms, questions will take. Survey questions can be classified broadly into two forms: open and closed. **Open-ended questions** ask for a reply in the respondents’ own words – no answers are suggested. **Close-ended questions** (e.g., multiple choice questions) ask respondents to choose one of two or more response alternatives suggested to them.

Open-ended Questions: questions designed to allow respondents to answer in their own words; no answers are suggested

Close-ended Questions: questions which offer a series of response alternatives among which respondents must choose

Open-ended questions that do not provide answers allow respondents to answer according to their own frames of reference, without having to choose among specific alternatives suggested by the interviewer. Open-ended questions generally reveal what is most salient to respondents, what things are foremost in their minds. However, open-ended questions can also elicit a great deal of repetitious, irrelevant material. Respondents may miss the point of the question or engage in long, awkward silences as they try to organize and articulate their thoughts. The interviewer must then skillfully probe to bring respondents back to the subject, to clarify responses, and to encourage elaboration (the appropriate use of probes will be discussed later in this chapter). Individuals also differ a great deal in their ability to articulate their thoughts, with the result that differences in responses may reflect differences in ability to express opinions as much as real differences in shades of opinion. Close-ended questions ensure that respondents will choose among alternatives of interest to the investigator; but the list of alternatives might suggest answers that respondents had not thought of before, or force respondents into what may be an unnatural frame of reference, and they generally do not permit respondents to express their exact meaning.

The value of any open- or close-ended question depends on the information the question 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 close-ended questions are suitable for assessing the choices between well-identified options or obtaining ratings on a clear set of alternatives.

◆ **Rating Frequencies of Behavior**

Often surveys will ask respondents to indicate how frequently they engage in a particular behavior during a specified period of time. This is a common type of close-ended question in which the respondent is given alternatives of, for example, "once a week," "twice a week," "one week a month," etc. How the response alternatives are presented in the close-ended format influences the respondent's judgment of frequency. For example, the response categories may provide implicit cues to the respondent about how rare or common the

researcher expects the event to be. Researchers often use response categories of “sometimes” or “frequently,” but the meaning of those terms may differ depending upon the issue being addressed and the person providing the answer (e.g., does the respondent define “sometimes” in the same way that the researcher does?). The influence of response categories on reporting of frequencies of behavior tend to be more pronounced when respondents have difficulty recalling behaviors, either due to poor memory or because the behavior is not very distinctive.

◆ **Rating Scales**

Rating scales are frequently used in survey research. For example, respondents might be asked to provide a rating of the amount of experience they have had with some event on a scale from 0 (no experience) to 10 (a great deal of experience).

Research has shown that the scale used can influence the answer given (e.g., a scale of 0-10 vs. a scale of -5 to +5). For example, respondents in a research study were asked the following question: “How successful would you say you have been in life?”⁴ Respondents were then asked to rate their answer on a scale ranging from “not at all successful” to “extremely successful.” One group of respondents were told to rate their answer on a scale of 0 (not at all successful) to 10 (extremely successful), while the other group was told to rate their answers on a scale of -5 (not at all successful) to +5 (extremely successful). The research demonstrated that differences between the endpoints of the scales influenced the respondents’ interpretation of what is meant by “not at all successful.” That is, when “not at all successful” was assigned a value of 0 on the 0 to 10 scale, respondents interpreted “not at all successful” to mean the absence of outstanding achievements. By contrast, when “not at all successful” was assigned a value of -5 on the -5 to +5 scale, respondents interpreted “not at all successful” to mean the presence of explicit failures.⁵

These findings, as well as findings from other research on survey methodology, indicate that answers given are influenced by the manner in which the questions and the response categories are designed. That is, respondents draw on implicit cues in the question and response categories when providing answers. This is especially true if questions are difficult or ambiguous. It is therefore important that researchers pay particular attention to how questions and response categories are designed and to acknowledge the potential influence of implicit cues on answers given.

◆ **The Use of Probes**

When questions allow respondents to express their opinions in their own words, some of the respondents may give ambiguous or incomplete answers. In such cases, interviewers may be instructed to record any answer the respondent gives and move on to the next question, or they may be instructed to probe to obtain a more complete response or to clarify the meaning of a response. If probes are used, the wording of probes should be clearly defined and they should be used consistently across all interviewers. If probes are not used systematically across all interviewers, the probes themselves may introduce bias into the results (e.g., some interviewers may probe more than others and elicit more complete responses while other interviewers may under-probe or the probes themselves may provide implicit cues about how the respondent should answer the question).

◆ **The Importance of Question Order and Context**

It is important to realize that the order of questions may influence the responses given. Each question-response provides context for the next question-response. That is, how a respondent interprets the meaning of a question is influenced by questions that come before it. Thus, the wording of questions may be a confounding variable. Care should be taken in how questions are ordered and according to what logical sequence. Particular care should be taken in the placement of sensitive questions (e.g., questions about personal issues that might make the respondent uncomfortable should not be the initial questions as such placement of sensitive questions might interfere with the respondent's willingness to continue).

Often survey researchers will develop two versions of a survey instrument. Each version reflects a different ordering of questions, or “blocks” of questions. Statistical checks can then be conducted to determine whether there were any **order effects** – that is, whether differences in responses can be attributed to the way in which questions were ordered.

Order Effects: responses to survey questions may be influenced by the order in which questions are asked; order effects introduce a confound into the research and draw into doubt conclusions drawn about the respondents' answers

???

Questions to consider when evaluating survey evidence ...

- ◆ Were questions constructed appropriately (e.g., clear, concise, consistent meaning)?
- ◆ Were filter questions used appropriately?
- ◆ Were open- and close-ended questions used? How was the use of each type of question justified?
- ◆ Were the potential biasing influences of question construction and response categories adequately acknowledged and minimized?
- ◆ Were steps taken to guard against order and context effects?

Designing Good Questions: Issues to Consider

- ◆ How the wording of questions may influence responses given
- ◆ Whether the respondent interpreted questions in the manner intended by the researcher
- ◆ Whether respondents with no opinions or no knowledge of specific issues were screened through filter questions and, if not, whether the consequences of not filtering respondents were recognized
- ◆ How response categories, including rating scales, provide implicit cues to the respondent about how to answer the question
- ◆ How probes, if used improperly or inconsistently, can influence answers given
- ◆ How question order may influence responses

◆ **The Importance of Pilot-Testing**

In a pilot, or pretest, the proposed survey is administered to a small sample of individuals who are the same as, or very similar to, the individuals who would be eligible to participate in a full-scale survey. During the pilot test, the researchers observe the respondents for any difficulties they may have with questions and probe the source of any difficulties so that questions can be rephrased if confusion or other problems arise. The length of the survey, both in terms of the number of questions and the time it takes to complete, can also be reviewed during the pilot test. If a survey instrument is overly long and cumbersome, or if it takes a long time to complete, respondents are less likely to agree to participate, or, if they agree, they are less likely to complete the survey. For self-administered surveys, pilot tests are often conducted with a focus or discussion group of individuals who each complete the survey and then provide feedback to the researchers regarding question clarity, areas of confusion, and so forth.

???

Questions to consider when evaluating survey evidence ...

- ◆ Were appropriate pilot tests conducted and the feedback incorporated into the final survey instrument?

III. Interviewing Techniques

Although not all surveys involve interviewing (some surveys have respondents answer self-administered questions), it is certainly common to use an interviewer to ask questions and record answers. When interviewers are used, it is important to ensure that the interviewer does not influence the answers respondents give, while at the same time maximizing the accuracy with which questions are answered.

Interviewers must be properly trained on the overall purpose and goals of the survey, the question-specific objectives, and proper interview techniques (e.g., reading questions as stated, the appropriate use of probes). The more complex the survey, the more highly trained the interviewers should be. Inadequately trained and supervised interviewers can be a serious source of inaccuracy and uncertainty in survey research.

Researchers can use a variety of validation techniques to ensure that the survey is administered in such a way as to minimize error and bias. For example:

- researchers can, and in fact should, monitor interviews as they occur and closely supervise interviewers;
- researchers can contact a small sample of respondents to ensure that the interview took place and that they were qualified to participate; and
- researchers can compare the work done by each individual interviewer – for example, by reviewing the interviews and individual responses recorded by each interviewer, any response patterns and inconsistencies can be identified and addressed.

???

Questions to consider when evaluating survey evidence ...

- ◆ Were interviewers appropriately selected, trained, and supervised?
- ◆ What procedures were used to ensure that the survey was administered to minimize error and bias?

CRITICAL QUESTIONS REVIEWED

- ◆ Were the experts who designed, conducted, and/or analyzed the survey appropriately skilled and experienced?
- ◆ Were the experts who presented the results of the surveys conducted by others appropriately qualified by skill and experience?
- ◆ Was the target population identified appropriately and properly defined?
- ◆ Was the target population relevant to the issue in question?
- ◆ Did the population members constitute individuals whose attitudes and/or behaviors are relevant to the issue in question?
- ◆ Did the sample population adequately reflect the target population; that is, the individuals whose attitudes and behaviors are relevant to the issue in question?
- ◆ Was the sampling frame appropriately specified; including, sources of information, sampling procedures to be used, probability of an element being selected?
- ◆ Was the sampling frame appropriately comprehensive; that is, did it include all necessary population elements?
- ◆ Was the sample drawn using specifiable probabilities for selection?
- ◆ Was the sample free of bias, or were some categories of people likely to be omitted or over- or under-represented?
- ◆ Was the mode of data collection selected appropriate (e.g., in-person, telephone, mail)?
- ◆ Were the disadvantages of the selected collection mode acknowledged?
- ◆ Were questions constructed appropriately (e.g., clear, concise, consistent meaning)?
- ◆ Were filter questions (e.g., “don’t know’s”) used appropriately?
- ◆ Were open- and close-ended questions used? How was the use of each type of question justified?

- ◆ Were the potential biasing influences of question construction and response categories adequately acknowledged and minimized?
- ◆ Were steps taken to guard against order and context effects?
- ◆ Were appropriate pilot tests conducted and feedback incorporated into the final survey instrument?
- ◆ Were interviewers appropriately selected, trained, and supervised?
- ◆ What procedures were used to ensure and determine that the survey was administered in such a way as to minimize bias?

Endnotes:

1. For a discussion of the unique legal issues created by the use of survey results in legal proceedings, see Diamond, S.S. (1994). "Reference Guide on Survey Research." In *Reference Manual on Scientific Evidence*. Federal Judicial Center. Washington, D.C.: Government Printing Office, pgs. 221-272; Shapiro, M. et al. (1997). "Guide to Survey Research." In Black, B. and Lee, P.W. (Eds.). *Expert Evidence: A Practitioner's Guide to Law, Science and the FJC Manual*. West Publishing Co., pgs. 159-194.
2. Dillman, D.A. (1978). *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley.
3. For a good overview of some of the problems and challenges of designing good survey questions see Schwartz, N. (1999). "Self-Reports: How the Questions Shape the Answers." *American Psychologist*, Vol. 54(2), pgs. 93-105. This article presents an overview of the issues and provides many illustrative examples, and it does so in a readable and accessible format.
4. Schwartz, N., Knauper, B., Hippler, H.J., Noelle-Neuman, E., and Clark, F. (1991). "Rating Scales: Numeric Values May Change the Meaning of Scale Labels." *Public Opinion Quarterly*, Vol. 49, pgs. 388-395, in Schwarz (1997), *Supra* note 3.
5. *Ibid.*

GLOSSARY

close-ended questions	questions which offer a series of response alternatives (answers) among which the respondent must choose
cross-sectional survey	administering the survey to a group of people at a given point in time, yielding data on the measured characteristics as they exist at the time of the survey; the information can be completely descriptive or it can involve testing relationships among different characteristics of the sample
differential probabilities of selection	when a sample population is stratified and the probability of selecting any given element within a stratum differs across strata
filter question	a screening that asks the respondent if he or she has an opinion about, or knowledge of, the issue; only those with an opinion of, or knowledge about, the issue are then asked the follow-up question
full-filter question	ask respondents if they have an opinion on the issue and then only ask the follow-up question of those that have an opinion

longitudinal survey	administering the survey to the same group of people at different points in time; makes it possible to assess changes over time within individuals; often difficult, however, to obtain subjects who are willing to be surveyed several times and often large numbers of people drop out of the study before it is completed
open-ended questions	questions designed to allow respondents to answer in their own words, no answers are suggested
order effects	responses to survey questions may be influenced by the order in which questions are asked; order effects introduce a confound into the research and draw into doubt conclusions drawn about the respondents' answers
population element	a single member of the population
probability sampling	every member of the population has a known and specified probability of being included in the sample
quasi-filter question	providing the respondent with a "don't know" or "no opinion" response option; reduces the demand for an answer and reduces guessing on part of respondent
sampling frame	the set of objects, events, or people that has a chance to be selected; includes an identification of the sources (e.g., telephone book) from which elements will be drawn, a specification of the probability for selection, and a detailed discussion of the sampling process
sample population	the smaller population of elements from whom information will actually be gathered; information gathered from the sample population will be used to infer information about the target population
simple random sampling	elements of a population are selected one at a time, independent of the other and without replacement; once an element unit is selected, it has no further chance to be selected
stratified sampling	when at least a few characteristics of the population can be identified at the time of sampling, the sampling process can be structured so that the population is organized or stratified according to known characteristics, thereby producing a sample that is more likely to reflect the total population
stratum	a population characteristic (e.g., gender, geographic location) the basis on which the population can be divided; each population element can only belong to one stratum, and each stratum is mutually exclusive (i.e., a population element can only exist in one stratum)
target population	the larger population that consists of all the elements (e.g., objects, individuals, or other social units) whose attitudes, perceptions, behaviors, or knowledge the survey results are intended to represent; the larger population of interest

SUGGESTED READINGS

- Converse, J. and Presser, S. (1986). *Survey Questions: Handcrafting the Standardized Questionnaire*. Newbury Park: Sage Publications.
- Diamond, S.S. (1994). "Reference Guide on Survey Research." In *Reference Manual on Scientific Evidence*. Federal Judicial Center. Washington, D.C.: Government Printing Office.
- Dillman, D.A. (1978). *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley
- Fowler, F. J. (1993). *Survey Research Methods, 2nd Edition*. Newbury Park: Sage Publications.
- Schwartz, N. (1999). "Self-Reports: How the Questions Shape the Answers." *American Psychologist*, Vol. 54(2), pgs. 93-105.
- Shapiro, M., Chase, J.L., Eshelman, R.L., Bode, H.J., Apjohn, N.G., and Leibensperger, E.P. (1997). "Guide to Survey Research." In Black, B. & Lee, P.W. (Eds.) *Expert Evidence: A Practitioner's Guide to Law, Science, and the FJC Manual*. West Publishing Co., pgs. 159-194.

Judge's Notes:

CHAPTER 7

An Introduction to Qualitative Methods

Learning Objectives for Chapter 7

Upon completion of this chapter, the reader should be able to:

- ◆ Understand the differences between quantitative and qualitative research;
- ◆ Discuss when qualitative methods might be the most appropriate methods and why;
- ◆ Understand the general method of qualitative research;
- ◆ Understand both the limits and power of qualitative methods; and
- ◆ Critically evaluate research utilizing qualitative methodologies.

Research based on qualitative methods may appear in a variety of court cases as proffered evidence. For example, a sociologist might offer testimony about the experiences of members of new religious movements (often referred to as “cults”) or gang members based upon his participant observations of these groups. A social worker or social anthropologist might offer testimony in a child protection case based on field studies he conducted with different cultural groups about child-rearing practices (e.g., different culturally derived definitions of appropriate discipline or differences in the acceptable ages for children to contribute to the family income). The list of possible examples of how qualitative research might appear in court is endless.

Qualitative research is grounded in a philosophical tradition that is broadly interpretivist; the focus is on how the social world is interpreted, understood, experienced, or produced. It is based upon methods of data generation and collection which are flexible and sensitive to the social context within which the data are produced and on methods of analysis and explanation building which involve understandings of complexity, detail, and context.¹

Qualitative research should be:

- systematically and rigorously conducted;
- strategically conducted, yet flexible and contextual;
- involve self-scrutiny by the researcher; active reflexivity; and
- produce social explanations that are generalizable in some form.

Qualitative research should not be seen as necessarily in opposition to, and uncomplimentary with, quantitative research.

Qualitative research ... is about “ways of seeing.” The goal of the research, whatever the methodology, is understanding gained through a process of discovery. What is expressed in qualitative research is a process of discovery that asserts particular assumptions of how knowledge is perceived and acquired. Qualitative research generally focuses on complex human social interactions.²

Reflecting on the Assumptions of Qualitative Research

- qualitative researchers are concerned primarily with *process*, rather than outcomes or products;
- qualitative researchers are interested in *meaning* - how people make sense of their lives and experiences;
- qualitative research involves *fieldwork* - the researcher physically goes to the people, setting, site, or institute to observe or record behavior in natural settings;
- qualitative research is *descriptive*; and
- qualitative research is *inductive* - the researcher builds abstractions, concepts, hypotheses, and theories from details

Adapted from Creswell (1994). *Research Design: Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage Publications.

◆ **Developing a Plan**

A well developed plan for a qualitative study addresses data collection, analysis, and report-writing. A plan includes: a well articulated definition of the research design; the unit of analysis used in the design (i.e., what will be studied -- A specific individual? A specific group of people? A particular situation?); specification of the various data sources; an outline of data analysis procedures; formats for reporting the information; and any other special characteristics of the design.

In Chapter 5, the importance of a well articulated statement of the problem and research design was discussed. Although problem statements and their ultimate development into one or more research hypotheses are most formalized and best developed in experimental research, the articulation of a well developed research problem is important in all types of research, both quantitative and qualitative.

◆ **Acknowledging and Identifying the Role of the Researcher**

Qualitative research is interpretative research. As such, the biases, values, and judgment of the researcher must be explicitly acknowledged. A discussion of the researcher's role should include statements about past experiences of the researcher that provide familiarity with the topic, setting, or the informants; his attitudes toward the research problem; why he is interested in the problem; and the like. The steps taken by the researcher to gain access to the setting and to secure permission to study the informants or situation must also be specified.

It is important to recognize that researchers conducting quantitative, experimental, or quasi-experimental research are subject to many of the same biases as qualitative researchers.

However, in qualitative research, personal biases, values, attitudes, and so forth are explicitly acknowledged and, to a greater or lesser degree, reflected in the research.

◆ **Specification of Data Collection Procedures**

Data collection steps involve:

- i. Setting the Boundaries for the Study;
- ii. Collecting Data; and
- iii. Establishing the Protocol for Recording Data.

i. *Setting the Boundaries*

Four general study parameters should be considered:

- *the setting*: where did the research take place?
- *the actors*: who will be observed or interviewed?
- *the events*: what were the actors interviewed about or observed doing?
- *the process*: what was the evolving nature of events undertaken by the actors within the setting?

ii. *Collecting Data*

Data collection procedures differ considerably in how the behaviors or observations of interest are recorded and coded. At one end of the continuum are methods that are relatively unstructured and open-ended. On the other end of the continuum are more structured and pre-defined methods, which itemize, count, and categorize behavior.

Qualitative Data Collection

Observations (as an observer or as a participant)

Unobtrusive observer: the researcher tries to avoid responding in any way to the subject who is under observation

Participant observer: the researcher becomes part of the situation and, to a greater or lesser degree, contributes to the situation

Advantages: researcher has first-hand experience with informant; researcher can record information as it is observed; unusual aspects can be noticed and recorded immediately; useful in exploring topics that may be uncomfortable for informants

Disadvantages: may be seen as intrusive; private information may be observed that researcher cannot record; researcher may not have good observation skills; certain informants (e.g., children) may present problems with building rapport

Interviews (face-to-face, telephone, mail, or focus (delphi) groups)

Advantages: useful when informants cannot be directly observed; informants can provide historical information; allows researchers a degree of "control" over the line of questioning

Disadvantages: provides "indirect" information filtered through the view of those interviewed; information is removed from the immediate natural setting

Qualitative Data Collection (Continued)

Documents (public or private)

Advantages: researcher can obtain the language of informant; more time convenient; represents data that are thoughtful to the extent that informants have given attention to compiling the information (e.g., personal journal)

Disadvantages: may be protected information unavailable to the researcher; materials may be incomplete; documents may not be authentic or accurate

Visual Materials (photographs, videotapes, art objects, film)

Advantages: may be unobtrusive; provides an opportunity for informant to share directly his or her "reality"

Disadvantages: may be difficult to interpret; may not be accessible publicly or privately

iii. Establishing Protocols for Recording Data

Standard forms are often used for observing, collecting, and organizing information. Data protocols can vary in the extent to which they provide structured events to watch for, specific questions to ask, and particular information to uncover.

◆ **Data Analysis**

In qualitative research, data analysis often occurs simultaneously with data collection, data interpretation, and narrative report-writing. In this respect, qualitative analysis clearly differs from the quantitative approach of engaging in the separate activities of data collection, analysis, and report-writing. In qualitative analysis several simultaneous activities engage the attention of the researcher: collecting information from the field; sorting information into categories; formatting the information into a story or picture of the event; and actually writing the qualitative narrative report.

In qualitative research, the researcher takes a voluminous amount of information and reduces it into meaningful categories, patterns, or themes and then interprets the information. Flexible rules govern how one goes about sorting through interview transcripts, observational notes,

documents, and visual material. However, the researcher generally forms categories of information and attaches codes to these categories. These categories and codes form the basis of the emerging story to be told by the qualitative researcher. Although the rules of this process are flexible, the process should be systematic and well articulated as part of the qualitative plan and final report. While much of the work of analysis is taking apart and organizing information, the final goal is the emergence of a larger, consolidated picture of some social reality.

◆ **Reliability and Validity**

The internal validity of qualitative results can be assessed in a number of ways. For example:

- through **triangulation** - that is, by assessing the amount of convergence (or agreement) among different sources of information, different investigators, or different methods of data collection;
- through feedback from informants (subjects) to assess whether interpretations by the researcher, and conclusions drawn, accurately reflect the perspectives of the informants; and
- through involvement of key informants in the interpretation and development of results.

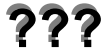
<p>Triangulation: the convergence of different sources of information, different investigators, or different methods of data collection</p>
--

The intent of qualitative research is not to generalize findings, but rather to form a unique interpretation of events for a given group of individuals or institutions, within a given context, at a particular point in time. Like the issue of generalizability, the uniqueness of the study within a particular context mitigates against replicating the research exactly in another context. However, the inclusion of statements about the researcher's position, the central assumptions of the study, the method by which informants are selected, the biases and values of the researcher, all enhance the ability to replicate the research in other settings.

Combining Quantitative and Qualitative Approaches

Often researchers will combine methodologies from both the quantitative and qualitative research approaches. For example, qualitative research is often done in the initial exploratory phases of research. Once tentative hypotheses emerge from the initial qualitative research they are tested through experimental or quasi-experimental designs. Qualitative research may also be conducted at the end of a research project in an attempt to determine how well experimental findings generalize to field settings.

When quantitative and qualitative approaches are combined, each phase of the research should be evaluated according to the standard practices and agreed upon principles of the particular approach.



Questions to consider when evaluating qualitative evidence ...

- ◆ Was the basic type of qualitative design adequately specified in enough detail?
- ◆ Was the research design consistent with the underlying assumptions of qualitative research?
- ◆ Was the researcher explicit about her background, values, experiences, and so forth, and why this particular research issue and research method were selected?
- ◆ Was a detailed description provided of the steps taken to gain entry and approval for data collection at the research site?
- ◆ Were the viewpoints of all participants considered?
- ◆ Were the procedures for collecting data adequately discussed?
- ◆ Was the analysis plan adequately discussed?
- ◆ Was there an appropriate level of inference regarding the generalizability of the results?

CRITICAL QUESTIONS REVIEWED

- ◆ Was the basic type of qualitative design adequately specified in enough detail?
- ◆ Was the research design consistent with the underlying assumptions of qualitative research?
- ◆ Was the researcher explicit about his background, values, experiences, and so forth, and why this particular research issue and research method were selected?
- ◆ Was a detailed description provided of the steps taken to gain entry and approval for data collection at the research site?
- ◆ Were the viewpoints of all participants considered?
- ◆ Were the procedures for collecting data adequately discussed?
- ◆ Was the analysis plan adequately discussed?
- ◆ Was there an appropriate level of inference regarding the generalizability of the results?

Endnotes:

1. Mason, J. (1996). *Qualitative Researching*. Sage Publications, pg. 4
 2. McCombs, G.M.Y., Maylone, T.M. (1998). "Qualitative Research." *Library Trends*, Vol. 46(4), Introduction.
-

GLOSSARY

data protocols	standard forms developed to assist in observations and help to organize data systematically
participant observation	the researcher becomes part of the situation being observed and, to a greater or lesser extent, contributes to the situation
qualitative research	research grounded in a philosophical tradition that is broadly interpretivist; the focus is on how the social world is interpreted, understood, experienced, or produced; context sensitive methods of data collection are flexible
triangulation	the convergence of difference sources of information, different investigators, or different modes of data collection; type of validity check
unobtrusive observation	the researcher avoids becoming involved in the situation or event being observed; the researcher tries to avoid responding in any way to the subject who is under observation

SUGGESTED READINGS

- Creswell, J.W. (1994). *Research Design: Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage Publications.
- Denzin, N.K. and Lincoln, Y.S. (Eds.) (1994). *Handbook of Qualitative Research*. Thousand Oaks: Sage Publications.
- Kirk, J. and Miller, M.L. (1986). *Reliability and Validity in Qualitative Research*. Beverly Hills, CA: Sage Publications.
- Mason, J. (1996). *Qualitative Research*. Thousand Oaks: Sage Publications.
- Robinson, C. (1993). *Real World Research: A Resource for Social Scientists and Practitioner-Researchers*. Oxford: Blackwell.

Judge's Notes:

CHAPTER 8

Peer Review, Publication, and General Acceptance

Learning Objectives for Chapter 8

Upon completion of this chapter, the reader should be able to:

- ◆ Provide an overview of the peer review process;
- ◆ Identify and discuss critiques of the peer review process; and
- ◆ Discuss the differences between referred and unreferred journals and books, and presentations at professional conferences.

In determining whether theory or technique is scientific knowledge that will assist the trier of fact, and, thus, whether expert testimony is admissible, is whether the theory or technique has been subject to peer review and publication. Fact of publication of theory or technique, or lack thereof, in a peer-reviewed journal will be relevant, though not dispositive, consideration in assessing scientific validity of a particular technique or methodology on which expert opinion is premised; submission to scrutiny of scientific community is a component of “good science,” in part because it increases the likelihood that substantive flaws in methodology will be detected.¹

‘General acceptance’ of scientific theory or technique can have bearing in determining admissibility of expert testimony ... Widespread acceptance of scientific theory or technique can be an important factor in ruling particular evidence admissible, a known technique that has been able to draw only minimal support within the community may properly be viewed with skepticism.²

The Peer Review Process

Though pre-publication review is important, it does not by itself establish validity. The review by other scientists that comes after publication is far more significant. Valid scientific knowledge results from mutual criticism and intellectual cooperation. This process does not merely reflect or accompany the scientific method, it is the scientific method. Sharing new discoveries with other scientists, and establishing priority, lies at the heart of science.

Peer review is the process by which scientific papers submitted to journals for publication are reviewed by experts in the field. These experts, or peer reviewers, are familiar with the topic area from their own work and thus can be considered peers of the authors. Reviewers are asked to evaluate the importance and usefulness of the research and judge whether it was performed carefully and accurately. Journal editors consider the reviewers' opinions in deciding whether to publish the submissions. It is important to note that in most cases reviewers do *not* have access to the raw data. Rather, the reviewers must rely on the narrative description of the research design, including how the data were collected, and from what source, and how the data were interpreted.

Peer Review: the process by which scientific papers submitted to journals for publication are reviewed by experts in the field

Institutional Review Mechanisms

Because the process of formulating and testing hypotheses is far from simple and trivial, a probing and careful review is an indispensable part of the scientific enterprise. The review of scientific ideas takes place in a variety of contexts. Informal review can occur when scientists discuss their work with one another at the laboratory bench, during seminars, and at scientific meetings. Formal peer review is generally an integral part of the scientific publication process and the process by which funds are allocated for the conduct of research. Any claim that would significantly add to or change a body of scientific knowledge must be regarded skeptically if it has not been subjected to some form of peer scrutiny, preferably by submission to a reputable journal.

Critiques of the Peer Review Process

Commentators have recently criticized the peer review process on several grounds:

- it lacks explicit standards for review;
- it lacks objectivity and fairness in the selection of manuscripts for review;
- it favors conservative, traditional research and excludes innovative work by independent researchers;
- editors are biased in their selection of reviewers, sending manuscripts only to those individuals who share the opinions of the editorial board; and
- reviewers may not spend enough time critically reviewing the methods and procedures of the research.

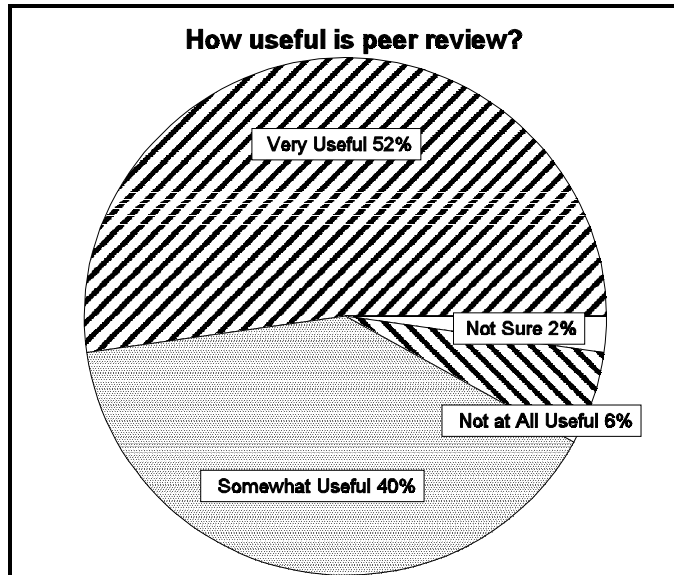


To what extent do judges around the country find the concept of peer review a useful criterion for critically evaluating scientific evidence?

All of the judges surveyed (N=400) were asked how useful the concept of “peer review” is for determining the admissibility of scientific evidence.

The majority of judges surveyed felt that it was a useful concept. Just over half (52%) felt that it was “very useful” and 40% felt that it was “somewhat useful.” Only 6% reported that peer review was “not at all useful” when determining the admissibility of scientific evidence.

When the responses of judges in FRE/*Daubert* states (n=205) were examined, ratings were the same, with half (53%) reporting that peer review was “very useful” and 40% reporting that it was “somewhat useful.” 7% of FRE/*Daubert* judges felt that peer review was “not at all useful” to determinations of admissibility.



When asked a question about how they might apply the concept of peer review to a determination of the admissibility of proffered evidence, 71% of all of the judges surveyed (N=400) provided responses that demonstrated a clear understanding of the scientific peer review process. In 26% of responses, it was not clear that the judge fully understood the concept. In only 3% of the responses, however, was it clear that the judge did *not* understand the concept of peer review. When the responses of judges in FRE/*Daubert* states (n=205) were examined, the results were almost identical. Some sample comments related to the peer review process included:

“You have to look at the motivations behind publication – generally, you have to assume that articles are published in good faith and not just for the purposes of litigation – you have to examine why the study was done”

“I would give greater weight to a criticism of the technique or procedure if it appeared in a significant number of high status journals – or if the prestige of the criticizer was high – nevertheless, practically speaking, it would be difficult for me to evaluate the prestige of the critic”

“Peer review gives you an idea of whether this is a scientific idea that has been debated in the field – it would be important to have the experts describe and debate the peer-reviewed literature to determine the acceptability of the evidence”

➡ *The Evaluation of Research Papers*

Scientific evaluation is made on the basis of criteria established by the journal to which a research paper is submitted. Typically, the more prestigious the journal, the more rigorous the evaluation, and the more difficult it is to get a paper published in that journal. While some journals may accept almost all manuscripts they receive, others may reject 90% or more. Especially high rejection rates develop as a consequence of a journal's high prestige. That is, the most prestigious journals receive many more papers than they can publish, thus allowing the journal to set even higher standards for acceptance.

◆ *Selection of Referees*

Managing editors initially evaluate submissions and select the individuals who will act as referees or peer reviewers. Peer reviewers are usually selected on the basis of their own research experience in the field and their expertise with the methodologies or statistical procedures used. Any given manuscript is typically reviewed by more than one referee. Depending on the scientific discipline and the journal, referees may or may not know the identity and professional affiliation of the authors. When the identity of the author or authors is not known, the process is referred to as a **blind review**. The identity of the referees, however, is not disclosed to the authors.

Blind Review: when identity of the author is not known by the reviewer

After the referees return their reviews of the manuscripts, the managing editor or an editorial board decides whether the comments are reasonable. A preliminary decision regarding suitability for publication is sent to the authors, accompanied by the referees' comments. The author(s) may or may not respond to the comments. Depending on the nature and scope of the referees' comments, the author(s) may address the reviewers comments and resubmit the paper for a second review.

Typical Questions Addressed by Referees When Evaluating a Manuscript for Publication

Is the subject matter appropriate for readers of the journal?

Different journals cover different subjects and target different audiences. The referee must determine whether the work described in the paper is consistent with the stated intent of the journal.

Are the methods used suitable for the questions under study?

The referee must determine whether the research design was appropriate, whether appropriate controls were used, whether the data were presented adequately, and whether the interpretation of the results was reasonable. The referee must consider whether the methods and analysis plan were presented clearly enough, and in enough detail, and whether or not the research could be replicated.

Are the data analyses convincing?

Do the statistical analyses support the conclusions drawn?

Are there other possible explanations for the findings?

Does the paper make an important contribution to the field?

The referee's evaluation may go beyond the adequacy of the work and also consider the potential impact of the research on the relevant field or discipline. More prestigious journals want to publish research that is most likely to have a significant impact on the field. The referee must evaluate the impact of the research within the context of the relevant field – the current problems and questions of interest, currently accepted methodologies, and so forth.

Will the work be of broad-reaching working in the field?

Is the paper in the top 1%, 5%, 10%, etc. of papers written in this field or for this journal?

➡ *Refereed vs. Unrefereed Journals*

Articles published in refereed journals are typically given more weight and afforded more status than articles published in unrefereed journals -- that is, in journals that do not send out submitted articles for independent peer review.

Referred Journals: articles submitted for publication are sent for independent peer review

Unreferred Journals: articles submitted for publication are not independently peer reviewed

Peer Review in the Computer Age

The underlying assumptions and procedures of peer review are receiving increasing attention within scientific disciplines. One development giving urgency to this particular topic is the revolution taking place in scholarly communication. As technology advances it becomes possible to make every work available electronically to all scholars, obviating the need for printed volumes and potentially eliminating the peer review process. This eliminates practical constraints that until now have made it impossible for journals to publish huge amounts of material and for libraries to acquire and store them. As a consequence of advances in technology, much more can be "published" without long delays and backlogs. This increased capacity for communication is also accompanied by relaxed standards of peer review (most electronic journals are not refereed).

➡ *Determining the Status of a Journal*

High status journals are generally considered to be those journals that have a high rejection rate for submitted manuscripts (i.e., the journal only accepts for publication a very small percentage of the number of manuscripts submitted for review). Other factors taken into account in determining the prestige of a journal include: its prominence in the relevant field; its circulation; who subscribes to the journal (e.g., researchers, practitioners, legal professionals); the target audience; and the editorial board membership.

Things to consider when determining the status of a scientific journal:

- whether or not articles published in the journal are peer-reviewed;
- rejection rate of submitted manuscripts;
- level of prominence in the relevant field;
- circulation record;
- subscribers;
- target audience; and
- editorial board membership.

It is important to recognize that while authors may count a journal's reputation more than any other factor in deciding where to submit an article, the appearance of an article in a lower-status journal does not necessarily mean that the article was rejected by a higher-status journal. There are a variety of legitimate reasons why an author may choose to submit an article to a lower-status journal. For example, the author may have decided to submit an article to a journal with higher acceptance rates and quicker response times in an effort to ensure that the research reaches the relevant community in a relatively timely manner. In

addition, the higher-status journal may have recently published an article on a similar topic and would be likely to reject a similar submission on that ground alone. Also, many journals are becoming increasingly specialized and target increasingly specialized audiences. Thus, while a journal may not be defined as prestigious in terms of a high circulation rate, it may be considered a high status, prestigious journal within its specialized, narrow field.

The prestige of books is often measured by who published them. Typically, books published by "prestigious" university presses are counted more heavily than books published by other publishing companies. However, many of the criticisms leveled at journal publications also apply to status determinations about books.

Informal Peer Review

The process of publishing a research paper can often take six months to a year between submission of a paper and its appearance in a journal, although there is great variation among different scientific disciplines. If investigators want to make their research known to the relevant scientific community, they must find a means of reporting the work in a more preliminary form.

Preliminary research reports are often presented at meetings of professional societies or associations. In most cases, researchers request the opportunity to make a presentation at a meeting. By communicating the major conclusions of the study, the work becomes known by the relevant community, and the researcher establishes a claim of ownership and priority over the research. When a researcher is invited by the organizers of a professional society to present her current work at a meeting or conference, it is an indication that the researcher's work is expected to be of interest to the professional audience and that the researcher is an authority whose recent work is probably reflective of the current thinking in the relevant field.

Unlike formal research papers, oral presentations are reviewed minimally or not at all and the details of the methods and materials are generally not presented. Other researchers can learn the thought style behind the work and the general methodologies utilized, but not the precise methods by which the research was accomplished.

[W]hile peer review improves the overall quality of research, as well as the likely integrity of individual publications, it does not mechanically guarantee reliability. Courts therefore should neither be wholly dismissive of peer review nor accord it more respect than it deserves.³

???

Questions to consider when evaluating scientific evidence ...

- ◆ In what ways has the research been peer-reviewed?
- ◆ Was the research published in a referred or unreferred journal?
- ◆ If the research was published in a referred journal, was it a blind review?
- ◆ What is the relative status and reputation of the journal in which the research was presented?
- ◆ Was the disciplinary focus of the journal appropriate for the research?

