Psychology, Public Policy, and Law
2002, Vol. 8, No. 4, 339–372
Copyright 2002 by the American Psychological Association, Inc.
1076-8971/02/$5.00 DOI: 10.1037//1076-8971.8.4.339

THE EFFECTS OF *DAUBERT* ON THE ADMISSIBILITY OF EXPERT TESTIMONY IN STATE AND FEDERAL CRIMINAL CASES

Jennifer L. Groscup
University of Nebraska—Lincoln

Steven D. Penrod
John Jay College of Criminal Justice

Christina A. Studebaker
Federal Judicial Center

Matthew T. Huss
Creighton University

Kevin M. O'Neil
University of Nebraska—Lincoln

Appellate opinions were evaluated on variables related to expert admissibility to assess the effects of *Daubert v. Merrell Dow Pharmaceuticals, Inc.* in criminal cases. Analysis reveals changes in appellate courts' consideration of *Frye v. United States*, the 4 *Daubert* criteria, and several Federal Rules of Evidence. The importance of *Frye* and the general acceptance criterion decreased over time, and the importance of the *Daubert* criteria increased over time. However, these changes were not consistent for all types of testimony. Overall, there is greater reliance on *Daubert* when determining the admissibility of a scientific expert. However, only criteria related to the Federal Rules of Evidence are reliably related to admissibility decisions. Details of appellate court application of the 4 *Daubert* criteria, the Federal Rules of Evidence, and other related factors are discussed.

Scientific and technical advances are being made daily; therefore, keeping up to date on all, or even most, of these advancements is nearly impossible. In courts, these advancements affect the evidence that is presented, including expert testimony. Determining which expert evidence is admissible and which is "junk" is a daunting task, particularly for the trial judges who are responsible for making these decisions. Not only must judges decipher complex issues in fields of study that are possibly unfamiliar to them in order to determine admissibility, but their decisions to admit or exclude expert testimony may affect the outcome of the case. In civil cases, this could result in an injured plaintiff who is denied compensation for some grievous harm; in criminal cases, expert testimony could convince a jury to convict a defendant, depriving him or her of freedom or even of life. Recent

---

Jennifer L. Groscup and Kevin H. O'Neil, Law/Psychology Program, University of Nebraska—Lincoln; Steven D. Penrod, John Jay College of Criminal Justice, City University of New York; Christina A. Studebaker, Research Division, Federal Judicial Center; Matthew T. Huss, Department of Psychology, Creighton University.

Christina A. Studebaker is now at the Chicago School of Professional Psychology. Kevin H. O'Neil is now at the John Jay College of Criminal Justice.

This research was supported by National Science Foundation Grants 9618580 and 9974577 to Steven D. Penrod, and is in partial fulfillment of Jennifer L. Groscup's dissertation. Portions of this research were presented at the 2nd Annual National Conference on Science and the Law, San Diego, California, October 13, 2000. We are grateful for the research assistance of Justin D. Gill on this project.

Correspondence concerning this article should be addressed to Jennifer L. Groscup, who is now at the John Jay College of Criminal Justice, Department of Psychology, 445 West 59th Street, New York, New York 10019-1199. E-mail: jgroscup@jjay.cuny.edu

changes in admissibility standards have made expert testimony admissibility an important field of study. Since 1923, the predominant standard for evaluating expert testimony was derived from the opinion in *Frye v. United States*. In *Frye*, the D.C. Circuit stated that for scientific evidence to be admissible the method used "must be sufficiently established to have gained general acceptance in the particular field in which it belongs" (p. 1014).

In 1976, the Federal Rules of Evidence were adopted. Article VII of the Rules governs the admissibility of expert and opinion testimony. During the period in which the cases in this study were decided, Rule 702 stated that "[i]f 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" (Mueller & Kirkpatrick, 1999). The Advisory Committee's Note to Rule 702 stated that the rule is phrased broadly to include a wide variety of witnesses and types of testimony such that "within the scope of the rule are not only experts in the strictest sense of the word, e.g. physicians, physicists, and architects, but also the large group sometimes called '*skilled*' witnesses, such as bankers or landowners testifying to land values" (Mueller & Kirkpatrick, 1999). The Note also suggests that, within the requirements of Rule 702, assistance provided to the trier of fact is of paramount importance to admissibility. The adoption of these Rules did not specifically preclude the use of the *Frye* standard to evaluate expert testimony.

In 1993, the Supreme Court opinion in *Daubert v. Merrell Dow Pharmaceuticals, Inc*. modified the standard for the admissibility of expert testimony in federal courts. Some commentators had noted that the *Frye* standard was too inflexible. New science and new technology were developing at a rapid rate, and it was taking time for these advances to be accepted by the relevant community. Therefore, the general acceptance standard was suggesting the exclusion of some novel evidence even though it may have been reliable (Saks, 1998). On the other hand, the Rules provided for a more liberal admission of expert testimony. The *Daubert* Court determined that the Rules superceded the general acceptance test established in *Frye* as the standard for determining expert testimony admissibility because there was no indication that the Rules were intended to incorporate the *Frye* test. The Court also reasoned that the *Frye* test was a rigid, "austere standard" that "would be at odds with the 'liberal thrust' of the Federal Rules" (*Daubert*, 1993, p. 588), and that the main focus of a court's admission decision should be the requirements set forth in Rule 702. When evaluating scientific knowledge, the reliability of the scientific methods must be assessed. The *Daubert* opinion reemphasized the trial judge's duty to evaluate the quality of the evidence presented, including the underlying quality of the science presented via expert testimony (*Daubert*, 1993). The *Daubert* Court was lauded for this aspect of its opinion by those who argued that the Rule should require an assessment of evidentiary reliability in the context of expert testimony (Slobogin, 1998). Rule 702 was later amended to require an assessment of reliability, consistent with *Daubert* and its supporters.

The main focus of the *Daubert* opinion was to encourage courts to evaluate scientific evidence based on the methodology used and not the conclusions derived by the researcher (*Daubert*, 1993). However, most judges are not trained

in scientific methodology. To assist judges who lack scientific training in their "gatekeeping" task of evaluating the reliability of scientific evidence, the Court created a nonexclusive list of factors to be considered. First, judges may consider whether the knowledge "can be (and has been) tested" or is falsifiable. Second, "whether the theory or technique has been subjected to peer review or publication" can be considered as a means of assessing the validity of the methodology. Third, the known or potential error rate of the technique should be considered. Fourth, although no longer a necessary precondition for admission, general acceptance may also be used by judges as a criterion (*Daubert*, 1993). These four factors were designed to assist judges in evaluating scientific and complex expert evidence. The Court reasoned that this type of evaluation, in conjunction with Rule 702, would provide a more flexible standard than that derived from *Frye*.

The *Daubert* decision and its potential effects on the general admissibility of expert testimony created an explosion of discussion in both legal and scientific communities. There was little agreement concerning the likely future effects of the *Daubert* decision, and several questions were raised. Some scholars argued that there would be no dramatic effects on admissibility as a result of the modified standard (Friedman, 1994). Others predicted that the courts would continue to utilize the familiar standard set forth in *Frye* while labeling the analysis as "*Daubert*esque" (Allen, 1994; Gottesman, 1994; Mack, 1994). Some scholars argued that by excluding "junk science" more effectively, admission rates would decrease (Black, Ayala, & Saffran-Brinks, 1994). Others argued that the effects of the *Daubert* opinion on admissibility would depend on the type of testimony offered (Faigman, Porter, & Saks, 1994). For example, "soft" sciences, such as psychology, might be admitted less frequently after *Daubert*, but "hard" sciences, such as chemistry, might not be affected (Kobayashi, 1995).

Although the Court attempted to provide guidance to judges on evaluating scientific reliability, the practical applicability of the guidelines established in *Daubert* has been questioned (Harvard Law Review, 1995). It has been argued that judges may have difficulties applying the four *Daubert* criteria because of their general unfamiliarity with scientific theories and methods (Faigman, 1995). Recent research has shown that judges may be unable to understand and apply these criteria appropriately (Gatowski et al., 2001). The use of court-appointed experts under Rule 706 has been suggested as a potential aid to judges in this regard (Cecil & Willging, 1994). Others have argued that judges are able to perform the gatekeeping function on scientific evidence, that they are actually performing it, and that they are taking the duty seriously (Clark, 1996).

Another important issue with regard to the admissibility of expert evidence in appellate courts was the standard under which trial court decisions would be reviewed. The standard of review has the potential to affect admissibility because it determines what appellate judges can consider in their decisions and the amount of deference given to the trial court. For several years following the *Daubert* decision, appellate courts were using different standards of review in their decision making about admissibility. Prior to *General Electric Co. v. Joiner* (1997), legal scholars argued that different levels of review should be applied to the decision to admit expert testimony depending on the issue raised. For instance, a de novo standard should be used to review aspects of expert testimony which have broad ramifications for subsequent cases, including scientific theory and

procedures. An abuse of discretion standard should be used to review other aspects relating to the specific case in question (Faigman, 1995, 1997). In *General Electric Co. v. Joiner* (1997), the Supreme Court determined that abuse of discretion was the appropriate standard of review to be applied to all expert testimony issues.

Although *Daubert* concerned a civil matter, its ruling extends to criminal as well as civil cases. The influence of *Daubert* on the admissibility of expert testimony in criminal cases is especially important because of the significance of a potentially unfavorable outcome for the defendant. In a criminal case, the outcome of the decision to admit or exclude expert testimony could affect the defendant's freedom, liberty, and life. The expert evidence proffered can range from testimony about complex procedures such as DNA testing to common events such as the observations of a police officer. Much expert evidence in criminal cases, such as testimony about fingerprint identification, has been offered and accepted in courts for many years in countless cases. Because this evidence is not novel to criminal courts, its reliability may not be reexamined under *Daubert*. Prosecution experts are frequently employees of the state and are crucial to the adjudication of criminals. As such, they could be perceived as inherently reliable. This could result in differential treatment of these experts in the criminal system. Determining the effects of *Daubert* on criminal cases could prove especially important given the recent Supreme Court decision in *Kumho Tire Co., Ltd. v. Carmichael* (1999). *Kumho* determined that *Daubert* can be applied to nonscientific testimony, such as technical and specialized testimony. Some commentators argue that *Kumho* could exert its greatest influence on criminal cases because much of the expert evidence proffered in these cases is specialized in nature (Rovella, 1999). Therefore, it is particularly important to determine the impact of *Daubert* in these cases.

Our investigation of *Daubert* (and continuing investigation of *Kumho*) attempts to shed some light on these issues. In the present study, we address the questions raised by *Daubert* as they pertain to criminal cases. Are judges in appellate courts discussing *Daubert* or paying attention to the decision's suggestions? When they discuss admissibility, what is the content of that discussion? How influential are *Frye*, *Daubert*, and the Rules? Are they applied similarly to all types of expert testimony? Finally, how has the admissibility of expert testimony been affected by the *Daubert* decision?

Method

In our efforts to determine the actual effects of the *Daubert* decision, we studied appellate court opinions concerning expert testimony. Appellate court decisions were selected because of their potential to demonstrate trends in judicial decision making about expert testimony. Decisions involving expert testimony were located with the Westlaw database using the search term "admiss! 5 expert & witness," which garnered the most relevant cases. Searches were limited to those cases decided within the 5½ years prior to *Daubert* (decision issued on June 28, 1993) and those decided within the 5½ years following *Daubert*, totaling 11 years of appellate court opinions. Cases that were included in the study contained substantive discussion of expert testimony admissibility. Cases that were excluded dealt with issues other than admissibility, (e.g., procedural errors and ineffective assistance of counsel). All of the federal cases were evaluated and coded if they met the selection

criteria. Because many more state cases than federal cases were identified (over 2,000), an equivalent number of state cases were randomly selected for evaluation and coding.

Once a case was selected for evaluation, it was coded for content by one of three independent raters.[1] If a case contained substantive admissibility discussion of more than one expert, then each expert was coded separately. All of the coders were graduate students in both psychology and law, who had taken at least 1 year of classes in law school, including evidence law. Cases were randomly assigned to raters and were coded in a random order. The total number of experts coded was 1,184, drawn from both civil and criminal appellate cases. There were 693 criminal appellate cases. The 491 experts discussed in civil appellate cases are not included in the analysis reported in this article.

The coding scheme consisted of over 100 variables relating to the admissibility of expert testimony, only some of which are reported here. The categories of variables used in these analyses included information about the case in general, the expert, the admission decision, and courts' use of *Frye*, *Daubert*, and selected Federal Rules. General information about the case was recorded to provide information about the types of criminal cases that are appealed based on issues related to expert testimony. General case information included jurisdiction (federal or state), the state in which the decision was rendered (when the decision was from a state jurisdiction), the type of case (criminal or civil), descriptions of the parties involved, the type of counsel retained by the parties, and when the case was decided relative to *Daubert* (number of months prior to or post-*Daubert*).

Because *Daubert* focused on expert testimony issues, the characteristics of the experts in appellate cases are of interest, such as the qualifications of the experts proffered. Coded information about the expert included the domain of the expert's testimony, the number of experts testifying for each of the parties, the party for whom the key expert testified, and the basis for the expert's knowledge (i.e., theory, body of research, case specific research, experience, education, and case specific experience). Some of the more intense post-*Daubert* controversy centered around rates of admission. Coded information about the admission decision included the trial court admission decision and the appellate court admission decision. Because of its potential to affect appellate admissibility, we recorded the standard of review and when the case was decided with respect to *Joiner*.

Among the most interesting effects of the *Daubert* decision are the potential decrease in courts' use of *Frye* and the application of *Daubert* and the four *Daubert* criteria. Additionally, courts' use of various Rules is of interest because of their potential importance in the admission decision, as mentioned in the *Daubert* opinion. They could have procedural influence, could become more important after *Daubert*, or could be applied differently to various types of testimony. Courts' use and evaluation of *Frye*, *Daubert*, and the Rules was recorded in three ways. First, the length of the discussion in words devoted to several variables was recorded as a measure of the attention paid to these concepts by the courts. Second, although it is interesting to note how much attention is focused on different factors relevant to admissibility, the more interesting question is how these factors affect the admission decision. The influence on the admission decision of several potential judgment criteria was recorded. Ratings of influence were determined by the rater and were recorded on a 9-point Likert-type scale, with 0 indicating that the variable was never mentioned, 1 indicating that it was not influential, and 9 indicating that it was highly influential. Third, courts' determinations of whether the testimony "met" or "did not meet" several potential judgment criteria were recorded. One of these factors could have a rating of influence on the decision and not have been mentioned as a criterion that was specifically met or not met.

---

[1]Interrater reliability was computed for each variable in the dataset. Cases were double coded, and reliability was calculated by comparing pairs of coder responses. Kappas were computed for categorical variables, and correlations were used to compare continuous variables. Kappas ranged from .324 to 1.0, with 80% of the variables having a value over the recommended value of .8. Correlations ranged from .006 to 1.0, with 87.3% having a significance of $p < .05$. Given the number of variables in the dataset, we considered these results to indicate satisfactory interrater reliability.

On the foregoing variables, each of the four *Daubert* criteria and the *Daubert*-related concepts were evaluated separately. These included *Frye*, falsifiability, peer review, error rate, general acceptance and reliability. The general acceptance standard was evaluated separately from the discussion of the *Frye* opinion because it was also one of the four specified *Daubert* criteria. Courts could discuss the *Frye* standard in general, perhaps discussing its use in the past or how it was overruled in *Daubert* without applying the general acceptance standard to the testimony in question. Reliability also was recorded separately from the four *Daubert* criteria themselves, even though the criteria were intended to indicate reliability, because it is possible that courts could make judgments about this more general concept without making reference to any of the four *Daubert* criteria. The Rules evaluated included Rule 104 (judge can determine preliminary questions, such as admissibility), 403 (exclude prejudicial evidence), and 702 (expert testimony may be admitted if it assists the trier of fact, the expert is qualified, and the knowledge is scientific, technical, or specialized). Potential judgment criteria drawn from the Rules that were evaluated included relevance, expert qualifications, assisting the trier of fact, and prejudicial impact.

Results

Description of the Cases

Of the 693 criminal cases examined, 372 were from federal appellate courts, and 321 were from state appellate courts. Analyses of these data are described in terms of the timing of the case (before vs. after the *Daubert* decision) and in terms of the type of testimony presented. Because federal courts are bound by the *Daubert* decision, the actual date of the Court's decision was used to divide the federal cases in the sample into cases decided either before or after *Daubert*. State courts were not required to adopt the *Daubert* decision as their standard for the admissibility of expert testimony. Some state courts chose to adopt *Daubert*, whereas others kept *Frye* as their standard or developed their own standard. Those states that did adopt *Daubert* did so at different times (Faigman, Kaye, Saks, & Sanders, 1997). To distinguish between cases decided before versus after *Daubert* among the state cases in the sample, the date that each state adopted *Daubert* was used. If a state did not adopt *Daubert*, all cases from that state were considered to be before *Daubert*. Using these distinctions, 458 (66.1%) of the criminal decisions were rendered before *Daubert*, and 235 (33.9%) were rendered after *Daubert*. To assess more detailed trends in the data, the cases decided after *Daubert* were subdivided into three groups: cases decided 1–22 months after *Daubert* ($n = 100$, 14.5%), cases decided 23–44 months after *Daubert* ($n = 92$, 13.3%), and cases decided 45–66 months after *Daubert* ($n = 42$, 6.1%). Only results that demonstrated significant differences across these three post-*Daubert* time periods are described.

Not surprisingly, the vast majority of appellants in criminal cases consisted of private parties ($n = 675$, 97.4%), and the mean number of appellants was 1.45 ($SD = 1.33$), indicating that appeals regarding expert testimony are most often initiated by single defendants. The vast majority of respondents in the criminal appellate cases were representatives of the government ($n = 674$, 97.3%), and the mean number of respondents was 1.01 ($SD = 0.14$), indicating that the prosecution is often the appellee. The expert who was the focus of the appeal most often testified for the prosecution ($n = 510$, 76.1%), but appeals focusing on an expert proffered by the defendant–appellant were also frequent ($n = 160$, 23.9%).

The total number of experts involved in the case on behalf of the appellant

was most often 0 ($n = 456$, 66%), further indicating that the appeal focused on a prosecution expert who had testified. In 179 (25.9%) cases, there was one expert testifying on behalf of the defendant–appellant, and more than one expert testified for the defendant–appellant in 56 cases (8.1%). The maximum number of experts testifying for defendant–appellants was 5. In contrast, there were 431 cases (62.3%) in which there was one expert testifying for the prosecution–respondent, and only 135 cases (19.5%) in which no expert testified for the prosecution–respondent. The maximum number of experts testifying on behalf of a prosecution–respondent was 8 and there were 126 cases (18.2%) in which more than one expert testified on behalf of the prosecution–respondent.

Domains of Expert Testimony

To determine whether the type of testimony presented affected admissibility, the specific domain of the testifying expert was recorded. Over 70 categories of expert domains were organized into four "types" of testimony: medical–mental health, technical–engineering experts, scientific experts, and business experts. Of all the types of experts who testified in the criminal cases on appeal, technical experts testified most often, followed by medical–mental health experts, scientific experts, and business experts. The most prevalent topic of testimony concerned police procedures, which was placed in the "technical" category. This type of testimony typically concerned officers' observations of drug dealing activities and the procedures used by the police in apprehending this type of perpetrator. Table 1 presents the topics of testimony comprising each of the four general categories and their frequencies.

Rates of Admission and the Appellate Standard of Review

One of the main concerns raised about the *Daubert* decision was its potential effect on the admissibility of expert testimony. Trial court and appellate court admission decisions were recorded. It is noteworthy to recall that these rates of admission are only for cases that were appealed, not for all cases in which an expert was presented in a trial court. Of the cases that were appealed, 74.3% ($n = 513$) of the experts were admitted at the trial court level. At the criminal appellate court level for those cases, 69.1% ($n = 465$) of the experts were admitted. We computed chi squares to determine the patterns of relationship between (a) admissibility and timing and (b) admissibility and type of testimony at the trial and appellate court levels. Contrary to the predictions of most commentators, the basic rates of admission at the trial and the appellate court levels did not change significantly after *Daubert* in criminal cases on appeal. There were also no significant changes in admission rates across the three post*Daubert* time periods at either adjudicative level. However, different rates of admission at the trial court, $\chi^2(3, N = 669) = 59.58$, $p < .001$, and the appellate court, $\chi^2(3, N = 652) = 52.33$, $p < .001$, levels occurred depending on the type of testimony presented. At both the trial court and appellate court levels, technical testimony was admitted at a higher-than-average rate. Medical–mental health, scientific, and business testimony were all admitted at a lower-than-average rate. Within each type of testimony, there were no significant changes in admission over time at either the

Table 1
Type of Expert Testimony in Criminal Appeals from January 1988 to December 1998

Type of expert testimony	Total no. (n = 693)	% of type	% of total
Medical/mental health	224	100.0	32.3
Medical examiner	8	3.6	1.2
Pediatrician	12	5.4	1.7
Social worker	31	13.8	4.5
Psychiatrist	36	16.1	13.9
Clinical psychologist	96	42.9	13.9
Other medical	41	18.3	5.9
Technical	256	100.0	36.9
Police procedures	239	93.4	34.5
Accident Reconstruction	6	2.3	0.9
Fire/arson	5	2	0.7
Other technical	6	2.3	0.9
Scientific	136	100.0	19.6
Chemists	29	21.3	4.2
Biologists/geneticists	29	21.3	4.2
Experimental psychologists	25	18.4	3.6
Social–behavioral scientists	32	23.5	4.6
Other scientific	21	15.4	3.0
Business	56	100.0	5.0
Accountants	7	12.5	1.0
Business practices	7	12.5	1.0
Attorney	18	32.1	2.6
Securities	10	17.9	1.4
Other business	14	25.0	2.0

trial or the appellate court levels. Admission rates at the trial and appellate levels before and after *Daubert* are presented in Table 2.

One explanation for the lack of any changes in the observed rates of admission before versus after *Daubert* is that admissibility depends on the party offering the testimony. The party for whom the key expert testified was significantly related to admission at both the trial court, $\chi^2(1, N = 686) = 512.70$, $p < .001$, and the appellate court levels, $\chi^2(1, N = 670) = 251.79$, $p < .001$. At both adjudicative levels, experts proffered by the prosecution were more likely to be admitted than experts proffered by defendants. At the trial court level, prosecution experts were admitted 95.8% ($n = 497$) of the time, and defendant–appellant experts were admitted only 7.8% ($n = 13$) of the total number of times they were offered. This pattern was slightly less pronounced at the appellate level, with prosecution experts admitted 85.1% ($n = 434$) of the time and defense experts admitted 18.8% ($n = 30$) of the total number of times they were offered. This indicates the reversal of some trial court decisions by the appellate court. Appellate courts do not simply affirm trial court judgments regarding expert testimony.

Another potential explanation for the stability of admission decisions before and after *Daubert* is the type of counsel retained. Of particular interest in criminal cases is whether an individual, private party appellant retains a private attorney or is assigned a public defender. Popular perceptions of public defenders may be that

Table 2
Trial and Appellate Rates of Admission of Experts Before and After Daubert and Across Testimony Type

Type of testimony	No. (and %) of experts admitted					
	Trial court			Appellate court		
	Before	After	Total	Before	After	Total
Medical	121 (69.1%)	32 (66.7%)	153 (68.6%)	103 (59.9%)	29 (61.7%)	132 (60.3%)
Technical	140 (90.3%)	89 (89.9%)	229 (90.2%)	126 (84.0%)	85 (88.5%)	211 (85.8%)
Scientific	57 (66.3%)	25 (50.0%)	82 (60.3%)	53 (61.6%)	22 (50.0%)	76 (57.6%)
Business	14 (43.8%)	18 (56.3%)	136 (57.1%)	12 (46.2%)	19 (65.5%)	31 (56.4%)
Total	334 (75.4%)	169 (72.2%)	513 (74.3%)	303 (67.8%)	162 (71.7%)	465 (69.1%)

they provide an inferior legal service to their clients compared with privately retained attorneys because of a lack of resources. The significant question is whether this perception is reflected in the admission decision. Of the appellants in criminal cases, 538 (77.6%) were private parties represented by privately retained attorneys, and 113 (16.3%) of the private party appellants were represented by public defenders. The type of counsel retained by private party appellants did not have a significant relationship with either the admission decision at the trial court levels, $\chi^2(2, N = 690) = 1.76$, $p = .416$, or the admission decision at the appellate level, $\chi^2(2, N = 673) = 0.63$, $p = .729$. Therefore, the type of counsel retained by the parties does not explain expert admissibility at the appellate level. This finding also may indicate that, insofar as the admissibility of expert testimony is concerned, popular fears about inferior representation delivered by public defenders are unfounded.