➔ ***General Acceptance***

[G]eneral acceptance can yet have a bearing on the inquiry. A reliability assessment does not require, although it does permit, explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community. Widespread acceptance can be an important factor in ruling particular evidence admissible and a known technique which has been able to attract only minimal support within the community may properly be viewed with skepticism. ... The focus, of course, must be solely on principles and methodology, not on the conclusions they generate.⁴

Just when a scientific principle or discovery crosses the line between the experimental and demonstrable tests is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts go along way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.⁵



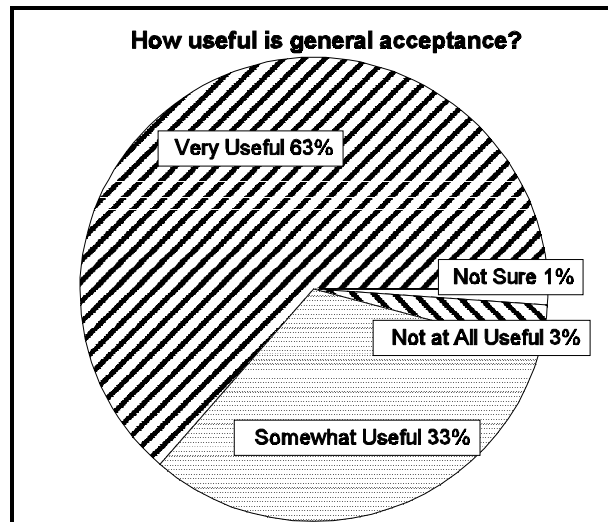
To what extent do judges around the country find the concept of general acceptance a useful criterion for critically evaluating scientific evidence?

All of the judges (N=400) were asked how useful the concept of general acceptance is when determining the admissibility of scientific evidence.

The vast majority of judges indicated that general acceptance was a useful criterion, with 63% rating it as “very useful” and 33% rating it as “somewhat useful.”

When judges in *FRE/Daubert* states (n=205) were examined, the results were much the same, with 61% reporting that general acceptance was “very useful,” 36% reporting that it was “somewhat useful,” and 2% reporting that general acceptance was “not at all useful” to admissibility decision-making.

When asked a question about how they might apply the concept of general acceptance to a determination of the admissibility of proffered evidence, the majority (82%) of all of the judges surveyed (N=400) provided responses that demonstrated a clear understanding of the concept of general acceptance. In 17% of responses, understanding of general acceptance was questionable. It was clear that the judge did *not* understand the scientific concept of general acceptance in only 1% of responses.



Before going any further, stop and reflect ...

- ◆ How might new or novel theories or methods defend themselves against a lack of general acceptance?
- ◆ Does a lack of general acceptance in the relevant scientific community necessarily draw into question the validity and reliability of the research principles and methods? Why or why not?



Questions to consider when evaluating scientific evidence ...

- ◆ Are the theories or principles underlying the research generally accepted in the relevant scientific field?
- ◆ If the theory or principles underlying the research are not generally accepted, how are they defended?
- ◆ What is the relevant scientific field?
- ◆ Are the methods employed in the research generally accepted in the relevant scientific field?
- ◆ Are there relevant but unidentified scientific disciplines that may have a different perspective on the issue?
- ◆ If the methods employed are not generally accepted, how is their use defended?

CRITICAL QUESTIONS REVIEWED

- ◆ In what ways has the research been peer-reviewed?
- ◆ Was the research published in a referred or unreferred journal?
- ◆ If the research was published in a referred journal, was it a blind review?
- ◆ What is the relative status and reputation of the journal in which the research was presented?
- ◆ Was the disciplinary focus of the journal appropriate for the research?
- ◆ Are the theories or principles underlying the research generally accepted in the relevant scientific field?
- ◆ If the theory or principles underlying the research are not generally accepted, how are they defended?
- ◆ What is the relevant scientific field?
- ◆ Are the methods employed in the research generally accepted in the relevant scientific field?
- ◆ Are there relevant but unidentified scientific disciplines that may have a different perspective on the issue?
- ◆ If the methods employed are not generally accepted, how is their use defended?

Endnotes:

1. *Daubert vs. Merrell Dow Pharmaceuticals* 509 U.S. 579, 113 S.Ct., at 2786.
2. *Ibid.*
3. Sheila Jasanoff (1997). *Judging Science: Issues, Assumptions, Models.*
4. *Supra* note 1, at 2797.
5. *Frye*, at 1014.

GLOSSARY

blind review	when identity of the author is not known by the reviewer
peer review	the process by which scientific papers are submitted to journals for publication are reviewed by peers in the field; peer reviewers typically evaluate the importance and usefulness of the research and assess whether the research was carried out in a methodologically appropriate manner and supports the conclusions drawn
referred journal	articles submitted for publication are sent for independent peer review
unreferred journals	articles submitted for publication are not independently peer reviewed

SUGGESTED READINGS

- Bell, R. (1992). "Examining the Peer Review Process." In *Impure Science: Fraud, Compromise and Political Influence in Scientific Research*. New York: Wiley.
- DuRant, R.H. (1994). "Checklist for Evaluation of Research Articles." *Journal of Adolescent Health*, Vol. 15, pgs. 4-8.
- Imwinkelried, E.J. (1997). "*Frye*'s General Acceptance Test v. *Daubert*'s Empirical Validation Standard - 'Either ... Or' or 'Both ... And?'" *Criminal Law Bulletin*, Vol. 33, pg. 72.
- Lock, S. and Smith, D. (1990). "What do Peer Reviewers Do?" *JAMA*, Vol. 263, No. 10 (9 March), pgs. 1341-1343.

Judge's Notes:

CHAPTER 9

Data Analysis: An Introduction to Statistics

Learning Objectives for Chapter 9

Upon completion of this chapter, the reader should be able to:

- ◆ Understand the difference between descriptive and inferential statistics;
- ◆ Understand the difference between different levels of measurement;
- ◆ Understand the measures of central tendency, measures of variability, and measures of relationship;
- ◆ Understand the concepts of "population" and "sample" in inferential statistics;
- ◆ Understand parameter estimates and hypothesis-testing;
- ◆ Understand the concept of error rate, including Type I and Type II, false positive and false negative errors; and
- ◆ Identify common problems associated with the presentation of statistics in court.

Statistics is the science and art of gaining information from data -- of collecting, organizing, and interpreting numerical facts. The field of statistics includes the methods and procedures used to summarize, analyze, and draw inferences from data.

There are two main branches of statistical methods:

Descriptive Statistics

- statistics that summarize, describe, and make understandable the numbers generated by a research study

Inferential Statistics

- statistics used to draw conclusions and inferences which are based upon, but go beyond, the numbers generated by a research study

Statistics: A Practical History of Craps and Beer

During the seventeenth century, the birth of statistics finally took place. It happened one night in France. The scene was a gambling table, and the main character was the Chevalier de Mere, a noted gambler of his time. He had been having a disastrous run of losing throws. To find out whether his losses were indeed the product of bad luck or simply of unrealistic expectations, he sought the advice of the great French mathematician and philosopher Blaise Pascal (1623-1662). Pascal worked out the probabilities for the various dice throws, and the Chevalier de Mere discovered that he had been making some very bad bets indeed. Thus, the father of probability theory was Pascal.

Another milestone for statistics occurred at the turn of the century in Ireland at the famous Guinness brewery, now known worldwide for the record books of the same name. In 1906, to produce the best beverage possible, the Guinness Company decided to select a sample of people from Dublin to do a little beer tasting. Since there turned out to be no shortage of individuals willing to participate in this taste test, the question of just how large a sample would be required became financially crucial to the brewery. They turned the problem over to the mathematician William Sealy Gossett. In 1908, under the pen name "Student," Gossett produced the formula for specifying how large a sample must be to generalize the results to the entire beer-drinking population.

So that's the history - craps and beer. ... The point is that the hallmark of statistics is the very practicality that gave rise to its existence in the first place. The field is not an area of mysticism or sterile speculations. It is a no-nonsense area of here-and-now pragmatism.

Richard C. Sprinthall, *Basic Statistical Analysis*, 5th Edition. Allyn and Bacon (1997), pg. 13.

➔ *Statistics and Measurement Scales*

Measurement is essentially the assigning of numbers to observations according to certain rules. The way in which the numbers are assigned to observations determines the scale measurement being used. The choice as to which statistical test can legitimately be used for data analysis rests largely on which scale of measurement has been employed. Further, the inferences that can be drawn from a study cannot, or at least should not, outrun the data being used.

- **Nominal Scale - Categorical Data:** The single property of nominal measurement is *classification* -- that is, the sorting of observations into different classes or categories; using numbers to label categories, sorting observations into these categories, and then noting their frequencies of occurrence. Nominal measurement reflects only differences in kind, *not* differences in degree or amount.

- **Ordinal Scale - Ranked Data:** The distinctive property of ordinal measurement is *order*. Differences on an ordinal scale not only reflect differences in kind (as with nominal measurement), but also differences in degree. Information regarding greater-than or less-than status is contained in ordinal data, but information as to how much greater or less than is not -- that is, no claim can be made about the amount of difference between adjacent categories.
- **Interval Scale - Measurement Data:** The distinctive property of interval measurement is *equal intervals* -- that is, scales in which distances between successive scale points are assumed to be equal. Interval data do contain information as to 'how much greater than' or 'how much less than.'
- **Ratio Scale:** The distinctive property of ratio measurement is a *true zero*. This is a special form of interval scale for which an absolute zero can be determined. Ratio data allow for such ratio comparisons as 'one measure is twice as great as another.'

Levels of Measurement			
Level	Properties	Observations Reflect	Examples
Ratio	true zero equal intervals order classification	measurable differences in total amount	weight income
Interval	equal intervals order classification	measurable differences in magnitude	Fahrenheit temp. IQ score GPA
Ordinal	order classification	differences in degree	attitudes letter grade movie ratings
Nominal	classification	differences in kind	ethnicity political affiliation

I. Descriptive Statistics

Descriptive statistics refers to a set of procedures used to describe and summarize samples of data.

Descriptive Statistics: statistical procedures used to describe and summarize samples of data

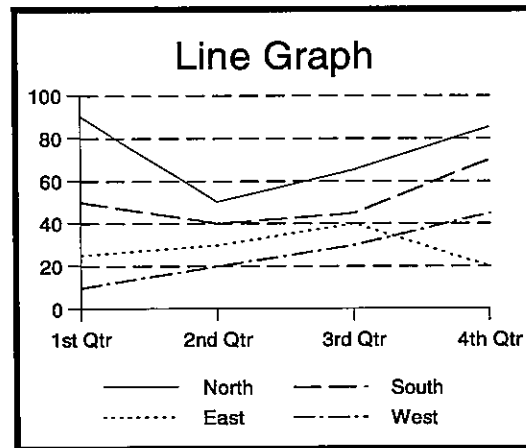
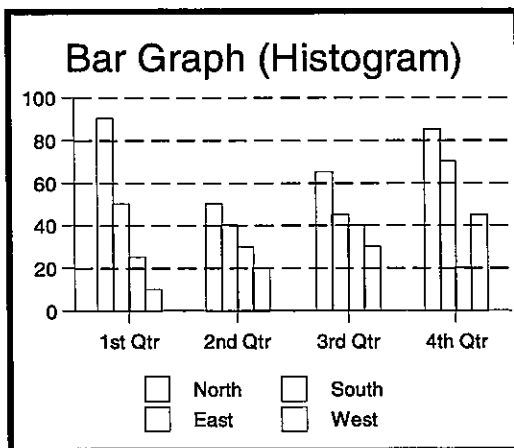
➡ Graphing Data

The process of graphing data usually begins with the creation of a distribution. To extract some meaning from the original data, the researcher begins by bringing order to the data. The first step is to form a distribution of scores. A **distribution** is the arrangement of any set of scores in order of magnitude.

Distribution: arrangement of any set of scores in order of magnitude

Frequency Distribution: a listing, in order of magnitude, of each score achieved, and the number of times each score occurred

Frequency distributions allow the researcher to see general trends more readily than does an ordered set of raw data. A frequency distribution is a listing, in order of magnitude, of each score achieved, together with the number of times that score occurred. Frequency distributions can be presented in both tabular and graphic form (e.g., bar graphs or line graphs).



➔ *Measures of Central Tendency*

Measures of central tendency are designed to give information concerning the average, or typical, score of a large number of scores - that is, which single score best represents an entire set of scores. There are three methods for obtaining a measure of the central tendency:

1. The Mean
2. The Median
3. The Mode

Measures of Central Tendency: measures that provide information about the average, or typical, score of a large number of scores; which single score (mean, median, mode) best represents an entire set of scores

◆ **The Mean**

The mean is the *arithmetic average* of all the scores. It is calculated by adding all the scores together and then dividing by the total number of scores involved. It is

Mean: arithmetic average of a set of scores

important to realize that in some cases the mean can give a very distorted picture of the average value of a distribution of scores. That is, when there are extreme scores (called outliers) the average score will give a distorted picture of the distribution of scores.

◆ **The Median**

The median is the *exact midpoint* of any distribution. The median is a much more accurate representation of central tendency than is the mean. To calculate the median, the

Median: exact midpoint of a distribution

scores must first be arranged in order of magnitude (e.g., from lowest to highest), the middle score is the median. In certain cases, the median is better than the mean as a typical or representative value for a group of scores. This happens when there are a few extreme scores (called outliers) that would strongly affect the mean but would not affect the median.

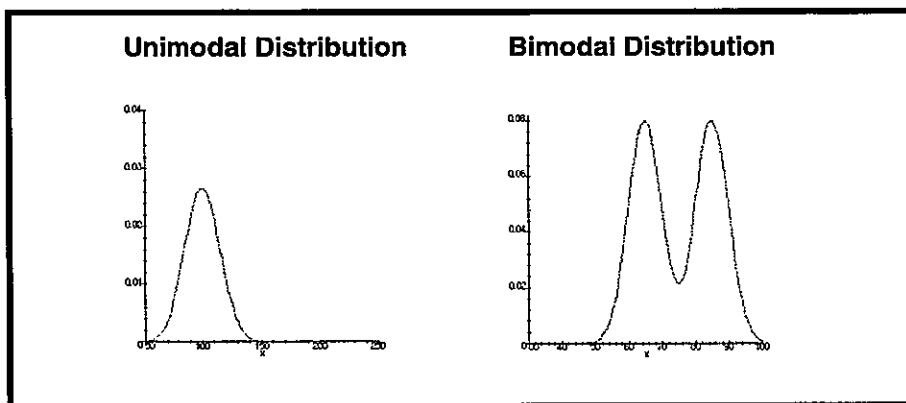
◆ **The Mode**

The mode is the *most common single number* in the distribution; in a perfectly symmetrical unimodal distribution, the mode is the same as the mean. However, when it is not the same, the mode is not really a good representative value of the distribution. A distribution having a single mode is called a **unimodal distribution**. A distribution having two or more modes is called a **bimodal distribution**.

Mode: most frequently occurring score

Unimodal Distribution: a distribution of scores with a single mode

Bimodal Distribution: a distribution of scores having two modes



Measures of Central Tendency and the Effects of the Scale of Measurement Used

- ◆ **Interval and Ratio Data:** Because with interval and ratio data the difference between scores is equal, interval and ratio data allow for the calculation of the *mean*, *median*, and *mode*.
- ◆ **Ordinal Data:** Since ordinal data provides no information regarding the distance between the scale points, calculating an ordinal mean is inappropriate and misleading. When ordinal data is used a *median* should be calculated-- that is, ordinal data can be ranked and the median is the middle score.
- ◆ **Nominal Data:** With nominal data, neither the mean nor the median can be used, since each of these measures implies comparisons of greater than and less than. The only measure of central tendency permissible for nominal data is the *mode*, the most frequently occurring score.

➔ *Measures of Variability*

A measure of central tendency (i.e., mean, median, or mode) is a single number that describes a hypothetical, typical person.

A statistic that describes the extent to which scores differ from one another in a distribution, and the extent to which they differ from the mean, is called a **measure of variability**. Just as measures of central tendency give information about

similarity among scores, measures of variability give information about how scores differ or vary.

Measures of Variability: procedures used to describe the extent to which scores differ from one another in a distribution; range, standard deviation, and variance statistics

There are three major measures of variability:

1. The Range
2. The Standard Deviation
3. The Variance

◆ **The Range**

The range is the measurement of the width or spread of an entire distribution and is found simply by calculating the difference between the highest and lowest scores. The range is a limited measure of variability. For example,

distributions can have identical means and ranges and yet vary widely in terms of other important measures of variability.

Range: measure of the width or spread of an entire distribution

◆ **The Standard Deviation**

The standard deviation is one of the most important measures of variability and it takes into account all scores in a distribution. The standard deviation is defined as a measure of the variability that indicates by how much all of

Standard Deviation: measure of the variability that indicates by how much all of the scores in a distribution typically vary from the mean

the scores in the distribution typically deviate or vary from the mean. Since the standard deviation is always calculated with reference to the mean, its calculation demands the use of interval or ratio data. The standard deviation is the typical deviation of a given distribution. The larger the value of the standard deviation, the more the scores are spread out around the mean; the smaller the value of the standard deviation, the less the scores are spread out around the mean. That is, a distribution with a small standard deviation indicates that the group being measured is homogeneous; their scores are clustered very close to the mean. A distribution with a large standard deviation indicates that the group is heterogeneous; their scores are more widely dispersed from the mean.

◆ **The Variance**

The variance of a distribution is the square of the standard deviation. It is a useful term because it reflects how much of the variability between people on one characteristic (e.g., income) can be explained by knowing where they stand on another characteristic (e.g., education).

➔ ***The Normal Curve and Z-Scores***

The normal curve is a theoretical distribution. However, many distributions of people-related measurements come close to approximating the normal curve and thus it is of crucial significance for describing data.

The **normal curve** is a unimodal frequency distribution with scores plotted on the X axis (the horizontal axis) and frequency plotted on the Y axis (the vertical axis). In a normal curve, most of the scores cluster around the middle of the distribution (where the curve is at its highest). As

<p>Variance: measures how much of the variance between people on one characteristic can be explained by knowing where they stand on another characteristic</p>
<p>Normal Curve: a theoretical distribution on a unimodal frequency distribution with scores plotted on the X axis (the horizontal axis) and frequency plotted on the Y axis (the vertical axis); most of the scores cluster around the middle of the distribution; curve is symmetrical and all three measures of central tendency (mean, median, mode) fall precisely at the middle of the distribution</p>
<p>Skewed Distribution: a distribution of scores where the majority of scores in the distribution bunch up at one end of the distribution</p>

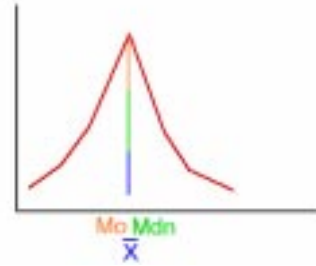
the distance from the middle increases, in either direction, there are fewer and fewer scores. The normal curve is symmetrical - both sides are mirror images of the other - and all three measures of central tendency (the mean, median, and mode) fall precisely at the same point, the exact middle of the distribution. In a **skewed distribution**, scores tend to pile up at one end or the other. The direction of skewness is indicated by the “tail” of the curve. The curve is **positively skewed** when most of the scores pile up near the bottom (the tail points toward the high or positive end). The curve is **negatively skewed** when most of the scores pile up near the top (the tail points toward the low or negative end).

Positively Skewed Distribution: distribution in which scores are concentrated near the bottom of the distribution; tail of the distribution points to the top or positive end

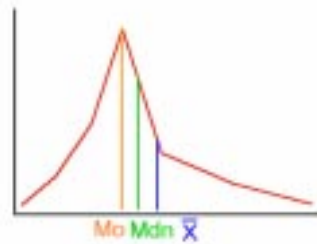
Negatively Skewed Distribution: distribution in which scores are concentrated near the top of the distribution; tail of the distribution points to the low or negative end

Standard Normal Curve (Z-Distribution): the normal curve is marked off in units of standard deviation; a normally distributed set of scaled scores whose mean is always equal to zero and whose standard deviation equals 1.00

The Normal Curve

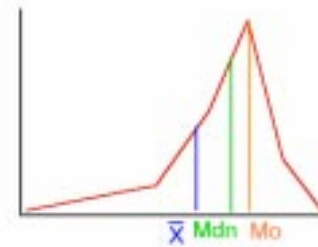


Positively Skewed Distribution



Negatively

Skewed Distribution

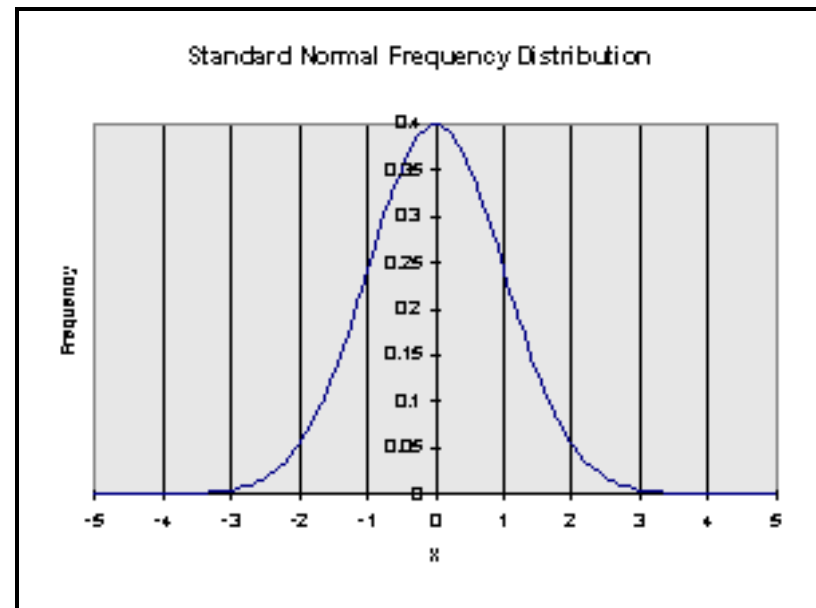


The normal curve has a constant relationship with the standard deviation. When the normal curve is marked off in units of standard deviation, a series of constant percentages under the normal curve are formed. Once the curve is plotted according to standard deviation units, it is called the **standard normal curve**, or z-distribution.

A **z-distribution** is a normally distributed set of specially scaled scores whose mean is always equal to zero and whose standard deviation must equal 1.00. Z-scores take into account both the mean of the distribution and the amount of variability, the standard deviation. Thus, z-scores can be used to assess an individual's relative performance compared to the performance of the entire group being measured. The z-score is the number of standard deviations the observed value is from the mean.

Z-Distribution: a normally distributed set of scaled scores whose mean is always equal to zero and whose standard deviation must equal 1.00; takes into account both the mean of the distribution and the amount of; the z-score is the number of standard deviations the observed value is from the mean

II. *Inferential Statistics*



The primary goal of inferential statistics is to measure a few and generalize to many. That is, observations are made of a small segment of the group, and then, from these observations, the characteristics of the entire group are inferred. Inferential statistics are procedures used to reach conclusions (generalizations) about larger populations from a small sample of data with a minimal degree of error.

Inferential Statistics: statistical procedures used to draw conclusions and inferences which are based upon, but go beyond, the numbers generated by a research study

There are usually two issues to be explored:

1. Does the mean of a sample actually reflect the mean of the larger population of interest?
2. Is a difference found between two means (e.g., between an experimental group and a control group) a real and important difference, or is it merely the result of chance?

Key Concepts of Inferential Statistics

Population (or universe): an entire group of persons, things, or events having at least one trait in common

Sample: a smaller number of observations taken from the total number making up the population; in typical applications of inferential statistics, the sample size is small relative to the population size

To make accurate predictions, the sample should be representative of the population. In a sense, a good representative sample provides the researcher with a miniature mirror with which to view the entire population. *Recall that you have seen these concepts before in the chapter on surveys.*

➡ **Measures of Relationship: Correlation**

Measures of central tendency and variability are basic descriptive statistics that tell us something about the distribution of a variable.

Measures of relationships provide information

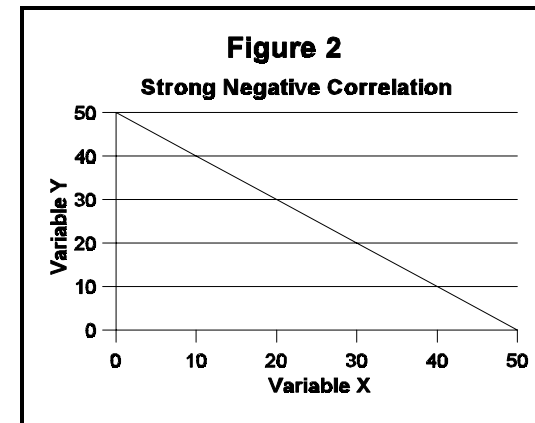
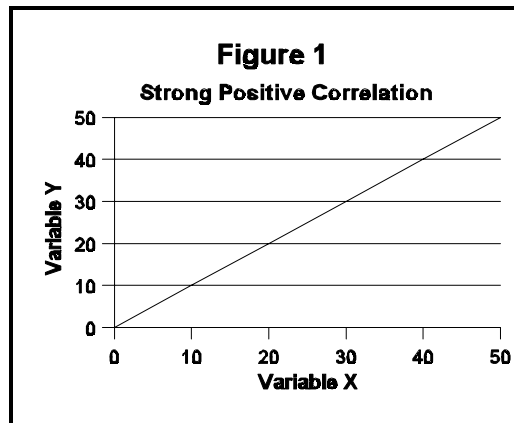
about what relationship the variable has to other variables. The association between one variable and any other variable is described as a **correlation**.

Correlation: an association between two variables; can be positive or negative

Correlation does not equal Causation

If two variables have a perfect correlation (their data points fall along a straight line), then $r = 1.0$

(Fig 1) or $r = -1.0$ (Fig 2) (“r” is the **correlation coefficient**). The positive and negative values simply show the direction of the relationship. When two variables are positively correlated, as one increases, the other also increases. When they are negatively correlated, as one increases, the other decreases. Two variables with less than a perfect correlation will have an “r value” between 0 and 1.0 or 0 and -1.0. If no relationship exists between two variables, $r = 0$. Figure 1 depicts a positive correlation between Variable X and Variable Y. That is, as Variable X increases, Variable Y also increases. Figure 2 depicts a negative correlation between the two variables. That is, as Variable X increases, Variable Y decreases.



Correlation Coefficient: a number between -1 and 1 which measures the degree to which two variables are linearly related. If there is a perfect positive linear relationship, $r = 1$ (i.e., an increase (or decrease) in one variable is associated with an increase (or decrease) in the other variable); if there is a perfect negative linear relationship, $r = -1$ (i.e., an increase (decrease) in one variable is associated with a decrease (increase) in the other variable); If $r = 0$ there is no linear relationship between the variables

Pearson's Product Moment Correlation Coefficient: Pearson's product moment correlation, usually denoted by r , is one example of a correlation coefficient; a measure of the linear association between two variables that have been measured on the interval or ratio scales (e.g., the relationship between height in inches and weight in pounds)

Regression: predicts the extent to which the value of one or more variables can be predicted by knowing the value of other variables

➔ *Measures of Relationship: Regression*

Regression analysis predicts the extent to which the value of one or more variables can be predicted by knowing the value of other variables. A **linear regression** predicts the magnitude of the expected change in variable Y given a change in variable X. A **simple linear regression** is designed to determine whether there is a linear relationship between a response variable and a possible predictor variable. A **multiple linear regression** is designed to examine the relationship between a response variable and several possible predictor variables. **Nonlinear regression** is designed to describe the relationship between a response variable and one or more explanatory variables in a non-linear fashion.

Linear Regression: predicts the magnitude of the expected change in variable Y given a change in variable X

Simple Linear Regression: designed to determine whether there is a linear relationship between a response variable and a possible predictor variable

Multiple Linear Regression: designed to examine the relationship between a response variable and several possible predictor variables

Nonlinear Regression: designed to describe the relationship between a response variable and one or more explanatory variables in a non-linear fashion

➡ *Sampling Revisited*

Sampling techniques were briefly discussed in the chapter on survey methodology. They are briefly revisited here.

◆ **Random Sampling**

Random sampling demands that each member of the entire population has an equal chance of being included and that no member of the population may be systematically excluded. It is important to note that *randomness describes the selection process*, (i.e., the procedures by which the sample is selected), and not the particular pattern of observations in the sample.

◆ **Stratified Sampling**

To obtain this kind of sampling, the researcher must know beforehand what some of the major population characteristics are and, then, deliberately select a sample that shares these same characteristics in the same proportions.

Whenever the sample differs systematically from the population of interest, a **bias** has occurred.

Bias is a constant difference, in one direction, between the mean of the sample and the mean of the population. Bias occurs when most of the sampling error loads up on one side, so that the sample means are constantly either over- or under-estimating the population mean.

Bias: a constant difference, in one direction, between the mean of the sample and the mean of the population; occurs when most of the sampling error loads up on one side, so that the sample means are constantly either over- or under-estimating the population mean

Sampling Error

Whenever a sample is selected, it must be assumed that the sample measures will not precisely match those that would be obtained if the entire population were measured. The sampling error reflects, or is an index of, the difference between the sample value and the population value.

Sampling error is not a mistake. Any sample mean should be expected to deviate from the mean of the whole population, but the deviation will hopefully be random and should not be large.

◆ Sampling Distributions

Each distribution discussed so far has been a distribution of individual scores - each point in the distribution represents a measure of a characteristic or performance of an individual. In sampling distributions, each point represents a measure of a characteristic or performance of a sample of individuals. The mean increase of a sample of U.S. adults is an example; it would be one data point in the sampling distribution of mean income. Sampling distributions are important in testing hypotheses.

III. Parameter Estimates and Hypothesis-Testing

Hypothesis-Testing in Statistical Terms

The purpose of a **hypothesis test** is to determine the likelihood that a particular sample could have originated from a population with a hypothesized characteristic.

The **null hypothesis** supplies the value about which the hypothesized sampling distribution is centered. It always makes a statement about a characteristic of the population, never about a characteristic of the sample. The null hypothesis always makes the claim about a single numerical value, never a range of values.

The **experimental hypothesis**, asserts the opposite of the null hypothesis. A decision to accept the null hypothesis (or a failure to reject the null hypothesis) implies a lack of support for the experimental or research hypothesis, and a decision to reject the null hypothesis implies support for the experimental or research hypothesis.

A **decision rule** specifies precisely when the null hypothesis should be rejected.

Criminal suspects are presumed innocent until proven guilty. Under hypothesis-testing procedures, the null hypothesis is presumed to be true until proven false. Once all the evidence has been considered, a verdict is reached, and the null hypothesis is either retained (failure to reject) or it is rejected.

Evidence for testing an hypothesis about a sample statistic is based on the relationship between the observed sample statistic and the sampling distribution of that statistic. For example, if a researcher predicts that the mean weight of rats in an experimental group is greater than the mean weight in a control group, then the statistic at issue is the difference between the two means. The experimental or research hypothesis is that the two means represent different populations and that the difference between them is dependable. The null hypothesis is that the two means come from the same population and that the difference between them would not hold up under repeated replications of the experiment. The difference between the means is compared to the sampling distribution of such differences, the mean of which is usually zero (no difference). If a difference as large as or larger than the

obtained difference is very unlikely for groups coming from the same population, then the difference will be judged to be an improbable outcome under the null hypothesis of no dependable difference and the null hypothesis will be rejected. On the other hand, if the observed difference is not so large as to be highly improbable, the null hypothesis will be accepted (or the null hypothesis will not be rejected).

An observed sample statistic will qualify as a probable outcome if the difference between its value and that of the hypothesized population statistic is small enough to be attributed to chance. For example, a sample mean will qualify as a probable outcome if the difference between its value and that of the hypothesized population mean is *small* enough to be attributed to chance. Under these circumstances, because there is no compelling reason to reject the hypothesis, the null hypothesis is tentatively accepted.

An observed sample statistic will qualify as an improbable outcome if the difference between its value and the hypothesized value is too *large* to be attributed to chance. That is, a sample mean will qualify as an improbable outcome if it deviates too far from the hypothesized mean and appears to emerge from the sparse concentration of possible sample means in either "tail" of the sampling distribution. Under these circumstances, because there are grounds for suspecting the hypothesis, the hypothesis is rejected.

The decision to reject the null hypothesis involves a degree of risk. Having rejected a null hypothesis, we can never be absolutely certain whether the decision is correct or incorrect, unless, of course, the entire population was surveyed. Even if the null hypothesis is true, there is a slight possibility that just by chance, the one observed sample mean really originates from rejection regions (the tails) of the hypothesized sample distribution, thus causing the true null hypothesis to be erroneously rejected.

IV. Error Rates

In determining the admissibility of expert opinion regarding a particular scientific technique, the court ordinarily should consider known or potential rates of error, and existence and maintenance of standards controlling the technique's operation.¹

To assess known or potential rates of error, the judiciary must be prepared to carefully and critically evaluate the methodology and underlying assumptions of proffered scientific evidence. Such an evaluation would entail examination of whether the research hypothesis was appropriately articulated and tested, whether appropriate controls were utilized, whether threats to validity were controlled for, or at least severely minimized, and so forth.

The likelihood with which a measurement device or a technological procedure leads to an incorrect classification is the **error rate**. Whereas formal testing of hypotheses usually relies on theoretical sampling distributions for estimating the likelihood that the decision based on the data is erroneous (especially Type I error), the likelihood of an incorrect classification is usually assessed in terms of error rates. Several rates should be taken into account, typically termed “true positive,” “true negative,” “false positive,” and “false negative” rates. For example, if a laboratory claims that a particular test reliably identifies the existence of a serious disease, it is necessary to consider the proportion of people with the disease who were correctly identified as having it (**true positive**) and those who were correctly identified as not having it (**true negative**). It is also important to consider the proportion of individuals without the disease who were incorrectly identified as having it (**false positive**) and the proportion of individuals with the disease who were incorrectly identified as not having it (**false negative**). False positives could lead to unnecessary further expense and painful medical interventions; false negatives could lead to further and perhaps fatal progression of the disease. It usually is

<p>False Positive Error: incorrectly classifying someone without a particular characteristic as possessing that characteristic (e.g., person does not have disease but incorrectly classified as having disease)</p> <p>True Positive Error: correctly classifying someone as possessing a particular characteristic or falling into a particular category (e.g., person has disease and is classified as having disease)</p> <p>False Negative Error: incorrectly classifying someone who has a particular characteristic as someone who does not possess that characteristic (e.g., person has disease but is incorrectly identified as not having it)</p> <p>True Negative Error: correctly classifying someone who does not possess a particular characteristic or who does not fall into a particular category (e.g., person does not have disease and is classified as not having the disease)</p>
--

essential to examine both types of erroneous classification rates; if proffered evidence does not include both error rates, it is likely to be of little value.

Error rates are generally stated as percentages or proportions. In the above case, for example, the data might have been drawn from people who visited their physicians because of certain bothersome symptoms, and when the physicians conducted the diagnostic test, the results for 104 patients might have been:

	Actually has disease	Actually free of disease	Total
Test says has disease	90	10	100
Test says free of disease	2	2	4
Total	92	12	104

The true positive rate is .98 (90/92), with only two diseased patients mis-diagnosed (2/92, a false negative rate of .02). There were 12 patients without the disease, 10 of whom were mis-diagnosed as having the disease for a false positive rate of .83 (10/12).

This example illustrates two points. First, the rate of correct classifications has to be compared to the rates of both false positive and false negative classifications. The relative importance of the two types of errors will depend on what they lead to—false security, expensive or painful further intervention, and so on. Second, although proportions and percentages are very useful modes of presenting data, sometimes the raw numbers underlying the percentages are equally important. In the example above, only 12 of the 104 patients were actually free of the disease, and that base of 12 is too small to draw firm conclusions about the false positive

rate. We would be much more confident if the number of disease-free patients who were tested was larger. In general, if we were told that 50% of people held a certain opinion, we would want to know if the reference was 50% of 2 people or 50% of 2,000.

↳ *Type I and Type II Errors*

The decision to reject the null hypothesis is based on probabilities rather than on certainties. The decision is made without direct knowledge of the true state of affairs in the population.

There are two possible decisions: (1) reject the null hypothesis, or (2) fail to reject (accept) the null hypothesis. There are also two possibilities that may be true in the population: (1) the null hypothesis is true, or (2) the experimental hypothesis is true. Thus, there are two kinds of correct decisions and two kinds of errors.

Type I Error: when the researcher rejects the null hypothesis but the null hypothesis is actually true (e.g., the researcher claims that there is a causal relationship between variable A and variable B when, in fact, there is not)

Type II Error: when the researcher fails to reject the null hypothesis (i.e., accepts the null hypothesis) when in actuality the experimental hypothesis is true (e.g., the researcher claims there is no causal relationship between variable A and variable B when, in fact, there is one)

		Researcher's Decision	
		Reject the null	Failure to reject
hypothesis	the null hypothesis	<hr/>	
True state of nature	Null hypo. is true	Type I Error	Correct decision
	Null hypo. is false	Correct decision	Type II Error

Consider the decision made by a juror in a criminal trial. As is the case with statistics, a decision must be made on the basis of evidence: Is the defendant innocent or guilty? However, the decision is the juror's and does not necessarily reflect the true state of affairs that the person really is innocent or guilty. Assume the null hypothesis is that the defendant is innocent. The *rejection* of the null hypothesis is to decide, based upon the evidence, that the defendant is guilty. *Acceptance* of the null hypothesis is to decide, based upon the evidence, that the defendant is innocent.

		Juror's Decision		the null
		Reject the null (find guilty)	Failure to reject (find innocent)	
True State	Null is true (innocent)	Type I Error	Correct Decision	<hr/>
	Null is false (guilty)	Correct Decision	Type II Error	

Most scientists begin with the assumption that the phenomenon they are studying does not cause the effect they expect -- the null hypothesis. In other words, the standard method of science is to presume 'innocence' and only with strong proof reject that assumption.