The standard of review used by appellate courts in reviewing the trial court admission decisions may also account for the lack of changes in the appellate rates of admission. Nine separate standards of review were recorded. The standard of review was significantly related to appellate admissibility, $\chi^2(7, N = 636) = 21.11$, $p = .004$. When plain error was the standard, admission of experts occurred at a higher than average rate. In 93.2% ($n = 55$) of the cases in which plain error was the standard, the evidence was admitted. A somewhat different relationship was observed when abuse of discretion was the standard of review. The testimony was admitted in the majority of the cases in which abuse of discretion was the standard ($n = 319$, 65.6%), but abuse of discretion was most likely to be the standard of review when the testimony was excluded. Of the excluded experts, 84.8% ($n = 167$) were reviewed under an abuse of discretion standard. Plain error and abuse of discretion were evenly split between cases decided before and after *Daubert*, and the remaining standards of review were not related to admissibility. Therefore, it is unlikely that the standard of review fully explains the observed admission rates pre- and post-*Daubert*.

In *General Electric Co. v. Joiner* (1997), the Supreme Court determined that

an abuse of discretion is the appropriate appellate standard of review to be applied to expert testimony admissibility. Using the date of the *Joiner* decision to distinguish between before and after *Joiner*, 646 criminal cases (93.2%) were decided before *Joiner*, and 47 criminal cases (6.8%) were decided after *Joiner*. Surprisingly, no significant shift toward an abuse of discretion as the standard of review was observed after *Joiner*, but some trends deserve mention. Before *Joiner*, 75% ($n = 455$) of the appellate courts used abuse of discretion as the standard of review, in addition to the 6 other standards. After *Joiner*, 87.2% ($n = 41$) of the appellate courts used abuse of discretion as their standard of review, and the major change was that only two other standards were used. Plain error accounted for half of the remaining post-*Joiner* cases ($n = 3$, 6.4%), indicating procedural problems such as improperly preserving the issue for appeal. Manifest error (substantially similar to abuse of discretion) accounted for the remaining cases ($n = 3$, 6.4%). Whether or not the case was decided before or after *Joiner* was not significantly related to admissibility at the appellate level, $\chi^2(1, N = 673) = 1.29$, $p = .256$. Therefore, it is unlikely that *Joiner's* change in the appellate standard of review for expert testimony explains the lack of change in the observed appellate rates of admission before and after *Daubert*.

Discussion Devoted to Potential Judgment Criteria

Several measures of appellate court judges' decision making about expert testimony were recorded. One of these measures was the length of discussion (in number of words) devoted to several potential judgment criteria, including some relevant Rules, *Frye*, and the four *Daubert* criteria. Both general discussion of these criteria and discussion specifically applying the criteria to the facts of the case were included in this measure. Results of this measure are presented as the mean number of words devoted to the discussion of the selected criteria. For simplicity, the mean length of discussion devoted to a criterion is hereinafter referred to as *length of discussion* of that criterion. These measures are intended to provide some objective information about the amount of attention paid by appellate court judges to different decision factors. Although these measures indicate the *amount* of discussion devoted to these factors, they do not provide information about the *content* of that discussion, nor do they provide information about the influence of these criteria on admissibility. The content of the discussion and the influence of the selected judgment criteria are addressed by variables discussed in subsequent sections.

We computed an analysis of variance (ANOVA) to determine mean differences in the length of discussion of the criteria by comparing several conditions: before versus after *Daubert*, differences across the four types of testimony, and the interaction between timing and testimony type. We performed post hoc tests using Fisher's least significant difference. ANOVAs testing the differences across the four more refined time periods (before, 1–22 months after, 23–44 months after, and 45–66 months after *Daubert*) are only reported where those results were significant. To further uncover the effects of *Daubert*, we also conducted ANOVAs comparing the *Daubert*-adopting jurisdictions (state and federal) to the nonadopting states on all of the measures. Differences between these jurisdictions

after *Daubert* indicate potential effects of *Daubert* on judicial decision making, but only the statistically significant differences are described.

When computing the mean length of discussion devoted to the potential judgment criteria, all cases in the database were included. A large percentage of the cases never mention the criterion for which the discussion was measured. We chose to include in the analysis those cases containing no discussion in order to provide a measure of the attention paid to each criterion overall, not just the attention paid to them when they were discussed. We retained the data in their natural and more interpretable metric rather than using a transformation that renders the data incomparable across variables. Although this results in an attenuation of statistical power, given the large sample sizes with which we are working (and the fact that we included the entire population of federal cases), we opted for interpretability over transformation.

The length of discussion of expert admissibility in general in criminal appellate court cases comprised 22% of the length of the total opinion. The length of the portion of the opinion that discussed admissibility did not change significantly over time. However, the discussion of expert testimony issues was significantly longer in cases involving scientific and medical/mental health experts than in cases involving technical and business experts, $F(3, 667) = 6.67$, $p < .001$. This overall length of expert discussion provides a baseline against which to compare the length of discussion of different criteria within the portion of the opinions dealing with expert testimony. These data are presented in Table 3 for comparison with the discussion of the selected criteria.

Discussion of the Federal Rules of Evidence

Because they provide the standards for admissibility and because the *Daubert* Court emphasized their import, the Rules have the potential to exert a significant impact on the admissibility of expert testimony. The Rules selected for evaluation cover topics such as pretrial determinations of admissibility, qualifications, helpfulness, and prejudicial impact. The overall means, significant changes across time, and significant differences among testimony type for each of the Rules discussed below are presented in Table 3.

The *Daubert* Court highlighted the importance of trial judges' roles as the gatekeepers of scientific evidence. Rule 104(a) states that "[p]reliminary questions concerning the qualification of a person to be a witness . . . or the admissibility of evidence shall be determined by the court" (Mueller & Kirkpatrick, 1999). Judges can make a preliminary assessment of the quality of the testimony when determining its admissibility before to its presentation to the trier of fact. Increased discussion of these preliminary determinations under Rule 104 could indicate an increase in judicial gatekeeping.

Rule 104 was discussed less often compared with the other selected Federal Rules, with a mean length of discussion of 3.54 words ($SD = 28.16$; means are presented in Table 3). The length of discussion of Rule 104 increased significantly after *Daubert*, $F(1, 690) = 8.29$, $p = .004$. However, the increase in the discussion of Rule 104 did not immediately follow the *Daubert* decision. There was no significant increase in the discussion of Rule 104 during the first 22 months after *Daubert*. Discussion of Rule 104 increased significantly during the

Table 3
Mean Number (and Standard Deviation) of Words Devoted to the Discussion of the Federal Rules of Evidence, Frye, Daubert, and the Daubert Criteria

		Timing of case		Type of testimony				
Criterion discussed	Grand mean	Before *Daubert* (n = 458)	After *Daubert* (n = 235)[a]	Med/mental health (n = 224)	Technical (n = 256)	Scientific (n = 136)	Business (n = 56)[b]	Timing × Type interaction
Rule 104	3.54 (28.16)	1.35 (23.44)	7.83** (35.30)	1.81_b (17.46)	0.83_b (12.23)	9.26_a (49.82)	0_b** (0)	*ns*
Rule 403	36.75 (129.88)	32.87 (131.76)	44.29 (126.08)	18.56_b (68.55)	$46.20_{a,c}$ (160.07)	59.43_a (165.05)	$13.71_{b,c}$** (47.67)	*ns*
Rule 702	246.31 (474.21)	201.21 (425.84)	333.63** (546.53)	240.17 (523.34)	228.13 (401.41)	229.75 (540.78)	194.73 (265.02)	*ns*
Frye	97.62 (507.51)	136.00 (612.98)	22.50** (135.33)	44.91_b (199.14)	78.63_b (510.85)	267.97_a (847.70)	0_b** (0)	*ns*
Daubert	97.87 (472.81)	36.71 (407.57)	217.58** (561.63)	48.14_b (282.06)	60.73_b (521.81)	266.07_a (626.71)	7.86_b** (58.80)	**
Falsifiability	4.31 (59.09)	5.05 (71.49)	2.86 (18.42)	0.15 (1.61)	4.69 (70.93)	12.17 (90.69)	0 (0)	*ns*
Peer review/ publication	15.29 (263.53)	20.66 (323.92)	4.8 (17.07)	10.72 (129.68)	0.70 (7.02)	57.75 (570.64)	0 (0)	*ns*
Error rate	6.94 (87.72)	7.08 (103.80)	6.67 (41.37)	0.35_b (4.00)	2.35_b (26.03)	29.95_a (193.53)	0_b** (0)	*ns*
General acceptance	56.44 (270.19)	79.36 (328.62)	11.77** (44.17)	39.33_b (171.30)	22.73_b (140.18)	175.36_a (518.69)	0.14_b** (1.07)	**
Expert (general) discussion	1,161.69 (1,356.32)	1,196.39 (1,390.81)	1,093.78 (1,286.33)	$1,322.77_a$ (1,448.17)	967.89_b (1,216.28)	$1,407.60_a$ (1,508.20)	705.02_b (681.31)	*ns*
Length of entire opinion	5,256.93 (4,768.94)	5,352.23 (4,947.89)	5,069.59 (4,400.27)	4,772.15 (3,352.50)	5,315.89 (5,540.61)	6,016.88 (5,500.68)	4,891.46 (3,125.06)	*ns*

Note. Subscripts appearing next to means indicate the locations of the mean differences where applicable. Means include cases in which these reasons were not discussed, having a rating of zero. Med = medical.
[a]Asterisks indicate a significant mean difference for the timing of the case. [b]Asterisks indicate a significant mean difference for the type of testimony.
**$p < .01$.

23–44-month period after *Daubert* but decreased after that, $F(3, 688) = 5.71$, $p = .001$. Means for the differences in the amount of discussion across these time periods are presented in Table 4. The discussion of Rule 104 was significantly more extensive in cases using scientific experts than in cases using any other type of expert (means are presented in Table 3), $F(3, 667) = 3.69$, $p = .012$. However, there was no significant interaction between the timing of the case and the type of the testimony on this measure.

Courts were instructed in *Daubert* to rely heavily on the evidentiary rules requirements in their admission decisions. Rule 702 should be an important factor in cases dealing with expert testimony because it is the Rule that most directly governs the admissibility of expert opinion testimony in federal courts and because most states have adopted a similar standard. Consistent with this instruction, the results indicate that more attention is paid to Rule 702 than to all other evaluated judgment criteria, including *Daubert* and *Frye* (means are presented in Table 3). The length of discussion of Rule 702 was 246.31 words ($SD = 474.21$), and the length of discussion of Rule 702 increased significantly after *Daubert*, $F(1, 688) = 12.28$, $p < .001$. That increase did not occur immediately, but the discussion increased during the 23–44-month period after *Daubert* and remained at this increased length thereafter, $F(3, 686) = 6.25$, $p < .001$. Means for the length of discussion devoted to Rule 702 across these time periods are presented in Table 4. This pattern of mean differences indicates that *Daubert* had some influence on the length of discussion of Rule 702. Further evidence of this influence emerged in a comparison of *Daubert*-adopting jurisdictions and nonadopting jurisdictions after *Daubert*. Significantly more lengthy discussion of Rule 702 occurred in adopting ($M = 333.63$, $SD = 556.53$) than in nonadopting jurisdictions after *Daubert* ($M = 171.80$, $SD = 369.57$), $F(1, 342) = 7.88$, $p = .005$. However, the length of discussion of Rule 702 was equivalent across all types of testimony (means are presented in Table 3), and there was no significant interaction between the timing of the case and the type of testimony in relation to the length of discussion of Rule 702.