Scientific conventions have developed regarding the strength of this presumption; that is, how much evidence is needed before rejecting the null hypothesis and accepting an alternative hypothesis that the experimental manipulation caused the observed effect (this will be discussed further in this chapter). It is important to realize, however, that an attempt to decrease one type of error results in an increased likelihood of making the other type of error.

To what extent do judges around the country find the concept of error rate a useful criterion for critically evaluating scientific evidence?



All judges in the survey sample, even those not in FRE/*Daubert* states, were asked how useful they thought the concept of error rate is for admissibility decision-making (N=400).

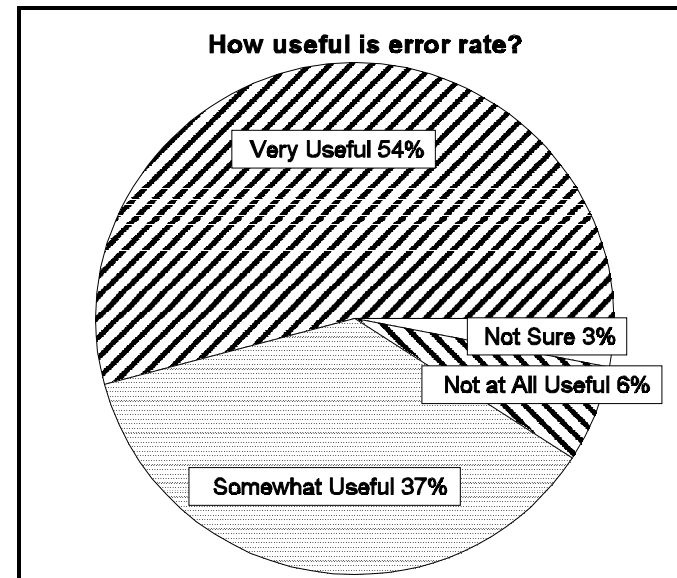
The majority (91%) indicated that a consideration of error rate was a useful when determining the admissibility of scientific evidence, with 54% of those judges rating error rate as “very useful.”

Focusing just on responses from judges in states which follow the FRE/*Daubert* standards, the vast majority of judges rated error rate as a useful guideline for evaluating the admissibility of scientific evidence.

Even though the vast majority of judges rated error rate as a useful guide, the results of the survey indicate that judges do not fully understand the scientific meaning of error rates and that, as a result, they are unsure how to utilize the concept as a guideline for determining admissibility.

When asked a question about how they would apply the concept of error rate, the majority of judges expressed some hesitancy or uncertainty. In order for a response to be coded as “judge understands concept” the response had to include reference to an evaluation of the variety of sources of error, or refer to a number or percent of instances in which the classification procedure was mis-classified. From the answers provided, the researchers could only infer a true understanding of the concept in 4% of the responses (N=400).

In 86% of the responses the judges’ understanding of the concept was questionable. In 10% of the responses, the judge relied solely upon a low error vs. high error heuristic (or rule of thumb) when explaining how the concept of error rate is applied to admissibility (i.e., if there is a high rate of error then the judge is more likely to exclude the evidence than if there is a low rate of error).



➡ *Confidence Level*

An alternative indicator of the probability of a Type I error is the confidence level. It specifies the range of values around the empirically obtained result within which the “true” or population value is likely to lie. Confidence levels are frequently reported in sample surveys. For example, it might be reported that the 95% confidence level for an obtained percentage of 20% is 20% plus or minus 3%. The higher the confidence level, the lower the probability of a Type I error, but the broader the range of values within which the “true” or population value might actually lie.

Confidence Level: specifies a range of values around the empirically obtained result within which the “true” or population value is likely to lie

Both types of error are important. For example, in a toxic tort case, a Type I error could mean that the frequency of occurrence of a symptom among workers could be accepted as indicating that the symptom was caused by a toxic substance found in the plant environment, whereas in fact that frequency of occurrence was just at the outer extreme of random fluctuation and was not a reflection of a causal link. The firm might improperly be held accountable. But with a Type II error, the frequency of occurrence of the symptom would be taken as well within the normal range of fluctuation, and no causal link between substance and symptom frequency would be inferred. The firm could be erroneously exonerated. The less likely a Type I error, the more likely a Type II error.

➡ *Significance and P-Values*

In order to decide whether the difference in the observed score differs significantly from the null hypothesis, a standard or criterion

for deciding whether to accept or reject the null hypothesis must be established. Statisticians typically use two levels of significance: .05 and .01. These levels have been established by convention. When a significance level of $p < .05$ is chosen, the decision rule is that the null hypothesis will be rejected if the data are so unlikely that they could have occurred by chance less than 5 times out of 100. If a significance level of $p < .01$ is chosen, the probability of the observed value occurring by chance is less than 1 in 100.

The odds of making a Type I error (rejecting the null hypothesis when it is true) are exactly equal to the value chosen for the significance level. That is, if a researcher has chosen a significance value of .05, the probability of a Type I error is .05 -- 5 times out of 100 (5%) the researcher will reject the null hypothesis when it is true. That is, there will be 5 times out of 100 when extreme differences are due to chance and not to some experimental manipulation.

Can the odds of making a Type I error be minimized by choosing a more extreme significance level (e.g., $p < .01$)? Yes, but there is a trade-off: an increased likelihood of making a Type II error (failure to reject an hypothesis when it is false) -- the researcher concludes that the results were caused by chance and not by the experimental manipulation.

Statistical Significance and Legal Significance

The scientist's concept of statistical error does not translate directly into the judge's concept of legal error. It cannot be said, therefore, that a study that is statistically significant at the .05 level of confidence will lead judges, if they admit the evidence, to make only 5 errors (Type I errors) out of 100. There is no true correspondence between statistical confidence and legal burdens of proof.

Limitations of moving from statistical significance to legal significance:

- a confidence level is a statistical statement and does not incorporate the variety of factors that judges must take into account in making a decision
- most scientific research examines the general relationship between variables, while trial courts are usually concerned with specific effects on specific individuals

➡ ***Statistical Significance and Importance***

The significance of a finding (the probability of a Type I error) does not have a clear relationship to the importance of the finding, either. A small difference, or a small correlation, could still be highly significant statistically, if the sample were large enough. A finding of small magnitude would still be reliably replicated on repeated investigations, if large portions of the population were included in each investigation – but although dependable, the finding might not have practical or theoretical importance.

Common Problems with the Use of Statistical Evidence in Court

???

Questions to consider when evaluating scientific evidence

- ◆ Were the appropriate statistical tests conducted?
- ◆ Were proper statistical inferences drawn from the data?
- ◆ Were there methodological problems that undermined the scientific validity, reliability, or relevancy of the statistical results?

◆ **Statistics in court are not presented in their natural form**
 In scientific presentations, statistics are rarely presented in a single, complete presentation (i.e., one side presents statistical evidence that is challenged on cross-examination and then, at a later point in the trial, the other side proffers opposing statistical conclusions). In court, statistics are often presented in graphic form, and there is rarely a detailed discussion of the statistical techniques and models used, their assumptions and shortcomings.

◆ **Improper inferences are drawn**
 When statistics are presented in court, improper inferences are often drawn about what the data mean and what conclusions can be drawn. This problem typically occurs in three ways: (1) by extrapolating results of a statistical analysis to a population that is different from the population defined in the study; (2) by inferring, within the correct population, something beyond what is statistically correct given the available data and analysis; and (3) by misinterpreting statistical significance and the burden of proof.

◆ **Improper methodologies used**
 Methodological problems that undermine the scientific validity or relevance of statistical results occur at many stages of the research: study design, data collection, and data analysis.

Iancu, C.A., and Steitz, P.W. (1997). "Guide to Statistics," *Expert Evidence: A Practitioner's Guide to Law, Science, and the FJC Manual*. St. Paul, Minnesota: West, pp. 298-310.

Endnotes:

1. *Daubert vs. Merrell Dow Pharmaceuticals* 509 U.S. 579, 113 S.Ct., 2786 at 508.
-

GLOSSARY

bias	a constant difference, in one direction, between the mean of the sample and the mean of the population; occurs when most of the sampling error loads up on one side, so that the sample means are constantly either over- or under-estimating the population mean
bimodal distribution	a distribution of scores with two modal scores (two commonly occurring scores)
confidence level	specifies a range of values around the empirically obtained result within which the “true” or population value is likely to lie
correlation	an association between two variables; can be positive or negative; correlation does not equal causation
correlation coefficient	a number between -1 and 1 which measures the degree to which two variables are linearly related; if there is a perfect positive linear relationship, $r = 1$ (i.e., an increase in one variable is associated with an increase (or decrease) in the other variable); if there is a perfect negative linear relationship, $r = -1$ (i.e., an increase (decrease) in one variable is associated with a decrease (increase) in the other variable); if $r = 0$ there is no linear relationship between the variables
decision rule	specifies precisely when the null hypothesis should be rejected
descriptive statistics	statistics that summarize, describe, and make understandable the numbers generated in a research study
distribution	the arrangement of any set of scores or values in order of magnitude
error rate	the likelihood with which a measurement device or a technological procedure leads to an incorrect classification
false negative error	incorrectly classifying someone who has a particular characteristic as someone who does not possess that characteristic

(e.g., person has disease, but is incorrectly identified as not having it)

false positive error	incorrectly classifying someone without a particular characteristic as possessing that characteristic (e.g., person does not have disease, but incorrectly classified as having disease)
frequency distribution	a listing, or order of magnitude, of each score and how many times that score occurred
inferential statistics	statistics used to draw conclusions and inferences which are based upon, but go beyond, the numbers generated by a research study
interval scale	a unit of measurement characterized by equal intervals; measures differences in amount (e.g., I.Q. score)
linear regression	predicts the magnitude of the expected change in variable Y given a change in variable X
mean	the arithmetic average of all the scores; calculated by adding all the scores together and then dividing by the total number of scores involved
measures of central tendency	measures that provide information about the average, or typical, score of a large number of scores; which single score (mean, median, mode) best represents an entire set of scores
measures of variability	procedures used to describe the extent to which scores differ from one another in a distribution; range, standard deviation, and variance statistics
median	the exact midpoint of any distribution; much more accurate representation of central tendency than the mean; to calculate the median, the scores must first be arranged in order of magnitude (e.g., from lowest to highest), the middle score is the median
mode	a measure of central tendency; the most common single number in the distribution; in a perfectly symmetrical unimodal distribution, the mode is the same as the mean; when it is not the same, the mode is not really a good representative value of the distribution
multiple linear regression	designed to examine the relationship between a response variable and several possible predictor variables
negatively skewed distribution	distribution in which scores are concentrated near the top of the distribution; tail of the distribution points to the low or negative end
nominal scale	a unit of measurement based on classification; measures differences in kind (e.g., ethnicity)

nonlinear regression	designed to describe the relationship between a response variable and one or more explanatory variables in a non-linear fashion
normal curve	a theoretical distribution; a unimodal frequency distribution with scores plotted on the X axis (the horizontal axis) and frequency plotted on the Y axis (the vertical axis); most of the scores cluster around the middle of the distribution; curve is symmetrical and all three measures of central tendency (mean, median, mode) fall precisely at the middle of the distribution
ordinal scale	unit of measurement characterized by order and classification; measures differences in degree (e.g., attitudes)
pearson's product moment correlation coefficient	a measure of the linear association between two variables that have been measured on interval or ration scales (e.g., the relationship between height in inches and weight in pounds); usually denoted by r , is an example of a correlation coefficient
population	an entire group of persons, things, or events having at least one trait in common; the larger group of all people of interest from which the sample is selected
positively skewed distribution	distribution in which scores are concentrated near the bottom of the distribution; tail of the distribution points to the top or positive end
range	a measure of variability; the width or spread of an entire distribution; found simply by calculating the difference between the highest and lowest scores
regression	predicts the extent to which the value of one or more variables can be predicted by knowing the value of other variables
ratio scale	a unit of measurement characterized by a true zero and equal intervals; measures differences in total amount (e.g., income)
sample	a smaller number of observations taken from the total number making up the population; in typical applications of inferential statistics, the sample size is small relative to the population size
simple linear regression	designed to determine whether there is a linear relationship between a response variable and a possible predictor variable
skewed distribution	a distribution of scores where the majority of scores in the distribution bunch up at one end of the distribution
standard deviation	a measure of variability; a measure of the variability that indicates by how much all of the scores in the distribution typically deviate or vary from the mean
standard normal curve	the normal curve is marked off in units of standard deviation; a normally distributed set of scaled scores whose mean is always

equal to zero and whose standard deviation equals 1.00

true positive error	correctly classifying someone as possessing a particular characteristic or falling into a particular category (e.g., person has disease and is classified as having disease)
true negative error	correctly classifying someone who does not possess a particular characteristic or who does not fall into a particular category (e.g., person does not have disease, and is classified as not having the disease)
type I error	when the researcher rejects the null hypothesis when the null hypothesis is true
type II error	when the researcher fails to reject the null hypothesis when the null hypothesis is false
unimodal distribution	a distribution of scores with a single modal score
variance	measures how much of the variance between people on one characteristic can be explained by where they stand on another characteristic

SUGGESTED READINGS:

Barnes, D.W. (1983). *Statistics as Proof: Fundamentals of Quantitative Evidence*. Boston: Little, Brown and Company.

DeVore, J. and Peck, R. (1997). *Statistics: The Exploration and Analysis of Data, 3rd Edition*. San Francisco, CA: Duxbury.

Hagg, R.V. and Craig, A. T. (1995). *Introduction to Statistics, 5th Edition*. Englewood Cliffs, N.J.: Prentice Hall.

Iancu, C.A., and Steitz, P.W. (1997). "Guide to Statistics." In *Expert Evidence: A Practitioner's Guide to Law, Science, and the FJC Manual*. Washington, D.C.: Government Printing Office. pgs. 298-310.

Kaye, D.H., and Freedman, D.A. (1994). "Reference Guide to Statistics." In the *Federal Judicial Center's Reference Manual on Scientific Evidence*. St. Paul, Minnesota: West., pgs. 331-414.

Saville, D.J. and Wood, G.R. (1996). *Statistical Methods: A Primer*. New York: Springer

Judge's Notes:

CHAPTER 10

Psychological and Psychiatric Evidence: A Brief Overview

Learning Objectives for Chapter 10

Upon completion of this chapter, the reader should be able to:

- ◆ Discuss the differences between clinical and empirical methods in psychological and psychiatric research;
- ◆ Identify and discuss some of the strengths, weaknesses and criticisms of the Diagnostic and Statistical Manual of Mental Disorders (DSM);
- ◆ Discuss the differences between objective and projective personality tests and why this distinction is important;
- ◆ Identify elements of expertise, including differences in degrees, licensure, and professional affiliations;
- ◆ Identify different theoretical paradigms within psychology and discuss why recognition of these paradigms might be important; and
- ◆ Critically evaluate different types of psychological and psychiatric testimony, including prediction of future behavior and psychological syndromes.

The judges who participated in the national survey clearly had experience with a variety of psychological and psychiatric evidence testimony in their courtrooms. However, these judges also expressed considerable difficulty in dealing with such testimony. Typical comments provided when discussing psychological and psychiatric evidence revealed concerns about its “subjective nature,” its “lack of scientific rigor,” as well as concerns over the “wide variety and differing backgrounds of experts who testify about psychological and psychiatric matters.” Because most of the judges surveyed had at least some experience with psychological or psychiatric evidence in the courtrooms, and found this type of evidence to be problematic, a chapter on psychological and psychiatric evidence is included in this

Deskbook. The purpose of this chapter is not only to highlight some of the issues with respect to psychological and psychiatric evidence, but also to illustrate the concepts discussed in previous chapters.

Psychological and psychiatric evidence has become a significant part of many trials in the past few decades, and since courts continue to be faced with legal issues whose resolution may be informed by social and behavioral science evidence, it will probably continue to be so. Social and behavioral science evidence can take several forms, including the presentation of the results of empirical research or clinical assessments conducted for purposes of trial. For example, social and behavioral science evidence has been used by the court to:

- discover if an individual understood the information conveyed in the informed consent process;
- determine community beliefs and standards for obscenity cases;
- focus on the operation of the trial process (e.g., jury's understanding of instructions); and
- provide information about how parents respond to their child in the home for post-custody placement.

The value of social and behavioral science evidence is not limited to understanding mentally ill persons, dangerous persons, or other deviant populations of concern to behavioral, clinical science. [Social and behavioral science] evidence can be used to understand the behavior of all persons.¹



A national snapshot of judges' experience with psychological and psychiatric evidence ...

The judges surveyed had a variety of experiences with psychological and psychiatric evidence:

- 47% of the judges surveyed had at least "some" courtroom experience with *psychological syndrome evidence*
- 44% of the judges surveyed had at least "some" courtroom experience with *psychological evidence about insanity*
- 26% of the judges surveyed had at least "some" courtroom experience with *psychological evidence about the prediction of dangerousness*



A national snapshot of judge's experience with psychological and psychiatric evidence (Continued) ...

Judges surveyed were also asked about their experience with testimony about the psychiatric or psychological disorders contained within the *Diagnostic and Statistical Manual of Mental Disorders* (DSM):

- 17% of judges surveyed had encountered testimony about *bipolar disorder*
- 17% of judges surveyed had encountered testimony about *schizophrenia*
- 16% of judges surveyed had encountered testimony about a *personality disorder*
- 6% of judges surveyed had encountered testimony about *depression* generally
- 5% of judges surveyed had encountered testimony about *post-traumatic stress disorder*
- 4% of judges surveyed had encountered testimony about *attention-deficit hyperactivity disorder*

N=312 (Judges were given an option of completing these questions over the telephone or via mail; sample size for these questions does not equal 400 as not all mail-out questionnaires were returned).

➔ *The Social Science Challenge*

Some critics argue that the fields of psychiatry and psychology are merely pseudo or soft sciences (i.e., not derived from a strict scientific methodology involving the development of hypotheses that can be tested through objective, quantifiable observation). These critics also point to the diversity of psychiatric or psychological theories on the same topics and cite conflicting studies and instances of mis-diagnosis. However, even hard sciences, such as physics, incorporate a body of theories that are debatable or unverified. And even given the underlying uncertainties and discrepancies within the psychiatric communities, psychiatrists and psychologists, through their education and experiences, acquire special information and skills that are beyond that of the lay community to better understand and interpret human behavior.²

The subject matter of the social sciences is the human animal. As a result, many of the hypotheses of the social sciences are difficult to test. Experiments and direct observation, for example, can be complicated because of the complexity of the subject matter and because of ethical considerations. There is greater difficulty in controlling researchers' biases, and observation is more likely to affect what is being observed. Moreover, it is often difficult to account for, control, and operationalize all of the complex variables that go into human

behavior and choices. Moreover, ethical constraints on withholding treatment often make the random assignment of subjects to treatment and control conditions difficult. Despite the inherent challenges of social science research, however, it is important to recognize the agreed upon practices and standards in the relevant disciplines and to evaluate the evidence in accordance with those standards. It is also important to make a distinction between the therapeutic utility of a psychological or psychiatric concept and the legal utility of that concept.

➡ *Clinical vs. Empirical Methods*

Social and behavioral science disciplines may use both empirical and clinical methods to study human behavior. For instance, some experts who conduct empirical research are clinicians as well and many clinical methods include empirical information. **Empirical research** refers generally to methodologically sound studies, regardless of whether the researchers are primarily clinicians or empirical researchers. The **clinical method** is based on observation and relies primarily on personal examination, history-taking and testing. Clinicians make diagnoses using the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, 4th ed. (DSM-IV). Clinicians also make assessments using or incorporating information from multiple sources, including medical examinations, clinical interviews, psychological testing, and analyses of psycho-social histories.

Empirical Research: systematic gathering of information and study of problems in accordance with the agreed upon methodological practices of experimental, quasi-experimental or qualitative research

Clinical Method: based upon observation, history-taking, testing, and diagnosis; relies upon clinical interviews, psychological testing, psycho-social histories, and medical exams

[With respect to social science evidence,] it is just as important for judges to understand and exercise their gate-keeping authority over both empirical and clinical evidence and testimony.³

◆ **The Clinical Interview**

The basic clinical interview usually includes a systematic history-taking and a mental status examination (MSE). The MSE is an evaluation of the individual's current functioning as revealed through observations and responses to interview questions. The MSE, and the basic clinical interview, is often unstructured and can be highly subjective, frequently influenced by the interview setting and the relationship between the interviewer and the interviewee. There is a concern that a clinician's subjective impressions and prior experiences might inappropriately direct the interview process and its findings. However, despite these limitations, clinical interviews can provide important information about the mental and social functioning of an individual.

Things to take into account when evaluating the findings of a clinical interview

- The interview format and the purpose for which it was conducted (e.g., some clinical interviews may be open-ended while others may be semi-structured in format; clinical interviews may be conducted with the purpose of obtaining information about the issue before the court);
- The place, date, and duration of the interview;
- The possible need to interview other persons to obtain a more complete understanding of the subject's mental status;
- The possible limitations to the interviewer's knowledge of the relevant information; and
- The possible strengths associated with the interviewer's knowledge of the relevant information.

◆ **Diagnostic and Statistical Manual of Mental Disorders (DSM)**

The DSM, including the current DSM-IV, was developed by the American Psychiatric Association primarily to provide mental health professionals with a diagnostic classification system of mental disorders. The DSM includes the criteria necessary for diagnosing organic brain disorders, personality disorders, childhood disorders, relational disorders, associative disorders, and the like. Diagnostic criteria were developed following a systematic review of published literature, re-analysis of previously collected data, and issue-focused field trials

evaluating more than 6,000 subjects.⁴ Under the DSM-IV, an individual is diagnosed using five axes, each providing different information about the history of the individual's mental condition and functioning.

DSM-IV Multiaxial Diagnosis	
Axis I: Clinical Disorders and Other Conditions that May be a Focus of Clinical Attention	<ul style="list-style-type: none">records the individual's various mental disorders or conditions with the principal diagnosis first
Axis II: Personality Disorders and Mental Retardation	<ul style="list-style-type: none">records the individual's various personality disorders or mental retardation with principal diagnosis first
Axis III: General Medical Conditions	<ul style="list-style-type: none">records general medical conditions relevant to a better understanding of individual's mental disorder(s)
Axis IV: Psycho-Social and Environmental Problems	<ul style="list-style-type: none">records psycho-social and environmental problems that may affect mental disorder(s) such as history of sexual abuse, inadequate finances or family problems
Axis V: Global Assessment of Functioning	<ul style="list-style-type: none">records clinician's determination of the individual's overall level of functioning on a scale of 1-100 (e.g., 1 to 10 indicates danger to self and others; 91-100 indicates superior functioning)
See DSM-IV at 25-35.	

The *strength* of the DSM-IV is that it provides a standard, comprehensive diagnostic tool. It reflects a consensus about the diagnosis of mental conditions and categorizes information in a systematic manner. DSM-IV's major *limitation* is that it also reflects a consensus opinion at the time of publication.

While the DSM-IV provides an operational classification and official list of what is and is not considered a psychiatric condition, and there are clinically distinct and medically significant conditions classified within the DSM-IV, there are also intense controversies surrounding some of the lists of symptoms categorized as mental disorders (e.g., multiple personality disorder and attention-deficit hyperactivity disorder).

Common Criticisms of the DSM

- The DSM medicalizes normal every-day problems in living (e.g., with the inclusion of adjustment disorders and relational problems).
- The DSM reflects a cultural and social bias, and consequently categories may be in a state of flux dependent upon the social climate (e.g., in the DSM-II homosexuality was listed as a mental disorder whereas in the current manual it is no longer included).
- Many of the categories of mental disorder classified in the DSM lack reliability, such that two psychiatrists or clinical psychologists can diagnose the same client with the same presenting problems as having two different disorders.
- Critics argue that many of the categories (e.g., the personality disorders) are reflective of an underlying Freudian-based theory and are therefore inherently unfalsifiable.

Before going any further, stop and reflect ...

- ◆ Why do you think that the DSM is not 'scientific' according to some critics? Why do others claim it is scientific?
- ◆ Does your understanding of philosophy of science help to answer these questions?

◆ **Standardized Psychological Tests**

Standardized psychological tests allow clinicians to compare various aspects of an individual's behavior to those of a normative group with known characteristics. That is, they have been statistically normed or standardized, allowing clinicians to assess how a patient performs compared to normal individuals or patients with similar disorders. Standardized psychological tests are less susceptible to the subjectivity of the individual evaluator. However, a disadvantage of psychological tests is that they may be used inappropriately to draw a legal conclusion or to label a person on the basis of a test score alone.

Standardized Tests: tests that have been statistically normed or standardized, permitting clinicians to compare scores of test subjects with the scores of normal individuals or patients with similar disorders



Questions to consider when evaluating the results of psychological tests and how they are interpreted ...

- ◆ Was the appropriate test used for a specific individual (e.g., certain tests are more appropriate for children than adults)?
- ◆ Is the test being used appropriately for the specific legal issue at hand; and
- ◆ Can the data obtained be applied properly to the specific subject of that particular legal inquiry?

Objective vs. Projective Measures of Personality

◆ **Objective Personality Tests**

Objective measures generally consist of true-false or check the best answer questions. The aim of objective personality tests varies. *The term 'objective' means that the test scoring is objective, not that the interpretation of the test results is objective.*

The **Minnesota Multiphasic Personality Inventory-2 (MMPI-2)** is the most widely used and researched personality test. It is designed to measure major personality patterns and psychopathology on a variety of clinical and validity scales. The validity scales measure lying, defensiveness, and exaggeration. The norms for the MMPI-2 are based on the profiles of thousands of 'normal' people and a smaller group of psychiatric patients. The MMPI-A is a variation of the MMPI-2 for use with adolescents between the ages of 14 and 18. This version is shorter, contains a section that relates only to adolescent problems, and scores the test results by comparison only to adolescent averages.

◆ **Projective Personality Tests**

The fundamental assumption underlying projective personality tests is that a person's unconscious motivates and directs daily thoughts and behavior. To uncover those unconscious motivations, clinicians provide ambiguous stimuli to which the individual can provide responses that might reflect his unconscious. The assumption is that the individual 'projects' unconscious feelings, drives, and motives onto the ambiguous stimuli.

A widely known projective test is the **Rorschach Inkblot Test**. In the Rorschach, inkblots are shown, one at a time, to the individual being examined who then tells the clinician what he sees in the design. A detailed report of the response is made for later interpretation. Although norms are available for responses, skilled interpretation and good clinical judgment are necessary to place a subject's responses in a meaningful context.

Use and Misuse of the MMPI and MMPI-2

The major problem with the MMPI and MMPI-2 is that mental health professionals may give testimony that is far beyond what the test can assess. For example, professionals may interpret MMPI scores and draw conclusions about how an individual's MMPI is in some fashion typical or not typical of sex offenders. However, neither the MMPI nor the MMPI-2 have scales that determine whether or not an individual is a pedophile or a sex offender. There has been research on MMPI scale elevations in sex offenders, but there is no typical sex offender MMPI profile. Although mean profiles often involve scales 4 and 8, with 9 and 2 also sometimes elevated, these elevations were also found in murderers, arsonists, and property offenders in a forensic psychiatric facility. Elevations on scale 4 are also common in prison populations. Thus, the MMPI cannot establish whether an individual is a sex offender.

Nevertheless, the MMPI and MMPI-2 can provide information about personality characteristics that might be useful in the overall analysis of a case. It is most useful when there are allegations of highly deviant, low base rate, or sadistic abuse which the individual denies, and a valid, within 'normal limits MMPI' suggests the absence of psychopathology. In such cases, the clinician must pay attention to the discrepancy.

➡ *Who Has the Appropriate Expertise?*

Usually, psychiatrists (physicians specializing in mental disorders) and doctors of clinical psychology are qualified as experts on psychological matters, but courts have also permitted (and precluded) a wide range of other individuals to testify about psychological issues. Indeed, several courts have admitted testimony from social workers, police, and even lay witnesses about a variety of psychological issues. Other courts, however, have not been so willing to admit such testimony. Generally, courts should consider a proposed social and behavioral science expert's education and training, licensing and certification, professional work history, publications, and status in the profession.

A Matter of Degree

M.D. in Psychiatry: involves medical training, internship and residency with a specialization in psychiatry. As medical doctors, psychiatrists can prescribe medication.

Ph.D. (Doctor of Philosophy) in Psychology: involves training in research and theory within specific sub-disciplines and requires conducting original dissertation research (e.g., a Ph.D. in developmental psychology focuses on the emotional, physical, and intellectual changes over the life span; a Ph.D. in social psychology studies how other people affect individual behavior and thoughts, especially through social interaction).

A Matter of Degree

Ph.D. (Doctor of Philosophy) in Clinical Psychology: training can be described as the '*scientist-practitioner*' model. In addition to training in research and theory, clinical psychologists complete several clinical practica, a research dissertation, a pre-doctoral internship, and often a year of post-doctoral supervision. Despite lobbying efforts, clinical psychologists are not currently permitted to prescribe medication.

Psy.D. (Doctor of Psychology): training can be described as the '*practitioner model*.' This degree does not focus on research. Rather, the focus is on extensive training in assessment and treatment of psychological disorders.

Ed.D. (Doctor of Education): often specializing in counseling or educational psychology; Involves training in theory and scholarly consideration of a behavioral or educational problem or issue (e.g., through practical educational experiences, directed field experiences).

M.A., M.S., M.Ed, M.S.W., etc.: in many states persons with a master's degree may perform some aspects of psychological practice. However, this is often required to be conducted under the supervision of a doctoral level professional.

➡ *Licensure*

Licensure of psychiatrists, psychologists, and other mental health professionals is performed at the state level, with requirements varying from jurisdiction to jurisdiction. Licensure, which is state permission to perform certain functions (e.g., diagnosis, treatment) is distinct from **certification**, which focuses not on the function performed but on the use of a particular professional title (e.g., “psychologist”), and limits its use to individuals who have met specified standards for education, experience, and examination of performance.

Licensure: state granted permission to perform certain functions related to diagnosis, treatment, or dispensing of medication; legal permission to perform a specified function; typically requires an educational degree requirement and a specified number of practical training hours

Certification: a formal designation of someone to use a professional title based upon education, training, and experience; a statement authorizing an individual to officially practice in a certain profession

◆ **Board Certification in a Speciality**

Both the American Academy of Psychiatry and Law and the American Psychology-Law Society created boards for forensic certification in 1976. The American Board of Forensic Psychiatry governs board certification for psychiatrists and the American Board of Forensic

Psychology governs certification for psychologists. While board certification does not guarantee competence in all sub-disciplines of the relevant field, it does ensure that expert witnesses have demonstrated an advanced level of knowledge, skill, experience, training, and education, to the satisfaction of a well-credentialed board of peers.

◆ **Membership in Professional Associations**

Active membership, even elected office, in professional associations may not necessarily correspond to any particular degree of knowledge, skill, experience, training, or education. In fact, the American Psychological Association considers it an ethical violation for psychologists to represent their general membership in that organization as a form of certification of expertise. Exceptionally active involvement in professional associations may indicate professional scientific recognition within one's field, or it may signify little more than a reputation for administrative ability.

<p style="text-align: center;">Some Professional Associations</p> <p>For psychiatrists: American Psychiatric Association American Board of Psychiatry and Neurology American Psychoanalytic Association American Academy of Psychiatry and Law</p> <p>For psychologists: American Psychological Association American Board of Professional Psychology American Psychology-Law Society American Psychological Society (focuses on scientific rather than practice or applied interests)</p> <p>Most professional organizations have various divisions which reflect areas of specialization within the general field.</p>

<p>??? <i>Questions to consider when evaluating expert qualifications ...</i></p> <p>◆ Employment history</p> <ul style="list-style-type: none">• Does the expert actually practice in the field he intends to testify about? <p>◆ Educational history</p> <ul style="list-style-type: none">• Does the expert have sufficient education and training in the field?• What was the nature of his education (e.g., research, clinical training and supervision, etc.)?• Is the expert's educational history relevant to the field he intends to testify about?
--

??? *Questions to consider when evaluating expert qualifications (continued) ...*

- ◆ **Licensing and board certifications**
 - Is the expert appropriately certified or licensed within the profession?
- ◆ **Memberships in national associations**
 - What professional associations does the expert belong to and why?
 - To what extent are these professional memberships relevant to the expert's training and expertise?
 - To what extent are these professional memberships relevant to the issue at hand?
- ◆ **Areas of expertise/specialty**
 - What is the expert's particular area of expertise?
 - Is this area of expertise relevant to the issue at hand?
- ◆ **Advanced specialized training in that field of expertise, including research, studies, lectures attended or given**
 - Has the expert received any advanced specialized training or education (beyond the degree and licensure requirements) in his field of expertise (e.g., continuing education)?
- ◆ **Professional experience**
 - What relevant professional experience does the expert have since completing minimum degree, training, and licensure requirements (e.g., clinical experience, research experience)?
 - What professional positions has the expert held (e.g., employment history)?
- ◆ **Publications and professional contributions to the field**
 - Has the expert published within the subject about which he intends to testify?
 - If the expert has published, what was the nature of those publications (e.g., subject matter, which journals)?
 - To what extent is the expert a recognized authority in his particular speciality?
- ◆ **Prior testimony or employment as expert**
 - Has the expert served as a witness before?
 - If yes: On what topic? In what type of cases? In how many cases? For what side?
 - Has the expert ever testified to a contrary position in another proceeding?

Theoretical Paradigms in Psychology and Psychiatry

A variety of theoretical paradigms exist within psychology and psychiatry. Each reflects different underlying assumptions about human behavior and the causes of psychological/psychiatric disorders, and each proposes different treatment plans.

- ◆ **Medical-Biological Theories**
 - assume psychological disorders are the result of biological and physiological conditions
 - treatment usually involves drugs to change biochemical and other bodily functions
- ◆ **Psychodynamic Theories**
 - assume psychological disorders result from anxiety produced by unresolved conflicts and forces a person may not be aware of
 - treatment usually involves helping a patient become more aware of motivations, conflicts, and desires
- ◆ **Behavioral Theories**
 - assume psychological disorders are caused by faulty or ineffective learning and conditioning patterns
 - treatment usually involves reshaping disordered behavior through learning techniques such as reinforcement and modeling
- ◆ **Cognitive Theories**
 - assume psychological disorders are the result of faulty or unrealistic thinking
 - treatment usually involves developing new thought processes that instill new values
- ◆ **Social Psychological Theories**
 - assume that behavior and social interaction are setting dependent
 - not a discipline of psychology typically focused on treatment issues; a research sub-discipline

Remember that a particular paradigm reflects certain assumptions and values, and influences what areas and problems are selected for attention, how those problems are addressed, and how success will be measured.

Before going any further, stop and reflect ...

- ◆ Does psychological and psychiatric evidence constitute scientific evidence in the manner meant by *Daubert*? Might this differ according to which sub-discipline of psychology is at issue - clinical psychology vs. social psychology vs. neuro-psychology?
- ◆ What challenge does psychological and psychiatric evidence face from evidentiary standards that rely on “falsifiability?” Might this differ depending on the sub-discipline at issue?
- ◆ What do you think the future holds for the admissibility and utility of various types of psychological and psychiatric evidence?

➤ ***Assessments of Future Behavior***⁵

Courts frequently rely on psychiatrists and psychologists to assist them in assessing future behavior, especially violence or dangerousness. Recall that 26% of the judges surveyed had “some” courtroom experience with psychological evidence about the prediction of dangerousness. However, the ability of psychiatrists and psychologists to make accurate predictions of future dangerousness remains controversial. In particular, there remains a debate as to whether predictions of dangerousness are accurate enough to justify the deprivation of liberty or life.

Today, researchers and clinicians have shifted the debate from whether predictions of dangerousness are accurate enough to meet legal standards of proof, to how judges can use *risk assessment factors* to aid their decision-making. For example, in juvenile and family courts risk assessment instruments are sometimes offered in child protection cases to provide an estimate of the probability that in a case of child maltreatment there will be a future recurrence of maltreatment. The goal of such a risk assessment is to help the decision-maker determine whether a child should be removed from the home.

Before going any further, stop and reflect ...

- ◆ How much confidence do you have in the ability of psychologists and psychiatrists to assess risk or to predict dangerousness?
- ◆ As a judge, what factors would you consider if you were to predict the future behavior of a particular individual?
- ◆ Do you think psychologists and psychiatrists are any more accurate in their assessments of future behavior than are judges or lay people? Why or why not?

There is a key difference between the ability of clinicians to predict dangerousness and their ability to identify factors that increase the potential risk of violence. While research does not support the finding that clinicians can predict with certainty who will commit a violent act, they can identify those, who over the short term, may have an increased risk of doing so and who may respond to treatment.⁶

The Statistical Base-Rate Problem

The type of behaviors that would meet many of the applicable legal standards for dangerousness are rare -- this results in a statistical base-rate problem. For example, suppose that only 1 in 100,000 individuals is likely to commit a violent sexual act in the next six months. Even if a clinician is able to estimate with good accuracy that based on risk assessment factors a particular individual is 10 times more likely to commit a sexually violent act in the next six months than the average person in the community, this still means that there is only a 1 in 10,000 chance that this particular individual will commit a violent sexual act within that time frame.

◆ **Assessment Methods**

- **Clinical Method:** Based on observation and personal examination, history-taking, and testing. The clinician reviews the data obtained from the assessment and forms an opinion about the likelihood of the individual engaging in a particular future behavior.
- **Actuarial Method:** Based on assigning statistical probabilities of outcomes from combinations of a number of variables that correlate with the behavior at issue (future violence). The expert's opinion is a general probability based on given variable percentages. For example, assuming that males have a higher probability of committing violent crimes than females, an individual male has a higher probability for violence under the actuarial method.

??? *Questions to consider when evaluating assessments of future behavior ...*

- ◆ Is the clinical information that is being presented based upon an assessment (i.e., identification of risk factors associated with violence) or a prediction?
- ◆ What method or approach was used to gather information to make the assessment or prediction?
- ◆ What is the reliability and validity of the testimony or opinion?