Table 4
Mean Number and (Standard Deviation) of Words Devoted to the Discussion of the Federal Rules of Evidence

		Timing of the case			
Judgment criterion	Grand mean	Before *Daubert*	1–22 months after *Daubert*	23–44 months after *Daubert*	45–66 months after *Daubert*
Rule 104	3.54	1.35_b	5.44_b	14.09_a	0.4_b**
	(28.16)	(23.44)	(26.24)	(48.68)	(2.62)
Rule 702	246.31	204.21_a	267.42_a	398.29_b	388.44_b**
	(474.21)	(425.84)	(419.28)	(681.62)	(486.62)
Daubert	97.87	36.71_a	158.97_b	299.11_c	195.38_c**
	(472.81)	(407.57)	(528.00)	(656.42)	(379.42)

Note. Means include cases in which these reasons were not discussed, having a rating of zero. Subscript letters indicate the locations of the mean differences where applicable.
[a]Asterisks indicate a significant mean difference for the timing of the case.
**$p < .01$.

Expert testimony which would be found admissible may be excluded on the basis that it is more prejudicial than probative. Under Rule 403, testimony from an expert can be excluded "if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury" (Mueller & Kirkpatrick, 1999). The *Daubert* Court recognized the potential use of Rule 403 in expert testimony admission decisions (*Daubert*, 1993). Even though it was mentioned in *Daubert*, the length of discussion of Rule 403 did not significantly change after *Daubert* (means are presented in Table 3). However, the discussion of Rule 403 was significantly longer in cases with scientific and technical experts than in cases with either medical/mental health or business experts, $F(3, 667) = 3.85$, $p = .010$. There was no significant interaction between the timing of the case or the type of testimony presented and the length of discussion of Rule 403.

Frye and Daubert

Daubert signaled trial judges to refrain from using *Frye* as an admissibility standard in federal courts and in several state courts. Even though the general acceptance standard was incorporated into *Daubert*, it is possible that judges continued to rely explicitly on *Frye* outside of the context of the four *Daubert* criteria. Analyses were conducted to assess whether courts persisted in applying the *Frye* standard and, if there was a decrease in this application, whether courts supplanted it by reference to the four *Daubert* criteria. Additional analyses examined the length of courts' discussion of *Daubert*. The mean length of discussion, significant changes across time, and significant differences across types of testimony are discussed below for *Frye*, *Daubert*, and the four *Daubert* criteria (means are presented in Table 3).

The length of discussion of *Frye* was 97.62 words ($SD = 507.51$). The length of this discussion significantly decreased after *Daubert* (means are presented in Table 3), $F(1, 690) = 7.82$, $p = .005$. Discussion of *Frye* was significantly more abbreviated in *Daubert*-adopting jurisdictions ($M = 22.50$, $SD = 135.33$) than in nonadopting jurisdictions after *Daubert* ($M = 292.88$, $SD = 994.75$), $F(1, 342) = 16.68$, $p < .001$. The discussion of *Frye* was significantly longer for scientific experts than for all other types of experts (means are presented in Table 3), $F(3, 667) = 6.71$, $p < .001$. There was no significant interaction between the timing of the case and the type of testimony presented as they related to the length of discussion of *Frye*.

The length of discussion of *Daubert*, including the cases in state courts which did not adopt *Daubert*, was 97.87 words ($SD = 472.81$). Because state courts that did not adopt *Daubert* were classified as "before *Daubert*," the mean number of words devoted to *Daubert* before the *Daubert* decision did not equal zero. These cases discussed the *Daubert* decision in general but did not apply *Daubert* to the admission decision. Not surprisingly, there was a significant increase in the length of discussion of *Daubert* after *Daubert* (means are presented in Table 3), $F(1, 690) = 23.30$, $p < .001$. The length of discussion of *Daubert* increased immediately after *Daubert* (in the first 22 months), and it increased again significantly during the 23–44-month period after *Daubert*, remaining steady after that time. Although the discussion of *Daubert* during the 45–66-month period post*Daubert*

was significantly longer than pre-*Daubert* discussions, it was not significantly longer than discussions in the period just before that (means for these time periods are presented in Table 4), $F(3, 688) = 9.75$, $p < .001$. Significantly longer discussions of *Daubert* were observed in cases with scientific experts than in cases with any other type of expert (means are presented in Table 3), $F(3, 667) = 8.25$, $p < .001$. The interaction between the timing of the case and the type of testimony as they related to the length of discussion of *Daubert* was significant, $F(3, 663) = 10.54$, $p < .001$. On this variable, the increase was disproportionately large with respect to scientific testimony (means increased significantly from 69.91 to 603.48 words after *Daubert*) and medical/mental health testimony (mean increased significantly from 8.34 to 194.08 words after *Daubert*) as compared with technical and business testimony.

Given the importance of the *Daubert* decision and the attention paid to it, a dramatic increase in the courts' attention to the four *Daubert* criteria would be expected. Therefore, the most mysterious impact of the *Daubert* decision is the lack of discussion devoted to the four *Daubert* criteria themselves. Although there was a significant increase in the length of discussion of *Daubert* in general, there was not a significant increase over time in the length of discussion devoted to the first three *Daubert* criteria of falsifiability, peer review, and error rate (means are presented in Table 3). However, consistent with the reduction in the use of *Frye* over time, there was a significant decrease in the length of discussion of general acceptance, $F(1, 691) = 9.84$, $p = .002$, even though it is also one of the four *Daubert* criteria. Also consistent with the length of discussion of *Frye*, the length of discussion of general acceptance was significantly shorter in *Daubert*-adopting jurisdictions ($M = 11.77$, $SD = 44.17$) than in nonadopting jurisdictions after *Daubert* ($M = 88.72$, $SD = 314.91$), $F(1, 342) = 13.50$, $p < .001$.

Differences in the length of discussion of the four *Daubert* criteria were also observed across the types of testimony. The length of discussion about the two criteria of falsifiability and peer review was equivalent across types of testimony. However, the discussion of the other two criteria of error rate, $F(3, 668) = 3.82$, $p = .010$, and general acceptance, $F(3, 668) = 11.43$, $p < .001$, was significantly longer for cases involving scientific experts than for those involving all other types of experts (means are presented in Table 3). There was a significant interaction between the timing of the case and the type of expert in relation to the length of discussion of general acceptance, $F(3, 664) = 4.55$, $p = .004$. On this variable, the decrease was disproportionately large with respect to scientific testimony (mean decreased significantly from 257.64 to 33.88 words after *Daubert*) as compared with all other types of testimony.

Influence of the Judgment Criteria on the Admissibility of Expert Testimony in Criminal Cases

Judgment criteria that may impact on the admissibility of expert testimony in criminal cases were further evaluated for their influence on the admission decision. These criteria were drawn from the language in the *Daubert* decision and from the Rules. These variables included *Frye;* the four *Daubert* criteria (falsifiability, peer review, error rate, and general acceptance); other variables derived from the *Daubert* decision (general reliability); and selected requirements regarding expert evidence derived from the Rules (relevance, expert qualifications, assisting the trier of fact, and

prejudicial impact). For each of these criteria, a rating of its influence on the decision was made. This rating was on a 10-point scale, where 0 indicated that the criterion was not mentioned, 1 indicated that it was mentioned but was not at all influential, and 9 indicated that it was mentioned and was very influential.

The criteria with the highest ratings of influence were derived from the Federal Rules of Evidence, namely assisting the trier of fact, expert qualifications, relevance of the testimony, and potential for prejudicial impact (means are presented in Table 5). These and general reliability of the expert evidence were the only criteria for which the mean importance ratings exceeded 1.0. However, there were significant changes over time in the influence of many of the criteria on the admission decision, regardless of how low that influence was rated overall. ANOVAs were conducted on the ratings of influence of the criteria to determine whether there were significant changes over time (before vs. after *Daubert*), significant differences across the four types of testimony, and interactions between the timing and type of testimony. Additional ANOVAs were computed examining differences across the three post*Daubert* time periods and between the *Daubert*-adopting and nonadopting jurisdictions. Only test results that were statistically significant are described.

Several effects on the influence of the *Frye* standard were observed. First, as suggested by the *Daubert* decision, the importance of *Frye* significantly decreased after *Daubert* (means are presented in Table 5), $F(1, 690) = 11.60$, $p = .001$. However, this decrease in importance was only observed during the period 23–44 months after *Daubert* ($M = 0.11$, $SD = 0.56$). In all other time periods, the influence was rated as equivalent: before ($M = 0.67$, $SD = 2.11$); 1–22 months after ($M = 0.30$, $SD = 0.97$); and 45–66 months after ($M = 0.23$, $SD = 0.90$), $F(3, 688) = 3.54$, $p = .014$. These results provide some indication that *Daubert* exerted an impact on courts' use of *Frye*. Additionally, *Frye* was significantly more influential in nonadopting ($M = 1.03$, $SD = 2.48$) than in adopting jurisdictions ($M = 0.19$, $SD = 0.76$) after *Daubert*, $F(1, 342) = 22.21$, $p < .001$. This further indicates that *Daubert* influenced courts' use of *Frye* because *Frye* obviously remained influential in those jurisdictions not adopting *Daubert* but its influence decreased in adopting jurisdictions. Consistent with the discussion of *Frye*, the importance rating of *Frye* was affected by the type of testimony. The *Frye* standard was rated as significantly more important for scientific testimony than for all other types of testimony (means are presented in Table 5), $F(3, 667) = 44.23$, $p < .001$. There was a significant interaction between the timing of the case and the type of testimony in relation to the importance of *Frye*. Notably, the decrease in the importance of the *Frye* standard was significantly larger for scientific experts (mean decreased significantly from 1.86 to 0.54 after *Daubert*) than for all other types of experts, $F(3, 663) = 3.27$, $p = .021$.

With regard to the four *Daubert* criteria, there was a decrease in influence for general acceptance, $F(1, 691) = 6.39$, $p = .012$, and an increase in influence for falsifiability, $F(1, 691) = 7.96$, $p = .005$; peer review, $F(1, 691) = 10.55$, $p = .001$; and error rate, $F(1, 691) = 10.33$, $p = .001$ (means are presented in Table 5). Further evidence of *Daubert's* effects on the judgment criteria is provided by evident differences between adopting and nonadopting jurisdictions. For example, general acceptance was rated as more important in nonadopting jurisdictions after *Daubert* ($M = 1.14$, $SD = 2.48$) than in adopting jurisdictions ($M = 0.66$, $SD = 1.88$), $F(1, 342) = 3.90$, $p = .049$. The opposite effect was observed for peer

Table 5
Means (and Standard Deviations) for Influence of Selected Criteria on the Admission Decision

	Mean rating of influence on the decision							
		Timing of case		Type of testimony				
Reason for decision	Grand mean	Before *Daubert* (n = 458)	After *Daubert* (n = 235)[a]	Med/M.H. (n = 224)	Technical (n = 256)	Scientific (n = 136)	Business (n = 56)[b]	Timing × Type
Assist trier	5.26 (3.43)	5.18 (3.46)	5.42 (3.38)	5.32 (3.34)$_a$	5.6 (3.34)$_a$	4.36 (3.65)$_b$	6.13 (3.10)$_a$**	*ns*
Qualifications	3.41 (3.41)	3.53 (3.44)	3.17 (3.34)	3.49 (3.44)$_c$	3.80 (3.4)$_{a,c}$	2.70 (3.27)$_b$	3.27 (3.49)$_c$*	*ns*
Relevance	3.0 (3.37)	3.19 (3.36)	2.65 (3.37)*	3.23 (3.46)	2.88 (3.27)	2.98 (3.38)	3.05 (3.62)	*ns*
Prejudice	2.96 (3.53)	2.71 (3.46)	3.45 (3.61)*	2.56 (3.46)$_b$	3.61 (3.59)$_a$	2.77 (3.45)$_b$	2.4 (3.39)$_b$**	*ns*
Reliability	1.6 (2.66)	1.64 (2.83)	1.51 (2.90)	1.72 (2.85)$_b$	0.85 (2.13)$_c$	3.10 (3.62)$_a$	0.75 (2.05)$_c$**	*ns*
Gen. accept.	0.97 (2.34)	1.13 (2.54)	0.66 (1.88)**	1.12 (2.51)$_b$	0.36 (1.47)$_c$	2.25 (3.19)$_a$	0.16 (0.71)$_c$**	*ns*
Frye	0.51 (1.79)	0.68 (2.12)	0.19 (0.76)**	0.41 (1.59)$_b$	0.26 (1.23)$_b$	1.37 (2.81)$_a$	0 (0)$_b$**	*
Peer review	0.27 (1.23)	0.16 (0.93)	0.48 (1.63)**	0.20 (1.08)$_b$	0.13 (0.90)$_b$	0.74 (1.94)$_a$	0 (0)$_b$**	**
Error rate	0.16 (0.95)	0.08 (0.67)	0.32 (1.36)**	0.05 (0.52)$_b$	0.06 (0.55)$_b$	0.62 (1.37)$_a$	0 (0)$_b$**	**
Falsifiability	0.11 (0.75)	0.05 (0.48)	0.22 (1.09)**	0.02 (0.33)$_b$	0.04 (0.42)$_b$	0.39 (1.42)$_a$	0 (0)$_b$**	**