Empirical research has consistently demonstrated that psychiatric and psychological predictions of dangerousness generally prove to be inaccurate.⁷ In *Barefoot v. Estelle*, the Supreme Court ruled that special caution should be expected when considering the admissibility of predictions of dangerousness testimony because it often literally involves a life and death decision.⁸ Yet the Supreme Court permitted psychiatrists to testify about the dangerousness of the defendant even though they had never examined him. The continued reliance on dangerousness predictions and the lack of empirical validity has led task forces of psychiatric and psychological professional organizations to acknowledge that mental health professionals lack competence in this area. The American Psychiatric Association Task Force on Clinical Aspects of the Violent Individual, for example, has taken the position that “neither psychiatrists nor anyone else have demonstrated an ability to predict future violence or dangerousness”⁹ (American Psychiatric Association, 1974, p. 20). In *Barefoot v. Estelle* The American Psychiatric Association had even submitted an amicus brief that stated,

Contrary to the claims of the prosecution psychiatrists who testified in this case, psychiatric predictions of long-term future dangerousness -- even under the best of conditions and on the basis of complete medical data -- are of fundamentally low reliability.¹⁰

More recently, the American Psychological Association Task Force on the Role of Psychology in the Criminal Justice System concluded that “the validity of psychological predictions of violent behavior ... is extremely poor, so poor that one could oppose their use on the strictly empirical grounds that psychologists are not professionally competent to make such judgments.”¹¹ Yet despite cautions by the Supreme Court, the American Psychiatric Association, and the American Psychological Association, expert testimony about future dangerousness remains a significant means of persuading legal fact-finders that a defendant poses a threat to society.

Some Case Law on Future Dangerousness

The courts, including the United States Supreme Court, have consistently upheld the use of predictions of dangerousness. The logic underlying the courts' decisions in this area seems to be that predictions of dangerousness are essential to the functioning of the justice system. Indeed, courts have ruled that predictions of dangerousness are admissible even though they "lack scientific merit." In upholding the Texas capital sentencing statute which requires a finding of dangerousness, the U.S. Supreme Court in *Jurek v. Texas* specifically considered the claim that the statute was invalid because "it is impossible to predict dangerous behavior" (at 274). In addressing this argument the Court concluded:

"It is, of course, not easy to predict future behavior. The fact that such a determination is difficult, however, does not mean that it cannot be made. Indeed, prediction of future criminal conduct is an essential element in many of the decisions rendered throughout our criminal justice system" (at 276).

Before going any further, stop and reflect ...

- ◆ Why do you think courts have continually upheld the use of prediction of dangerousness testimony?
- ◆ What do you think the consequences would be if courts no longer used such testimony?
- ◆ How do risk assessments differ from predictions of future behavior? Do risk assessments have more or less value for the court?
- ◆ What information would you need about these procedures to determine their reliability and validity?

A CLOSER LOOK AT THE EVIDENCE: PSYCHOLOGICAL SYNDROMES

All types of syndromes have found their way not only into the courtroom but into the lexicon of most Americans – battered woman syndrome, child sexual abuse syndrome, rape trauma syndrome – and the list continues.

A **syndrome** is a group or constellation of symptoms that appear together regularly enough to become associated with each other. Unlike a disease, whose underlying cause is either known or relatively well understood, the etiology of a syndrome is usually not known or is poorly understood. Therefore, the symptoms of a syndrome do not clearly and convincingly result from a specific etiology. Some syndromes are descriptions of behaviors consistently exhibited by large numbers of people exposed to similar stimuli, such as those described as the battered woman syndrome. Some syndromes are an attempt to postulate a diagnostic tool based on observable behaviors, such as the child sexual abuse syndrome.

Syndrome: a group or constellation of symptoms that appear together regularly enough to become associated

Syndromes can be descriptions of behaviors consistently exhibited by large numbers of people exposed to similar stimuli (e.g., BWS or battered woman syndrome), or attempts to postulate a diagnostic tool based on observable behaviors (e.g., CSAAS or child sexual abuse accommodation syndrome)

Common Examples of Psychological Syndromes

Battered Woman (Spouse) Syndrome	Repressed Memory Syndrome
Rape Trauma Syndrome	False Memory Syndrome
Child Sex Abuse Accommodation Syndrome	Parental Alienation Syndrome

Both diseases and syndromes share the medically and forensically important feature of diagnostic value. Both point with varying degrees of certainty to particular causes. However, whereas with many diseases the relationship between symptoms and etiology is clear, with syndromes, this relationship is often unclear or unknown. The certainty with which a

syndrome points to a particular cause varies with the syndrome. Two syndromes often offered in expert testimony in cases of alleged child abuse are the Battered Child Syndrome (BCS) and Child Sexual Abuse Accommodation Syndrome (CSAAS). The battered child syndrome has high certainty since a child with the symptoms is very likely to have suffered non-accidental injury. Therefore, this syndrome has high probative value and, in fact, has been approved by every appellate court to consider it. This can be contrasted with CSAAS which does not point with any certainty to sexual abuse. The fact that a child shows behaviors of the CSAAS does not change the likelihood that the child was sexually abused.



A national snapshot of judges' experience with psychological syndrome evidence ...

Of those judges reporting that they had past experience with psychological syndrome evidence (n=255):

- 77% of the judges reported that they had experience with *Battered Woman Syndrome* (BWS) evidence; 33% of these judges reported their level of experience with BWS as a 3 or higher on a 5 point scale (with 0 being "no experience" and 5 being a "great deal of experience")
- 73% of the judges reported that they had experience with *Child Sex Abuse Accommodation Syndrome* (CSAAS) evidence; 35% of these judges reported their level of experience with CSAAS as a 3 or higher on a 5 point scale (with 0 being "no experience" and 5 being a "great deal of experience")
- 64% of the judges reported that they had experience with *Rape Trauma Syndrome* (RTS) evidence; 24% of these judges reported their level of experience with RTS as a 3 or higher on a 5 point scale (with 0 being "no experience" and 5 being a "great deal of experience")
- 41% of the judges reported that they had experience with *Repressed Memory Syndrome* (RMS) evidence; 7% of these judges reported their level of experience with RMS as a 3 or higher on a 5 point scale (with 0 being "no experience" and 5 being a "great deal of experience")

N=312 (Judges were given an option of completing these questions over the telephone or via mail; sample size for these questions does not equal 400 as not all mail-out questionnaires were returned).

➡ **Issue: Falsifiability and Syndrome Evidence**

Some syndromes may not be falsifiable or testable at all, especially to the extent they are derived from Freudian-based theories. Indeed, the application of *Daubert* criteria to Freudian-based theories may prove to be particularly problematic. For example, if the effects of an early trauma may be manifested in art ('sublimation'), in maladaptive presumptions about others ('projection'), or not manifested at all ('repression'), then there is no

Recall that an underlying theory is not falsifiable if there is no circumstance, behavior, or observation which could be used to conclude that an event did or did not occur – when virtually any type of behavior can be viewed as supportive of the hypothesis the underlying theory is unfalsifiable.

way to falsify the claimed causal sequence, unless conditions are specified that determine which of the consequences of the trauma will occur under which circumstances. The theory does not provide such specificity, so that no manifestation, an artistic manifestation, or a maladaptive interpersonal manifestation may all be accepted as evidence of the trauma. In addition, such theories are typically used post hoc to provide explanations for events but not to make discrete predictions. The result is an absence of falsification attempts. *While falsification is problematic for psychological syndrome testimony, it is important to stress that Daubert does not require that falsifiability, or lack of falsification attempts, serve as the definitive admissibility criteria or the definitive characteristic of science.*

Before going any further, stop and reflect ...

- ◆ To what extent do you think syndrome evidence is falsifiable? Why?
- ◆ Some critics argue that psychological syndrome evidence should not be classified as scientific evidence, and therefore the guideline of falsifiability is not applicable. Others argue that it is. What do you think? Why?
- ◆ What type of information would you expect to hear from an expert proffering testimony on a psychological syndrome?

➡ ***Issue: Error Rates and Syndrome Evidence***

A consideration of error rates with respect to syndrome evidence can be thought of in terms of weighing the risk of making a “false positive” error or a “false negative” error (see earlier chapters on experimental method and statistics). For example, widespread acceptance of questionable syndrome evidence that has, in actuality, no scientific reliability, could result in false positive errors – identifying an individual as abused or an abuser, for example, when it is not true. This could arise in part because of *Daubert’s* statement that the more liberal Federal Rules of Evidence supersede the *Frye* test. Conversely, the court should be concerned about wholesale rejection of valuable syndrome evidence that might have assisted the trier of fact in making a proper and just decision, resulting in more false negative errors. This could easily happen, given the considerable controversy about new syndromes.

Considerations when Assessing Known or Potential Error Rates with Syndrome Evidence:

- carefully evaluate the methodology used in corroborative testing carried out on the syndrome evidence;
- determine whether the method of data collection (e.g., clinical interview, actuarial methods, reliance on the DSM, empirical research) was well-accepted in the relevant discipline;
- consider whether there have been adequate replications of the findings or whether the claim of a syndrome rests on a narrow database; and
- determine whether the probative value of the syndrome outweighs its potential prejudicial value.

Before going any further, stop and reflect ...

- ◆ Some critics argue that psychological syndrome evidence is not scientific and therefore is not subject to an error rate analysis. Others argue that it is. What do you think? Why?
- ◆ To what extent can the potential error rate of syndrome evidence be determined?
- ◆ What information would you expect to hear from an expert proffering psychological testimony with respect to potential error rates?
- ◆ With respect to psychological syndrome evidence, what are the potential consequences of making a false positive error? A false negative error?

➡ ***Battered Woman Syndrome***

Battered woman syndrome (BWS) is described by Lenore Walker in *The Battered Woman* (1979)¹² as a combination of a three-phase cycle of abuse which is accompanied by “learned helplessness” on the part of the abused woman. This cycle is repeated again and again, with the victim developing a “learned helplessness” that precludes her from leaving the batterer. Specifically, Walker has identified three distinct periods within the course of a battering episode:

- I. *The Tension-Building Stage*
- II. *An Acute Battering Incident*
- III. *Contrite and Loving Behavior*

I. *The Tension-Building Stage*

This stage is characterized by a series of minor verbal and physical battering events which precede an acute battering incident. Typical characteristics of this stage include:

- (a) increases in the frequency and severity of the battering incidents;
- (b) decrease in the effectiveness of strategies used by the women to placate their batterers; and

- (c) feelings of responsibility on part of the women because they have been unable to control the batterer's behavior.

II. *An Acute Battering Incident*

During the tension-building stage, tension builds and is ultimately released in an acute battering incident. The characteristics of this stage are:

- (a) uncontrollable rage and destructiveness on the part of the batterer which usually does not end until the victim is severely beaten;
- (b) the triggering of rage by some external event or the internal state of the batterer, as opposed to the behavior of the victim; and
- (c) an acute battering incident that is briefer in duration than either the tension-building stage or the period of contrite and loving behavior.

III. *Contrite and Loving Behavior*

A period of unusual calm typically follows the acute battering incident, during which the batterer attempts to make up for the abusive behavior. Characteristics associated with this stage are:

- (a) admission of wrongdoing on the part of the batterer, accompanied by apologies and the batterer's assurances that the abusive behavior will not happen again;
- (b) intensive effort on the part of the batterer to "win back" the victim through gifts, enlistment of other family members, and appeals to the victim's guilt over the adverse consequences the batterer will suffer if the victim leaves;
- (c) victim's capitulation from anger, fear, hurt, and loneliness to happiness and confidence; and
- (d) a strengthening of the the bond between the batterer and victim as they both come to believe, through reinforcement, that their relationship can be made to work.

Contextual factors possibly affecting the battered woman's response to abuse:

- fear of retaliation
- available economic and other tangible resources
- worry for children
- emotional attachment to batterer
- personal emotional strengths such as hope
- culture
- personal vulnerabilities (emotional, mental, physical)
- perception of available social supports

From: Dutton, M.A. (1993). "Understanding Women's Responses to Domestic Violence: A Redefinition of Battered Woman Syndrome,." *Hofstra Law Review*, Vol 21,1191, pgs. 1205-06.

The battered woman syndrome (BWS) has been brought to the attention of nearly every court in the country during the last decade, partly as a result of the success defendants have had introducing the evidence in early cases (e.g., *Ibn-Tamas v. United States*,¹³ *State v. Anaya*,¹⁴ and *Smith v. State*¹⁵). Battered woman syndrome has been endorsed as a theory by the American Psychological Association and has been recognized and admitted by a majority of states.¹⁶ Nevertheless, the introduction of BWS as evidence has caused considerable debate.

Some Problems Associated With the Use of BWS in Court

- Definitional problems with BWS and variations of behaviors exhibited both by battering men and by battered women
- Significant variations among courts about whether BWS testimony should be admitted and for what purpose
- Whether BWS constitutes 'good science,' and under what standard of admissibility it should be judged



The problematic nature of deciding the admissibility of BWS testimony

Those judges who reported 'some' or 'a great deal' of experience with BWS evidence were asked what aspect of the evidence is most problematic when determining its admissibility. Identified problem areas include (in order of frequency of mention):

- the subjectivity of the diagnostic process
- determining its relevance to the facts at issue
- weighing its probative vs. prejudicial value
- determining its general acceptance
- determining the appropriate qualifications of BWS experts

Some Case Law on BWS

Early on, certain courts did not want to admit BWS testimony, determining that the jury could comprehend the facts about the woman's situation without an expert to interpret such testimony. Increasingly over the years, however, courts began accepting that battered women not only view the danger of the abuse differently than others, but that juries were having difficulty comprehending why women stay with their abusers. One reason that BWS has become admissible evidence in the court is that many jurors have inaccurate, stereotypic ideas about battered women. It is because facts about battered women and battering relationships are often considered to be 'beyond the ken of the jury' that they have been admitted into evidence (cf. *Arcoren v. U.S.*, 929 F.2d 1235 (8th Cir.); *State v. Allery*, 682 P.2d 312, 316 (Wash. 1984); *State v. Borelli*, 629 A.2d 1105 (Conn. 1993); *Smith v. State*, 277 S.E. 2d 678 (Ga. 1981); *State v. Gallegos*, 719 P.2d 1268 (N.M. App. 1986)).

Two-thirds of states have held expert testimony on BWS relevant to the issue of why the defendant did not leave the batterer, or to explain other conduct such as acts committed under duress from the abuser. Of the courts which have recognized the syndrome and permit testimony about it, there are many variations in the method of its admission. Some courts permit experts to testify in a self-defense scenario but refuse to permit the expert to render an opinion as to whether the woman was a battered woman (e.g., *People v. Wilson*, 487 N.W.2d 822 (Mich. App. 1992)). Some courts have permitted experts to render opinions on the ultimate issue, namely, whether the woman was in fear of death or great bodily harm at the time of the killing (e.g. *State v. Wilkins*, 407 S.E.2d 670 (S.C. App. 1991)).

The majority of courts that have allowed BWS have admitted it into evidence as part of the defense in cases in which the defendant is alleging self-defense in response to a murder or attempted murder charge. The testimony has been admitted because the expert's knowledge is beyond the ken of the juror and to assist the jury in understanding the state of mind of the defendant at the time of the killing (an essential element of self-defense). For example, in *Bechtel v. State* (840 P.2d 1, Okla. App. 1992) the court permitted the defendant to introduce the entire history of battering in the relationship in order to "put into context" the defendant's understanding of imminent bodily harm. Many other courts have combined the concepts of overcoming stereotypes and educating juries about the perceptions of a battered woman when permitting BWS evidence in self-defense cases (e.g., *State v. Allery*, 682 P.2d 312 (Wash. 1984); *Ex parte Hill*, 507 So. 2d 558 (Ala. 1987); *People v. Humphrey*, 921 P.2d 1 (Cal. App. 1996); *Ibn-Tamas v. United States*, 407 A.2d 626 (D.C. App. 1979)).

The use of BWS has been raised in other types of criminal cases, where a woman has raised duress or compulsion as a defense, or where she claims to have been unable to protect her children from abuse. In a few cases, courts have held that expert testimony about BWS might be helpful where the defendant claimed she was unable to form the requisite capacity to commit the crimes charged (e.g., *United States v. Brown*, 891 F. Supp. 1501 (D. Kan. 1995); *United States v. Marengi*, 893 F. Supp. 85 (D. Me. 1996)).

Several states have enacted statutes addressing the admissibility of expert testimony about BWS. While each statute is different, the states that have enacted statutes have uniformly permitted expert testimony in cases in which the accused alleges self defense to a charge of homicide (e.g., Cal. Evid Code § 1107; Mo Stat § 563.033; Ohio Rev Code § 2901.06).

The effects of battering and abusive relationships have been studied using a variety of methods, including clinical interviews, surveys, and the administration of psychological tests. National surveys have attempted to document the prevalence of abuse using large representative samples, but have failed to provide rich descriptions of the experience of battered women. Clinical interviews provide detailed information about battered women's experiences, but because they typically use small, selective samples, the representativeness of this research is questionable.

The early research regarding the battered woman syndrome has been directly attacked for methodological flaws. For example:

- (a) failure to employ an appropriate control group
- (b) failure to employ appropriate statistical tests
- (c) failure to employ controls to guard against researcher bias.¹⁷

Walker's application of learned helplessness has also been criticized. A review of the literature on learned helplessness found several studies in which battered women took action to end their abuse and employed a wide range of coping skills in their relationships.¹⁸ These self-help measures, however, conflict with learned helplessness theory, which would predict that battered women would not take positive action to change their situations. However, others found that the actions taken by these women are typically passive, such as fantasizing or obtaining social or spiritual support.¹⁹ More recent studies have attempted to address criticisms regarding lack of control groups in the study of battered women by comparing battered women who kill their abusers with those who do not and finding significant differences between these groups.²⁰

Before going any further, stop and reflect ...

- ◆ What purpose does BWS evidence serve for the court?
- ◆ Are there alternative ways in which the concerns of battered women could be addressed in court without reliance on BWS testimony?
- ◆ Can you think of a research design to study the effects of battering on victims? What ethical concerns are raised when conducting this type of research?
- ◆ To what extent, if any, do you think the introduction of BWS into court has been driven by political or social concerns?

➔ *Rape Trauma Syndrome*

Burgess and Holmstrom coined the term “rape trauma syndrome”(RTS) in 1974, and it refers to the acute phase and long-term reorganization process that is said to occur as a result of forcible rape or attempted forcible rape. Burgess and Holmstrom studied 92 patients who presented to the emergency room at Boston City Hospital during a one-year period with the complaint of rape. According to this study, rape victims generally go through a two phase reaction:

- I. *Acute Phase*
- II. *Long-term Reorganization Phase*

I. *Acute Phase*

In the acute phase, lasting several weeks after the rape, victims are said to experience physical trauma from the rape, skeletal muscle tension manifesting in headaches, sleep disturbances, and an elevated startle reaction, gastrointestinal disturbances, as well as intense fear and self-blame.

II. *Long-term Reorganization Phase*

In the long-term phase, victims experience disorganization in their lives and must go through a "reorganization process." This phase is marked by increased motor activity (e.g., moving, traveling, and contacting support systems), nightmares, and phobic reactions to the circumstances of the rape, such as fears of being indoors or outdoors, and the fear of being alone or in crowds.

In the legal arena, evidence about RTS is usually introduced against a male criminal defendant charged with forcible rape. Sometimes it is offered as social framework evidence to put the victim's action into context. For instance, the prosecution may offer an expert on RTS to explain the victim's actions around the time of her alleged rape, especially if they seem inconsistent with the expected behavior of someone who has just been raped.²¹ Other times, testimony regarding RTS is offered to support the contention that the victim did not consent to intercourse²² or to show that intercourse did, in fact, take place.²³ RTS evidence has also been proffered by the defense who argue that if the victim had actually been raped she would exhibit symptoms of RTS. That is, the absence of RTS has been used as a defense for rape.²⁴

A 1992 review of case law concluded that the reliability and validity of scientific research on RTS was not often a concern of the courts.²⁵ When it was an issue, however, two basic themes emerged: (1) problems recognizing a single syndrome given the broad range of symptoms and responses exhibited by different victims; and (2) problems ruling out alternate causes of post-traumatic stress reactions.

Burgess and Holmstrom's original work, as well as other early studies of rape trauma syndrome, have been criticized as suffering serious methodological flaws, such as:

- lack of a control group;
- problems with sample size and sample characteristics;
- failure to operationalize important definitions and concepts; and

- potential selection bias, inconsistent interviewing methods, and inadequate long-term follow-up of victims

However, the methodology of later studies on rape trauma syndrome has significantly improved. Though not error-free, these studies have “assessed victim recovery at several points after the assault using standardized assessment measures and ... have employed carefully matched control groups.”²⁶ The newer studies confirm those documented by Burgess and Holmstrom (1974), with victims exhibiting fear, depression, guilt, sleep disturbances, and emotional disorders. Indeed, soon after post-traumatic stress disorder (PTSD) was added to the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM) in 1980, “the authoritative treatise interpreting PTSD immediately recognized RTS as a quintessential example.”²⁷ Despite this endorsement, some critics maintain that recent studies still suffer from inadequate operational definitions and biased research samples, and that individual reactions to rape are so dependent upon various factors that no causal relationship can be established.²⁸



The problematic nature of deciding the admissibility of RTS testimony

Those judges who reported ‘some’ or ‘a great deal’ of experience with RTS evidence were asked what aspect of the evidence is most problematic when determining its admissibility. Not surprisingly, the problems identified with respect to RTS testimony were the same as those identified for BWS testimony, and included (in order of frequency of mention):

- the subjectivity of the diagnostic process
- determining its relevance to the facts at issue
- weighing its probative vs. prejudicial value
- determining its general acceptance
- determining the appropriate qualifications of RTS experts

Some Case Law on RTS

Since the first RTS case in 1982 (*State v. Saldana*, 324 N.W. 2d 227 (Minn. 1982)), there have been a variety of approaches to the admissibility of such evidence. Only one jurisdiction has permitted RTS testimony as substantive evidence that the victim was, in fact, raped, although some states' decisions could arguably be interpreted as doing so. In *State v. Allewalt* (517 A.2d 741, Md. 1986) the Maryland Supreme Court held that it was not an abuse of discretion to admit testimony of an expert who testified about post-traumatic stress disorder (PTSD) and his belief that the witness suffered from PTSD as a result of the rape in question. Some states have permitted evidence of RTS to be admitted for the purpose of helping to determine whether the woman consented to the sexual intercourse or whether such intercourse was, in fact, against her will (e.g., *State v. Marks*, 647 P.2d 1292, Kan. 1982; *State v. McQuillen*, 689 P.2d 822, Kan. 1984; *State v. Huey*, 699 P.2d 1290, Ariz. 1985).

Many courts have permitted admission of RTS evidence concerning the typical behaviors of women who have been raped – that is, an explanation of the symptomology of RTS (e.g., *People v. Bledsoe*, 681 P.2d 291, Cal. 1984; *Commonwealth v. Mamay*, 553 N.E.2d 945, Mass. 1990; *People v. Pullins*, 378 N.W. 2d 502, Mich. App. 1985; *People v. Hampton*, 746 P.2d 947, Colo. 1987; *People v. Reid*, 475 N.Y.S. 2d 741, N.Y. Sup. Ct. 1984; *People v. Taylor*, 552 N.E. 2d 131, N.Y. 1990)

Although the majority of states addressing the subject have admitted RTS evidence, there are a few states that have disallowed any evidence of RTS to be introduced. In *State v. Saldana* (324 N.W. 2d 227, Minn. 1982) for example, the court ruled that “rape trauma syndrome is not a fact-finding tool, but a therapeutic tool useful in counseling” (at 230). Accordingly, such testimony was held to be inadmissible. *Saldana* has been followed in several jurisdictions for the proposition that RTS is not admissible as proof that a rape occurred (e.g., *Commonwealth v. Gallagher*, 547 A.2d 355, Pa. 1988).

Courts applying the *Daubert* standard to rape trauma syndrome in rape cases engage in what appears, on its face, to be a *Daubert* analysis. In *State v. Alberico*, for example, the Supreme Court of New Mexico found that “rape trauma syndrome testimony [wa]s grounded in valid scientific principles,” since it had been catalogued in the diagnostic and statistical manual of psychiatrists and psychologists and “appear[ed] to be grounded in basic behavioral psychology.” The court did not, however, discuss any of the factors articulated in *Daubert*, nor did they consider any other factors which are critical to “good science.”

Before going any further, stop and reflect ...

- ◆ What purpose does RTS evidence serve in the court?
- ◆ Are there alternative ways in which the concerns of victims of rape and sexual assault could be addressed in court without reliance on RTS testimony?
- ◆ Can you think of a research design to study the effects of experiencing a traumatic event such as rape? What ethical concerns are raised by this type of research?
- ◆ To what extent, if any, do you think that the introduction of RTS into court was driven by political and social concerns?

CRITICAL QUESTIONS REVIEWED

- ◆ Was the appropriate test used for a specific individual (e.g., certain tests are more appropriate for children than adults)?
- ◆ Is the test being used appropriately for the specific legal issue at hand and can the data obtained be applied properly to the specific subject of that particular legal inquiry?
- ◆ Is the expert appropriately certified or licensed within the profession?
- ◆ What professional associations does the expert belong to and why?
- ◆ To what extent are these professional memberships relevant to the expert's training and expertise?
- ◆ To what extent are these professional memberships relevant to the issue at hand?
- ◆ What is the expert's particular area of expertise?
- ◆ Is this area of expertise relevant to the issue at hand?
- ◆ Has the expert received any advanced specialized training or education (beyond the degree and licensure requirements) in his field of expertise (e.g., continuing education)?
- ◆ What relevant professional experience does the expert have since completing minimum degree, training, and licensure requirements (e.g., clinical experience, research experience)?
- ◆ What professional positions has the expert held (e.g., employment history)?
- ◆ Has the expert published within the subject about which he intends to testify?
- ◆ If the expert has published, what was the nature of those publications (e.g., subject matter, which journals)?
- ◆ To what extent is the expert a recognized authority in his particular speciality?
- ◆ Has the expert served as a witness before?
- ◆ If yes: On what topic? In what type of cases? In how many cases? For what side?

- ◆ Has the expert ever testified to a contrary position in another proceeding?
- ◆ Is the clinical information that is being presented based upon an assessment (i.e., identification of risk factors associated with violence) or a prediction?
- ◆ What method or approach was used to gather information to make the assessment or prediction?
- ◆ What is the reliability and validity of the testimony or opinion?

Endnotes:

1. Sales, B.D. (1994). "Social and Behavioral Science Evidence in Litigation: Is it a Special Type of Scientific Evidence?" *Shepard's Expert and Scientific Evidence Quarterly*, Vol. 2(2), pg. 321.
2. *National Benchbook on Psychiatric and Psychological Evidence and Testimony* (1998), Chp. 3, pg. 47. State Justice Institute.
3. *Ibid*, Chp. 1, pg. 19.
4. See DSM-IV at xiii-xxv.
5. This section is adapted from *National Benchbook on Psychiatric and Psychological Evidence and Testimony*, *Supra* note 2, Chp. 3, pgs. 48-49. State Justice Institute.
6. Monahan, J. (1997). "Clinical and Actuarial Predictions of Violence." In David Faigman et al. (eds.), *Modern Scientific Evidence: The Law and Science of Expert Testimony*. West Publishing Co.
7. See for example Cocozza, J.J. and Steadman, H.J. (1976). "The Failure of Psychiatric Predictions of Dangerousness: Clear and Convincing Evidence." *Rutgers Law Review*, Vol. 29, pg. 1084; Kozol, P., Boucher, A., and Garofolo, M. (1972). "The Diagnosis and Treatment of Dangerousness." *Crime and Delinquency*, Vol. 18, pg. 371; Steadman, H.J., and Morrissey, L. (1981). "The Statistical Prediction of Violent Behavior." *Law and Human Behavior*, Vol. 5, pg. 263; Wenk, B., Robison, R., and Sineth, C. (1972). "Can Violence be Predicted?" *Crime and Delinquency*, Vol. 18, pg. 393.
8. *Barefoot v. Estelle*, 463 U.S. 880 (1983).
9. American Psychiatric Association. (1974). *Task Force on Clinical Aspects of the Violent Individual: Report No. 8*.
10. *Ibid*.
11. American Psychological Association. (1990). "Report of the Task Force on the Role of Psychology in the Criminal Justice System." *American Psychologist*, Vol. 73, pg. 1099.
12. Lenore Walker (1979) developed BWS based upon commonalities she observed among battered women in her psychotherapy practice and 120 interviews with battered women.
13. 407 A.2d 626 (D.C. 1979).
14. 438 A.ed 892 (Me. 1981).
15. 277 S.E. 2d 678 (Ga. 1981).
16. See cases collected in *State v. Bechtel*, 840 P.2d 1, 7 n.5 (Okla. App. 1992).

17. Faigman, D.L. (1986). "The Battered Woman Syndrome and Self-Defense: A Legal and Empirical Disent." *Virginia Law Review*, Vol. 72, pg. 619; Faigman, D.L. and Wright, A.. (1997). "The Battered Woman in the Age of Science." *Arizona Law Review*, Vol. 39, pg. 67; Schuller R.A. and Vidmar, N. (1992). "Battered Woman Syndrome Evidence in the Courtroom." *Law and Human Behavior*, Vol. 16, pg. 273.
18. Schuller, R.A. and Vidmar, N. (1992). "Battered Woman Syndrome Evidence in the Courtroom: A Review of the Literature." *Law and Human Behavior*, Vol. 16, pg. 273.
19. Barnett and Laviolette. (1993). *It Could Happen to Anyone: Why Battered Women Stay*. New York: Plenum.
20. Browne. (1987). *Why Battered Women Kill*. Chicago: University of Chicago Press.
21. cf., *People v. Taylor*, 552 N.E.2d 131 (N.Y. 1990) allowing such testimony to aid in explaining the victim's behavior.
22. *State v. Huey*, 699 P.2d 1290 (Ariz. 1985); *State v. Marks*, 647 P2d 1292 (Kan. 1982); *State v. Saldana*, 324 N.W. 2d 227 (Minn. 1982).
23. *State v. Bressman*, 689 P.2d 901 (Kan. 1984); *People v. Taylor*, 552 N.E.2d 131.
24. *Henson v. State*, 535 N.E. 2d 1189 (Ind. 1989).
25. Frazier, P., and Borgida, E. (1992). "Rape Trauma Syndrome: A Review of Case Law and Psychological Research." *Law and Human Behavior*, Vol. 16, pg. 301.
26. *Ibid.*
27. McCord, D. (1985). "Syndromes, Profiles, and Other Mental Exotica: A New Approach to the Admissibility of Nontraditional Psychological Evidence in Criminal Cases." *Oregon Law Review*, Vol 66, pg. 19.
28. Lauderdale, H.J. (1984). "The Admissibility of Expert Testimony on Rape Trauma Syndrome." *Journal of Criminal Law and Criminology*, Vol. 75, pg. 1366; McCord (1985), *Supra* note 27.

GLOSSARY

actuarial method	based on assigning statistical probabilities of outcomes from combinations of a number of variables that correlate with the behavior at issue; expert's opinion is a general probability based on given variable percentages
certification	a formal designation of someone to use a professional title based upon education, training, and experience; a statement authorizing an individual to officially practice in a certain profession
clinical interview	a systematic history-taking and a mental status exam
clinical methods	based upon observation, history-taking, testing, and diagnosis; relies upon clinical interviews, psychological testing, psycho-social histories, and medical exams
DSM	Diagnostic and Statistical Manual of Mental Disorders; developed by the American Psychiatric Association to provide mental health practitioners with a diagnostic classification system of mental disorders; includes diagnostic criteria necessary for diagnosing organic brain disorders, personality disorders, childhood disorders, relational disorders, associative disorders, and the like
empirical research	systematic gathering of information and study of problems in accordance with the agreed upon methodological practices of experimental, quasi-experimental, or qualitative research
licensure	state granted permission to perform certain functions related to diagnosis, treatment, or dispensing of medication; legal permission to perform a specified function; typically requires an educational degree requirement and a specified number of practical training hours
mental status exam	an evaluation of an individual's current functioning as revealed through observations and responses to interview questions; often unstructured and highly subjective
objective personality test	generally consist of true-false or check the best answer questions; test scoring is objective, but not necessarily the interpretation of the results (e.g., MMPI)
projective personality test	based upon the assumption that a person's unconscious motivates and directs daily thoughts and behavior; to uncover those unconscious motivations, clinicians provide ambiguous stimuli to which the individual can provide responses that might reflect his unconscious; that is, the individual projects unconscious feelings, drives, and motives onto the ambiguous stimuli (e.g., Rorschach Inkblot Test)
standardized tests	tests that have been statistically normed or standardized, permitting clinicians to compare scores of test subjects with the scores of normal individuals or patients with similar disorders
syndrome	a group or constellation of symptoms that appear together regularly enough to become associated

SUGGESTED READINGS:

General Overview

National Benchbook on Psychiatric and Psychological Evidence and Testimony. (1998). State Justice Institute.

Psychological Assessment Generally

Anastasi, A. and Urbina, S. (1997). *Psychological Testing, 7th Edition.* New York: Prentice-Hall.

Shuman, D.W. and Sales, B. (1994). *Psychiatric and Psychological Evidence, 2d Edition.* Shepard's McGraw-Hill.

Assessments of Future Behavior

Barefoot v. Estelle, 463 U.S. 880 (1983).

McCann, T. (1997). "Risk Assessment and the Prediction of Violent Behavior." *Federal Law Review*, Vol. 44, pg. 18.

Monahan, J. (1997). "Clinical and Actuarial Predictions of Violence." In David L. Faigman et al., (eds.), *Modern Scientific Evidence: The Law and Science of Expert Testimony.* West Publishing Co.

Psychological Syndromes Generally

McCord, D. (1987). "Syndromes, Profiles, and Other Mental Exotica: A New Approach to the Admissibility of Nontraditional Psychological Evidence in Criminal Cases." *Oregon Law Review*, Vol. 66, pg. 19.

Myers, M. (1993). "Expert Testimony Describing Psychological Syndromes." *Pacific Law Journal*, Vol. 24, pg. 1449.

Richardson, J.T., Ginsburg, G.P., Gatowski, S.I., and Dobbin, S.A. (1995). "The Problems of Applying *Daubert* to Psychological Syndrome Evidence." *Judicature*, Vol. 79, pg. 81.

Battered Woman Syndrome

Faigman, D.L., and Wright, A. (1997). "The Battered Woman Syndrome in the Age of Science." *Arizona Law Review*, Vol. 39, pg. 67.

Schuller, R., and Vidmar, N. (1992). "Battered Woman Syndrome Evidence in the Courtroom." *Law and Human Behavior*, Vol. 16, pg. 273.

Rape Trauma Syndrome

Dobbin, S.A., and Gatowski, S.I. (1998). "The Social Production of Rape Trauma Syndrome as Science and as Evidence." In M.D.A. Freeman and H. Reece, (eds.), *Science in Court.* London: Dartmouth Press.

Frazier, P., and Borgida, E. (1992). "Rape Trauma Syndrome: A Review of Case Law and Psychological Research." *Law and Human Behavior*, Vol. 16, pg. 293.

Judge's Notes:

CHAPTER 11

of *A Judge's Deskbook on the Basic Philosophies and Methods of Science*,
by Shirley A. Dobbin, Ph.D, and Sophia I. Gatowski, Ph.D

DNA Evidence

For all the diversity of the world's five and a half billion people, full of creativity and contradictions, the machinery of every human mind and body is built and run with fewer than 100,000 kinds of protein molecules. And for each of these proteins, we can imagine a single corresponding gene (though there is sometimes some redundancy) whose job it is to ensure an adequate and timely supply. In a material sense, then, all of the subtlety of our species, all of our art and science, is ultimately accounted for by a surprisingly small set of discrete genetic instructions. More surprisingly still, the differences between two unrelated individuals, between the man next door and Mozart, may reflect a mere handful of differences in their genomic recipes - perhaps one altered word in five hundred. We are far more alike than we are different. At the same time, there is room for near-infinite variety. It is no overstatement to say that to decode our 100,000 genes in some fundamental way would be an epochal step toward unraveling the manifold mysteries of life. (1)

Advances in molecular biology have spawned a host of new tests for criminal identification and paternity assessment. Because they examine variations in human DNA, the new tests are

Learning Objectives for Chapter 11

Upon completion of this chapter, the reader should be able to:

- Understand the basic theory of molecular biology;
- Articulate some of the social, moral, and legal questions surrounding breakthroughs in genetic research;
- Understand DNA testing concepts;
- Identify the controversies surrounding laboratory protocols, population databases, and the possibility of errors; and
- Understand the courts' reaction to DNA evidence.

Mapping the Human Genome

The Human Genome Project (HGP) is a multi-national project involving sixteen

generally called "DNA tests" or "DNA typing procedures." These labels cover a variety of different methods, however, and different methods for "typing" DNA can vary in their reliability, specificity, and appropriateness for particular applications. (2)

\The fundamental science underlying forensic DNA tests is not in dispute. There is no question, for example, that genetic differences exist among individuals and that DNA tests can uncover those differences. Disputes have arisen, however, about whether particular procedures have been adequately validated, whether laboratories are using adequate standards and controls to ensure the reliability of tests, and whether the results in particular cases have been interpreted correctly. Disputes have also arisen about the validity and appropriateness of methods used to characterize the statistical meaning of a DNA match and, more generally, about the appropriate way to present DNA test results to juries. Moreover, as genetic research advances, and geneticists and biochemists come increasingly closer to mapping the human genome, the court, and society as a whole, must come to terms with challenging social, moral, and legal issues.



The origins of genetic research can be traced back centuries when farmers began breeding domestic animals and crops to achieve the desired characteristics for productivity and durability. Observations of how physical

characteristics (phenotype) could be manipulated led to the work of Gregor Mendel, an Austrian monk, who came to be known as the "Father of Genetics" for his work on genetics in the nineteenth century. In the first half of the twentieth century, through the efforts of James Watson and Francis Crick, genetics was linked to DNA. By discovering the double helix structure of DNA (i.e., two DNA strands are twisted together into an entwined spiral), Watson and Crick revolutionized genetics.