Note. Subscripts appearing next to means indicate the locations of the mean differences where applicable. Ratings were made on a 0 to 9-point scale. Means include cases in which these reasons were not discussed, having a rating of zero. Med/M.H. = Medical/mental health; Gen. accept. = general acceptance.
[a]Asterisks indicate a significant mean difference for the timing of the case. [b]Asterisks indicate a significant mean difference for the type of testimony.
*$p < .05$. **$p < .01$.

review, which was rated as less important in nonadopting jurisdictions after *Daubert* ($M = 0.13$, $SD = 0.82$) than in adopting jurisdictions ($M = 0.48$, $SD = 1.63$), $F(1, 342) = 4.43$, $p = .036$. Surprisingly, the influence of general reliability did not change significantly after *Daubert.*

The influence of several Rules requirements also changed significantly over time (means are presented in Table 5). The influence of the criterion of relevance, $F(1, 690) = 4.02$, $p = .045$, decreased after *Daubert,* whereas the influence of the potentially prejudicial impact of the expert evidence, $F(1, 691) = 6.93$, $p = .009$, increased after *Daubert.* Further evidence of *Daubert's* impact on prejudicial impact is reflected in a comparison of influence ratings of this criterion in adopting and nonadopting jurisdictions. Consideration of the potentially prejudicial impact was more influential in adopting ($M = 3.45$, $SD = 3.61$) than in nonadopting jurisdictions after *Daubert* ($M = 2.50$, $SD = 3.50$), $F(1, 342) = 5.16$, $p = .024$. Neither the influence of general reliability, expert qualifications, nor the requirement that the testimony assist the trier of fact changed significantly as a function of time.

Significant differences in the influence of the judgment criteria were related to the type of expert testimony proffered (means are presented in Table 5). Ratings of influence on the criteria of falsifiability, $F(3, 668) = 9.43$, $p < .001$, peer review, $F(3, 668) = 8.99$, $p < .001$, error rate, $F(3, 668) = 12.95$, $p < .001$, general acceptance, $F(3, 668) = 23.89$, $p < .001$, and general reliability, $F(3, 668) = 22.07$, $p < .001$, were significantly higher for scientific testimony than for all other types of expert testimony. Ratings of the importance of the criteria such as prejudicial impact, $F(3, 668) = 4.44$, $p = .004$, and the qualifications of the expert, $F(3, 668) = 3.21$, $p = .023$, were significantly higher for technical testimony than for all other types of expert testimony. Assistance provided to the trier of fact was rated as significantly less influential for the admissibility of scientific testimony than it was regarding all other types of testimony, $F(3, 667) = 5.25$, $p = .001$. The influence of the relevance of the expert evidence did not change significantly as a function of the type of testimony presented.

In addition, there were significant interactions between the timing of the case and the type of testimony offered for the ratings of influence of falsifiability, peer review, and error rate on the admission decision. For each of these variables, the increase in importance was disproportionately large with respect to scientific testimony versus all other types of testimony. The mean importance rating of falsifiability for scientific experts increased significantly from 0.13 to 0.84 after *Daubert, F*(3, 664) = 8.91, $p < .001$. The mean importance rating of peer review for scientific experts increased significantly from 0.40 to 1.32 after *Daubert, F*(3, 664) = 4.17, $p = .006$. The mean importance rating of error rate for scientific experts increased significantly from 0.30 to 1.16 after *Daubert, F*(3, 664) = 6.21, $p < .001$. There were no significant interactions between the timing of the case and the type of testimony on any other judgment criteria.

Expert Knowledge and Qualifications

Although the influence of expert qualifications in general did not change significantly because of *Daubert,* it is possible that *Daubert* affected courts' evaluations of expert qualifications. Ninety-eight percent ($n = 671$) of the experts

derived some of their expertise from experience, and 74.8% ($n = 513$) derived some expertise from case specific experience. Education was the next most frequently cited source of expert knowledge ($n = 424$, 62%). Experts also derived expert knowledge from a body of research ($n = 117$, 17.1%), case specific research ($n = 89$, 13%), and theory ($n = 35$, 5.1%). Many experts gained expertise from more than one or all of these sources.

The importance of these potential sources of expertise was rated as they related to the qualifications of the expert. ANOVAs were conducted on the importance ratings of the knowledge sources to determine if there were significant changes over time (before and after *Daubert*), significant differences across the four types of testimony, and interactions between the timing and type of testimony. Overall, knowledge gained by experience was rated as the most important source of expertise, closely followed by knowledge gained by case specific experience and by education (see Table 6). Knowledge gained by experience was significantly more important for technical testimony than for all other types of testimony, $F(3, 662) = 13.93$, $p < .001$. The relationship was reversed with respect to knowledge gained by education; formal education was significantly less important for technical testimony than for all other types of testimony, $F(3, 661) = 256.46$, $p < .001$. Case specific experience was equally important for all types of expert evidence. Knowledge gained by theory, $F(3, 663) = 6.68$, $p < .001$, a body of research, $F(3, 663) = 41.73$, $p < .001$, or case specific research, $F(3, 664) = 57.20$, $p < .001$, were all rated as relatively unimportant to the expert's qualifications. Although relatively unimportant, each of these sources of knowledge had the highest importance rating in scientific testimony. However, only the rating of importance of knowledge derived from a body of research changed significantly over time. Completely contrary to the thrust of the *Daubert* opinion, this source of knowledge was rated as less important after *Daubert*, $F(1, 684) = 11.42$, $p = .001$. There were no significant interactions between the timing of the opinion and the type of testimony proffered on any of the measures relating to the sources of expertise, except again for expertise derived from a body of research, $F(3, 659) = 5.63$, $p = .001$. On this variable, the decrease in importance was disproportionately large with respect to scientific testimony (mean decreased significantly from 3.06 to 1.19 after *Daubert*) as compared with all other types of expert testimony.

These results are perplexing. *Daubert* seemed to suggest that an expert's methods, not only his or her credentials, should be evaluated. However, the most important ratings on sources of expertise were the experience and education of the proffered expert. Sources of knowledge which indicate methodological reliability were rated as less important, although they were more influential for scientific experts. Perhaps the most perplexing of these results is the disproportionate decrease in the importance of research as a source of expertise for scientific experts. This is exactly the type of knowledge (and the type of expert) at which the *Daubert* criteria are targeted. This finding indicates that judges may be influenced more by credentials rather than by the reliability of the methodology used. Unfortunately, it also indicates that judges may be unable to determine what factors are important in assessing scientific reliability, particularly when research is the basis for an expert opinion.

Table 6
Mean Ratings and (Standard Deviations) of Importance of the Expert's Qualifications and Knowledge Sources

Source of expert knowledge	Grand mean	Timing of case		Type of testimony				Timing × Type
		Before *Daubert* (n = 458)	After *Daubert* (n = 235)[a]	Med/mental health (n = 224)	Technical (n = 256)	Scientific (n = 136)	Business (n = 56)[b]	
Experience	5.99	6.0	5.97	5.74_b	6.47_a	5.58_b	5.95_b**	*ns*
	(1.60)	(1.65)	(1.51)	(1.63)	(1.49)	(1.44)	(1.22)	
Case-specific experience	4.55	4.49	4.66	4.80	4.47	4.42	4.47	*ns*
	(2.77)	(2.84)	(2.64)	(2.66)	(2.78)	(2.93)	(2.63)	
Education	3.30	3.38	3.13	4.68_a	0.70_b	5.07_a	4.80_a**	*ns*
	(2.80)	(2.77)	(2.84)	(1.93)	(1.87)	(1.78)	(2.26)	
Theory	0.27	0.31	0.18	0.47_a	0.05_b	0.51_a	0_b**	*ns*
	(1.28)	(1.40)	(1.02)	(1.68)	(0.56)	(1.71)	(0)	
Body of research	1.03	1.25	0.62**	1.55_b	0.11_c	2.39_a	0_c**	**
	(2.34)	(2.50)	(1.91)	(2.65)	(0.82)	(3.19)	(0)	
Case-specific research	0.83	0.87	0.77	0.36_b	0.36_b	2.87_a	0.22_b**	*ns*
	(2.22)	(2.29)	(2.07)	(1.49)	(1.50)	(3.37)	(1.13)	

Note. Subscripts appearing next to means indicate the locations of the mean differences where applicable. Ratings made on a 0 to 9-point scale. Means include cases in which these reasons were not discussed, having a rating of zero. Med/mental health = Medical/mental health. [a]Asterisks indicate a significant mean difference for the timing of the case. [b]Asterisks indicate a significant mean difference for the type of testimony.
**$p < .01$.

Predicting Admissibility With Courts' Evaluations of the Judgment Criteria

To provide some insight into the manner in which appellate court judges apply the criteria to expert evidence in making admissibility decisions, we coded courts' applications of each of the potential judgment criteria to the proffered testimony. Whether the court specifically stated that a judgment criterion was met or not met by an expert's testimony was recorded, in addition to the rating of influence for each judgment criterion. Courts' statements that these criteria were or were not met were generally consistent with the ratings of the importance of those criteria to the admission decision. For example, the criteria representing the requirements of the Rules were the most frequently cited as having been met or not met (see Table 7). Judges rarely made statements that the four *Daubert* criteria were or were not met.

A critical question to be answered about the effects of *Daubert* is whether or not the judgment criteria are associated with appellate admissibility when they are considered together. Logistic regression was used to assess the utility of the legal criteria used in the courts' judgments, the timing of the case, the type of the testimony, and the interaction between the timing and testimony type.[2] The timing of the case was coded as 1 for "before *Daubert*" and as 2 for "after *Daubert*." The category "before *Daubert*" was used as the reference category. Type of testimony (medical–mental health, technical, scientific, business) was dummy coded to create three dichotomous variables (medical/mental health, technical, and scientific testimony). The category of business testimony was used as the comparison condition because business testimony was admitted at the appellate level at the lowest rate and had the lowest ratings on a majority of the measurements.[3] The legal judgment criteria selected for entry in the models included the requirements specified in the Rules (relevance, qualifications, assisting the trier of fact, prejudicial impact) and the four *Daubert* criteria (general acceptance, peer review/publication, falsifiability, and error rate), and general reliability. All judgment criteria were coded as 0 when not mentioned by the court and 1 when the court specifically stated that the criterion was met. On a separate variable, all judgment criteria were coded as 0 when not mentioned and 1 when the court specifically stated that the criterion was not met. The outcome variable was whether the evidence was admitted or excluded by the appellate court, which was coded as a "1" for "excluded" and as a "2" for "admitted." Therefore, a positive relationship between a predictor and the outcome variable indicates the predictor is related to admissibility, and a negative relationship between a predictor and the outcome

[2]Models were tested including variables expressing the interaction between the timing of the case and the type of testimony (Timing × Type), but these interactions never contributed significantly to the models. Therefore, they were removed from the analysis reported here.

[3]Individual models for each of the types of testimony were also created by first entering the timing of the case and then entering the judgment criteria. Results were generally consistent with the full model. The timing of the case was not a significant predictor of admissibility for any type of expert. None of the *Daubert* criteria were significant predictors in any of the models. Some Federal Rules of Evidence were significant predictors of admissibility for medical (qualifications, relevance, assisting the trier of fact) and technical testimony (relevance, assisting the trier of fact, and prejudicial impact), but not for scientific testimony. No model for business testimony could be tested because courts did not use several of the judgment criteria when determining the admissibility of business experts.

Table 7
Frequencies of Courts Stating Selected Criteria Were Specifically Met or Not Met by the Testimony

	Criterion was met		Criterion was not met		Not specified in opinion[a]	
	No. of times cited (n = 693)		No. of times cited (n = 693)		No. of times not cited	
Criterion	Before *Daubert*	After *Daubert*	Before *Daubert*	After *Daubert*	Met	Not met
Assist trier	200 (43.7%)	114 (48.5%)	97 (21.2%)	51 (21.7%)	379	544
Qualifications	199 (43.4%)	96 (40.9%)	43 (9.4%)	11 (4.7%)	398	638
Relevance	123 (26.9%)	54 (23.0%)	58 (12.7%)	23 (9.8%)	516	612
Prejudice	102 (22.3%)	69 (29.4%)*	54 (11.8%)	34 (14.5%)	521	604
Reliability	61 (13.3%)	24 (10.2%)	27 (5.9%)	16 (6.8%)	608	649
General acceptance	52 (11.4%)	13 (5.5%)*	18 (3.9%)	6 (2.6%)	628	668
Peer review	5 (1.1%)	8 (3.4%)	2 (0.4%)	8 (3.4%)**	680	682
Error rate	4 (0.9%)	2 (0.9%)	1 (0.2%)	3 (1.3%)	687	688
Falsifiability	3 (0.7%)	1 (0.4%)	0 (0%)	3 (1.3%)*	689	689

[a]Not specified in the opinion means that the court did not specifically state that the particular criterion was met or not met by the testimony.
*$p < .05$. **$p < .01$.

variable suggests that the predictor is related to exclusion. To create the full model, variables were entered stepwise.

First, we investigated whether the timing of the case or the type of expert testimony presented was associated with appellate decisions to admit or exclude the evidence. A logistic regression analysis was performed on admissibility as outcome and the timing of the case (before and after *Daubert*) and the type of testimony (dummy coded with "business testimony" as the comparison condition) as predictors to determine if these factors are related to admissibility. A test of this model indicates that these predictors, as a set, reliably distinguish between admissible and inadmissible testimony, $\chi^2(4, N = 650) = 56.13$, $p < .001$, with an overall correct classification rate of 69.08%. In this model, technical expert testimony was the only reliable predictor of the admission decision, and technical expert testimony was positively related to admissibility ($R = .16$, Wald = 22.30, $p < .001$). That is, technical testimony was admitted at a significantly higher rate than business testimony. This result is consistent with the high rate of admission observed for technical testimony. Neither the timing of the case ($R = .00$, Wald = 0.21, $p = .644$), medical/mental health testimony ($R = .00$, Wald = 0.34, $p = .560$), nor scientific testimony significantly predicted admissibility.

Although it is interesting that the timing of the case relative to *Daubert* did not reliably distinguish between admissible and inadmissible testimony and that only technical testimony predicted admissibility, these legal decisions are not made in a lawless vacuum. Judges are bound by certain decision rules when presented with expert testimony. Therefore, the influence of the type of testimony and the timing of the case on admissibility may be affected by other factors. When considered together, we would like to determine if these factors are associated with admissibility, how judicial use of the judgment criteria changes over time, how the type of testimony proffered influences the criteria used to evaluate it, and whether the judgment criteria predict admissibility. A multivariate path model using logistic regression was developed to assess (a) the predictive utility of the legal criteria used in the courts' judgments, (b) the timing of the case, and (c) the type of the testimony on the decision to admit or exclude the expert evidence (see Figure 1). Admissibility was used as the outcome variable. The timing of the case and the dummy coded testimony type variables were entered into the model first. Each of the judgment criteria was regressed onto the timing of the case and the type of testimony variables to determine their impact on the use of the judgment criteria. Then, the judgment criteria were entered into the model. A test of the full model indicates that the combination of all the predictors reliably distinguishes between admissible and inadmissible testimony, $\chi^2(22, N = 650) = 509.78$, $p < .001$, with an overall correct reclassification rate of 92.46%.

The full model addresses several issues. First, it examines whether different types of expert testimony are evaluated differently using the same judgment criteria and whether those criteria are used differently after *Daubert*. Scientific and technical expert testimony appear to be treated differently by appellate courts under the judgment criteria. Technical testimony is more likely to be viewed as helpful to the trier of fact ($R = .05$, Wald = 4.13, $p = .042$), less likely not to assist the trier of fact ($R = -.14$, Wald = 15.25, $p < .001$), and is less likely to be found irrelevant ($R = -.09$, Wald = 6.44, $p = .011$). These findings imply that courts find technical testimony, primarily that of police officers, to be relevant and

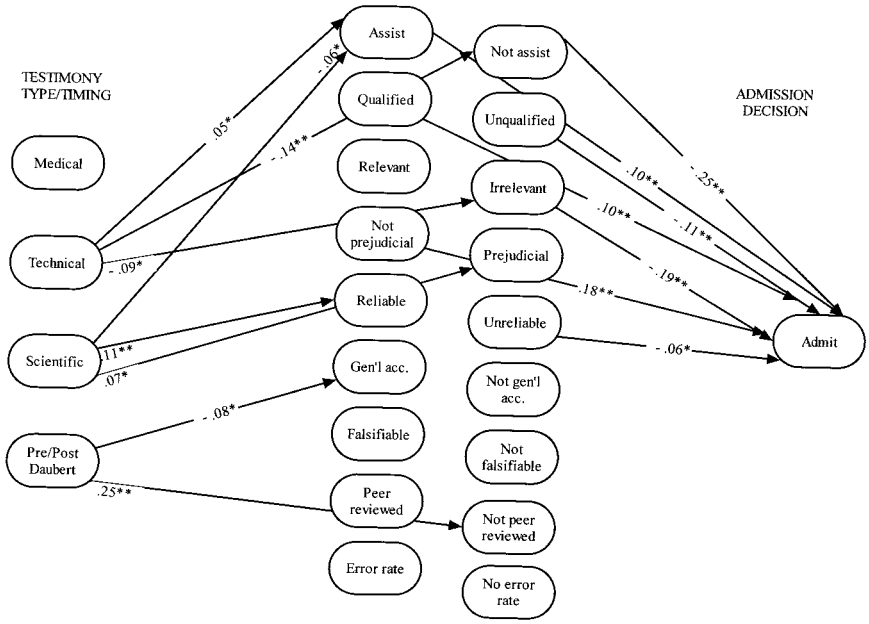

Figure 1. Multivariate path model using logistic regression to assess the influences of the type of testimony and the timing of the case on admissibility. Numbers are standard coefficients. Paths not shown were nonsignificant. Gen'l acc. = general acceptance. $*p < .05$. $**p < .01$.

helpful in criminal cases. However, courts were less likely to find that scientific testimony can assist the trier of fact ($R = -.06$, Wald = 5.68, $p = .017$), were more likely to regard scientific testimony as potentially prejudicial ($R = .07$, Wald = 4.59, $p = .032$), and were more likely to conclude that scientific evidence was reliable ($R = .11$, Wald = 8.43, $p = .004$). Overall, these findings imply that although courts regard scientific expert testimony as reliable, it is also viewed as unhelpful and potentially prejudicial to jurors. These findings confirm that appellate courts follow the *Daubert* Court's mandates to scrutinize the reliability of scientific expert testimony. However, they do so without overtly applying the four *Daubert* criteria, as none of the four *Daubert* criteria were related to the type of expert testimony proffered. For cases decided after *Daubert*, courts were less likely to conclude that the expert testimony was generally accepted ($R = -.08$, Wald = 4.47, $p = .034$) and were more likely to conclude that the content of the testimony was not peer reviewed ($R = .25$, Wald = 8.43, $p = .004$). These findings are consistent with a decrease in the amount of discussion devoted to the criterion of general acceptance after *Daubert*. There were no other significant relationships between judgment criteria and the timing of the case or the type of testimony.

Second, the model addressed whether admission of expert evidence can be predicted by the use of certain judgment criteria, by the type of expert testimony, or by the timing of the case relative to *Daubert*. Overall, several of the requirements of the Rules were reliably related to the admission or exclusion of expert

evidence in criminal cases. When testimony is found to assist the trier of fact ($R = .10$, Wald = 9.82, $p = .002$), is not regarded as prejudicial ($R = .18$, Wald = 26.92, $p < .001$), and/or when the expert is deemed qualified ($R = .10$, Wald = 10.73, $p = .001$), it is likely that the expert testimony will be admitted. Conversely, expert testimony is likely to be excluded if the expert is deemed unqualified ($R = -.11$, Wald = 12.25, $p = .001$), when the testimony is not regarded as helpful to the trier of fact ($R = -.25$, Wald = 52.99, $p < .001$), when it is deemed irrelevant ($R = -.19$, Wald = 32.25, $p < .001$), and/or when it is regarded as unreliable ($R = -.06$, Wald = 4.85, $p = .028$). None of the other judgment criteria, including the four *Daubert* criteria, were significantly related to decisions to admit or exclude expert testimony in criminal cases at the appellate level.

Although neither testimony type nor timing of the case was significantly associated with admissibility when evaluated in the full model, whether the expert testimony was technical or scientific in nature did exert some indirect influence on the admission decision. Technical testimony was indirectly related to admissibility through two mediating variables, namely the extent to which the evidence was seen as helpful and the determination that the evidence was irrelevant. Scientific testimony was indirectly related to the admission decision through one mediating variable, namely that the testimony is not helpful to the trier of fact. However, the timing of the case was unrelated to the admission decision, either directly or indirectly—these patterns did not differ pre- versus post-*Daubert*.

Results indicating that testimony is excluded when it fails to conform to the standards elucidated in the Rules and is admitted when it conforms to them are hardly shocking. The more interesting finding is the striking absence of any significant relationships between the four *Daubert* criteria and decisions to admit or exclude expert evidence. Although courts may discuss these criteria and may indicate that the criteria are influential in the admission decision, they clearly are not driving admission decisions in criminal cases. The only tested variables strongly associated with the admission decision are the requirements of the Rules. It appears that judges rely faithfully on the Rules in determining whether to admit or exclude expert evidence.

Discussion and Conclusions

The *Daubert* decision and subsequent Supreme Court decisions on expert testimony created a flurry of scholarship predicting the effects of those landmark decisions. The purpose of this study was to assess the effects of these decisions on criminal appellate courts' treatment of expert evidence and on the admissibility of expert testimony. Given the importance of the *Daubert* decision, we expected to find increased discussion of the four *Daubert* criteria, increased importance attributed to the criteria, and some changes in admissibility as a result of *Daubert* following the *Daubert* decision. Unexpectedly, no change in the overall rate of admission for all types of expert evidence was observed. However, this finding is only the prelude to the story. *Daubert* called for increased scrutiny of expert evidence, and appellate courts appear to have answered that call. Many changes occurred in the way criminal appellate court judges evaluated evidence after *Daubert* and across different types of testimony.

Summary of Results

Overall, we observed some general trends and some effects of the *Daubert* decision in criminal appellate cases: Defendants generated the vast majority of criminal appeals. The experts who were the focus of a criminal appeal testified for the prosecution at trial. The majority of the experts testifying in these cases were experts on police procedures, and the vast majority of this expert testimony was admitted. It is hardly surprising that the majority of appeals in criminal trials are lodged by defendants who appeal the admission of a prosecution expert's evidence, given that the prosecution has a limited right to appeal a criminal conviction. It is also not surprising that police officers, who are integral to criminal prosecutions, frequently served as the prosecution experts.

A surprising result was the observed rates of admission of expert testimony in criminal cases before and after *Daubert*. One of the most pressing questions raised by the *Daubert* decision was its potential effect on the admissibility of expert evidence. However, the *Daubert* decision did not impact on the admission rates of expert testimony at either the trial or the appellate court levels. Some potential explanations for the rates of admission that we explored were the party offering the testimony, the type of counsel retained by the defendant, and the standard of review used by the appellate court. However, none of these factors provided a convincing explanation for this lack of a significant change in admissibility. Another surprising finding was that the standard of appellate review issued in *Joiner* did not change rates of admission of expert evidence. Although the appellate standard of review was related to appellate admissibility, cases using each of these standards were evenly spread across time. Plain error was associated with appellate admission because most cases were appealing the admission of a prosecution expert. Plain error is the applicable standard when the defendant failed to preserve admissibility as an issue for appeal, leaving the appellate court little choice other than to affirm the trial court's admission of the evidence. The abuse of discretion standard provides appellate courts with an opportunity for gatekeeping, which explains its association with exclusion of expert evidence.