DNA analysis is based on well-established principles of the wide genetic variability among humans and the presumed uniqueness of an individual's genetic makeup (identical twins excepted). Laboratory techniques for isolating and observing the DNA of human chromosomes have long been used in non-forensic scientific settings. The forensic application of the technique involves comparing a known DNA sample obtained from a suspect with a DNA sample obtained from the crime scene, and often with one obtained from the victim. Such

nations. The United States has worked most closely with Great Britain, France, and Japan. In the U.S., the project is funded by the National Institutes of Health (through the National Center for Human Genome Research) and the Department of Energy (through the Human Genome Project). The effort was expected to cost approximately \$3 billion and to take about 15 years, but the project is progressing ahead of schedule. It is now generally agreed that this international project will produce the complete sequence of the human genome by the year 2005. The first goal of the project is to discover and map all 50,000 to 100,000 human genes, which together make up the human genome, and to make the human genome itself accessible for biological study. Current and future applications of genome research will address national needs in molecular medicine, waste control and environmental cleanup, biotechnology, energy sources, and risk assessment. Breakthroughs in genetic research will change our understanding of ourselves and our families, and the meaning of health and illness.

analyses typically are offered to support or refute the claim that a criminal suspect contributed to a biological specimen (e.g., semen or blood) collected at the crime scene. Other considerations may arise where DNA is used to narrow the field of suspects by comparing a crime sample with samples from a blood bank, to establish the commission of a crime where no body is found, and to establish parentage.

Before going any further, stop and reflect ...

The social, moral, and legal issues raised by modern genome research are possibly among the most challenging ever faced — by the courts and by society as a whole. For example, one of the primary goals of genetic research is finding the genetic basis of diseases which will ultimately lead to therapies and interventions that will relieve human suffering. Emphasis will shift from treatment of the sick to a prevention-based approach. Researchers will be able to identify individuals predisposed to, or at risk for, particular diseases and will be able to provide novel therapeutic regimens based on new types of drugs, immunotherapy techniques, and the possible replacement of defective genes through gene therapy.

- How might increased genetic information change our current understanding of "health," "illness," and "disease"?
- How might increased genetic information change our current notion of "family" and "self"?
- What legal issues do you see arising from increased genetic information? Consider: privacy and confidentiality; informed consent; discrimination; insurance coverage; adoption; and fairness in risk assessment.
- Is the court, and the judiciary, ready to deal with these issues?
- Can human beings be entrusted with the control of their own inheritance?

For a good overview of some of these issues see: Furrow et al (1998). "Legal, Social and Ethical Issues in Human Genetics." *Bioethics and the Law*, West Publishing, pps 150-178.



A national snapshot of judges experience with DNA evidence ...

Although 34% of judges surveyed had no experience with DNA evidence in the courtroom, 76% of those judges who reported that they had had experience with DNA evidence were only somewhat familiar with the methods and procedures of DNA analysis.

Of those judges reporting experience with DNA evidence, 26% noted that the most common type of case involving a proffer of DNA was a sexual assault case. 24% reported that the most common cases proffering DNA evidence, in their experience, were cases where paternity of a child is at issue. 14% of judges saw DNA evidence proffered most often in criminal homicide cases.

What Comprises DNA?⁽³⁾

DNA, or deoxyribonucleic acid, is often called the blueprint of life. It contains all the genetic material that comprises an individual and makes that individual unique. DNA is a double-stranded molecule that contains the genetic code. DNA is composed of 46 rod-shaped chromosomes; 23 of which are inherited from the mother and 23 of which are inherited from the father. Each chromosome has the shape of a double-helix, as first described in 1952 by scientists James Watson and Francis Crick. A piece of a chromosome that dictates a particular trait is called a gene. Genes are the fundamental units of heredity. Collectively, the genes of an organism inform nearly every aspect of the development and formation of that organism.

A gene is found at a particular site, or locus, on a particular chromosome. Each distinctive sequence or configuration that may be found is an allele. Normal individuals have two copies of each gene at a given locus -- one from the father and one from the mother. Consequently, at each locus examined by DNA tests, a person typically has two alleles, one maternal and one paternal. This pair of alleles is a genotype. A locus on a DNA molecule where all humans have the same genetic code is called monomorphic. However, genes vary. An individual may receive the genetic code for blue eyes from her mother and the genetic code for brown eyes from her father. A locus where the allele differs among individuals is called polymorphic, and the difference is known as polymorphism. The set of genotypes possessed by a person at two or more loci is a multi-locus genotype or DNA profile.

Structurally, DNA is a double helix -- two strands of genetic material spiraled around each other. Each strand has a "backbone" made of sugar and phosphate groups and a sequence of nitrogenous bases, also called nucleotides, attached to the sugar groups. A base is one of four chemicals (adenine, guanine, cytosine and thymine). The two strands of DNA are connected at each base. Each base will only bond with one other base, as follows: Adenine (A) will only bond with thymine (T), and guanine (G) will only bond with cytosine (C). Code messages (genes) are stored along a chromosome in sequences of these chemical bases. These linkages, called **base pairs**, are

DNA (Deoxyribonucleic Acid): a double-stranded molecule that contains the genetic code; composed of 46 rod-shaped chromosomes; 23 of which are inherited from the mother and 23 of which are inherited from the father

Chromosome: a threadlike structure that carries genetic information arranged in a linear sequence; in humans, it consists of a complex of nucleic acids and proteins

Gene: the fundamental unit of heredity; an ordered sequence of nucleotide base pairs to which a specific production or function can be assigned

Locus: a specific, physical location on a chromosome
Allele: alternative form of a genetic locus (e.g., at a locus for eye color there might be alleles resulting in blue or brown eyes); alleles are inherited separately from each parent

Genotype: the genetic constitution of an organism
Monomorphic: a locus on a DNA molecule where all humans have the same genetic code

Polymorphic: the existence of more than one form of a genetic trait

DNA Profile: the set of genotypes possessed by a person at two or more loci is a multi-locus genotype or DNA

the identification guide for DNA. The human genome is comprised of about three billion base pairs.

The chemical structure of everyone's DNA is the same. The only difference between people (or any animal) is the order of the base pairs. There are so many millions of base pairs in each person's DNA that every person has a different sequence. Using these sequences, every person could be identified solely by the sequence of their base pairs. However, because there are so many millions of base pairs, the task would be very time-consuming. Instead, because of repeating patterns in DNA, scientists are able to use a shorter method.

These patterns do not, however, give an individual "fingerprint," rather, they are able to determine whether two DNA samples are from the same person, related people, or non-related people. Scientists use a small number of sequences of DNA that are known to vary among individuals a great deal, and analyze those to get a certain probability of a match.

DNA tests are useful for identification because DNA profiles are highly variable across different people, making it unlikely that two different people will happen to have exactly the same profile. Of course, DNA profiles are not necessarily unique because different individuals may, by chance, have the same genotypes in one or more loci. The likelihood of such a chance similarity depends on both the rarity of the matching genotype at each locus and the number of loci examined.

Forensic Use vs. Paternity Case Use of DNA Comparisons

DNA is used differently in forensic cases than in paternity cases. In the paternity case, there is no need to use a population database, as there is with forensic DNA testing. In criminal cases, the DNA is extracted from blood or tissue at the scene without the knowledge of whose DNA it is. The laboratory then uses statistics to determine the probability that the DNA found at the scene matches the known sample of the suspect's DNA. This is not done in paternity cases. Instead, the DNA from the child is compared to the parents' DNA to determine the likelihood of paternity by the probability that certain genetic markers would show up in the child's DNA.

profile

Nucleotide: the unit of DNA consisting of one of four bases - adenine, guanine, cytosine, or thymine - attached to a phosphate-sugar group

Suppose one strand of DNA looks like this:

A-A-C-T-G-A-T-A-G-G-T-C-T-A-G

The DNA strand bound to it will look like this:

T-T-G-A-C-T-A-T-C-C-A-G-A-T-C

Together, the section of DNA would be represented like this:

T-T-G-A-C-T-A-T-C-C-A-G-A-T-C

A-A-C-T-G-A-T-A-G-G-T-C-T-A-G

DNA testing in paternity cases involves comparing DNA from a child to the parents DNA to determine the likelihood of paternity by the probability that certain genetic markers would show up in the child's DNA; there is no need to use a population database

Making a DNA Print

There is currently more than one method of obtaining a DNA profile, but basically the process can be simplified as follows:

1. Collect the sample: DNA is extracted from traces of blood, semen, saliva, urine, hair roots, or bone - wherever nucleated cells are found.

2. Extract and purify DNA: The sample of DNA is treated with chemicals to break open the cells. In a centrifuge, DNA is separated from the cells and later purified.

3. Cut DNA into fragments: Enzymes that recognize certain sequences in the chemical base patterns are added to the DNA. These enzymes, proteins that cause a chemical reaction, act like molecular scissors and cut the DNA molecule at specific points, leaving fragments of various lengths. This process is known as **restriction enzyme analysis**. These restriction enzymes recognize a specific sequence normally of four to six bases in size. This target sequence will occur at various frequencies in different DNA molecules (i.e., different individuals) but will have the same number and location within identical DNA molecules (i.e., same individual). This allows DNA to be "cut" into repeatable and recognizable patterns of specific size fragments.

4. Sort fragments by length: The DNA fragments are placed on a bed of gel, and an electric current is applied. The DNA, which is negatively charged, moves towards the positive end. Several hours later the fragments have become arranged by length. This process is known as **gel electrophoresis**. Once the electrophoresis is complete the gel can be stained to allow visualization of the molecules.

5. Split and transfer DNA: Alkaline chemicals are introduced to split the DNA fragments apart. At the same time, a nylon sheet is placed over the gel and covered with layers of paper. The process, known as **Southern blotting** (named after its inventor) allows transfer of molecules (separated by size and/or charge) from a fragile semisolid gel to a membrane matrix. Following transfer to the membrane, the molecules can be subjected to a variety of different tests which can provide

Restriction Enzyme: an enzyme normally found in bacteria which cuts DNA at specific sites; because a restriction enzyme always acts upon DNA in the same manner, a map can be made of a restriction enzymes actions on a known set of molecules

Gel Electrophoresis: technique used to separate molecules such as DNA fragments or proteins; in forensic uses of DNA tests, electric current is passed through a gel, and the fragments of DNA are separated by size; smaller fragments will migrate farther than larger pieces

Southern Blotting: the transfer of molecules which have been separated by electrophoresis from the gel onto a membrane (named for the inventor)

Hybridization: the process of joining two complementary strands of DNA, or of DNA and RNA, together to form a double-stranded helix

Denaturing: a process by which the hydrogen bonds on the original double-stranded DNA are broken, leaving a single strand of DNA whose bases are available for hydrogen bonding

information concerning their quantity, genetic relatedness, size, etc.

6. Hybridization: Hybridization is the coming together, or binding, of two genetic sequences. The binding occurs because of the hydrogen bonds between base pairs. Between an A base and a T base, there are two hydrogen bonds; between a C base and a G base, there are three hydrogen bonds. When making use of hybridization in the laboratory, DNA must first be denatured (unzipped), usually by using heat or chemicals. **Denaturing** is a process by which the hydrogen bonds of the original double-stranded DNA are broken, leaving a single strand of DNA whose bases are available for hydrogen bonding. Once the DNA has been denatured, a single-stranded radioactive **probe** can be used to see if the denatured DNA contains a sequence similar to that on the probe. The denatured DNA is put into a plastic bag along with the probe and some saline liquid; the bag is then shaken to allow sloshing. If the probe finds a fit, it will bind to the DNA.

7. Make a print and analyze it: X-ray film is exposed to the nylon sheet containing the radioactive probes. Two dark bands develop at the probe sites. If an X-ray is taken of the blot after a radioactive probe has been allowed to bond with the DNA on the paper, only the areas where the radioactive probe binds will show up on the film. This picture is termed an **autoradiogram**, and allows researchers to identify, in a particular person's DNA, the occurrence and frequency of the particular genetic pattern contained in the probe.

The PCR Analysis Method

Polymerase Chain Reaction (PCR) is an amplification technique used when the DNA sample is small or when the sample is degraded by chemical impurities or damaged by environmental conditions. It is a process that yields copies of desired DNA through repeated cycling of a reaction that involves the enzyme DNA polymerase. This process mimics the DNA's own replication process in order to make up to millions of copies of the original genetic material in only a few hours.

Although PCR is only an amplification technique, PCR is commonly used to describe an alternative analysis technique for testing for the presence of matching alleles. Instead of measuring the length of DNA fragments, allele-specific probes

Probe: short segment of DNA that is labeled with a radioactive or other chemical tag and then used to detect the presence of a particular DNA sequence through hybridization to its complementary

Autoradiogram: an x-ray film image showing the position of radioactive substances; sometimes called an autorad

Polymerase Chain Reaction (PCR):

an amplification technique used when the DNA sample is small or when the sample is degraded by chemical impurities or damaged by environmental conditions

are used to determine whether a specific allele is present. The usual format for the use of these probes is to introduce them onto a nylon membrane where the PCR has already been placed. This method is called 'dot blotting' because the samples are spotted as 'dots' on the membrane. An alternative method is called 'reverse dot blotting' because instead of introducing probes onto a strip of PCR product strands, strands of probes are immobilized on a test strip and the test strip is then immersed in the PCR product solution, causing the PCR product to hybridize only to its complementary probe. In the allele-specific technique, the spots where DNA fragments have combined indicate a 'yes' answer - the targeted alleles are present. If the types do not match, the next step is to perform a statistical analysis of the population frequency of the allele types to determine the probability of such a match being observed by chance in a comparison of samples from different persons.

VNTR Profiling

Every strand of DNA has pieces that contain genetic information which informs an organism's development (exons) and pieces that, apparently, supply no relevant genetic information at all (introns). Although the introns may seem useless, it has been found that they contain repeated sequences of base pairs. These sequences, called Variable Number Tandem Repeats (VNTRs) can contain anywhere from twenty to one hundred base pairs. Every human being has some VNTRs. To determine if a person has a particular VNTR, a Southern blot is performed, and then the Southern blot is probed, through a hybridization reaction, with a radioactive version of the VNTR in question. The pattern which results from this process is what is often referred to as a DNA fingerprint.

A given person's VNTRs come from the genetic information donated by her parents; she could have VNTRs inherited from her mother or father, or a combination, but never a VNTR that neither of her parents had. Because VNTR patterns are inherited genetically, a given person's VNTR pattern is more or less unique. The more VNTR probes used to analyze a person's VNTR pattern, the more distinctive and individualized that pattern, or DNA fingerprint, will be. (4)

Variable Number Tandem Repeats (VNTRs): repeated sequences of base pairs found in every strand of DNA

Applications of DNA Testing

Paternity and Maternity: Because a person inherits his VNTRs from his parents, VNTR patterns can be used to establish paternity and maternity. The patterns are so specific that a parental VNTR pattern can be reconstructed even if only the children's VNTR patterns are known (the more children produced, the more reliable the reconstruction). Parent-child VNTR pattern analysis has been used to solve standard father-identification cases as well as more complicated cases of confirming legal nationality and, in instances of adoption, biological parenthood.

Criminal Identification and Forensics: DNA isolated from blood, hair, skin cells, semen, or other genetic evidence left at the scene of a crime can be compared, through VNTR patterns, with the DNA of a criminal suspect to determine guilt or innocence. VNTR patterns are also useful in establishing the identity of a homicide victim, either from DNA found as evidence or from the body itself.

Personal Identification: The notion of using DNA fingerprints as a sort of genetic bar code to identify individuals has been discussed, but this is not likely to happen anytime in the foreseeable future. The technology required to isolate, keep on file, and then analyze millions of very specified VNTR patterns is both expensive and impractical. Social security numbers, picture ID, and other more mundane methods are much more likely to remain the prevalent ways to establish personal identification.

Potential Sources of Error in DNA Testing

- Evidence samples are sensitive to mixture and contamination (e.g., samples might contain a mixture of DNA from multiple sources and might be contaminated with chemicals that can interfere with the analysis)
- 'Band Shifting': DNA samples may migrate at different speeds and yield shifted patterns – 'band shifting' could cause two DNA samples from the same person to show
- The possibility of human error exists at every stage of the process – if an analyst refuses to declare a match unless the DNA prints are identical in all respects, the declared non-match could result in false negatives (i.e., two samples of DNA may actually come from the same individual but the prints are interpreted as a negative match)

Using DNA Profiles for Comparison

Simply making a DNA profile is not useful without a means of comparison to other known (or unknown) samples. The essence of this process is to measure the bands from both samples and compare them. In order to declare a match, the bands do not need to line up exactly, but need to fall within a certain distance of one another. The acceptable deviation rate varies from one laboratory to another.

The issue of matching DNA profiles has caused some debate in the scientific community. There have been complaints that the standards by which a match is declared are not objective and that there is often too much interpretation of data when a match is declared. If no match has been declared between the known sample and the sample in question, then the inquiry ends there. Likewise, if the sample in question was contaminated, no match will be declared and the inquiry ends there. However, if a match is declared, the next step is to determine what, if any, meaning that match has. The question as to whether the match has any meaning arises because unless one knows how common such a match would be, the importance of the match is unknown.

If a profile match is declared, it means only that the DNA profile of the suspect is consistent with that of the source of the crime sample. The crime sample may be from the suspect or from someone else whose profile, using the particular probes involved, happens to match that of the suspect. Expert testimony concerning the frequency with which the observed alleles are found in the appropriate comparison population is necessary for the finder of fact to make an informed assessment of the incriminating value of this match.

The frequency with which an individual allele occurs in the comparison population is taken to be the probability of a coincidental match of that allele. These individual probabilities of a coincidental match are combined in an estimate of the probability of a coincidental match on the entire profile.

Statistical Calculations Used In DNA Testing

Population Genetics

In determining the probability of a certain DNA profile being the same as a profile in question, experts use a previously constructed database developed by the laboratory. Databases are generally divided into broad racial classification and each database is composed of the DNA of a few hundred people. In that database, the experts find and extract specific polymorphic alleles from known sites and calculate the odds from those alleles. Thus, allele A may be found in 1 in 10, allele B may be found in 1 in 20, allele C is 1 in 5, and allele D is one in 100. If there are sites A, B, C, and D, the sample in question must match the known sample in those four sites in order for there to be a match. The expert then calculates the odds of that match.

The Product Rule

In order to calculate the odds of the match being from the same sample, the product rule is employed. The odds of the first sample are multiplied with the odds of the second, the third and the fourth (e.g., $1/10 \times 1/20 \times 1/5 \times 1/100 = 1/100,000$). Thus, the odds of a match are one in 100,000 that the person

Problems with Determining Probability

VNTRs, because they are results of genetic inheritance, are not distributed evenly across all of the human population. A given VNTR cannot, therefore, have a stable probability of occurrence; it will vary depending on an individual's genetic background. The difference in probabilities is particularly visible across racial lines. Some VNTRs that occur very frequently among Hispanic Americans will occur very rarely among Caucasians or African Americans. Currently, not enough is known about the VNTR frequency distributions among ethnic groups to determine accurate probabilities for individuals within those groups; the heterogeneous genetic composition of interracial individuals, who are growing in number, presents an entirely new set of questions. Further experimentation in this area, known as population genetics, has been surrounded with and hindered by controversy, because the idea of identifying people through genetic anomalies along racial lines comes alarmingly close to the eugenics and ethnic purification movements of the recent past, and, some argue, could provide a scientific basis for racial discrimination.

Errors in the hybridization and probing process must also be figured into the probability, and often the idea of error is not acceptable. Most people will agree that an innocent person should not be sent to jail, a guilty person allowed to walk free, or a biological mother denied her legal right to custody of her children, simply because a lab technician did not conduct an experiment accurately. When the DNA sample available is minuscule, this is an important consideration, because there is not much room for error, especially if the analysis of the DNA sample involves amplification of the sample (creating a much larger sample of genetically identical DNA from what little material is available). That is, if the wrong DNA is amplified (e. g., a skin cell from the lab technician) the consequences can be profoundly detrimental.

in question has the same genetic profile as the sample taken from the scene.

Areas of Controversy in Statistical Calculations Used in DNA Testing

The possibility of false assumptions --What if the profile for allele D shows up in 1 out of 100 of the individuals in the database but in real life shows up in 1 out of 20 people? Problems like this can arise because the databases used by the laboratories are small. Also, the design of the database may create errors. For example, the Native American database referred to in the case of *Springfield v. State* (860 P.2d 435, Wyo. 1993) was composed of 200 individuals from the Navajo, Cherokee, and Cheyenne nations. The individual in question, however, was a member of the Crow tribe, which was not represented in the database at all.

The 'linkage equilibrium' problem -- Scientists must be sure when designing DNA studies that the alleles to be tested are truly unrelated, because the product rule is accurate only when the variables are independent of one

The 'Hardy-Weinberg equilibrium problem' -- The basis for statistical calculations

The Ceiling Principle and the Modified Ceiling Principle

In the late 1980s and early 1990s there was a debate over the significance of random mating in the construct of the database (i.e., the Hardy-Weinberg equilibrium problem). A compromise position commonly termed the 'ceiling principle' or 'modified ceiling principle' was proposed as a useful and conservative method for calculating statistical probabilities. The ceiling principle approach incorporates into the calculation "the greatest observed frequency of particular alleles within a given number of randomly selected population groups." (5) To determine ceiling frequencies, random samples of 100 persons from each 15-20 populations that represent groups relatively homogenous genetically are drawn. Then, either the largest frequency of those populations or 5%, whichever is larger, is taken as the ceiling frequency. (6) In some areas, the 'modified ceiling principle' has been used. This method uses a 10% frequency for each allele or, if it is higher, the highest frequency found for an allele among the laboratory's broad racial databases. (7) Beginning in 1994, the scientific landscape started shifting on the issues surrounding population restructuring. Now, scientists are concluding that these issues should not have a significant impact on final calculations. Stating that the "DNA wars are over," they have come out in favor of using the product rule and no longer requiring the ceiling or modified ceiling principle. (8) The areas of continuing concern with respect to DNA testing are on a case-by-case basis, where the validity of individual results will be challenged. Rather than the initial concerns about population substructure, the newest concerns with respect to DNA evidence center around the possibility of error in laboratories.

Who are the Relevant Scientific Communities?

Most courts have accepted molecular biologists as the relevant scientific community with respect to laboratory techniques of

in population genetics rests on the belief in a truly random population (i.e., a population that mates randomly and thus mixes the gene pool evenly).



Determining the admissibility of DNA evidence

Judges who reported that they had experience with DNA evidence in the courtroom were also asked what factors they considered when determining the admissibility of such evidence. These judges reported the following factors as important considerations (in order of frequency of mention):

- the way the DNA sample was obtained, gathered, or preserved
- the qualifications of the expert
- the status of the laboratory
- the general acceptance of DNA evidence generally
- the reliability of the testing procedure

When judges who reported that they had experience with DNA evidence in the courtroom were asked what aspect of such evidence is most problematic in determining its admissibility, the following factors were identified (in order of frequency of mention):

- determining how the sample was obtained,

isolating and probing the DNA. The statistical interpretation of the results has been more properly the province of population geneticists. Individuals from other fields, such as genetic epidemiology and biostatistics, also may have the requisite background to testify about analysis. Forensic scientists or laboratory technicians involved in the analysis often do not have a strong background in the relevant scientific discipline but may be knowledgeable about techniques of sample collection and preservation, forensic laboratory standards and procedures, and proficiency tests.

Issues Pertaining to Sample Quantity and Quality⁽⁹⁾

Several factors may affect a DNA sample's suitability for analysis. For each factor claimed to affect a particular analysis, the court may want to have the expert address whether its influence is likely to cause a false positive result (incorrect identification of the suspect as a potential source of the forensic DNA) or merely an inconclusive or uninterpretable result.

Did the Crime Sample Contain Enough DNA to Permit Accurate Analysis?

To be interpretable, the crime sample must contain enough DNA of sufficiently high molecular weight to allow isolation of longer DNA fragments, which are the most susceptible to degradation. Samples of blood, semen, or other DNA sources may be too small to permit analysis. To the extent that small sample size precludes a full series of tests (three to five probes), it can significantly diminish the power of the DNA testing procedure to distinguish between DNA samples obtained from different individuals. In addition, the unavailability of additional DNA precludes repeated testing that might verify or refute the initial test.

Was the Crime Sample of Sufficient Quality to Permit Accurate Analysis?

Exposure to heat, moisture, and ultraviolet radiation can degrade the DNA sample. Samples may also have been contaminated by exposure to chemical or bacterial agents that alter DNA, interfere with the enzymes used in the testing

- gathered or preserved
- the statistical basis of the conclusions drawn
- determining how the sample was tested
- determining the reliability of the testing procedure

Laboratory Standards

- Laboratories must maintain appropriate documentation on all significant aspects of the DNA analysis procedure, as well as any related documents or laboratory records that are pertinent to the analysis or interpretation of results.
- Laboratory procedures must be validated. Guidelines stress the importance of validation to acquire the necessary information to assess the ability of a procedure to reliably obtain a desired result, determine the conditions under which such results can be obtained, and determine the limitations of the procedure.
- Laboratories must have been subjected to appropriate proficiency testing, preferably by an independent institution.

process, or otherwise make DNA difficult to analyze. Such exposure is known as an 'environmental insult.' Although old samples of DNA may be analyzed successfully, attention must be given to possible sample degradation due to age.

How Many Sources of DNA are Thought to be Represented in the Crime Sample?

Often, the expected composition of a crime sample of DNA can be narrowed to a single perpetrator, or single victim, or both. However, a crime sample may be thought to include DNA from multiple sources, as when more than one person is thought to have contributed to the crime sample of blood or semen. Male and female DNA extracted from such a sample may be distinguished, as can same-sex DNA where the alternative sources are known and available for testing (e.g., a rape victim's husband). The presence of multiple, same-sex samples from unknown sources adds additional complications. Mixed samples can be difficult to interpret, although the intensity of the different bands can offer clues.

Questions to consider when evaluating DNA testing procedures

- Was the crime scene contaminated?
- Was the sample too small for proper testing?
- Was there destruction of the sample?
- What about the problem of shifting bands?
- What about the possibility of false assumptions?
- Were poor quality laboratory procedures used?
- Was the population database too small or not appropriately used?

Some Case Law on DNA

The science of DNA fingerprinting changes rapidly. As the science has changed, so have the courts decisions on whether to admit expert testimony. Following the 1992 NRC Report, a number of courts embraced the ceiling principle and modified ceiling principle. Since that time however, scientists appear to have reached agreement that the product rule provides a more accurate analysis and that there is no need to use the ceiling principle. This change in position has been recognized by a number of courts across the country (e.g., *State v. Johnson*, 922 P.2d 294, Ariz. 1996; *Clark v. State*, 679 So.2d 321, Fla. App. 1996; *State v. Marcus*, 683 A.2d221, N.J. Super. 1996). Furthermore, several courts have eliminated the use of any of the ceiling principles, deciding that the product rule provides a proper basis for statistical analysis (e.g., *People v. Pope*, 672 N. E.2d 1321, Ill. App. 1996; *Armstead v. State*, 673 A.2d 221, Md. 1996; *People v. Freeman*, 571 N. W.2d 276, Neb. 1997).

In March 1998 the Supreme Court of Arizona decided that PCR testing was admissible, after finding that it met the Frye standard of scientific admissibility (*State v. Tankersley*, 1998 WL 107864, Ariz. 1998). The court noted that the overwhelming consensus among scientists is that so long as proper procedures are followed, the results should be reliable (at 5). Arizona is not alone in its acceptance of either the PCR method of testing DNA or the admissibility of DNA evidence without the use of ceiling principles. In *State v. Stills*, the Supreme Court of New Mexico approved of the method, quoting a commentator who stated that PCR analysis has received overwhelming acceptance in the scientific community and the courts (957 P.2d 51, N.M., 1998 at 4). A number of

courts, both state and federal, have held PCR evidence admissible (e.g., *Brodine v. State*, 936 P.2d 545, Alaska App. 1997; *State v. Isley*, 936 P.2d 275, Kan. 1997; *State v. Harvey*, 699 A.2d 596, N.J. 1997; *State v. Begley*, 956 S.W.2d 471, Tenn. 1997; *United States v. Lowe*, 954 F. Supp. 401, D. Mass. 1996).

Since 1990, a number of states have enacted statutes governing the admissibility of DNA evidence. Virginia, for example, enacted a statute providing that DNA testing shall be deemed to be a reliable scientific technique and the evidence of a DNA profile comparison may be admitted to prove or disprove the identity of any person (Va Code 19.2-270.5; see also Md Cts & Jud Proc Code 10-915; Minn Stat 634.25, 634.26; Wash Rev Code 43.43.752 through 43.43.758; and La Rev Stat 15.441.1).

Before going any further, stop and reflect ...

How is the evaluation of DNA evidence and testing procedures like or unlike other forms of evidence?

Are Daubert-like criteria such as error rate and falsifiability more readily applied to evaluations of DNA evidence than other forms of evidence? Why or Why not?

Questions to consider when evaluating DNA evidence ...

What kind of DNA test is at issue?

- How was the DNA extracted?
- How much DNA was extracted?
- How were the autoradiogram bands scored and sized?
- What were the match criteria and how were they applied?
- How was the DNA typed?

What kind of tissue was being tested? (Blood, Semen, Other)

Was the sample mixed with tissue (e.g., blood) from several possible sources?

What kind of statistical estimates were made based on the match?

- What database was used for frequency statistics?
- Were there any allowances made for population structure?

Were there estimates of error rates?

- Were there possible errors in collecting and/or contaminating samples?
- Have laboratory procedures used been reviewed?
- Are data available on laboratory error rates?
- Were there proper controls?

What is the evidence being used for?

- Identification of a criminal suspect (Rape, Other)
- Identification of father in a paternity case
- Determination of lineage of plants or livestock for patent or trade secret purpose
- Other

What are the background and qualifications of the expert?

- Expert from an academic institution
- Expert from commercial testing laboratory
- Expert from government crime laboratory
- Other

How was the evidence presented?

- Were there invalid claims of uniqueness?
- Was there failure to consider genetic correlation among relatives?
- Were there invalid assertions regarding error rates?
- Were there any problems related to a mixed sample?

From the Practitioner's Guide to the Federal Judicial Center's Reference Manual on Scientific Evidence, pg. 265.

Critical Questions Reviewed

- What kind of DNA test is at issue?
- What kind of tissue is being tested?
- Was the sample mixed with tissue from several possible sources?
- Was the crime scene contaminated?
- Was the sample too small for proper testing?
- Was there destruction of the sample?
- What about the problem of shifting bands?
- What about the possibility of false assumptions
- Were poor quality laboratory procedures used?
- Was the population database used too small or not appropriately used?
- What kind of statistical estimates were made based on the match?
- Were there estimates of error rates?
- What is the evidence being used for?

What are the background and qualifications of the expert?

Endnotes:

1. "Introducing the Human Genome: The Recipe for Life" from *To Know Ourselves: The U.S. Department of Energy and The Human Genome Project* (July, 1996).
2. Significant articles and books relied upon for this chapter include: Wolfe, S.L. (1993). *Molecular and Cellular Biology*. Belmont, CA: Wadsworth Publishing Company; Kaye, D.H. and Sensabaugh, G. F. (1999). *Reference Guide on Forensic DNA Evidence*. Forthcoming, Federal Judicial Center; Federal Judicial Center (1994). *Reference Manual on Scientific Evidence*.; Thompson W. and Ford (1989). "DNA Typing: Acceptance and Weight of the New Genetic Identification Tests." *Virginia Law Review*, Vol. 75, pg. 45; Thompson, W. (1993). "Evaluating the Admissibility of New Genetic Identification Tests: Lessons From the 'DNA War.'" *Journal of Criminal Law and Criminology*, Vol. 84, pg. 22; Scheck, B. (1994). "DNA and Daubert." *Cardozo Law Review*, Vol. 15, pg. 1959; National Research Council Committee on DNA Forensic Science (1996). *The Evaluation of Forensic DNA Evidence*. Washington: National Academy Press.
3. This explanation of DNA and DNA testing is intended as an introduction to the subject for those who may have limited backgrounds in biological science. While accurate, this explanation involves somewhat of an over-simplification. Although intended to be informative, this is a brief and incomplete explanation of a complex subject.
4. Another category of repetitive non-coding DNA sequences is characterized by short core repeats (two to seven base pairs in length) and is known as 'short tandem repeats (STRs)'.
5. *Commonwealth v. Lanigan*, 596 N.E. 2d 311 (Mass. 1992).
6. NRC Report, *Supra* note 2, at 13.
7. NRC Report, *Supra* note 2, at 83 and 92.
8. Landar and Budowle. (1994). "DNA Fingerprinting Dispute Laid to Rest." *Nature*, Vol. 371, pg. 735.
9. This section is adapted from the Federal Judicial Center's

Reference Manual on Scientific Evidence. (1994)

Glossary

allele alternative form of a genetic locus (e.g., at a locus for eye color there might be alleles resulting in blue or brown eyes); alleles are inherited separately from each parent

autoradiogram an x-ray film image showing the position of radioactive substances; sometimes called an "autorad"

band shifting the phenomenon of DNA fragments in one lane of a gel migrating slower or faster than fragments in another lane; as visualized on an autoradiogram, the overall patterns would be the same, but out of register; factors responsible for band shift include contaminants, salt concentration, and DNA concentration

base pair two complementary nucleotides (adenine and thymine or guanine and cytosine) held together by weak bonds; two strands of DNA are held together in the shape of a double helix by the bonds between base pairs

chromosome a threadlike structure that carries genetic information arranged in a linear sequence; in humans, it consists of a complex of nucleic acids and proteins

denaturing a process by which the hydrogen bonds on the original double-stranded DNA are broken, or 'unzipped,' leaving a single strand of DNA whose bases are available for hydrogen bonding

DNA deoxyribonucleic acid; the molecule that encodes genetic information; double-stranded helix held together by weak bonds between base pairs of nucleotides

DNA profile set of genotypes possessed by a person at two or more loci is a "multi-locus genotype" or "DNA profile"

gel electrophoresis technique used to separate molecules such as DNA fragments or proteins; in forensic uses of DNA tests, electric current is passed through a gel, and the fragments of DNA are separated by size; smaller fragments will migrate farther than larger pieces

gene the fundamental unit of heredity; an ordered sequence of nucleotide base pairs to which a specific product or function can be assigned

genome all the genetic material in the chromosomes of a particular organism; its size is generally given as its total number of base pairs

genotype the genetic constitution of an organism

hybridization the process of joining two complementary strands of DNA, or of DNA and RNA, together to form a double-stranded molecule

locus a specific, physical position on a chromosome

monomorphic locus on a DNA molecule where all humans have the same genetic code

nucleotide the unit of DNA consisting of one of four bases - adenine, guanine, cytosine, or thymine - attached to a phosphate-sugar group

PCR Polymerase Chain Reaction; a process through which repeated cycling of the reaction reproduces a specific region of DNA, yielding millions of copies from the original

polymorphism the existence of more than one form of a genetic trait

probe short segment of DNA that is labeled with a radioactive or other chemical tag and then used to detect the presence of a particular DNA sequence through hybridization to its complementary sequence

restriction enzyme an enzyme normally found in bacteria which cuts DNA at specific sites (i.e., each time a specific nucleotide pattern occurs); because a restriction enzyme always acts upon DNA in the same manner, a map can be made of a restriction enzyme's actions on a known set of molecules

southern blotting the transfer of molecules which have been separated by electrophoresis from the gel onto a membrane (named for the inventor)

Variable Number Tandem Repeats repeated sequences of base pairs found in every strand of DNA; apparently do not supply any genetic information (VNTR)

Suggested Readings

Bell, R. B. (1997). "Instructing Juries in Genomic Evidence Cases." *Judges' Journal*, Vol. 36(3). pg. 42.

Coleman, H. and Swenson, E. (1994). *DNA in the Courtroom, A Trial Watcher's Guide*. Genelex Corporation.

Federal Judicial Center. (1994). *Reference Manual on Scientific Evidence*. Washington, D.C.: Government Printing Office.

National Research Council. (1996). *The Evaluation of Forensic DNA Evidence*. Washington: National Academy Press.

Inman, K. and Rudin, N. (1997). *An Introduction to Forensic DNA Analysis*. CRC Press.

Kaye, D.H., and Sensabaugh, G.F. (1999). *Reference Guide on Forensic DNA Evidence*. Forthcoming, Federal Judicial Center.

Lee, H.C., and Gaensslen, R.E. (1990). *DNA and Other Polymorphisms in Forensic Science*. Chicago: Year Book Medical.

Scheck, B. (1994). "DNA and Daubert." *Cardozo Law Review*, Vol. 15, pg. 1959.

Thompson, W.C. (1993). "Evaluating the Admissibility of New Genetic Identification Tests: Lessons from the 'DNA War.'" *Journal of Criminal Law and Criminology*, Vol. 84, pg. 22.

Thompson and Ford. (1989). "DNA Typing: Acceptance and Weight of the New Genetic Identification Tests." *Virginia Law Review*, Vol. 75, pg. 45.

Wolfe, S.L. (1993). *Molecular and Cellular Biology*. Belmont, CA: Wadsworth Publishing Company

For more information on the Human Genome Project

Human Genome Project Report, Part 1, Overview and Progress. (November, 1997). Human Genome Management Information System for the U.S. Department of Energy Research, Office of Biological and Environmental Research.

Copies available upon request from the Human Genome Management Information System (HGMIS); Oak Ridge National Laboratory, 1060 Commerce Park, Oak Ridge, TN 37830. Electronic versions are accessible via the following websites:

http://www.er.doe.gov/production/ober/HELSDRD_top.html

<http://www.ornl.gov/hgmis/research.html>

To Know OURSELVES: The U.S. Department of Energy and The Human Genome Project (July, 1996).

Copies available from Human Genome News, Oak Ridge National Laboratory, 1060 Commerce Park, Oak Ridge, TN 37830.

Information about the Human Genome Project can also be found at the following website:

<http://www.ornl.gov/hgmis>

--- CHAPTER 10 --- FRONT PAGE ---

The Judicial Role in Evidentiary Decision-Making

Learning Objectives:

Upon completion of this section, the judge-student should be able to:

- ◆ Describe his or her decision-making role with respect to the admissibility of scientific evidence;
- ◆ Describe some of the recent changes in evidence law; and
- ◆ Discuss strategies used to manage scientific evidence.

The Daubert Guidelines:

- **whether the theory or technique can be, and has been tested;**
- **whether the evidence has been subjected to peer review and publication;**
- **the known or potential error rate associated with applications of a theory; and**
- **the general acceptance of the theory or technique in question.**

Overhead 2.2

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 3: An Introduction to the Philosophy of Science, pgs.9-18).

© 1999

FRE 401 and 402: All relevant evidence is admissible, except as otherwise provided ... Evidence which is not relevant is not admissible. Relevant evidence is defined as that which has any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.

FRE 403: Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.

FRE 702: If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.

Some Strategies for Managing Scientific Evidence

Require or encourage early exchange of reports of prospective trial experts

Hold a pre-trial hearing on the admissibility of expert testimony

Ask the parties to provide special education or instruction to the court

Engage in your own research in specific area of expert testimony

Ask clarifying questions from the bench

Allow jurors to question experts directly or through the court

Limit the number of experts who testify on a particular issue

Permit expert testimony on videotape

Appoint an independent expert of the court

Overhead 2.4

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 3: An Introduction to the Philosophy of Science, pgs.9-18).
© 1999

Some Possible Problems with Expert Testimony

Delay in trial schedule caused by unavailability of experts

Indigent party unable to retain expert to testify

Failure of party(ies) to provide discoverable information concerning retained experts

Excessive expense of party-hired experts

Attorney(s) unable to adequately cross-examine expert(s)

Experts abandon objectivity and become advocates for side that hired them

Expert testimony appears to be of questionable validity

Expert testimony not comprehensible to trier of fact

Expert poorly prepared to testify

Overhead 2.5

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 3: An Introduction to the Philosophy of Science, pgs.9-18).
© 1999

An Introduction to the Philosophy of Science

Learning Objectives:

Upon completion of this section, the judge-student should be able to:

- ◆ Understand science as a particular way of thinking;
- ◆ Describe some of the basic differences between legal and scientific disciplines;
- ◆ Discuss the general tenets of fundamental perspectives in the philosophy and sociology of science;
- ◆ Understand why it is important that judges are informed about different philosophies of science; and
- ◆ Articulate a personal philosophy of science and how this personal philosophy might influence decision-making.

Overhead 3.1

From: "A Judge's *Deskbook* on the Basic Philosophies and Methods of Science." (Chapter 3: An Introduction to the Philosophy of Science, pgs. 19-40.). © 1999

Why is it important for judges to understand the various perspectives of the philosophy of science?

- **challenges the notion that objective scientific ‘truth’ is out there to be discovered;**
- **provides a more complete picture of the processes of science and scientific discovery;**
- **provides a more principled method for evaluating the validity, reliability, and applicability of scientific evidence;**
- **provides context for the analysis of the judicial role in admissibility decision-making; and**
- **makes explicit the implicit assumptions underlying various scientific methods and theories.**

Overhead 3.2

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 3: An Introduction to the Philosophy of Science, pgs. 19-40.).
© 1999

Science as a way of thinking ...

- ◆ **science seeks the systematic organization of information about the world and, in so doing, discovers new relationships among natural phenomena;**
- ◆ **science endeavors to explain why phenomena occur and how they are related; and**
- ◆ **scientific explanations must be formulated in a way that makes them subject to empirical testing**

Falsification Principle (Sir Karl Popper)

falsification

- ◆ the refutation, or potential refutation, of a theory
- ◆ testability
- ◆ a theory or technical procedure is falsifiable if it can be subjected to testing which could prove it to be incorrect or false

Overhead 3.4

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 3: An Introduction to the Philosophy of Science, pgs. 19-40.).

© 1999

Paradigms and Revolutions (Thomas Kuhn)

paradigm

- ◆ a world view that reflects assumptions about the social world and how science should be conducted, as well as what constitutes legitimate problems, methods, solutions, and standards of “proof”
- ◆ a world view sanctioned by the relevant scientific community
- ◆ falsification is not possible because it implies the existence of absolute standards of evidence which transcend any individual paradigm

***Science advances through a progression of paradigm shifts
— through a progression of revolutions.***

Overhead 3.5.1

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 3: An Introduction to the Philosophy of Science, pgs. 19-40.).
© 1999

Development of Paradigms: Series of Five Stages (Kuhn)

- 1. Pre-science or Immature Science**
- 2. Normal Science**
- 3. Crisis**
- 4. Revolutionary Science**
- 5. Resolution**

Overhead 3.5.2

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 3: An Introduction to the Philosophy of Science, pgs. 19-40.).

© 1999

Some Questions to Consider ...

- ▶ ***Does science progress in a rational way, or is it subject to historical and cultural factors, like other systems of belief?***
- ▶ ***Is the only difference between science and other disciplines the authority our society invests in science?***
- ▶ ***To what extent do the socially instilled biases of scientists influence the science they produce?***
- ▶ ***What role does politics play in science?***

Social Constructivism ...

- ◆ **the philosophical position that truth is contingent and conditional and there are multiple perspectives and multiple realities**
- ◆ **recognition that people in different geographic, cultural, or social locations may construe knowledge, truth, and relevance in different ways, and that each of these different ways of knowing are legitimate and worthy**

Overhead 3.7

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 3: An Introduction to the Philosophy of Science, pgs. 19-40.).

© 1999

Quantitative and Qualitative Research

Learning Objectives:

Upon completion of this section, the judge-student should be able to:

- ◆ Understand the differences between quantitative and qualitative research, including:
 - ◆ the differing assumptions underlying the two approaches;
 - ◆ the methods typical of each approach; and
- ◆ Understand and discuss how these two approaches to research differentially influence the scientific questions asked, the methodologies employed, and the conclusions drawn, and why this is important to consider.

Overhead 4.1

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 4: Quantitative and Qualitative Research, pgs. 41-47).
©1999

Quantitative Research

- ▶ a process of inquiry based on testing a theory composed of variables, measured with numbers, and analyzed using statistical techniques
- ▶ the goal of quantitative methods is to determine whether the predictive generalizations of a theory hold true

Overhead 4.2.1

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 4: Quantitative and Qualitative Research, pgs. 41-47).

©1999

Assumptions Underlying Quantitative Methods

- reality is objective, “out there” and independent of the researcher, therefore reality is something that can be studied objectively;
- the researcher should remain distant and independent of what is being researched;
- the values of the researcher do not interfere with, or become part of, the research -- research is value-free;
- research is based primarily on deductive forms of logic, and theories and hypotheses are tested in a cause-effect order; and
- the goal is to develop generalizations that contribute to theory that enable the researcher to predict, explain, and understand some phenomenon.

Overhead 4.2.2

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 4: Quantitative and Qualitative Research, pgs. 41-47).

©1999

Qualitative Research

- ▶ a process of building a complex and holistic picture of the phenomenon of interest, conducted in a natural setting
- ▶ the goal of qualitative research is to develop an understanding of a social or human problem from multiple perspectives

Overhead 4.3.1

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 4: Quantitative and Qualitative Research, pgs. 41-47).
©1999

Assumptions Underlying Qualitative Methods

- multiple realities exist in any given situation
- researcher interacts with those he studies and actively works to minimize the distance between the researcher and those being researched;
- researcher explicitly recognizes and acknowledges the value-laden nature of the research;
- research is context-bound;
- research is based on inductive forms of logic; categories of interest emerge from informants (subjects);
- the goal is to uncover and discover patterns or theories that help explain a phenomenon of interest; and
- determinations of accuracy involve verifying the information with informants or "triangulating" among different sources of information

Overhead 4.3.2

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 4: Quantitative and Qualitative Research, pgs. 41-47).
©1999

Qualitative vs. Quantitative Methods

Qualitative

- Involves the examination and interpretation of observations for the purpose of discovering underlying meanings and patterns of relationships
- Does not allow for the same level of objectivity and generalizability of quantitative approaches, but provides much richer, in-depth data which often provide insights into subtle nuances that quantitative approaches might miss
- Very useful for exploratory research and in the early stages of theory development

Quantitative

- Involves the numerical representation and manipulation of observations for the purpose of describing and explaining the phenomena that those observations reflect
- It is argued that quantification allows for more precision in analysis and ease in summarizing data and making inferences
- Attempts to be very objective, controlled, and value-free
- Can lack the depth of some qualitative approaches

These approaches are not necessarily mutually exclusive!

The method(s) selected to use in any research study should be the most appropriate to the questions being asked!

Quantitative and Qualitative Research

???

Questions to consider when evaluating scientific evidence ...

- ◆ Was a quantitative or qualitative research approach adopted?
- ◆ Was the research approach appropriate given the problem investigated and the goals of the research?
- ◆ Was the process of investigation consistent with the underlying assumptions of the research used?
- ◆ Were appropriate types of conclusions drawn given the research approach used?

Overhead 4.5

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 4: Quantitative and Qualitative Research, pgs. 41-47).

©1999

CRITICAL QUESTIONS REVIEWED

- ◆ Do you have a clear understanding of what the research was designed to study? Do you understand the nature of the predicted relationship? That is, did the researcher clearly articulate the statement of the problem?
- ◆ How were independent and dependent variables operationalized? That is, do you clearly understand what each variable means (as indicated by operational definition)?
- ◆ Do the operational definitions adequately capture the conceptual meaning of the variables?
- ◆ Do you have a clear understanding of how the researcher measured changes in the variables?
- ◆ Did the researcher clearly outline the research design? Do you have a good general overview of the major research steps involved?
- ◆ Did the researcher provide adequate justifications for why decisions were made, especially if alternative methods were also appropriate?
- ◆ Did the researcher use an experimental or quasi-experimental design?
- ◆ If an experimental design -- was random assignment used? Was an appropriate control group used?

Overhead 5.14.1

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 5: An Introduction to the Experimental Method, pgs. 49-84). © 1999

CRITICAL QUESTIONS REVIEWED (Continued)

- ◆ If a quasi-experimental design – were appropriate steps taken to avoid, or minimize or eliminate potential threats to validity? Was the appropriate level of care taken in expressing the strength or nature of the causal relationship?
- ◆ Did the researcher appropriately control for extraneous, or confounding, variables that might influence the nature and strength of the relationship between the variables?
- ◆ Did the researcher appropriately acknowledge and diminish threats to validity?
- ◆ How was the reliability of the measure determined?

Handout: The Danger of Inferring Causation from Correlation*

An article published in a newspaper discussed the stressfulness of a variety of occupations. A total of 130 job categories were rated for stressfulness by examining local hospital and death records for evidence of stress-related diseases such as heart attacks, ulcers, arthritis, and mental disorders. The 12 highest and the 12 lowest were as follows:

Most Stressful:

1. Unskilled Laborer
2. Secretary
3. Assembly-line Inspector
4. Clinical Lab Technician
5. Office Manager
6. Foreperson
7. Manager/Administrator
8. Waitress/Waiter
9. Factory Machine Operator
10. Farm Owner
11. Miner
12. House Painter

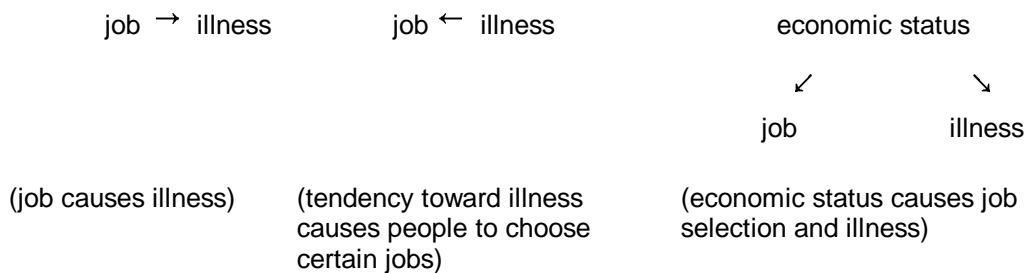
Least Stressful:

1. Clothing Sewer
2. Garment Checker
3. Stock Clerk
4. Skilled Craftsperson
5. Housekeeper
6. Farm Laborer
7. Heavy-Equipment Operator
8. Freight Handler
9. Child-care Worker
10. Factory Package Wrapper
11. College Professor
12. Personnel Worker

The article advises readers who want to remain healthy to avoid the “most stressful” job categories.

However, the evidence may not warrant the author’s advice. Although it is quite possible that diseases are associated with particular occupations, this does not necessarily mean that holding the jobs causes the illnesses. There are a variety of other explanations.

- people with a propensity for heart attacks and ulcers might be more likely to select jobs as unskilled laborers or secretaries — thus, the direction of causation might be that health condition causes job selection rather than the reverse
- some other variable might cause the apparent relationship between job and health
 - the income level might cause both life stress and illness
 - impoverished conditions might cause a person to accept certain jobs and also have more diseases



The possibility that job causes health condition is not be ruled out. In fact, it is quite plausible. However, because the nature of the evidence is correlational, it cannot be said with certainty that job causes illness.

*Adapted from: Kaplan, R.M. and Saccuzzo, D.P. (1982). *Psychological Testing: Principles, Applications, and Issues*. Monterey, CA: Brooks/Cole Publishing Company, pg. 142.

HANDOUT: HYPOTHESIS-TESTING

Statistical hypotheses are made about population parameters or population distributions.

- Do sample data nullify the experimental or research hypothesis?
 - The decision to reject or accept the statistical hypothesis is not made with certainty. Hypothesis-testing provides a means of controlling and assessing the risk of making an incorrect decision.
 - The researcher determines the probability of making a Type I error.
 - The researcher's strategy is to determine the probability that the observed sample statistic could have occurred by chance. The probability associated with sample statistics is determined by calculating a test statistic that has a known probability distribution.
- The decision rule relative to accepting or rejecting the experimental or research hypothesis is generally of the following form: if the test statistic calculated from the sample data indicates that the observed value of the sample statistic is –
 - an "extremely unlikely" chance occurrence, reject the statistical (null) hypothesis.
 - a "likely" chance occurrence, accept the statistical (null) hypothesis

An investigator's decision of what is an "extremely unlikely" chance occurrence or what is a "likely" chance occurrence is determined by the p value or significance level (probability of making a Type I error) selected.

- For example, if $p < .05$, a value of the test statistic is considered--an "extremely" unlikely chance occurrence if the probability associated with the test statistic is equal to or less than the significance level selected --a "likely" chance occurrence if the probability associated with the test statistic is greater than the significance level selected.

Decision Regarding H_0	H_0 is true	H_0 is False
H_0 is Rejected	Type I error	Correct Decision
H_0 is Accepted	Correct Decision	Type II error

The Ten Commandments of Question Construction*

*T*hou shalt not create double-barreled items.

*T*hou shalt not use "no" and "not" or words beginning with "un."

*T*hou shalt match the vocabulary used in items to the vocabulary of those who will respond to them.

*T*hou shalt not use complex grammatical forms.

*W*hen using "agree/disagree" or "true/false" items thou shalt avoid perception of bias toward one point of view

*T*hou shalt not use redundant or irrelevant items.

*T*hou shalt not permit any loaded questions to appear in your questionnaire.

*T*hou shalt not mix response formats within a set of questions.

Thou shalt not permit a non-committal response.

Thou shalt pretest questions before collecting data.

* Adapted from John H. Ellard and T. B. Rogers (1993). *Teaching Questionnaire Construction Effectively: The Ten Commandments of Question Writing. Contemporary Social Psychology, 17, 17-20.*

Overhead 6.9

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science.* (Chapter 6: An Introduction to the Scientific Methods of Survey Research, pgs. 85-112). © 1999.

The Ten Commandments of Question Construction* (Talking Points)

Further explanation to accompany corresponding overhead

(I) Thou shalt not create double-barreled items.

Double-barreled items ask respondents to provide a single answer to a question that contains two or more questions. Answers to such questions are impossible to interpret because there is no way of knowing which part of the question the respondent was answering. For instance, asking respondents to agree or disagree with the statement "I am shy and uncomfortable in social situations" is confronting them with a double-barreled question. Persons who are shy but comfortable or not shy but uncomfortable cannot answer this item except by making assumptions about what you mean by these words.

(II) Thou shalt not use "no" and "not" or words beginning with "un."

Negative qualifiers are to be avoided because they unnecessarily contribute to complexity. This is particularly problematic when they create double negatives such as in this statement: "Most people are not usually uncomfortable with not having to present their names on questionnaires." Insertion of one "not" or "un" can create complexities, which may lead to response errors.

(III) Thou shalt match the vocabulary used in items to the vocabulary of those who will respond to them.

Respondent difficulties with vocabulary can often be traced to the use of jargon (e.g., "I feel most kleptomaniacs are partly schizophrenic."). This item assumes knowledge of the term "kleptomaniac," which may not be the case for all respondents. Those not familiar with the term kleptomaniac will add significant error to the data because they will be responding to what they assume kleptomaniac means, which may not correspond to the intended meaning.

(IV) Thou shalt not use complex grammatical forms.

The time lost by respondents trying to recode statements such as this one into simple sentences causes frustration culminating in response errors: "Persons in this organization, by virtue of the pride they feel in their work, are likely to work harder during times of overload." Simple sentences are best. For instance, the following rewrite of the above sentence would be an improvement: "Persons in this organization are likely to work harder during times of overload."

(V) When using "agree/disagree" or "true/false" items thou shalt avoid perception of bias toward one point of view.

If all your items are agree- or true-keyed (i.e., an "agree" or "true" response shows the respondent favors whatever the scale is assessing), it will appear that the questionnaire is biased toward that point of view. Additionally, scales all keyed in the same direction are prone to response styles which decrease validity. The best strategy is to have one half your items true- or agree-keyed, the remainder false- or disagree-keyed.

* Adapted from John H. Ellard and T. B. Rogers (1993). *Teaching Questionnaire Construction Effectively: The Ten Commandments of Question Writing. Contemporary Social Psychology, 17, 17-20.*

(VI) Thou shalt not use redundant or irrelevant items.

Respondents' time is precious. Under ideal conditions it is desirable to have multiple items converging on the construct you wish to measure. It then becomes possible to calculate internal consistency measures of reliability, and, in general, reliability will be enhanced. However, often one well written and on-target item is better than several repeated and/or unrelated ones.

(VII) Thou shalt not permit any loaded questions to appear in your questionnaire.

If you create a question in which one response question is obviously more desirable than others, you have a loaded question. For example, it is not uncommon for questions such as this to be in voter surveys: "Do you agree or disagree that the government should stop wasting money on road construction?" The wording of this question makes obvious the investigator's bias and presumably the bias of most respondents (who would ever favor wasting money?). The major goal in the creation of any questionnaire is to make it possible for respondents to express freely and meaningfully their true views unencumbered by your predispositions and beliefs. Loaded questions clearly interfere with this goal.

(VIII) Thou shalt not mix response formats within a set of questions.

A most maddening and error-provoking practice in questionnaires is to change response format in mid-instrument. Asking a person to respond

"true" or "false" to one item, followed by an "agree/disagree" Likert scale, then a multiple choice question is likely to cause trouble -- respondents sometimes miss the changes and make errors and they also may become annoyed.

(IX) Thou shalt not permit a non-committal response.

It seems intuitively reasonable to assume that some respondents will be neutral or non-committal with respect to any given question. But, providing a neutral response alternative to accommodate these people creates other problems that are usually worth avoiding. Providing an odd number of response alternatives may not provide as much useful information because respondents "on the cusp" can opt out of making a decision by choosing the middle category (typically labeled "neutral," "?," or "Cannot Say"). An even number of response categories avoids the problem of overestimating the number of truly "neutral" respondent while providing meaningful information about respondents who are merely leaning one way or the other (Converse & Presser, 1986). Of course some respondents object to being "forced" to make a decision by the use of an even number of alternatives. If you know that respondents might object, then it might be wise to use an odd-number of alternatives. But, this decision should be accompanied by a strong and documentable rationale.

(X) Thou shalt pretest questions before collecting data.

No matter how carefully the first nine commandments are followed, there is no substitute for pretesting questions. Ideally, pretests should be done with respondents from the same population to be studied.

SURVEY RESEARCH: CRITICAL QUESTIONS TO CONSIDER

- ◆ Was the sample drawn using specifiable probabilities for selection?
- ◆ Was the sample free of bias, or were some categories of people likely to be omitted or over- or under-represented?
- ◆ Was the mode of data collection selected appropriate (e.g., in-person, telephone, mail)?
- ◆ Were the disadvantages of the selected collection mode acknowledged?
- ◆ Were questions constructed appropriately (e.g., clear, concise, consistent meaning)?
- ◆ Were filter questions (e.g., “don’t know’s”) used appropriately?
- ◆ Were open- and close-ended questions used? How was the use of each type of question justified?
- ◆ Were the potential biasing influences of question construction and response categories adequately acknowledged and minimized?
- ◆ Were steps taken to guard against order and context effects?
- ◆ Were appropriate pilot tests conducted and feedback incorporated into the final survey instrument?

Overhead 6.10.1

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 6: An Introduction to the Scientific Methods of Survey Research, pgs. 85-112). © 1999.

SURVEY RESEARCH: CRITICAL QUESTIONS TO CONSIDER (Continued)

- ◆ Were the experts who designed, conducted, and/or analyzed the survey appropriately skilled and experienced?
- ◆ Were the experts who presented the results of the surveys conducted by others appropriately qualified by skill and experience?
- ◆ Was the target population identified appropriately and properly defined?
- ◆ Was the target population relevant to the issue in question?
- ◆ Did the population members constitute individuals whose attitudes and/or behaviors are relevant to the issue in question?
- ◆ Did the sample population adequately reflect the target population; that is, the individuals whose attitudes and behaviors are relevant to the issue in question?
- ◆ Was the sampling frame appropriately specified; including, sources of information, sampling procedures to be used, probability of an element being selected?
- ◆ Was the sampling frame appropriately comprehensive; that is, did it include all necessary population elements?
- ◆ Were interviewers appropriately selected, trained, and supervised?

Overhead 6.10.2

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 6: An Introduction to the Scientific Methods of Survey Research, pgs. 85-112). © 1999.

Handout: A Survey to Assess Pretrial Prejudice*

The defendant is charged with murdering a 17-year old girl in a rural community. The prosecutor is not seeking the death penalty and will not plead guilty by reason of insanity. Although the killing occurred more than 3 years ago, the case is just now coming to trial. The defense lawyer believes that community sentiment against the defendant is still high. She is making an application for a change of venue. Even if she fails in the change of venue application she wants to convince the judge that the prospective jurors should be voir dired individually rather than in groups of six as is the practice in this jurisdiction and she also wants to convince the judge that questions should be allowed bearing on jurors' attitudes toward the insanity defense. She has commissioned an expert to provide survey evidence about the present state of public opinion.

The Basic Facts:

- The killing occurred in the rural community of Horton (population approximately 50,000 people)
- The victim was the 17 year old daughter of a prominent physician – the victim was followed for several blocks and attacked with a knife
- The town was stunned by the killing and memorial services were held and reported in the local newspapers
- The defendant, an African American aged 40 years old, is married with 3 children and has no prior criminal record – at the time of the killing he had just separated from his wife and had lost his job
- The defendant underwent a series of psychiatric assessments, which determined that he suffers from a form of epilepsy caused by a brain tumor – none of the experts can say with much certainty how the tumor affected his state of mind prior to and during the killing
- From contacts in the community the defense lawyer has learned that 'many' people are still incensed over the killing – furthermore, many people in the community are patients or acquaintances of the victim's father

The Survey:

- The expert hired to conduct the survey has a Ph.D. in Social Psychology, is a professor at the state university, and has served as an expert in a number of change of venue cases
- A random sample of jury-eligible persons was drawn from a list provided by the city's Department of Motor Vehicles
- Each jury-eligible person selected for inclusion in the survey sample received a telephone call from a survey interviewer between 10:00 a.m. and 5:00 p.m. weekdays
- Survey interviewers were experienced and well-trained on the survey instrument and were supervised
- The survey instrument included close-ended and open-ended questions about familiarity with the case

* Adapted from: Neil Vidmar (1995). *An Exercise on Designing A Survey: Assessing Pretrial Prejudice*. Duke Law School.

Handout: A Survey to Assess Pretrial Prejudice

Questions for Consideration:

- ▶ What information do you need about the way the survey was designed and conducted in order to make a determination about its validity?
- ▶ Was a random sample of jury-eligible persons drawn from another comparable community or district? What advantages would this bring?
- ▶ By using the DMV list, were all jury-eligible residents in the community sampled?
- ▶ Was the survey instrument pre-tested?
- ▶ Did the questionnaire used include questions concerning the level of publicity (e.g., awareness), case-specific issues, as well as general attitudes?
- ▶ Were appropriate screening questions used to ensure that only jury eligible persons responded to the telephone interview?
- ▶ Did conducting the interview during the hours of 9:00 a.m. to 5:00 p.m. bias the survey sample in any way?
- ▶ Was a validation procedure used?

Handout: The Survey - General Questions for Consideration*

Questions: Purpose and Design of the Survey –

- Was the survey designed to address the relevant questions?
- Was participation in the design, administration, and interpretation of the survey appropriately controlled to ensure the objectivity of the survey?
- Are the experts who designed, conducted or analyzed the survey appropriately skilled and experienced?
- Are the experts who will testify about surveys conducted by others appropriately skilled and experienced?

Questions: Population Definition and Sampling –

- Was an appropriate population identified?
- Did the sampling frame approximate the population?
- How was the sample selected to approximate the relevant characteristics of the population?
- Was the level of non-response sufficient to raise questions about the representativeness of the sample?
- What procedures were used to reduce the likelihood of a biased sample?
- What precautions were taken to ensure that only qualified respondents were included in the survey?

Questions: Survey Questions and Structure –

- Were questions on the survey clear, precise, and unbiased?
- Were filter questions provided to reduce guessing?
- Did the survey use open-ended or closed-ended questions, and how was the choice justified?
- If probes were used, what steps were taken to ensure that probes were not leading and were administered in a consistent fashion?
- What approach was used to avoid or measure potential order effects?
- If the survey was designed to test a causal proposition, did the survey include an appropriate control group or control question?
- What limitations are associated with the mode of data collection used (i.e., telephone interview, in-person interview, mail survey)

Questions: Surveys Involving Interviewers –

- Were interviewers appropriately selected and trained?
- What did the interviewers know about the survey and its sponsorship?
- What procedures were used to ensure and determine that the survey was administered to minimize error and bias?

Questions: Data Entry –

- What was done to ensure that data were recorded accurately?

Questions: Reporting –

- Does the survey report include complete and detailed information on all relevant characteristics?

* Adapted from: Diamond, S.S. (1994). "Reference Guide on Survey Research." In *Reference Manual on Scientific Evidence*. Federal Judicial Center. Washington, D.C.: Government Printing Office, pgs. 221-272.

SUGGESTED GROUP EXERCISES AND DISCUSSION QUESTIONS

Some suggested group



exercises ...

- ▶ Have students generate examples of qualitative research methods (**accompanying handout**) -- Have students discuss the primary strengths of qualitative research methods and the primary weaknesses of qualitative research methods

SAMPLE OVERHEADS

- 7.1 Learning Objectives
- 7.2 Definition of Qualitative Research
- 7.3 Reflecting on the Assumptions of Qualitative Data
- 7.4 The Qualitative Research Process
- 7.5.1 Data Collection Procedures: Observation
- 7.5.2 Data Collection Procedures: Interviews
- 7.5.3 Data Collection Procedures: Documents
- 7.5.4 Data Collection Procedures: Visual Materials

SAMPLE HANDOUTS

- 1. Some Examples of Qualitative Methods

An Introduction to Qualitative Methods

Learning Objectives:

Upon completion of this section, the judge-student should be able to:

- ◆ Understand the differences between quantitative and qualitative research;
- ◆ Discuss when qualitative methods might be the most appropriate methods and why;
- ◆ Understand the general method of qualitative research;
- ◆ Understand both the limits and power of qualitative methods; and
- ◆ Critically evaluate research utilizing qualitative methodologies

Qualitative research ... is about “ways of seeing.” The goal of the research, whatever the methodology, is understanding gained through a process of discovery. What is expressed in qualitative research is a process of discovery that asserts particular assumptions of how knowledge is perceived and acquired. Qualitative research generally focuses on complex human social interactions.

Overhead 7.2

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 7: An Introduction to Qualitative Methods, pgs. 113-122). © 1999

Reflecting on the Assumptions of Qualitative Research

- qualitative researchers are concerned primarily with ***process***, rather than outcomes or products
- qualitative researchers are interested in ***meaning*** - how people make sense of their lives and experiences
- qualitative research involves ***fieldwork*** - the researcher physically goes to the people, setting, site, or institute to observe or record behavior in natural settings
- qualitative research is ***descriptive***
- qualitative research is ***inductive*** - the researcher builds abstractions, concepts, hypotheses, and theories from details

The Qualitative Research Process ...

- 1. Developing a plan**
- 2. Acknowledging and identifying the role of the researcher**
- 3. Specification of data collection procedures**
 - setting boundaries
 - collecting data
 - establishing protocols for recording data
- 4. Data analysis**
- 5. Report writing**

Overhead 7.4

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 7: An Introduction to Qualitative Methods, pgs. 113-122). © 1999

Data Collection Procedures: Observation

Unobtrusive Observer

- the researcher tries to avoid responding in any way to the subject who is under observation

Participant Observer

- the researcher becomes part of the situation and, to a greater or lesser degree, contributes to the situation

Advantages

- researcher has first-hand experience with informant
- researcher can record information as it is observed
- unusual aspects can be noticed and recorded immediately
- useful in exploring topics that may be uncomfortable for informants

Disadvantages

- may be seen as intrusive
- private information may be observed that cannot be recorded
- researcher may not have good observation skills
- certain informants may present problems with building rapport

Data Collection Procedures: Interviews

Interviews

- face-to-face
- telephone
- mail
- focus (delphi) groups)

Advantages:

- useful when informants cannot be directly observed
- informants can provide historical information
- allows researchers a degree of "control" over the line of questioning

Disadvantages:

- provides "indirect" information filtered through the view of those interviewed
- information is removed from the immediate natural setting

Overhead 7.5.2

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 7: An Introduction to Qualitative Methods, pgs. 113-122). © 1999

Data Collection Procedures: Documents

Documents

- public
- private

Advantages:

- researcher can obtain the language of informant
- more time convenient
- represents data that are thoughtful to the extent that informants have given attention to compiling the information (e.g., personal journal)

Disadvantages:

- may be protected information unavailable to the researcher
- materials may be incomplete
- documents may not be authentic or accurate

Data Collection Procedures: Visual Materials

Visual Materials

- photographs
- videotapes
- art objects
- film

Advantages:

- may be unobtrusive
- provides an opportunity for informant to share directly his or her "reality"

Disadvantages:

- may be difficult to interpret
- may not be accessible publicly or privately

CHAPTER 8

Peer Review, Publication, and General Acceptance

Learning Objectives for Chapter 8

Upon completion of this chapter, the reader should be able to:

- ◆ Provide an overview of the peer review process;
- ◆ Identify and discuss critiques of the peer review process; and
- ◆ Discuss the differences between referred and unreferred journals and books, and presentations at professional conferences.

In determining whether theory or technique is scientific knowledge that will assist the trier of fact, and, thus, whether expert testimony is admissible, is whether the theory or technique has been subject to peer review and publication. Fact of publication of theory or technique, or lack thereof, in a peer-reviewed journal will be relevant, though not dispositive, consideration in assessing scientific validity of a particular technique or methodology on which expert opinion is premised; submission to scrutiny of scientific community is a component of “good science,” in part because it increases the likelihood that substantive flaws in methodology will be detected.¹

‘General acceptance’ of scientific theory or technique can have bearing in determining admissibility of expert testimony ... Widespread acceptance of scientific theory or technique can be an important factor in ruling particular evidence admissible, a known technique that has been able to draw only minimal support within the community may properly be viewed with skepticism.²

The Peer Review Process

Though pre-publication review is important, it does not by itself establish validity. The review by other scientists that comes after publication is far more significant. Valid scientific knowledge results from mutual criticism and intellectual cooperation. This process does not merely reflect or accompany the scientific method, it is the scientific method. Sharing new discoveries with other scientists, and establishing priority, lies at the heart of science.

Peer review is the process by which scientific papers submitted to journals for publication are reviewed by experts in the field. These experts, or peer reviewers, are familiar with the topic area from their own work and thus can be considered peers of the authors. Reviewers are asked to evaluate the importance and usefulness of the research and judge whether it was performed carefully and accurately. Journal editors consider the reviewers' opinions in deciding whether to publish the submissions. It is important to note that in most cases reviewers do *not* have access to the raw data. Rather, the reviewers must rely on the narrative description of the research design, including how the data were collected, and from what source, and how the data were interpreted.

Peer Review: the process by which scientific papers submitted to journals for publication are reviewed by experts in the field

Institutional Review Mechanisms

Because the process of formulating and testing hypotheses is far from simple and trivial, a probing and careful review is an indispensable part of the scientific enterprise. The review of scientific ideas takes place in a variety of contexts. Informal review can occur when scientists discuss their work with one another at the laboratory bench, during seminars, and at scientific meetings. Formal peer review is generally an integral part of the scientific publication process and the process by which funds are allocated for the conduct of research. Any claim that would significantly add to or change a body of scientific knowledge must be regarded skeptically if it has not been subjected to some form of peer scrutiny, preferably by submission to a reputable journal.

Critiques of the Peer Review Process

Commentators have recently criticized the peer review process on several grounds:

- it lacks explicit standards for review;
- it lacks objectivity and fairness in the selection of manuscripts for review;
- it favors conservative, traditional research and excludes innovative work by independent researchers;
- editors are biased in their selection of reviewers, sending manuscripts only to those individuals who share the opinions of the editorial board; and
- reviewers may not spend enough time critically reviewing the methods and procedures of the research.

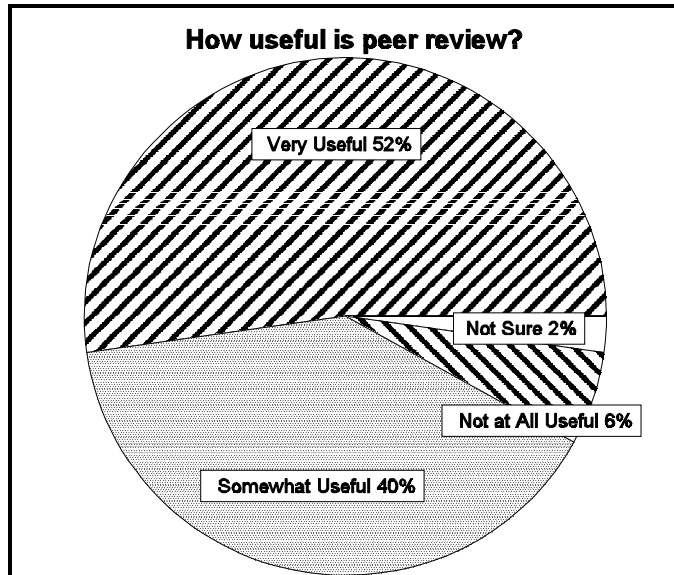


To what extent do judges around the country find the concept of peer review a useful criterion for critically evaluating scientific evidence?

All of the judges surveyed (N=400) were asked how useful the concept of “peer review” is for determining the admissibility of scientific evidence.

The majority of judges surveyed felt that it was a useful concept. Just over half (52%) felt that it was “very useful” and 40% felt that it was “somewhat useful.” Only 6% reported that peer review was “not at all useful” when determining the admissibility of scientific evidence.

When the responses of judges in FRE/*Daubert* states (n=205) were examined, ratings were the same, with half (53%) reporting that peer review was “very useful” and 40% reporting that it was “somewhat useful.” 7% of FRE/*Daubert* judges felt that peer review was “not at all useful” to determinations of admissibility.



When asked a question about how they might apply the concept of peer review to a determination of the admissibility of proffered evidence, 71% of all of the judges surveyed (N=400) provided responses that demonstrated a clear understanding of the scientific peer review process. In 26% of responses, it was not clear that the judge fully understood the concept. In only 3% of the responses, however, was it clear that the judge did *not* understand the concept of peer review. When the responses of judges in FRE/*Daubert* states (n=205) were examined, the results were almost identical. Some sample comments related to the peer review process included:

“You have to look at the motivations behind publication – generally, you have to assume that articles are published in good faith and not just for the purposes of litigation – you have to examine why the study was done”

“I would give greater weight to a criticism of the technique or procedure if it appeared in a significant number of high status journals – or if the prestige of the criticizer was high – nevertheless, practically speaking, it would be difficult for me to evaluate the prestige of the critic”

“Peer review gives you an idea of whether this is a scientific idea that has been debated in the field – it would be important to have the experts describe and debate the peer-reviewed literature to determine the acceptability of the evidence”

➡ *The Evaluation of Research Papers*

Scientific evaluation is made on the basis of criteria established by the journal to which a research paper is submitted. Typically, the more prestigious the journal, the more rigorous the evaluation, and the more difficult it is to get a paper published in that journal. While some journals may accept almost all manuscripts they receive, others may reject 90% or more. Especially high rejection rates develop as a consequence of a journal's high prestige. That is, the most prestigious journals receive many more papers than they can publish, thus allowing the journal to set even higher standards for acceptance.

◆ *Selection of Referees*

Managing editors initially evaluate submissions and select the individuals who will act as referees or peer reviewers. Peer reviewers are usually selected on the basis of their own research experience in the field and their expertise with the methodologies or statistical procedures used. Any given manuscript is typically reviewed by more than one referee. Depending on the scientific discipline and the journal, referees may or may not know the identity and professional affiliation of the authors. When the identity of the author or authors is not known, the process is referred to as a **blind review**. The identity of the referees, however, is not disclosed to the authors.

Blind Review: when identity of the author is not known by the reviewer

After the referees return their reviews of the manuscripts, the managing editor or an editorial board decides whether the comments are reasonable. A preliminary decision regarding suitability for publication is sent to the authors, accompanied by the referees' comments. The author(s) may or may not respond to the comments. Depending on the nature and scope of the referees' comments, the author(s) may address the reviewers comments and resubmit the paper for a second review.

Typical Questions Addressed by Referees When Evaluating a Manuscript for Publication

Is the subject matter appropriate for readers of the journal?