Even though the expected change in expert testimony admissibility was not observed, other effects of the *Daubert* decision were observed. First, we observed several differences in the amount of discussion devoted to various topics relevant to expert testimony. Although this measure does not provide information about the content of that discussion, it does provide a reasonably objective measure of the level of attention paid to both the Federal Rules and to the four *Daubert* criteria. Overall, the various Federal Rules were accorded significant consideration by the courts in their opinions on admissibility. Most attention was devoted to Rule 702 (namely requiring assistance to the trier of fact and a qualified expert), which generated the lengthiest discussion. The *Daubert* Court requested increased scrutiny of expert evidence under Rule 702, and appellate courts complied with this request, as more discussion was devoted to this Rule after *Daubert*. However, the increase in the discussion of Rule 702 did not occur right away and appears to have reached a plateau after 44 months post-*Daubert*. A partial answer to whether judges are gatekeeping is provided by the results of measures of the number of words devoted to the discussion of Rule 104 (permitting pretrial determinations), which followed a similar pattern to that of Rule 702. The

discussion of Rule 104 was more extensive after *Daubert*, did not occur immediately, and also leveled after 44 months post-*Daubert*. Courts may be attempting to dispose of expert testimony issues before trial under Rule 104, or there could be changes in the practices regarding the application of this Rule as a result of *Daubert*.

Changes over time were also observed in the amount of discussion devoted to *Frye*, *Daubert*, and the four *Daubert* criteria. The length of discussion devoted to general acceptance and to *Frye* decreased, as expected. These decreases were even more pronounced in jurisdictions adopting *Daubert*. Also, as expected, general discussion of *Daubert* increased. In addition, increased discussion of *Daubert* occurred immediately, but may have reached its maximum level. After increasing immediately, the increase appears to have reached a plateau after 22 months post*Daubert*. These increases over time may reflect courts' adjustment to the new standards. The *Daubert* decision in general was an obvious choice for discussion, so the attention paid to it by appellate courts increased immediately. As courts began applying *Daubert*, the Federal Rules of Evidence pertaining to expert evidence became more prominent, and more discussion of the Rules was generated as time passed. However, there was no comparable increase in the discussion of the four *Daubert* criteria in evaluating expert testimony. Discussion of *Daubert* was lengthy, but the discussion devoted to the three new criteria was relatively abbreviated. These findings suggest that judges understand the importance of the *Daubert* decision, but they pay only passing attention to the suggested criteria. Given the extremely low mean number of words discussing the four criteria, it is clear that judges briefly quote the *Daubert* opinion and infrequently devote substantive discussion to the criteria.

As for the content of the discussion in appellate court decisions on expert evidence, several effects were observed based on the importance ratings of the various topics, confirming the courts' reliance on the Rules over the four *Daubert* criteria. Overall, the Federal Rules requirements that the expert evidence assists the trier of fact, that the expert is qualified, that the evidence is relevant, and that the evidence is nonprejudicial were rated as the most important to determining admissibility. The four *Daubert* criteria were rated least important. Of the selected Federal Rules requirements, the importance of relevance decreased over time, while the potentially prejudicial impact of the testimony increased in importance, particularly in jurisdictions that adopted *Daubert*. The importance of both *Frye* and the general acceptance standard decreased overall, and both were less important in jurisdictions adopting *Daubert*. In addition, the importance of the remaining three *Daubert* criteria increased, namely falsifiability, peer review, and error rate. Judges' efforts to follow the *Daubert* guidelines may account for changes over time in the appellate courts' use of Rule 104, Rule 702, *Frye*, general acceptance, and *Daubert*.

Another finding that emerged in the content analysis of the courts' discussion of admissibility issues was that courts treated scientific expert evidence differently from medical–mental health, technical, or business evidence. On several of the measures, the discussion was longer and the importance of these measures was greater when the expert testimony was scientific in nature. For example, discussions of Rules 104 and 403 were more extensive for scientific experts. Discussions of *Frye* and general acceptance were both longer and more influential to the

admissibility decision when the expert was scientific. Discussion of *Daubert* in general was also longer for scientific experts. Additionally, the error rate, falsifiability, and peer review were all more important for scientific experts. However, assisting the trier of fact was rated as the least important for determining the admissibility of scientific experts, while expert qualifications and the prejudicial impact of the evidence were most important for technical experts. The length and content of the discussion suggest that courts scrutinize scientific evidence more closely after *Daubert* or that the distinctive treatment of scientific evidence is attributable to the complex nature of that type of testimony. Courts may be more motivated to provide a more detailed explanation of the rationale for their reliability assessments and their admission decisions in cases involving scientific evidence.

Significant interactions between the timing of the case and the type of testimony offered further highlighted the differential treatment of scientific experts. Decreases in the length of the discussion devoted to general acceptance and in the importance of *Frye* were both disproportionately large for scientific testimony. The increases in the discussion of *Daubert* in general and in the importance ratings of criteria such as falsifiability, error rate, and peer review, were all disproportionately large for scientific testimony. This indicates not only that scientific testimony is treated differently by courts in general, but also that they may view the four *Daubert* criteria and the mandates of the *Daubert* opinion as applicable primarily to scientific evidence.

Finally, although the importance ratings of the various criteria provides information about how influential they are in appellate decision making, it does not provide information about the positive or negative nature of the courts' judgments about the evidence using these criteria nor how these evaluative judgments relate to admissibility. Investigating courts' evaluative judgments provides information about whether evidence is excluded because it fails to meet the four *Daubert* criteria. None of the four *Daubert* criteria reliably distinguished between admissible and inadmissible evidence when placed in a model with other possible judgment criteria. However, the requirements of the Rules reliably distinguished between admitted and excluded evidence in the model. Specifically, when relevance, assistance to the trier of fact, expert qualifications, and lack of prejudicial impact were judged to have been met, expert testimony was admitted. When these requirements were not met, evidence was excluded. Therefore, contrary to the *Daubert* Court's suggestion that the four criteria be used to evaluate expert testimony, judges are not applying the *Daubert* criteria to determine admissibility and are instead applying the Federal Rules of Evidence.

Implications of the impact of Daubert. Overall, it appears courts are attempting to follow the mandates of the *Daubert* decision. First, *Daubert* stated that *Frye*, as a standard for admissibility, was obsolete, and criminal courts appear to have followed this suggestion with respect to *Frye* and general acceptance. Attention paid to *Frye* and the general acceptance standard significantly decreased after *Daubert*, as demonstrated by the decrease in the extent to which these topics were discussed and in the importance ratings assigned to these topics. These decreases were more pronounced in jurisdictions adopting *Daubert*. Because *Daubert* permitted the continued use of general acceptance, this result is somewhat unexpected. Possibly, judges continue to use general acceptance to evaluate evidentiary reliability because they believe it is an important indicator of reliabil-

ity (Gatowski et al., 2001) or because it was one of the four *Daubert* criteria. However, the decrease in discussion devoted to general acceptance and in its importance after *Daubert* may be because judges are less willing to discuss it in an opinion as a reason for admission due to its association with *Frye*.

Second, the Court said reliability of expert evidence should be assessed and suggested the four *Daubert* criteria as ways to accomplish this task. Although general discussion of *Daubert* occurred following this guidance, this discussion often did not include mention of falsifiability, peer review, or error rate. The four *Daubert* criteria themselves were rarely given more than a cursory mention in criminal appellate court opinions, as compared with general discussion of *Daubert*. These four criteria were rated as far less important than the Federal Rules requirements, and they were not related to admissibility. Scientific testimony was treated differently from medical/mental health, technical, and business testimony when criminal courts do apply *Daubert*. When evaluating a scientific expert, more attention was paid to *Daubert*, and the four *Daubert* criteria were rated as having more influence on the admission decision. Courts are more likely to apply the *Daubert* standards explicitly in cases involving scientific experts than in cases involving other fields of expert testimony.

Even though there are some effects observed with *Daubert* and scientific experts, we have observed that appellate courts devoted little discussion to the four *Daubert* criteria, that the *Daubert* criteria are not as influential in their admission decisions, and that the *Daubert* criteria do not predict appellate admissibility. One explanation for these results is that judges simply lack understanding of these criteria and of scientific reliability in general to apply them to their admission decision making. This explanation is consistent with results of a recent survey probing trial judges' understanding of the four *Daubert* criteria (Gatowski et al., 2001). The survey indicated that judges felt it was appropriate for them to act as evidentiary gatekeepers and that the four *Daubert* criteria created a useful decision-making framework for this role. Judges reported that they did understand the four *Daubert* criteria. However, the judges' actual understanding of some of the four *Daubert* criteria was very limited. Although the judges demonstrated an understanding of general acceptance and peer review, the vast majority of judges did not understand the meaning or correct application of error rate and falsifiability. This level of understanding did not vary significantly between *Daubert*-adopting and nonadopting jurisdictions (Gatowski et al., 2001).

Even though it was suggested that judges use the criteria in their decision making, it is clear that they are not using them as extensively as anticipated. The survey on judges provides some insight into why this might be. Poor comprehension of the four *Daubert* criteria and their application may account for the lack of appellate court attention to the four *Daubert* criteria. It is possible that judges would like to apply these criteria in their admission decision making, but they may lack the skill to do so. Therefore, they mention the criteria briefly, acknowledging that the criteria are useful and important, but do not apply the criteria to determine admissibility because they do not understand them fully or do not feel confident in basing their decisions solely on these poorly understood criteria.

There are several possible explanations for the finding that appellate courts apply the *Daubert* criteria most stringently in cases involving scientific expert evidence. Before *Kumho*, it was unclear whether the reliability analysis of expert evidence set forth in *Daubert* applied to nonscientific expert testimony. Although

all the cases in this study were decided before *Kumho*, the trends observed indicate that scientific testimony was in fact treated differently from nonscientific testimony. It is possible that some judges took the valid pre-*Kumho* stance that the four *Daubert* criteria only apply to scientific experts and, therefore, only applied them in those cases. It is also possible that some judges thought that it was inappropriate or impossible to apply the four *Daubert* criteria to the admissibility decisions for all types of experts. Judges may have determined that the four *Daubert* criteria were only applicable to scientific experts. Although it would be consistent with *Kumho* for courts to continue to focus on the four *Daubert* criteria only when appropriate, perhaps only for scientific experts, it would be inconsistent with *Kumho* for courts to ignore reliability in general for nonscientific experts. Because the *Daubert* criteria were so infrequently applied before *Kumho*, it is possible that the four *Daubert* criteria will not even be applied to scientific experts after *Kumho*.

Third, the *Daubert* Court signaled a new focus on the Rules instead of on *Frye* to screen expert evidence. *Daubert* suggested that courts focus on the Rules as part of the gatekeeping duty. Courts appear to be following this suggestion, as several significant changes occurred in criminal appellate courts' use and application of the Rules. These results demonstrate that Rule 702 is the most important of the Federal Rules in determining admissibility, and it is discussed at greatest length, even more so after *Daubert*. However, several other Rules were important to admission decisions. Courts have paid increasing attention to Rule 104, permitting the pretrial assessment of evidence. The Rules requiring scrutiny of assisting the trier of fact, qualifications, relevance, and prejudicial impact are the most important factors in the admission decision. In fact, they are the only judgment criteria that reliably distinguish between admissible and inadmissible testimony, with Rule 702's requirements leading the charge.

Although at first glance it would appear that judges were not following *Daubert*, it is possible that they were trying to follow *Daubert*, not by their use of the four *Daubert* criteria, but by their application of the Rules. First, it is highly likely that judges have a thorough understanding of the meaning and the application of the Federal Rules to expert testimony admissibility, and they should be quite capable of gatekeeping with these Rules. Second, the *Daubert* opinion itself was a call for courts to renew focus on the Rules, instead of relying on *Frye*, and courts may have attempted to assess reliability through a more thorough application of Rule 702. This would be consistent with both the *Daubert* opinion and with the later amendment to Rule 702 that specifically requires courts to consider reliability as part of admissibility. Instead of assessing reliability by applying the four *Daubert* reliability criteria, courts may have been attempting to assess reliability with a more rigorous application of the Rules requirements.

If it is important in the proper execution of a reliability analysis of expert evidence that lower courts use the four *Daubert* criteria as indicia of reliability, then the intent of the Court has only been partially realized. If we assume that judges lack the understanding to accomplish this task successfully, then some additional solutions are indicated. The most obvious of these potential solutions is to continue to provide training for judges in scientific methods. However, if the Court placed an equal value on evaluation of reliability by the use of the four scientific criteria and on the evaluation of reliability by the Rules requirements, then appellate courts are complying with the intent of *Daubert*. Lower courts are

trying to determine the reliability of expert evidence mainly by a rigorous application of the Rules. Additionally, reliance on the Federal Rules to assess the reliability may account for the lack of change in the rates of admission for expert evidence as a result of *Daubert*. The Rules requirements existed before *Daubert*, and judges were using them to evaluate the admissibility of expert evidence long before the *Daubert* decision was issued, reemphasizing these standards. Therefore, we would expect more attention being paid to the Rules requirements by the courts, which we do observe, but the ultimate decision to admit or exclude the testimony should be the same as before *Daubert* because the requirements on which the decision were based were the same as before *Daubert*, which we also observe. In the future, we might observe some change in admission rates as a result of Rule 702 being amended to include a reliability requirement, which is beyond the scope of this research.

The trends observed in this research have implications for both lawyers and experts. After *Daubert*, many lawyers and experts assumed they needed to present evidence about falsifiability, peer review, error rate, and general acceptance in order for the expert's testimony to be admitted. The findings in this study indicate that this course of action is advisable when the expert's testimony is scientific in nature. For most other experts, the presentation of information about the four *Daubert* criteria is probably unnecessary. The results of this research indicate that the most important information an expert must be prepared to present is that his or her testimony assists the trier of fact, is relevant, and is not prejudicial. In addition, experts should be more prepared to discuss their experience and education, rather than their own research, or a body of research, in order to be considered qualified.

Limitations of the foregoing research. As is true with all research and research methods, there are some limitations to this study, and we wish to highlight two. First, appellate opinions represent a subset of all cases brought to trial in which experts testify. Appellate courts do not review every trial court decision, only those cases that are appealed. Therefore, there is a selection bias in our database. The decision to appeal can be based on a variety of reasons unrelated to the quality of the expert testimony at issue, such as lack of funds to cover the cost of appeal or a perceived low probability of success on appeal. By examining only appellate court cases, our research excluded all of the decisions about expert testimony which were not appealed. These could include experts for whom admissibility is well settled and therefore not subject to appeal, and cases in which the decision to forgo any appeal bore no relationship to the admissibility of expert evidence.

The nature of appellate decision making must also be considered. Specifically, appellate courts apply a deferential standard of review in examining all of the trial court decisions, cases may be resolved at a higher level of abstraction in the appellate court than in the trial court, and issues raised may represent only what was contentious between the parties at the trial court. *Daubert* itself was directed toward judges at the trial court level. At the trial court level, it is possible that the four *Daubert* criteria were and are discussed extensively. That discussion, if it did occur, may not be reflected in the appellate court opinion. However, in the majority of these cases, an abuse of discretion standard was applied. The appellate court necessarily must discuss the actions taken by the trial court, including its application of *Daubert*. Therefore, it is likely that the degree of the use and

application of the Rules and the four *Daubert* criteria at the appellate level is reflective of the extent to which those issues arose at the trial court level as well, at least for the cases that are appealed.

Use of appellate court cases as opposed to trial court cases most directly impacts the interpretation of the length of discussion measures used in this study. Differences observed over time may simply reflect issues that were the most contentious between the parties at the trial court level. If so, those issues might require more discussion or assume more prominence in the appellate opinion even if they would not be influential in other expert appeals. An additional confound on this measure includes the writing style of the individual judges, which could be quite brief or verbose. Results on the length of discussion measures should not be interpreted without regard to the content of the discussion, as reflected in the importance ratings and statements about criteria being met or not. However, results on the content ratings are generally consistent with the results on the length of discussion measures, indicating that these measures are at least in part reflective of the attention paid by appellate court judges to the various judgment criteria.

Notwithstanding these limitations, a study of appellate court decisions yields three notable benefits. First, appellate opinions are published to a much larger extent than are trial court opinions, providing us and the legal community with a more convenient sample. Second, appellate court opinions discuss the rulings of the trial court regarding expert testimony, providing information about both courts' approaches to the topic. Third, trial courts should follow the guidance of appellate courts, so appellate court opinions may offer better information about admissibility "trends" and problems in various district and circuit court jurisdictions.

A second overall limitation of these results is that they are only informative about criminal cases. Among the differences between criminal and civil cases are the types of claims made, the type of experts proffered by the parties, and the level of risk incurred by the parties. *Daubert* concerned the admissibility of expert testimony in a civil case. Therefore, it is possible that appellate courts evaluate expert testimony in civil cases in a very different manner from that observed in this study.

Future Directions and Conclusions

At the appellate court level, we were unable to detect any major changes in the admission of expert testimony as a result of the *Daubert* opinion. Nonetheless, judges are gatekeeping in their own way. This gatekeeping is not necessarily accomplished by applying the suggested four *Daubert* criteria, but is instead accomplished by increased and differential application of the Rules to different types of testimony. This result has implications regarding the usefulness of the four *Daubert* criteria. If judges lack a true understanding of the criteria, they will likely be applied only infrequently, as observed. The results also have important implications for reliability assessments in a post-*Kumho* world. *Kumho* required a reliability analysis for all types of expert evidence. Before *Kumho*, courts were clearly treating scientific evidence differently from other types of evidence. This pattern may change after *Kumho*, as more focus shifts to the reliability of nonscientific expert evidence. *Kumho*'s recommendation that courts do not necessarily have to use the four *Daubert* criteria if they do not pertain to the type of

expert evidence in issue may encourage judges to use these criteria even less than they already do. If this is the case, *Kumho* could signal the further demise of the four *Daubert* criteria in judicial decision making about expert testimony.

Because of the implications for admissibility after *Kumho*, future research should focus on the effects of *Kumho* on the admissibility of expert testimony. In addition to an investigation of appellate court decision making, future examinations of trial court decisions would provide direct information about trial court decision making. This would be informative for the effects of both *Daubert* and *Kumho*, and results could be compared with those obtained from appellate court opinions. Another area of future research suggested by the results of this study is on the effects of the amendment to Rule 702. Judges were already relying heavily on the requirements of Rule 702, and it will be important to investigate if they also rely heavily on the amendment adding a reliability requirement. These lines of research would provide a more complete picture of expert testimony admissibility.

References

Allen, R. J. (1994). Expertise and the *Daubert* decision. *Journal of Criminal Law, 84,* 1157–1175.

Black, B., Ayala, F. J., & Saffran-Brinks, C. (1994). Science and the law in the wake of *Daubert:* A new search for scientific knowledge. *Texas Law Review, 72,* 715–802.

Cecil, J. S., & Willging, T. E. (1994). Accepting *Daubert*'s invitation: Defining a role for court-appointed experts in assessing scientific validity. *Emory Law Journal, 43,* 995–1070.

Clark, M. W. (1996). The impact of *Daubert* on the admissibility of expert testimony. *Advocate, 39,* 10–15.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 113 S.Ct. 2795 (1993).

Faigman, D. L. (1995). Mapping the labyrinth of scientific evidence. *Hastings Law Journal, 46,* 555–579.

Faigman, D. L. (1997). Appellate review of scientific evidence under *Daubert* and *Joiner. Hastings Law Journal, 48,* 969–979.

Faigman, D. L., Kaye, D. H., Saks, M. J., & Sanders, J. (1997). *Modern scientific evidence: The law and science of expert testimony* (Vol. 1, Sect. 1-3.4, 29–37, 1999 pocket part). St. Paul, MN: West Publishing Company.

Faigman, D. L., Porter, E., & Saks, M. J. (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, 15,* 1799–1834.

Friedman, R. D. (1994). The death and transfiguration of *Frye. Jurimetrics Journal, 34,* 133–148.

Frye v. United States, 293 F. 1013 (D. C. Cir. 1923).

Gatowski, S. I., Dobbin, S. A., Richardson, J. T., Ginsburg, G. P., Merlino, M. L., & Dahir, V. (2001). Asking the gatekeepers: A national survey of judges on judging expert evidence in a post-*Daubert* world. *Law and Human Behavior,* 25, 433–458.

General Electric Co. v. Joiner, 522 U.S. 136, 118 S. Ct. 512 (1997).

Gottesman, M. H. (1994). Should federal evidence rules trump state tort policy? The federal values of *Daubert* ignored. *Cardozo Law Review, 15,* 1837–1873.

Harvard Law Review (Note) (1995). Confronting the new challenges of scientific evidence, judicial responsibilities in assessing expert testimony. *Harvard Law Review, 108,* 1509.

Kobayashi, J. M. (1995). "Scientific" expert opinion testimony: Qualification and admissibility standards upon and after *Daubert v. Merrell Dow Pharmaceuticals, Inc. American Law Review, 32,* 27–82.

Kumho Tire Co., Ltd. v. Carmichael, 119 S. Ct. 1167 (1999).
Mack, T. J. (1994). Scientific testimony after *Daubert:* Some early returns from lower courts. *Trial, 23.*
Mueller, C. B., & Kirkpatrick, L. C. (1999). *Federal Rules of Evidence, with Advisory Committee Notes, Legislative History, and Cases* (1999 Ed.). Gaithersburg, New York: Aspen Law & Business.
Rovella, D. E. (April 12, 1999). *'Kumho'* could affect criminal cases. *The National Law Journal, 21,* A5, Col. 1.
Saks, M. (1998). Merlin and Solomon: Lessons from the law's formative encounters with forensic identification science. *Hastings Law Journal, 49,* 1069–1141.
Slobogin, C. (1998). Psychiatric evidence in criminal trials: To junk or not to junk? *William & Mary Law Review, 40,* 1–56.

New Editors Appointed, 2004–2009

The Publications and Communications Board of the American Psychological Association announces the appointment of five new editors for 6-year terms beginning in 2004.

Except where noted, as of January 1, 2003, manuscripts should be directed to the following individuals:

- For *Psychology and Aging* (http://www.apa.org/journals/pag.html), submit manuscripts to **Rose T. Zacks, PhD,** Department of Psychology, Michigan State University, East Lansing, MI 48824-1117.
- For *Psychological Assessment* (http://www.apa.org/journals/pas.html), submit manuscripts to **Milton E. Strauss, PhD,** Department of Psychology, Case Western Reserve University, Cleveland, OH 44106-7123.
- For *Journal of Family Psychology* (http://www.apa.org/journals/fam.html), submit manuscripts to **Anne Kazak, PhD,** ABPP, Oncology Psychosocial Research, The Children's Hospital of Philadelphia, Room 1486 (Market Street), 34th and Civic Center Boulevard, Philadelphia, PA 19104. For overnight couriers: Room 1486, 3535 Market Street, Philadelphia, PA 19104.
- For *Journal of Experimental Psychology: Animal Behavior Practices* (http://www.apa.org/journals/xan.html), submit manuscripts as of January 15 to **Nicholas Mackintosh,** Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, United Kingdom.
- For *Journal of Personality and Social Psychology: Personality Processes and Individual Differences* section (http://www.apa.org/journals/psp.html), submit manuscripts to **Charles S. Carver, PhD,** Department of Psychology, University of Miami, P.O. Box 248185, Coral Gables, FL 33124-2070.

Electronic submission: As of January 1, 2003, authors will be expected to submit manuscripts electronically through the journal's Manuscript Submission Portal (see the Web site listed above with each journal title). Authors who are unable to do so should correspond with the editor's office about alternatives.

Manuscript submission patterns make the precise date of completion of the 2003 volumes uncertain. Current editors Leah L. Light, PhD, Stephen N. Haynes, PhD, Ross D. Parke, PhD, Mark E. Bouton, PhD, and Ed Diener, PhD, respectively, will receive and consider manuscripts through December 31, 2002. Should 2003 volumes be completed before that date, manuscripts will be redirected to the new editors for consideration in 2004 volumes.