Different journals cover different subjects and target different audiences. The referee must determine whether the work described in the paper is consistent with the stated intent of the journal.

Are the methods used suitable for the questions under study?

The referee must determine whether the research design was appropriate, whether appropriate controls were used, whether the data were presented adequately, and whether the interpretation of the results was reasonable. The referee must consider whether the methods and analysis plan were presented clearly enough, and in enough detail, and whether or not the research could be replicated.

Are the data analyses convincing?

Do the statistical analyses support the conclusions drawn?

Are there other possible explanations for the findings?

Does the paper make an important contribution to the field?

The referee's evaluation may go beyond the adequacy of the work and also consider the potential impact of the research on the relevant field or discipline. More prestigious journals want to publish research that is most likely to have a significant impact on the field. The referee must evaluate the impact of the research within the context of the relevant field – the current problems and questions of interest, currently accepted methodologies, and so forth.

Will the work be of broad-reaching working in the field?

Is the paper in the top 1%, 5%, 10%, etc. of papers written in this field or for this journal?

➡ *Refereed vs. Unrefereed Journals*

Articles published in refereed journals are typically given more weight and afforded more status than articles published in unrefereed journals -- that is, in journals that do not send out submitted articles for independent peer review.

Referred Journals: articles submitted for publication are sent for independent peer review

Unreferred Journals: articles submitted for publication are not independently peer reviewed

Peer Review in the Computer Age

The underlying assumptions and procedures of peer review are receiving increasing attention within scientific disciplines. One development giving urgency to this particular topic is the revolution taking place in scholarly communication. As technology advances it becomes possible to make every work available electronically to all scholars, obviating the need for printed volumes and potentially eliminating the peer review process. This eliminates practical constraints that until now have made it impossible for journals to publish huge amounts of material and for libraries to acquire and store them. As a consequence of advances in technology, much more can be "published" without long delays and backlogs. This increased capacity for communication is also accompanied by relaxed standards of peer review (most electronic journals are not refereed).

➡ *Determining the Status of a Journal*

High status journals are generally considered to be those journals that have a high rejection rate for submitted manuscripts (i.e., the journal only accepts for publication a very small percentage of the number of manuscripts submitted for review). Other factors taken into account in determining the prestige of a journal include: its prominence in the relevant field; its circulation; who subscribes to the journal (e.g., researchers, practitioners, legal professionals); the target audience; and the editorial board membership.

Things to consider when determining the status of a scientific journal:

- whether or not articles published in the journal are peer-reviewed;
- rejection rate of submitted manuscripts;
- level of prominence in the relevant field;
- circulation record;
- subscribers;
- target audience; and
- editorial board membership.

It is important to recognize that while authors may count a journal's reputation more than any other factor in deciding where to submit an article, the appearance of an article in a lower-status journal does not necessarily mean that the article was rejected by a higher-status journal. There are a variety of legitimate reasons why an author may choose to submit an article to a lower-status journal. For example, the author may have decided to submit an article to a journal with higher acceptance rates and quicker response times in an effort to ensure that the research reaches the relevant community in a relatively timely manner. In

addition, the higher-status journal may have recently published an article on a similar topic and would be likely to reject a similar submission on that ground alone. Also, many journals are becoming increasingly specialized and target increasingly specialized audiences. Thus, while a journal may not be defined as prestigious in terms of a high circulation rate, it may be considered a high status, prestigious journal within its specialized, narrow field.

The prestige of books is often measured by who published them. Typically, books published by "prestigious" university presses are counted more heavily than books published by other publishing companies. However, many of the criticisms leveled at journal publications also apply to status determinations about books.

Informal Peer Review

The process of publishing a research paper can often take six months to a year between submission of a paper and its appearance in a journal, although there is great variation among different scientific disciplines. If investigators want to make their research known to the relevant scientific community, they must find a means of reporting the work in a more preliminary form.

Preliminary research reports are often presented at meetings of professional societies or associations. In most cases, researchers request the opportunity to make a presentation at a meeting. By communicating the major conclusions of the study, the work becomes known by the relevant community, and the researcher establishes a claim of ownership and priority over the research. When a researcher is invited by the organizers of a professional society to present her current work at a meeting or conference, it is an indication that the researcher's work is expected to be of interest to the professional audience and that the researcher is an authority whose recent work is probably reflective of the current thinking in the relevant field.

Unlike formal research papers, oral presentations are reviewed minimally or not at all and the details of the methods and materials are generally not presented. Other researchers can learn the thought style behind the work and the general methodologies utilized, but not the precise methods by which the research was accomplished.

[W]hile peer review improves the overall quality of research, as well as the likely integrity of individual publications, it does not mechanically guarantee reliability. Courts therefore should neither be wholly dismissive of peer review nor accord it more respect than it deserves.³

???

Questions to consider when evaluating scientific evidence ...

- ◆ In what ways has the research been peer-reviewed?
- ◆ Was the research published in a referred or unreferred journal?
- ◆ If the research was published in a referred journal, was it a blind review?
- ◆ What is the relative status and reputation of the journal in which the research was presented?
- ◆ Was the disciplinary focus of the journal appropriate for the research?

➔ ***General Acceptance***

[G]eneral acceptance can yet have a bearing on the inquiry. A reliability assessment does not require, although it does permit, explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community. Widespread acceptance can be an important factor in ruling particular evidence admissible and a known technique which has been able to attract only minimal support within the community may properly be viewed with skepticism. ... The focus, of course, must be solely on principles and methodology, not on the conclusions they generate.⁴

Just when a scientific principle or discovery crosses the line between the experimental and demonstrable tests is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts go along way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.⁵



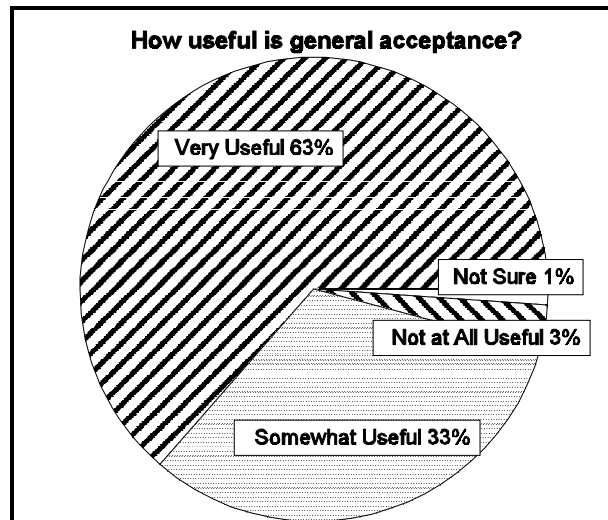
To what extent do judges around the country find the concept of general acceptance a useful criterion for critically evaluating scientific evidence?

All of the judges (N=400) were asked how useful the concept of general acceptance is when determining the admissibility of scientific evidence.

The vast majority of judges indicated that general acceptance was a useful criterion, with 63% rating it as “very useful” and 33% rating it as “somewhat useful.”

When judges in *FRE/Daubert* states (n=205) were examined, the results were much the same, with 61% reporting that general acceptance was “very useful,” 36% reporting that it was “somewhat useful,” and 2% reporting that general acceptance was “not at all useful” to admissibility decision-making.

When asked a question about how they might apply the concept of general acceptance to a determination of the admissibility of proffered evidence, the majority (82%) of all of the judges surveyed (N=400) provided responses that demonstrated a clear understanding of the concept of general acceptance. In 17% of responses, understanding of general acceptance was questionable. It was clear that the judge did *not* understand the scientific concept of general acceptance in only 1% of responses.



Before going any further, stop and reflect ...

- ◆ How might new or novel theories or methods defend themselves against a lack of general acceptance?
- ◆ Does a lack of general acceptance in the relevant scientific community necessarily draw into question the validity and reliability of the research principles and methods? Why or why not?



Questions to consider when evaluating scientific evidence ...

- ◆ Are the theories or principles underlying the research generally accepted in the relevant scientific field?
- ◆ If the theory or principles underlying the research are not generally accepted, how are they defended?
- ◆ What is the relevant scientific field?
- ◆ Are the methods employed in the research generally accepted in the relevant scientific field?
- ◆ Are there relevant but unidentified scientific disciplines that may have a different perspective on the issue?
- ◆ If the methods employed are not generally accepted, how is their use defended?

CRITICAL QUESTIONS REVIEWED

- ◆ In what ways has the research been peer-reviewed?
- ◆ Was the research published in a referred or unreferred journal?
- ◆ If the research was published in a referred journal, was it a blind review?
- ◆ What is the relative status and reputation of the journal in which the research was presented?
- ◆ Was the disciplinary focus of the journal appropriate for the research?
- ◆ Are the theories or principles underlying the research generally accepted in the relevant scientific field?
- ◆ If the theory or principles underlying the research are not generally accepted, how are they defended?
- ◆ What is the relevant scientific field?
- ◆ Are the methods employed in the research generally accepted in the relevant scientific field?
- ◆ Are there relevant but unidentified scientific disciplines that may have a different perspective on the issue?
- ◆ If the methods employed are not generally accepted, how is their use defended?

Endnotes:

1. *Daubert vs. Merrell Dow Pharmaceuticals* 509 U.S. 579, 113 S.Ct., at 2786.
2. *Ibid.*
3. Sheila Jasanoff (1997). *Judging Science: Issues, Assumptions, Models*.
4. *Supra* note 1, at 2797.
5. *Frye*, at 1014.

GLOSSARY

blind review	when identity of the author is not known by the reviewer
peer review	the process by which scientific papers are submitted to journals for publication are reviewed by peers in the field; peer reviewers typically evaluate the importance and usefulness of the research and assess whether the research was carried out in a methodologically appropriate manner and supports the conclusions drawn
referred journal	articles submitted for publication are sent for independent peer review
unreferred journals	articles submitted for publication are not independently peer reviewed

SUGGESTED READINGS

- Bell, R. (1992). "Examining the Peer Review Process." In *Impure Science: Fraud, Compromise and Political Influence in Scientific Research*. New York: Wiley.
- DuRant, R.H. (1994). "Checklist for Evaluation of Research Articles." *Journal of Adolescent Health*, Vol. 15, pgs. 4-8.
- Imwinkelried, E.J. (1997). "*Frye*'s General Acceptance Test v. *Daubert*'s Empirical Validation Standard - 'Either ... Or' or 'Both ... And?'" *Criminal Law Bulletin*, Vol. 33, pg. 72.
- Lock, S. and Smith, D. (1990). "What do Peer Reviewers Do?" *JAMA*, Vol. 263, No. 10 (9 March), pgs. 1341-1343.

Judge's Notes:

the scores in the distribution typically deviate or vary from the mean. Since the standard deviation is always calculated with reference to the mean, its calculation demands the use of interval or ratio data. The standard deviation is the typical deviation of a given distribution. The larger the value of the standard deviation, the more the scores are spread out around the mean; the smaller the value of the standard deviation, the less the scores are spread out around the mean. That is, a distribution with a small standard deviation indicates that the group being measured is homogeneous; their scores are clustered very close to the mean. A distribution with a large standard deviation indicates that the group is heterogeneous; their scores are more widely dispersed from the mean.

◆ **The Variance**

The variance of a distribution is the square of the standard deviation. It is a useful term because it reflects how much of the variability between people on one characteristic (e.g., income) can be explained by knowing where they stand on another characteristic (e.g., education).

➡ ***The Normal Curve and Z-Scores***

The normal curve is a theoretical distribution. However, many distributions of people-related measurements come close to approximating the normal curve and thus it is of crucial significance for describing data.

The **normal curve** is a unimodal frequency distribution with scores plotted on the X axis (the horizontal axis) and frequency plotted on the Y axis (the vertical axis). In a normal curve, most of the scores cluster around the middle of the distribution (where the curve is at its highest). As

<p>Variance: measures how much of the variance between people on one characteristic can be explained by knowing where they stand on another characteristic (e.g., income).</p>
<p>Normal Curve: a theoretical distribution with scores plotted on the X axis (the horizontal axis) and frequency plotted on the Y axis (the vertical axis); most of the scores cluster around the middle of the distribution; curve is symmetrical and all three measures of central tendency (mean, median, mode) fall precisely at the middle of the distribution</p>
<p>Skewed Distribution: a distribution of scores where the majority of scores in the distribution bunch up at one end of the distribution</p>

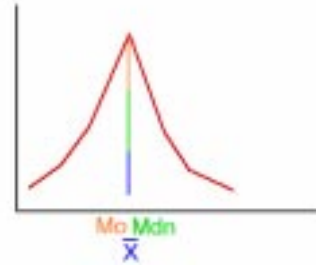
the distance from the middle increases, in either direction, there are fewer and fewer scores. The normal curve is symmetrical - both sides are mirror images of the other - and all three measures of central tendency (the mean, median, and mode) fall precisely at the same point, the exact middle of the distribution. In a **skewed distribution**, scores tend to pile up at one end or the other. The direction of skewness is indicated by the “tail” of the curve. The curve is **positively skewed** when most of the scores pile up near the bottom (the tail points toward the high or positive end). The curve is **negatively skewed** when most of the scores pile up near the top (the tail points toward the low or negative end).

Positively Skewed Distribution: distribution in which scores are concentrated near the bottom of the distribution; tail of the distribution points to the top or positive end

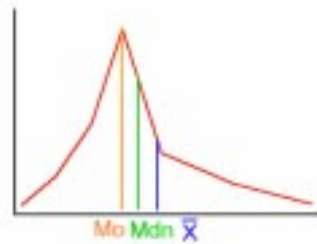
Negatively Skewed Distribution: distribution in which scores are concentrated near the top of the distribution; tail of the distribution points to the low or negative end

Standard Normal Curve (Z-Distribution): the normal curve is marked off in units of standard deviation; a normally distributed set of scaled scores whose mean is always equal to zero and whose standard deviation equals 1.00

The Normal Curve

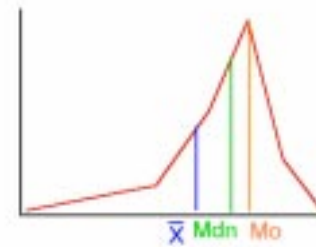


Positively Skewed Distribution



Negatively

Skewed Distribution

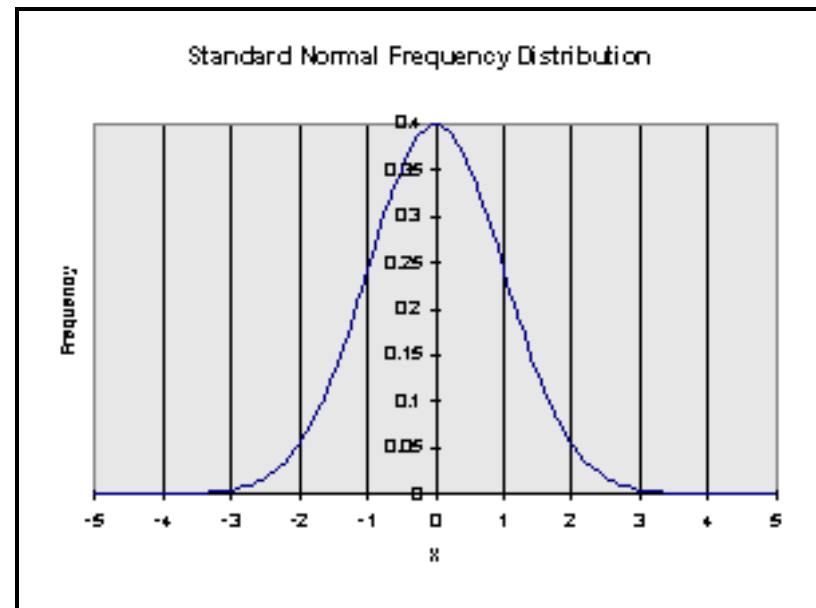


The normal curve has a constant relationship with the standard deviation. When the normal curve is marked off in units of standard deviation, a series of constant percentages under the normal curve are formed. Once the curve is plotted according to standard deviation units, it is called the **standard normal curve**, or z-distribution.

A **z-distribution** is a normally distributed set of specially scaled scores whose mean is always equal to zero and whose standard deviation must equal 1.00. Z-scores take into account both the mean of the distribution and the amount of variability, the standard deviation. Thus, z-scores can be used to assess an individual's relative performance compared to the performance of the entire group being measured. The z-score is the number of standard deviations the observed value is from the mean.

Z-Distribution: a normally distributed set of scaled scores whose mean is always equal to zero and whose standard deviation must equal 1.00; takes into account both the mean of the distribution and the amount of; the z-score is the number of standard deviations the observed value is from the mean

II. *Inferential Statistics*



The primary goal of inferential statistics is to measure a few and generalize to many. That is, observations are made of a small segment of the group, and then, from these observations, the characteristics of the entire group are inferred. Inferential statistics are procedures used to reach conclusions (generalizations) about larger populations from a small sample of data with a minimal degree of error.

Inferential Statistics:
statistical procedures used to draw conclusions and inferences which are based upon, but go beyond, the numbers generated by a research study

There are usually two issues to be explored:

1. Does the mean of a sample actually reflect the mean of the larger population of interest?
2. Is a difference found between two means (e.g., between an experimental group and a control group) a real and important difference, or is it merely the result of chance?

Key Concepts of Inferential Statistics

Population (or universe): an entire group of persons, things, or events having at least one trait in common

Sample: a smaller number of observations taken from the total number making up the population; in typical applications of inferential statistics, the sample size is small relative to the population size

To make accurate predictions, the sample should be representative of the population. In a sense, a good representative sample provides the researcher with a miniature mirror with which to view the entire population. *Recall that you have seen these concepts before in the chapter on surveys.*

➡ **Measures of Relationship: Correlation**

Measures of central tendency and variability are basic descriptive statistics that tell us something about the distribution of a variable.

Measures of relationships provide information

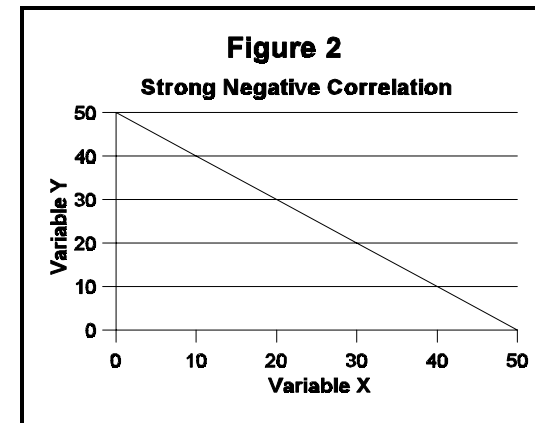
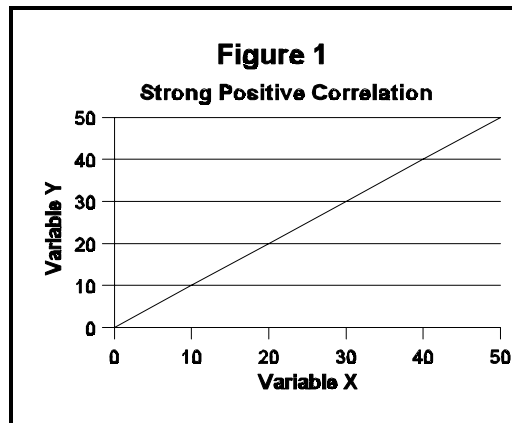
about what relationship the variable has to other variables. The association between one variable and any other variable is described as a **correlation**.

Correlation: an association between two variables; can be positive or negative

Correlation does not equal Causation

If two variables have a perfect correlation (their data points fall along a straight line), then $r = 1.0$

(Fig 1) or $r = -1.0$ (Fig 2) (“r” is the **correlation coefficient**). The positive and negative values simply show the direction of the relationship. When two variables are positively correlated, as one increases, the other also increases. When they are negatively correlated, as one increases, the other decreases. Two variables with less than a perfect correlation will have an “r value” between 0 and 1.0 or 0 and -1.0. If no relationship exists between two variables, $r = 0$. Figure 1 depicts a positive correlation between Variable X and Variable Y. That is, as Variable X increases, Variable Y also increases. Figure 2 depicts a negative correlation between the two variables. That is, as Variable X increases, Variable Y decreases.



Correlation Coefficient: a number between -1 and 1 which measures the degree to which two variables are linearly related. If there is a perfect positive linear relationship, $r = 1$ (i.e., an increase (or decrease) in one variable is associated with an increase (or decrease) in the other variable); if there is a perfect negative linear relationship, $r = -1$ (i.e., an increase (decrease) in one variable is associated with a decrease (increase) in the other variable); If $r = 0$ there is no linear relationship between the variables

Pearson's Product Moment Correlation Coefficient: Pearson's product moment correlation, usually denoted by r , is one example of a correlation coefficient; a measure of the linear association between two variables that have been measured on the interval or ratio scales (e.g., the relationship between height in inches and weight in pounds)

Regression: predicts the extent to which the value of one or more variables can be predicted by knowing the value of other variables

➡ *Measures of Relationship: Regression*

Regression analysis predicts the extent to which the value of one or more variables can be predicted by knowing the value of other variables. A **linear regression** predicts the magnitude of the expected change in variable Y given a change in variable X. A **simple linear regression** is designed to determine whether there is a linear relationship between a response variable and a possible predictor variable. A **multiple linear regression** is designed to examine the relationship between a response variable and several possible predictor variables. **Nonlinear regression** is designed to describe the relationship between a response variable and one or more explanatory variables in a non-linear fashion.

Linear Regression: predicts the magnitude of the expected change in variable Y given a change in variable X

Simple Linear Regression: designed to determine whether there is a linear relationship between a response variable and a possible predictor variable

Multiple Linear Regression: designed to examine the relationship between a response variable and several possible predictor variables

Nonlinear Regression: designed to describe the relationship between a response variable and one or more explanatory variables in a non-linear fashion

➡ *Sampling Revisited*

Sampling techniques were briefly discussed in the chapter on survey methodology. They are briefly revisited here.

◆ **Random Sampling**

Random sampling demands that each member of the entire population has an equal chance of being included and that no member of the population may be systematically excluded. It is important to note that *randomness describes the selection process*, (i.e., the procedures by which the sample is selected), and not the particular pattern of observations in the sample.

◆ **Stratified Sampling**

To obtain this kind of sampling, the researcher must know beforehand what some of the major population characteristics are and, then, deliberately select a sample that shares these same characteristics in the same proportions.

Whenever the sample differs systematically from the population of interest, a **bias** has occurred.

Bias is a constant difference, in one direction, between the mean of the sample and the mean of the population. Bias occurs when most of the sampling error loads up on one side, so that the sample means are constantly either over- or under-estimating the population mean.

Bias: a constant difference, in one direction, between the mean of the sample and the mean of the population; occurs when most of the sampling error loads up on one side, so that the sample means are constantly either over- or under-estimating the population mean

Sampling Error

Whenever a sample is selected, it must be assumed that the sample measures will not precisely match those that would be obtained if the entire population were measured. The sampling error reflects, or is an index of, the difference between the sample value and the population value.

Sampling error is not a mistake. Any sample mean should be expected to deviate from the mean of the whole population, but the deviation will hopefully be random and should not be large.

◆ Sampling Distributions

Each distribution discussed so far has been a distribution of individual scores - each point in the distribution represents a measure of a characteristic or performance of an individual. In sampling distributions, each point represents a measure of a characteristic or performance of a sample of individuals. The mean increase of a sample of U.S. adults is an example; it would be one data point in the sampling distribution of mean income. Sampling distributions are important in testing hypotheses.

III. Parameter Estimates and Hypothesis-Testing

Hypothesis-Testing in Statistical Terms

The purpose of a **hypothesis test** is to determine the likelihood that a particular sample could have originated from a population with a hypothesized characteristic.

The **null hypothesis** supplies the value about which the hypothesized sampling distribution is centered. It always makes a statement about a characteristic of the population, never about a characteristic of the sample. The null hypothesis always makes the claim about a single numerical value, never a range of values.

The **experimental hypothesis**, asserts the opposite of the null hypothesis. A decision to accept the null hypothesis (or a failure to reject the null hypothesis) implies a lack of support for the experimental or research hypothesis, and a decision to reject the null hypothesis implies support for the experimental or research hypothesis.

A **decision rule** specifies precisely when the null hypothesis should be rejected.

Criminal suspects are presumed innocent until proven guilty. Under hypothesis-testing procedures, the null hypothesis is presumed to be true until proven false. Once all the evidence has been considered, a verdict is reached, and the null hypothesis is either retained (failure to reject) or it is rejected.

Evidence for testing an hypothesis about a sample statistic is based on the relationship between the observed sample statistic and the sampling distribution of that statistic. For example, if a researcher predicts that the mean weight of rats in an experimental group is greater than the mean weight in a control group, then the statistic at issue is the difference between the two means. The experimental or research hypothesis is that the two means represent different populations and that the difference between them is dependable. The null hypothesis is that the two means come from the same population and that the difference between them would not hold up under repeated replications of the experiment. The difference between the means is compared to the sampling distribution of such differences, the mean of which is usually zero (no difference). If a difference as large as or larger than the

obtained difference is very unlikely for groups coming from the same population, then the difference will be judged to be an improbable outcome under the null hypothesis of no dependable difference and the null hypothesis will be rejected. On the other hand, if the observed difference is not so large as to be highly improbable, the null hypothesis will be accepted (or the null hypothesis will not be rejected).

An observed sample statistic will qualify as a probable outcome if the difference between its value and that of the hypothesized population statistic is small enough to be attributed to chance. For example, a sample mean will qualify as a probable outcome if the difference between its value and that of the hypothesized population mean is *small* enough to be attributed to chance. Under these circumstances, because there is no compelling reason to reject the hypothesis, the null hypothesis is tentatively accepted.

An observed sample statistic will qualify as an improbable outcome if the difference between its value and the hypothesized value is too *large* to be attributed to chance. That is, a sample mean will qualify as an improbable outcome if it deviates too far from the hypothesized mean and appears to emerge from the sparse concentration of possible sample means in either "tail" of the sampling distribution. Under these circumstances, because there are grounds for suspecting the hypothesis, the hypothesis is rejected.

The decision to reject the null hypothesis involves a degree of risk. Having rejected a null hypothesis, we can never be absolutely certain whether the decision is correct or incorrect, unless, of course, the entire population was surveyed. Even if the null hypothesis is true, there is a slight possibility that just by chance, the one observed sample mean really originates from rejection regions (the tails) of the hypothesized sample distribution, thus causing the true null hypothesis to be erroneously rejected.

IV. Error Rates

In determining the admissibility of expert opinion regarding a particular scientific technique, the court ordinarily should consider known or potential rates of error, and existence and maintenance of standards controlling the technique's operation.¹

To assess known or potential rates of error, the judiciary must be prepared to carefully and critically evaluate the methodology and underlying assumptions of proffered scientific evidence. Such an evaluation would entail examination of whether the research hypothesis was appropriately articulated and tested, whether appropriate controls were utilized, whether threats to validity were controlled for, or at least severely minimized, and so forth.

The likelihood with which a measurement device or a technological procedure leads to an incorrect classification is the **error rate**. Whereas formal testing of hypotheses usually relies on theoretical sampling distributions for estimating the likelihood that the decision based on the data is erroneous (especially Type I error), the likelihood of an incorrect classification is usually assessed in terms of error rates. Several rates should be taken into account, typically termed “true positive,” “true negative,” “false positive,” and “false negative” rates. For example, if a laboratory claims that a particular test reliably identifies the existence of a serious disease, it is necessary to consider the proportion of people with the disease who were correctly identified as having it (**true positive**) and those who were correctly identified as not having it (**true negative**). It is also important to consider the proportion of individuals without the disease who were incorrectly identified as having it (**false positive**) and the proportion of individuals with the disease who were incorrectly identified as not having it (**false negative**). False positives could lead to unnecessary further expense and painful medical interventions; false negatives could lead to further and perhaps fatal progression of the disease. It usually is

<p>False Positive Error: incorrectly classifying someone without a particular characteristic as possessing that characteristic (e.g., person does not have disease but incorrectly classified as having disease)</p> <p>True Positive Error: correctly classifying someone as possessing a particular characteristic or falling into a particular category (e.g., person has disease and is classified as having disease)</p> <p>False Negative Error: incorrectly classifying someone who has a particular characteristic as someone who does not possess that characteristic (e.g., person has disease but is incorrectly classified as not having it)</p> <p>True Negative Error: correctly classifying someone who does not possess a particular characteristic or who does not fall into a particular category (e.g., person does not have disease and is classified as not having the disease)</p>
--

essential to examine both types of erroneous classification rates; if proffered evidence does not include both error rates, it is likely to be of little value.

Error rates are generally stated as percentages or proportions. In the above case, for example, the data might have been drawn from people who visited their physicians because of certain bothersome symptoms, and when the physicians conducted the diagnostic test, the results for 104 patients might have been:

	Actually has disease	Actually free of disease	Total
Test says has disease	90	10	100
Test says free of disease	2	2	4
Total	92	12	104

The true positive rate is .98 (90/92), with only two diseased patients mis-diagnosed (2/92, a false negative rate of .02). There were 12 patients without the disease, 10 of whom were mis-diagnosed as having the disease for a false positive rate of .83 (10/12).

This example illustrates two points. First, the rate of correct classifications has to be compared to the rates of both false positive and false negative classifications. The relative importance of the two types of errors will depend on what they lead to—false security, expensive or painful further intervention, and so on. Second, although proportions and percentages are very useful modes of presenting data, sometimes the raw numbers underlying the percentages are equally important. In the example above, only 12 of the 104 patients were actually free of the disease, and that base of 12 is too small to draw firm conclusions about the false positive

rate. We would be much more confident if the number of disease-free patients who were tested was larger. In general, if we were told that 50% of people held a certain opinion, we would want to know if the reference was 50% of 2 people or 50% of 2,000.

➡ *Type I and Type II Errors*

The decision to reject the null hypothesis is based on probabilities rather than on certainties. The decision is made without direct knowledge of the true state of affairs in the population.

There are two possible decisions: (1) reject the null hypothesis, or (2) fail to reject (accept) the null hypothesis. There are also two possibilities that may be true in the population: (1) the null hypothesis is true, or (2) the experimental hypothesis is true. Thus, there are two kinds of correct decisions and two kinds of errors.

Type I Error: when the researcher rejects the null hypothesis but the null hypothesis is actually true (e.g., the researcher claims that there is a causal relationship between variable A and variable B when, in fact, there is not)

Type II Error: when the researcher fails to reject the null hypothesis (i.e., accepts the null hypothesis) when in actuality the experimental hypothesis is true (e.g., the researcher claims there is no causal relationship between variable A and variable B when, in fact, there is one)

		Researcher's Decision	
		Reject the null	Failure to reject
hypothesis	the null hypothesis	<hr/>	
True state of nature	Null hypo. is true	Type I Error	Correct decision
	Null hypo. is false	Correct decision	Type II Error

Consider the decision made by a juror in a criminal trial. As is the case with statistics, a decision must be made on the basis of evidence: Is the defendant innocent or guilty? However, the decision is the juror's and does not necessarily reflect the true state of affairs that the person really is innocent or guilty. Assume the null hypothesis is that the defendant is innocent. The *rejection* of the null hypothesis is to decide, based upon the evidence, that the defendant is guilty. *Acceptance* of the null hypothesis is to decide, based upon the evidence, that the defendant is innocent.

		Juror's Decision		the null
		Reject the null (find guilty)	Failure to reject (find innocent)	
True State	Null is true (innocent)	Type I Error	Correct Decision	<hr/>
	Null is false (guilty)	Correct Decision	Type II Error	

Most scientists begin with the assumption that the phenomenon they are studying does not cause the effect they expect -- the null hypothesis. In other words, the standard method of science is to presume 'innocence' and only with strong proof reject that assumption.

Scientific conventions have developed regarding the strength of this presumption; that is, how much evidence is needed before rejecting the null hypothesis and accepting an alternative hypothesis that the experimental manipulation caused the observed effect (this will be discussed further in this chapter). It is important to realize, however, that an attempt to decrease one type of error results in an increased likelihood of making the other type of error.

To what extent do judges around the country find the concept of error rate a useful criterion for critically evaluating scientific evidence?



All judges in the survey sample, even those not in FRE/*Daubert* states, were asked how useful they thought the concept of error rate is for admissibility decision-making (N=400).

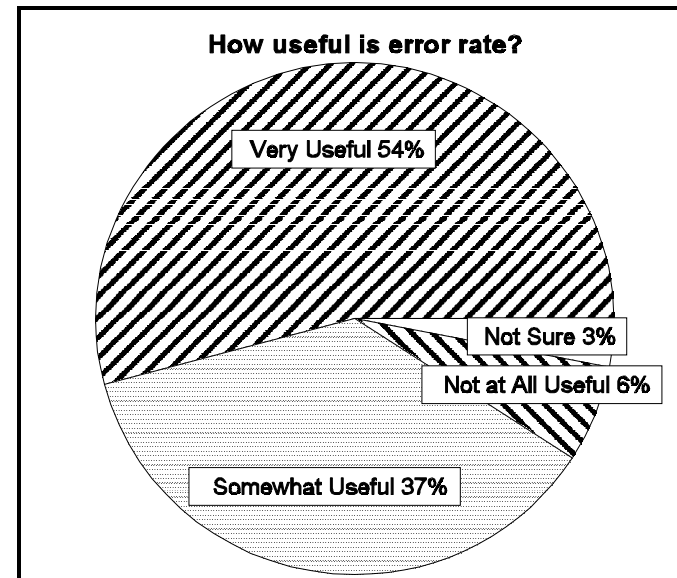
The majority (91%) indicated that a consideration of error rate was a useful when determining the admissibility of scientific evidence, with 54% of those judges rating error rate as “very useful.”

Focusing just on responses from judges in states which follow the FRE/*Daubert* standards, the vast majority of judges rated error rate as a useful guideline for evaluating the admissibility of scientific evidence.

Even though the vast majority of judges rated error rate as a useful guide, the results of the survey indicate that judges do not fully understand the scientific meaning of error rates and that, as a result, they are unsure how to utilize the concept as a guideline for determining admissibility.

When asked a question about how they would apply the concept of error rate, the majority of judges expressed some hesitancy or uncertainty. In order for a response to be coded as “judge understands concept” the response had to include reference to an evaluation of the variety of sources of error, or refer to a number or percent of instances in which the classification procedure was mis-classified. From the answers provided, the researchers could only infer a true understanding of the concept in 4% of the responses (N=400).

In 86% of the responses the judges’ understanding of the concept was questionable. In 10% of the responses, the judge relied solely upon a low error vs. high error heuristic (or rule of thumb) when explaining how the concept of error rate is applied to admissibility (i.e., if there is a high rate of error then the judge is more likely to exclude the evidence than if there is a low rate of error).



➡ *Confidence Level*

An alternative indicator of the probability of a Type I error is the confidence level. It specifies the range of values around the empirically obtained result within which the “true” or population value is likely to lie. Confidence levels are frequently reported in sample surveys. For example, it might be reported that the 95% confidence level for an obtained percentage of 20% is 20% plus or minus 3%. The higher the confidence level, the lower the probability of a Type I error, but the broader the range of values within which the “true” or population value might actually lie.

Confidence Level: specifies a range of values around the empirically obtained result within which the “true” or population value is likely to lie

Both types of error are important. For example, in a toxic tort case, a Type I error could mean that the frequency of occurrence of a symptom among workers could be accepted as indicating that the symptom was caused by a toxic substance found in the plant environment, whereas in fact that frequency of occurrence was just at the outer extreme of random fluctuation and was not a reflection of a causal link. The firm might improperly be held accountable. But with a Type II error, the frequency of occurrence of the symptom would be taken as well within the normal range of fluctuation, and no causal link between substance and symptom frequency would be inferred. The firm could be erroneously exonerated. The less likely a Type I error, the more likely a Type II error.

➡ *Significance and P-Values*

In order to decide whether the difference in the observed score differs significantly from the null hypothesis, a standard or criterion

for deciding whether to accept or reject the null hypothesis must be established. Statisticians typically use two levels of significance: .05 and .01. These levels have been established by convention. When a significance level of $p < .05$ is chosen, the decision rule is that the null hypothesis will be rejected if the data are so unlikely that they could have occurred by chance less than 5 times out of 100. If a significance level of $p < .01$ is chosen, the probability of the observed value occurring by chance is less than 1 in 100.

The odds of making a Type I error (rejecting the null hypothesis when it is true) are exactly equal to the value chosen for the significance level. That is, if a researcher has chosen a significance value of .05, the probability of a Type I error is .05 -- 5 times out of 100 (5%) the researcher will reject the null hypothesis when it is true. That is, there will be 5 times out of 100 when extreme differences are due to chance and not to some experimental manipulation.

Can the odds of making a Type I error be minimized by choosing a more extreme significance level (e.g., $p < .01$)? Yes, but there is a trade-off: an increased likelihood of making a Type II error (failure to reject an hypothesis when it is false) -- the researcher concludes that the results were caused by chance and not by the experimental manipulation.

Statistical Significance and Legal Significance

The scientist's concept of statistical error does not translate directly into the judge's concept of legal error. It cannot be said, therefore, that a study that is statistically significant at the .05 level of confidence will lead judges, if they admit the evidence, to make only 5 errors (Type I errors) out of 100. There is no true correspondence between statistical confidence and legal burdens of proof.

Limitations of moving from statistical significance to legal significance:

- a confidence level is a statistical statement and does not incorporate the variety of factors that judges must take into account in making a decision
- most scientific research examines the general relationship between variables, while trial courts are usually concerned with specific effects on specific individuals

➡ ***Statistical Significance and Importance***

The significance of a finding (the probability of a Type I error) does not have a clear relationship to the importance of the finding, either. A small difference, or a small correlation, could still be highly significant statistically, if the sample were large enough. A finding of small magnitude would still be reliably replicated on repeated investigations, if large portions of the population were included in each investigation – but although dependable, the finding might not have practical or theoretical importance.

Common Problems with the Use of Statistical Evidence in Court

???

Questions to consider when evaluating scientific evidence

- ◆ Were the appropriate statistical tests conducted?
- ◆ Were proper statistical inferences drawn from the data?
- ◆ Were there methodological problems that undermined the scientific validity, reliability, or relevancy of the statistical results?

◆ **Statistics in court are not presented in their natural form**
 In scientific evidence, statistics are rarely presented in a single, complete presentation (i.e., one side presents statistical evidence that is challenged on cross-examination and then, at a later point in the trial, the other side proffers opposing statistical conclusions). In court, statistics are often presented in graphic form, and there is rarely a detailed discussion of the statistical techniques and models used, their assumptions and shortcomings.

◆ **Improper inferences are drawn**
 When statistics are presented in court, improper inferences are often drawn about what the data mean and what conclusions can be drawn. This problem typically occurs in three ways: (1) by extrapolating results of a statistical analysis to a population that is different from the population defined in the study; (2) by inferring, within the correct population, something beyond what is statistically correct given the available data and analysis; and (3) by misinterpreting statistical significance and the burden of proof.

◆ **Improper methodologies used**
 Methodological problems that undermine the scientific validity or relevance of statistical results occur at many stages of the research: study design, data collection, and data analysis.

Iancu, C.A., and Steitz, P.W. (1997). "Guide to Statistics," *Expert Evidence: A Practitioner's Guide to Law, Science, and the FJC Manual*. St. Paul, Minnesota: West, pp. 298-310.

Endnotes:

1. *Daubert vs. Merrell Dow Pharmaceuticals* 509 U.S. 579, 113 S.Ct., 2786 at 508.
-

GLOSSARY

bias	a constant difference, in one direction, between the mean of the sample and the mean of the population; occurs when most of the sampling error loads up on one side, so that the sample means are constantly either over- or under-estimating the population mean
bimodal distribution	a distribution of scores with two modal scores (two commonly occurring scores)
confidence level	specifies a range of values around the empirically obtained result within which the “true” or population value is likely to lie
correlation	an association between two variables; can be positive or negative; correlation does not equal causation
correlation coefficient	a number between -1 and 1 which measures the degree to which two variables are linearly related; if there is a perfect positive linear relationship, $r = 1$ (i.e., an increase in one variable is associated with an increase (or decrease) in the other variable); if there is a perfect negative linear relationship, $r = -1$ (i.e., an increase (decrease) in one variable is associated with a decrease (increase) in the other variable; if $r = 0$ there is no linear relationship between the variables
decision rule	specifies precisely when the null hypothesis should be rejected
descriptive statistics	statistics that summarize, describe, and make understandable the numbers generated in a research study
distribution	the arrangement of any set of scores or values in order of magnitude
error rate	the likelihood with which a measurement device or a technological procedure leads to an incorrect classification
false negative error	incorrectly classifying someone who has a particular characteristic as someone who does not possess that characteristic

(e.g., person has disease, but is incorrectly identified as not having it)

false positive error	incorrectly classifying someone without a particular characteristic as possessing that characteristic (e.g., person does not have disease, but incorrectly classified as having disease)
frequency distribution	a listing, or order of magnitude, of each score and how many times that score occurred
inferential statistics	statistics used to draw conclusions and inferences which are based upon, but go beyond, the numbers generated by a research study
interval scale	a unit of measurement characterized by equal intervals; measures differences in amount (e.g., I.Q. score)
linear regression	predicts the magnitude of the expected change in variable Y given a change in variable X
mean	the arithmetic average of all the scores; calculated by adding all the scores together and then dividing by the total number of scores involved
measures of central tendency	measures that provide information about the average, or typical, score of a large number of scores; which single score (mean, median, mode) best represents an entire set of scores
measures of variability	procedures used to describe the extent to which scores differ from one another in a distribution; range, standard deviation, and variance statistics
median	the exact midpoint of any distribution; much more accurate representation of central tendency than the mean; to calculate the median, the scores must first be arranged in order of magnitude (e.g., from lowest to highest), the middle score is the median
mode	a measure of central tendency; the most common single number in the distribution; in a perfectly symmetrical unimodal distribution, the mode is the same as the mean; when it is not the same, the mode is not really a good representative value of the distribution
multiple linear regression	designed to examine the relationship between a response variable and several possible predictor variables
negatively skewed distribution	distribution in which scores are concentrated near the top of the distribution; tail of the distribution points to the low or negative end
nominal scale	a unit of measurement based on classification; measures differences in kind (e.g., ethnicity)

nonlinear regression	designed to describe the relationship between a response variable and one or more explanatory variables in a non-linear fashion
normal curve	a theoretical distribution; a unimodal frequency distribution with scores plotted on the X axis (the horizontal axis) and frequency plotted on the Y axis (the vertical axis); most of the scores cluster around the middle of the distribution; curve is symmetrical and all three measures of central tendency (mean, median, mode) fall precisely at the middle of the distribution
ordinal scale	unit of measurement characterized by order and classification; measures differences in degree (e.g., attitudes)
pearson's product moment correlation coefficient	a measure of the linear association between two variables that have been measured on interval or ration scales (e.g., the relationship between height in inches and weight in pounds); usually denoted by r , is an example of a correlation coefficient
population	an entire group of persons, things, or events having at least one trait in common; the larger group of all people of interest from which the sample is selected
positively skewed distribution	distribution in which scores are concentrated near the bottom of the distribution; tail of the distribution points to the top or positive end
range	a measure of variability; the width or spread of an entire distribution; found simply by calculating the difference between the highest and lowest scores
regression	predicts the extent to which the value of one or more variables can be predicted by knowing the value of other variables
ratio scale	a unit of measurement characterized by a true zero and equal intervals; measures differences in total amount (e.g., income)
sample	a smaller number of observations taken from the total number making up the population; in typical applications of inferential statistics, the sample size is small relative to the population size
simple linear regression	designed to determine whether there is a linear relationship between a response variable and a possible predictor variable
skewed distribution	a distribution of scores where the majority of scores in the distribution bunch up at one end of the distribution
standard deviation	a measure of variability; a measure of the variability that indicates by how much all of the scores in the distribution typically deviate or vary from the mean
standard normal curve	the normal curve is marked off in units of standard deviation; a normally distributed set of scaled scores whose mean is always

equal to zero and whose standard deviation equals 1.00

true positive error	correctly classifying someone as possessing a particular characteristic or falling into a particular category (e.g., person has disease and is classified as having disease)
true negative error	correctly classifying someone who does not possess a particular characteristic or who does not fall into a particular category (e.g., person does not have disease, and is classified as not having the disease)
type I error	when the researcher rejects the null hypothesis when the null hypothesis is true
type II error	when the researcher fails to reject the null hypothesis when the null hypothesis is false
unimodal distribution	a distribution of scores with a single modal score
variance	measures how much of the variance between people on one characteristic can be explained by where they stand on another characteristic

SUGGESTED READINGS:

- Barnes, D.W. (1983). *Statistics as Proof: Fundamentals of Quantitative Evidence*. Boston: Little, Brown and Company.
- DeVore, J. and Peck, R. (1997). *Statistics: The Exploration and Analysis of Data, 3rd Edition*. San Francisco, CA: Duxbury.
- Hagg, R.V. and Craig, A. T. (1995). *Introduction to Statistics, 5th Edition*. Englewood Cliffs, N.J.: Prentice Hall.
- Iancu, C.A., and Steitz, P.W. (1997). "Guide to Statistics." In *Expert Evidence: A Practitioner's Guide to Law, Science, and the FJC Manual*. Washington, D.C.: Government Printing Office. pgs. 298-310.
- Kaye, D.H., and Freedman, D.A. (1994). "Reference Guide to Statistics." In the *Federal Judicial Center's Reference Manual on Scientific Evidence*. St. Paul, Minnesota: West., pgs. 331-414.
- Saville, D.J. and Wood, G.R. (1996). *Statistical Methods: A Primer*. New York: Springer

Judge's Notes:

CHAPTER 10
Psychological and Psychiatric Evidence: A Brief Overview

Learning Objectives:

Upon completion of this section, the judge-student should be able to:

- ◆ Discuss the differences between clinical and empirical methods in psychological and psychiatric research;
 - ◆ Identify and discuss some of the strengths, weaknesses, and criticisms of the Diagnostic and Statistical Manual of Mental Disorders (DSM);
 - ◆ Discuss the differences between objective and projective personality tests and why this distinction is important;
 - ◆ Identify elements of expertise, including differences in degrees, licensure, and professional affiliations;
 - ◆ Identify different theoretical paradigms within psychology and discuss why recognition of these paradigms might be important; and
 - ◆ Critically evaluate different types of psychological and psychiatric testimony, including prediction of future behavior and psychological syndromes.
-

OUTLINE

- I. The Social Science Challenge**
- II. Clinical vs. Empirical Methods**
 - 1. The Clinical Interview
 - 2. Diagnostic and Statistical Manual of Mental Disorders (DSM)
 - 3. Standardized Psychological Tests
 - 4. Objective and Projective Personality Tests
- III. Who Has the Appropriate Expertise**
 - 1. Licensure
 - 2. Board Certification in a Speciality
 - 3. Membership in Professional Associations
 - 4. Questions to Consider When Evaluating Expert Qualifications
- IV. Theoretical Paradigms in Psychology and Psychiatry**
- V. A Closer Look at the Evidence: Predicting Future Behavior**
 - 1. Assessments of Future Behavior
 - 2. Assessment Methods
 - a. Clinical method
 - b. Actuarial method

VI. A Closer Look at the Evidence: Psychological Syndromes

1. Issue: Falsifiability and Syndrome Evidence
2. Issue: Error Rates and Syndrome Evidence
3. Battered Woman Syndrome
4. Rape Trauma Syndrome

KEY POINTS

- psychological and psychiatric evidence has become a significant part of many trials in the past few decades
- judges who participated in the national survey expressed considerable difficulty in dealing with psychological and psychiatric evidence, due to its “subjective nature,” “lack of scientific rigor,” and the “wide variety and differing backgrounds of experts who testify about psychological and psychiatric matters”
- **the social science challenge**
 - many of the hypotheses of the social sciences are difficult to test because of the complexity of the subject matter and ethical considerations associated with studying human behavior
 - it is important to recognize the agreed upon practices and standards of psychological and psychiatric research and to evaluate the evidence in accordance with those standards
 - it is important to distinguish between the therapeutic utility of a psychological or psychiatric concept and the legal utility of that concept
- **clinical vs. empirical methods**
 - empirical methods: systematic gathering of information and study of problems in accordance with the agreed upon methodological practices of experimental, quasi-experimental or qualitative research
 - clinical methods: based upon observation, history-taking, testing, and diagnosis; relies upon clinical interviews, psychological testing, psycho-social histories, and medical exams
 - clinical interview: includes a systematic history-taking and mental status examination
 - diagnostic and statistical manual of mental disorders (DSM)
 - strengths: provides a standard, comprehensive diagnostic tool; categorizes diagnostic information in a systematic manner
 - limitations: reflects a consensus opinion at the time of publication; controversy surrounds some of the lists of symptoms categorized as mental disorders
 - standardized psychological tests: tests that have been statistically normed or standardized, permitting clinicians to compare scores of test subjects with the scores of normal individuals or patients with similar disorders
 - objective personality tests: generally consist of true-false or check the best answer questions – term objective means that the test scoring is objective, not that the interpretation of the test results is objective (e.g., Minnesota Multiphasic Personality Inventory - MMPI)
 - projective personality tests: assume that a person’s unconscious motivates and directs daily thoughts and behavior; to uncover these motivations, clinicians

provide ambiguous stimuli to which test-takers provide responses that might reflect their unconscious (i.e., Rorschach Inkblot Test)

- **who has the appropriate expertise**

- licensure: state granted permission to perform certain functions related to diagnosis, treatment, or dispensing of medication; legal permission to perform a specified function; typically requires an educational degree requirement and a specified number of practical training hours
- certification: a formal designation of someone to use a professional title based upon education, training, and experience; a statement authorizing an individual to officially practice in a certain profession
- membership in professional associations: active membership, even elected office, in professional associations may not necessarily correspond to any particular degree, knowledge, skill, experience, training, or education
- questions to consider when evaluating expert qualifications
 - employment history
 - educational history
 - licensing and board certifications
 - membership in national associations
 - areas of expertise/specialty
 - advanced specialized training in that field of expertise, including research, studies, lectures attended or given
 - professional experience
 - publications and professional contributions to the field
 - prior testimony or employment as an expert

- **theoretical paradigms in psychology and psychiatry**

- particular paradigms reflect certain assumptions and values and influence what areas and problems are selected for attention
 - medical-biological theories – focus on biological and physiological conditions; treatment usually involves drugs
 - psychodynamic theories – focus on anxiety produced by unresolved, unconscious conflicts; treatment usually involves helping a patient become more aware of motivations, conflicts, desires
 - behavioral theories – focus on faulty or ineffective learning and conditioning patterns; treatment usually involves reshaping disordered behavior through learning techniques
 - cognitive theories – focus on faulty or unrealistic thinking; treatment usually involves developing new thought patterns
 - social psychological theories – focus on social interactions and the influence of context; not a discipline of psychology typically focused on treatment

- **a closer look at the evidence: predicting future behavior**

- there is a key difference between the ability of clinicians to predict dangerousness and their ability to identify factors that increase the potential risk of violence
- assessment methods
 - clinical method: based on observation, personal examination, history-taking, and testing
 - actuarial method: based on assigning statistical probabilities of outcomes from combinations of a number of variables that correlate with the behavior at issue (future violence)

- **a closer look at the evidence: psychological syndromes**
 - syndrome: a group or constellation of symptoms that appear together regularly enough to become associated
 - falsifiability and syndrome evidence
 - some syndromes may not be falsifiable or testable, especially to the extent they are derived from Freudian-based theories
 - while falsification is problematic for psychological syndromes, it is important to stress that *Daubert* does not require that falsifiability, or lack of falsification attempts, serve as the definitive admissibility criteria or the definitive characteristic of science
 - error rates and syndrome evidence
 - a consideration of error rates with respect to syndrome evidence can be thought of in terms of weighing the risk of making a “false positive” error or a “false negative” error
 - battered woman syndrome (BWS)
 - 3 distinct periods within the course of a battering episode (Lenore Walker):
 - the tension-building stage
 - an acute battering incident
 - contrite and loving behavior
 - some problems associated with the use of BWS in court:
 - definitional problems with BWS and variations of behaviors exhibited both by battering men and by battered women
 - significant variations among courts about whether BWS testimony should be admitted and for what purpose
 - whether BWS constitutes “good science,” and under what standard of admissibility it should be judged
 - rape trauma syndrome (RTS)
 - 2 phase reaction process that is said to occur as a result of forcible rape or attempted forcible rape (Burgess and Holmstrom)
 - acute phase
 - long-term reorganization phase

KEY TERMS

actuarial method
clinical method
DSM

licensure
objective personality tests
projective personality tests

standardized tests
syndrome

GLOSSARY

actuarial method
certification
clinical interview
clinical methods

DSM
empirical research
licensure
mental status exam

objective personality test
projective personality test
standardized tests
syndrome

SUGGESTED GROUP EXERCISES AND DISCUSSION QUESTIONS



Some suggested group discussions and exercises ...

- ▶ Break students into groups and have them work through an hypothetical case scenario involving the proffer of rape trauma syndrome evidence (**scenario and discussion guide provided**).
 - ▶ Have groups report back the results of their discussion.
- ▶ Break students into groups and have them discuss:
 - ▶ Does psychological and psychiatric evidence constitute scientific evidence in the manner meant by *Daubert*? Might this differ according to which sub-discipline of psychology is at issue - clinical psychology vs. social psychology vs. neuro-psychology?
 - ▶ What challenge does psychological and psychiatric evidence face from evidentiary standards that rely on “falsifiability?” Might this differ depending on the sub-discipline at issue?
 - ▶ What do you think the future holds for the admissibility and utility of various

AMPLE OVERHEADS

- 10.1 Learning Objectives
- 10.2 Things to take into account when evaluating the findings of a clinical interview
- 10.3.1 DSM-IV Multiaxial Diagnosis
- 10.3.2 DSM-IV Multiaxial Diagnosis (Continued)
- 10.3.3 Strengths and Limitations of the DSM
- 10.4.1 Objective Personality Tests
- 10.4.2 Projective Personality Tests
- 10.5 Questions to consider when evaluating expert qualifications
- 10.6 Syndromes
- 10.7 Battered Woman Syndrome
- 10.8 Rape Trauma Syndrome
- 10.9.1 Psychological and Psychiatric Evidence: Questions to Consider
- 10.9.2 Psychological and Psychiatric Evidence: Questions to Consider

SAMPLE HANDOUTS

1. Hypothetical Case Scenario
2. Hypothetical Case Scenario: Instructor’s Discussion Guide

PSYCHOLOGICAL AND PSYCHIATRIC EVIDENCE: QUESTIONS TO CONSIDER

- ◆ Was the appropriate test used for a specific individual (e.g., certain tests are more appropriate for children than adults)?
- ◆ Is the test being used appropriately for the specific legal issue at hand?
- ◆ Can the data obtained be applied properly to the specific subject of that particular legal inquiry?
- ◆ What is the expert's employment history (e.g., does he actually practice in the field he intends to testify about)?
- ◆ What is the expert's educational history (e.g., is the expert's educational history relevant to the field he intends to testify about)?
- ◆ Is the expert appropriately certified or licensed within the profession?
- ◆ To what extent are the expert's professional memberships relevant to his training, expertise, and the issue at hand?
- ◆ What is the expert's particular area of expertise and is it relevant to the issue at hand?
- ◆ What specialized training, including research, studies, lectures attended or given does the expert have?
- ◆ What relevant professional experience does the expert have?

Overhead 10.9.1

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 10: Psychological and Psychiatric Evidence: A Brief Overview, pgs. 167-204) © 1999

**PSYCHOLOGICAL AND PSYCHIATRIC EVIDENCE:
QUESTIONS TO CONSIDER (Continued)**

- ◆ Is the clinical information being presented based on an assessment or a prediction?
- ◆ What method or approach was used to gather information to make the assessment or prediction?
- ◆ What is the reliability or validity of the testimony or opinion?

Overhead 10.9.2

From: *A Judge's Deskbook on the Basic Philosophies and Methods of Science*. (Chapter 10: Psychological and Psychiatric Evidence: A Brief Overview, pgs. 167-204) © 1999

**Post-*Daubert* Science in the Courtroom
Hypothetical: Rape Trauma Syndrome**

Facts

On the night of July 15, 1997, Christopher Dean Rogers was arrested for the alleged rape of Joanne Renee Miller, which the State claims occurred at approximately 2:45 a.m. the morning of June 27, 1997, in Joanne's apartment. Ms. Miller claims that Rogers introduced himself to her at a local dance club, where she regularly goes with friends, and asked her to dance. She states that he seemed "very nice and real polite," so she danced with him several times and accepted when he offered to buy her drinks. She reports "feeling a little drunk" by the time the club closed, and mentioned that she planned to leave her car and take a taxi home. Rogers told her that he didn't live far from her, and suggested that she let him drive her home instead of taking a taxi. When they reached her home, Rogers claimed that he was feeling a little sleepy, and asked Ms. Miller if he might come in for a cup of coffee before going home. Ms. Miller states "I didn't really want to ask him to come in, but he was being so nice about giving me a ride and everything that I felt that it would be rude to say no." Once inside, Rogers immediately asked to see her bedroom. She grew nervous and told him that she wanted him to leave. She claims that Rogers then told her "you've been leading me on all night, and I'm not going home until I get what you've been promising me." He then shoved her onto the couch and held her down with his arm across her throat while he pulled down her pants. He kept his hand around her throat to keep her from struggling while he raped her. During this time, he repeatedly threatened that if she told anybody about what he was doing he would come back and kill her.

Rogers admits that he gave her a ride home. According to Rogers, when they arrived at Ms. Miller's home, she invited him in for a cup of coffee and made it a point to show him her bedroom. They sat on the couch and talked for a while, and finally had spontaneous consensual sex there on the couch rather than going into the bedroom. Rogers claims that he told Ms. Miller that he would like to go out with her again, but forgot to ask for her phone number before he left. He suggests that she was probably upset with him for failing to call her after he said he would.

Ms. Miller did not report the rape until several weeks later, after a friend, Rebecca Sanchez, noticed that Ms. Miller was behaving differently than she usually did. Ms. Sanchez states that she and Ms. Miller used to go to the dance club several times a week, but Ms. Miller suddenly stopped dating altogether and refused to go anywhere. When she finally got Ms. Miller to come to her home, Ms. Miller had checked the locks on her doors and windows several times during the evening. After Ms. Sanchez asked her a number of times why she seemed so worried, Ms. Miller finally told her what had happened with Rogers. Ms. Sanchez then advised her to report the rape at once.

Ms. Miller reports that since the incident she has had an intense aversion to the clothes she was wearing at the time of the incident and that she has a need to feel clean by constant bathing. She also claims to suffer from severe stress when recollecting events surrounding the incident, and has outbursts of anger and irritability when discussing the incident. Ms. Miller is presently looking for a different place to live. She has stopped going out, fears being home alone, and regularly has nightmares in which she dreams about the attack.

The district attorney argues that Ms. Miller was raped by the defendant and moves to call an expert on rape trauma syndrome, who will testify that Ms. Miller's behaviors are consistent with the behaviors of women who have been raped. The defense attorney argues that no medical examination was done, nor was there any physical evidence gathered at the time of the alleged assault to support the claim that the rape in fact occurred. She further argues that Ms. Miller's several-week delay in reporting the alleged rape suggests that Mr. Rogers and Ms. Miller had consensual sex, and that Ms. Miller is using the rape charge as a form of revenge. Defense also contends that Ms. Miller's testimony should be given little regard, as she admits to drug and alcohol use, which diminishes her credibility. The Defense moves to exclude expert testimony about rape trauma syndrome.

Before trial, pursuant to the defense's motion to exclude expert testimony about rape trauma syndrome, the court holds a hearing to determine whether such testimony would be admitted.

Prosecution Expert

- ▶ Ms. Yvonne Joelson, MSW, will testify for the prosecution.

Yvonne Joelson, is a social worker who specializes in the treatment of rape victims. Ms. Joelson has been employed as a counselor at a rape crisis center for three years, has counseled many rape victims, and has testified several times as an expert in cases involving the common characteristics of rape victims and the

Discussion Point: What are the appropriate qualifications for an expert on rape trauma syndrome?

Discussion Point: Peer Review - What disciplines are relevant? Who are appropriate peers? Psychiatry? Psychology, as a general discipline (including experimental psychology, social psychology, etc.) or only clinical psychology? What about counseling psychology? Sociology? Social work?

behaviors which are consistent with rape trauma syndrome. She has received training in rape crisis counseling and has attended conferences on the subject. Ms. Joelson will discuss the available literature on rape trauma syndrome and symptoms associated with the disorder. This literature appears in peer-reviewed academic and clinical journals. In discussing this literature, Ms. Joelson will explain that rape trauma syndrome was first identified in 1974 and has been accepted by national organizations of both psychiatrists and psychologists. Rape trauma syndrome is associated with three phases or stages, the "acute phase", the "outward adjustment stage", and the "resolution phase." The first stage, or "acute phase" begins about the time the rape victim is released or escapes from her attacker and is characterized by disorganization. Typically the

victim develops one of two styles of coping with her emotions: i) an expressed style (where the victim displays feelings by crying, sobbing, smiling, and becoming restless or tense; and ii) a controlled style (where the victim masks her feelings behind a calm, composed, or subdued appearance. The second phase of rape trauma syndrome, or the "outward adjustment stage," typically begins 2-3 weeks after the attack, and is characterized by what is called "reorganization." The victim tries to reassemble her lifestyle, integrating the rape experience. The primary symptoms expressed by someone suffering from rape trauma

syndrome are recurrent and intrusive recollections of the event, and sudden acting or feeling as if the traumatic event were recurring because of an association with an environmental or ideational stimulus. During the last phase, or the "resolution phase," the rape is no longer the central focus in the victim's life. The rape has been accepted as part of the victim's life experience. Some of the behaviors associated with the second stage may flare up at times, but they do so less frequently and with less intensity. Ms. Joelson will note, however, that rape trauma syndrome is not mentioned specifically as a diagnostic category in the fourth edition of the Diagnostic and Statistical Manual (DSM-IV). The American Psychiatric Association does recognize rape, however, as one potential cause of Post-Traumatic Stress Disorder, which is an official diagnostic category listed in the manual. The essential feature of Post-traumatic Stress Disorder is the development of characteristic symptoms following a psychologically traumatic event that is generally outside the range of usual human experience.

*Discussion Points: Is Daubert applicable to rape trauma syndrome evidence?
Which of the Daubert guidelines are most applicable?
Does rape trauma syndrome constitute "scientific" knowledge
or "specialized" knowledge?*

Ms. Joelson has not seen Ms. Miller in treatment, but has interviewed her on two occasions. She will testify that in general, women are reluctant to report incidents of sexual assault. Ms. Joelson will testify that when victimized women do disclose the incidents to friends or family, they are encouraged to report it, but in such instances the reporting is obviously already delayed. On the basis of her training and experience, Ms. Joelson will conclude that women delay reporting because they have suffered an incredible trauma, they are shocked and embarrassed and "feel very dirty." Furthermore, based on her knowledge and experience, Ms. Joelson will testify that Ms. Miller's behavior is consistent with the behaviors associated with rape trauma syndrome, a cluster of symptoms exhibited by a significant number of rape victims.

Defense Expert

- Dr. Mary Evans, Ph.D, Professor of Sociology.

Dr. Evans, a professor of sociology, will testify for the Defense. Ms. Evans teaches sociology at a prominent university, and has written numerous articles on sexual assault, violence against women, and medical sociology. Ms. Evans will testify that there is a large body of literature in peer-reviewed journals that questions the scientific validity of rape trauma syndrome as a psychiatric

Discussion Point: Is this expert appropriately qualified?

*Discussion Points: Peer Review -
What disciplines are relevant? Who
are the appropriate peers?
Publication: What types of journals
are most acceptable?*

designation, and that the prosecution's expert fails to consider literature that is critical of rape trauma syndrome. This counter literature raises serious questions about the general acceptance of rape trauma syndrome. This literature appears in peer-reviewed, but perhaps less widely distributed academic and clinical journals.

Dr. Evans will also testify that the existence of rape trauma syndrome is based upon clinical observation, and that the few empirical studies of rape trauma syndrome that have been conducted have serious methodological flaws. For example, research on rape trauma syndrome, particularly the early work outlining and defining the syndrome, has been criticized for failing to describe sampling procedures, not addressing potential sample bias, not using control or comparison groups, not using standardized psychometric testing instruments, and not documenting the

Discussion Point: How is general acceptance demonstrated?

Discussion Point: Is a reliance on the DSM justification for admissibility?

reliability of the measuring devices used. Ms. Evans will testify that "rape trauma syndrome" is not part of the specialized manual DSM-IV, although it is mentioned as a form of post-traumatic stress disorder. She will also discuss peer-reviewed academic literature that critically evaluates the DSM itself as lacking a solid scientific basis, noting that many of the psychiatric categories contained within the DSM are themselves unfalsifiable.

Furthermore, Dr. Evans will testify about the problematic nature of syndromes in general (e.g., that they are inherently unfalsifiable and typically use post hoc analysis to promote an explanation for post-event behavior), as well as the problems associated with rape trauma syndrome in particular (e.g., that symptoms associated with rape trauma syndrome are mutually contradictory and not subject to disproof).

The Prosecution maintains that expert testimony on rape trauma syndrome is important to explain Ms. Miller's delay in reporting the incident, and to counteract numerous myths and stereotypes that affect common understanding of rape and the behavior of rape victims. The prosecution will also point out that more recent studies on rape trauma syndrome have used control groups from among women whose race, age, and economic status matched those of the victims, and standardized instruments. Despite the critiques of the earlier studies, the results of recent, methodologically sound, empirical research both confirm and reinforce the findings of the initial studies. This research concludes that rape has an immediate, disruptive impact on the victim's emotional and psychological state.

The Defense moves to exclude the prosecution's expert, citing that she should not be qualified as an expert witness and that the testimony concerning rape trauma syndrome should not have been introduced. The Defense contention is that Ms. Joelson does not have a formal psychology degree and, thus, should be barred from being considered an expert on any psychological condition. The Defense also contends that Ms. Joelson will improperly testify that Ms. Miller was actually suffering from rape trauma syndrome. Furthermore, the Defense argues that there is a considerable literature in peer-reviewed journals that questions the scientific validity of rape trauma syndrome as a psychiatric designation.

*Discussion Points: Is rape trauma syndrome falsifiable?
Is psychological syndrome evidence such as rape trauma syndrome reliable and scientifically acceptable?*

➔ ***Should expert testimony on rape trauma syndrome be admitted in this case?***

Psychological Syndrome Evidence & *Daubert*: Instructor's Discussion Guide

➔ ***Is Daubert applicable to psychological syndrome evidence? (Should it be?)***

- ▶ Does rape trauma syndrome evidence constitute "scientific" knowledge or "specialized" knowledge?
- ▶ Is clinical psychology "science" according to *Daubert*? (Are other forms of psychology? Other forms of social and behavioral science?)
- ▶ Is psychological syndrome evidence such as rape trauma syndrome reliable and scientifically acceptable?

Courts around the country have reached contradictory conclusions. Some courts have applied the *Daubert* guidelines and have, as a consequence, rejected syndrome testimony. Other courts have applied *Daubert* and have admitted the testimony. Some courts, in defining psychological syndrome evidence as "specialized knowledge" rather than "science," have determined that the *Daubert* guidelines do not apply.

➔ ***Application of the Daubert Guidelines:***

(1) Falsification

- ▶ clinical syndromes are, by their very nature, inherently unfalsifiable and they are typically used post hoc to promote an explanation for post-event behavior
- ▶ syndrome symptoms are often mutually contradictory and not subject to disproof

Example: According to rape trauma syndrome theory, it is generally believed that a victim of rape may develop one of two styles of coping with her emotions: (1) an expressive style in which she displays her feelings by crying, sobbing, or smiling; or (2) a controlled style in which she masks her feelings behind a calm, composed, or subdued appearance. Thus, two contradictory coping styles are used to "predict back" as evidence of a prior event. In the absence of either style, one may conclude that the victim was not raped. If all behaviors, including exactly opposite behaviors, can be said to support a theory, then the theory cannot be falsified.

(2) Error Rates

- ▶ it is difficult, if not impossible, to calculate potential error rates with respect to psychological syndromes
 - ▶ to assess error rates, judges must evaluate the methodology used in the corroborative testing carried out

- ▶ Were proper control groups used and was the instrumentation used well-accepted in the relevant disciplines? Were the data gathered in such a way that allowed researcher preferences to influence the results? Have there been adequate replications of the findings, or does the claim of the syndrome rest on a narrow data base? (Richardson et al, 1995)

(3) Peer Review

- ▶ What disciplines are relevant? Who are appropriate peers? Psychiatry? Psychology, as a general discipline (including experimental psychology, social psychology, etc.) or only clinical psychology? What about counseling psychology? Sociology? Social work?

Example: Testimony on rape trauma syndrome has been offered through a variety of experts -- psychiatrists, psychologists, psychiatric nurses, psychology graduate students, social workers, directors or workers at rape crisis shelters, and police officers.

Disagreements about the validity of a theory or method may fall along disciplinary lines, even within the same general discipline, meaning that apparent peers can probably be found who share the perspective of the one offering the testimony. But, a clinical psychologist and a social psychologist may offer completely different testimony to explain the same event or circumstance.

Example: A clinical psychologist might offer testimony on rape trauma syndrome to explain why a woman delayed reporting her rape. A social psychologist might explain the same counter-intuitive behavior on the part of the woman by testifying about rape myths and stereotypes and the role they play in decisions to report a rape. What differences does this make in terms of admissibility?

(4) General Acceptance

- ▶ General acceptance within what discipline? Within how many relevant disciplines?
- ▶ Is reliance on the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM) justification for admissibility?
 - ▶ But, consider that many of the disorders in the DSM lack solid scientific basis and are themselves unfalsifiable. Consider that rape trauma syndrome is not explicitly listed as a diagnostic category in the DSM-IV, but rather rape is listed as an event that can lead to post-traumatic stress disorder.

➔ **General Evidentiary Issues -- Rape Trauma Syndrome**

- ▶ **Who is an appropriate expert? What makes an expert appropriately qualified to testify about rape trauma syndrome?**

Example: Testimony on rape trauma syndrome has been offered through a variety of experts -- psychiatrists, psychologists, psychiatric nurses, psychology graduate students, social workers, directors or workers at rape crisis shelters, and police officers.

- ▶ **Does the probative value of rape trauma syndrome evidence outweigh potential prejudice and confusion?**

- ▶ A number of courts have found that rape trauma syndrome evidence is probative of whether or not a rape occurred, has probative value, and is not overly prejudicial to the defendant or confusing to the jury. Other courts have rejected rape trauma syndrome evidence on the grounds that it was likely to mislead the jury by inferring that the witness was in fact raped because rape trauma syndrome by definition implies that the victim witness was raped ... otherwise why would she suffer from rape trauma syndrome? When rape trauma syndrome is offered as proof that a rape occurred, the evidence has generally been excluded. Some courts have argued that rape trauma syndrome evidence should be excluded because it was developed as a therapeutic tool to understand and treat rape victims, not as a method for assessing whether a rape has occurred.

- ▶ **Does rape trauma syndrome invade the providence of the jury or provide an opinion on the ultimate issue?**

- ▶ Some courts have ruled that when issues of consent are raised by defense, rape trauma syndrome testimony does not invade the providence of the jury since the expert is subject to cross-examination and the jury determines what weight, if any, the evidence should receive.

- ▶ **Does rape trauma syndrome enhance the credibility of the witness?**

- ▶ Some courts have excluded rape trauma syndrome evidence on the grounds that it carries with it an implied opinion that the victim was telling the truth. Others have ruled that although rape trauma syndrome evidence corroborates the victim's claim of rape, thereby supporting or bolstering the credibility of the victim, it nevertheless assists the jury in understanding some aspects of the behavior of rape victims and is therefore admissible as long as expert testimony does not include comment on the truthfulness of the victim's charge.

- ▶ **Does rape trauma syndrome evidence constitute hearsay?**

- ▶ Rape trauma syndrome evidence has been rejected on the grounds that it constitutes hearsay when victims see psychiatrists in preparation for trial rather than as an ongoing part of treatment. Other courts have found that rape trauma syndrome testimony does not constitute hearsay when based on data personally received by the testifying expert.

- ➡ ***If expert testimony on rape trauma syndrome was proffered in a civil case, such as the one described below, would the evaluation of the admissibility of***

the evidence differ in any way?

Civil Case

Christopher Dean Rogers is suing Joanne Renee Miller for slander. He claims that her charge of rape and the subsequent rape trial have defamed his character. The rape trial has received a lot of media attention in the community and Mr. Rogers' has been identified as the alleged rapist. Mr. Rogers is a real estate agent and he claims that the success of his business relies, in part, on his good name and reputation in the community. Since the rape charge was filed, Mr. Rogers claims that there has been in a dramatic decrease in the number of new listings with his company and him in particular, and that some of his current clients have removed him as their realtor. Mr. Rogers' company strongly recommended that Mr. Rogers take a leave of absence. Mr. Rogers fears that his reputation can never be fully recovered and that this will impact his long-term ability to conduct his business in the community in which he currently resides and in which he grew up.

Discussion Points: How might differing levels of proof in criminal and civil cases alter the evaluation of the evidence? Are there due process concerns or constitutional protections that might affect the admissibility decision?

Mr. Rogers maintains that he was falsely and maliciously accused of rape. He claims that the sex was consensual, and maintains that Ms. Miller is getting back at him for not telephoning her when he said he would. Mr. Rogers claims that a rape never occurred. Ms. Miller maintains that she was raped by Mr. Rogers on July 15, 1997 in her home. Mrs. Miller's therapist, Yvonne Joelson, will testify that Ms. Miller's behavior is consistent with the behavior of someone who suffers from rape trauma syndrome.

➔ ***Should testimony on rape trauma syndrome be admitted in this civil case?***

➔ **Would the evaluation of different types of psychological syndrome evidence differ from the evaluation of rape trauma syndrome in any significant way? If so, in what ways? If not, why not?**

▶ Would the evaluation of Battered Woman Syndrome evidence differ in any way?

Discussion Points: Rape trauma syndrome is typically proffered by the Prosecution and Battered woman syndrome is typically proffered by the Defense --does this alter the evaluation of the evidence in any way?

▶ What about Rape Trauma Syndrome proffered in a case where a man has been raped, or Battered Spouse Syndrome proffered by a battered husband?

Discussion Point: Does the novelty of the evidence alter its evaluation?

▶ What about Child Sexual Abuse Accommodation Syndrome?

Discussion Points: Child Sexual Abuse Accommodation Syndrome is often proffered in family court --how might this alter the evaluation of the evidence? How does the level of controversy surrounding the use of specific types of evidence, such as Child Sexual Abuse Accommodation Syndrome, alter the evaluation of the evidence?

▶ What about the profile evidence, such as Drug Courier Profile, Gang Member Profile, or Battering Parent profile evidence?

Discussion Points: Who is an appropriately qualified expert to testify about profile evidence --a police officer? a gang member? a psychologist? a sociologist? Does profile evidence constitute "scientific" knowledge or "specialized" or "technical" knowledge?

Discussion Point: When considering the admissibility of psychological syndrome evidence are there some general considerations that apply regardless of the specific type of syndrome, or profile, being proffered?

DNA EVIDENCE: CRITICAL QUESTIONS TO CONSIDER

- ◆ What kind of DNA test is at issue?
- ◆ What kind of tissue is being tested?
- ◆ Was the sample mixed with tissue from several possible sources?
- ◆ Was the crime scene contaminated?
- ◆ Was the sample too small for proper testing?
- ◆ Was there destruction of the sample?
- ◆ What about the problem of shifting bands?
- ◆ What about the possibility of false assumptions?
- ◆ Were poor quality laboratory procedures used?
- ◆ Was the population database used too small or not appropriately used?
- ◆ What kind of statistical estimates were made based on the match?
- ◆ Were there estimates of error rates?
- ◆ What is the evidence being used for?
- ◆ What are the background and qualifications of the expert?