EXPERT WITNESSES COMMITTEE
EXPERT ALERT
AMERICAN BAR ASSOCIATION SECTION OF LITIGATION
Volume 4, No. 3 Summer 2008

# LET DATA SPEAK EQUALLY TO ALL: THE PRODUCTION, PROTECTION, AND USE OF RAW DATA IN LITIGATION[1]

*by Laura Ellsworth, Chuck Moellenberg, and Neelie Simmons* [2]

## I. INTRODUCTION

Scientists seek truth from examining data. Litigators seek truth from examining witnesses. This article posits the thesis that litigators seeking truth from scientists should be able to get it from examining the same source as the scientists: the data.

Since the Supreme Court's decision in *Daubert v. Merrill-Dow Pharmaceuticals, Inc.*, trial judges in federal and many state courts have become "gatekeepers" of scientific testimony, "ensur[ing] that any and all scientific testimony . . . is not only relevant, but reliable.[3] As the Supreme Court later explained, the "reliability" component of the gatekeeping inquiry exists "to make certain that an expert … employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field."[4]

*continued on page 4*

***continued from page 1***

Meanwhile, Federal Rule of Civil Procedure 26 and many similar state rules mandate extensive pretrial disclosure of expert testimony and the bases for the opinions proffered.[5] This disclosure is consonant with the courts' desire to "make a trial less a game of Blind Man's Bluff and more a fair contest with the basic issues and facts disclosed to the fullest practical extent."[6] In the case of expert discovery, disclosure of materials on which the expert relied "better prepar[es] attorneys for cross-examination, minimize[s] surprise, and suppl[ies] a helpful focus for the court's supervision of the judicial process."[7] In other words, the more information a party has about the facts and analyses underlying an opposing expert's opinion, the better prepared the party will be to challenge that opinion, whether at trial or before. Full disclosure of the basis of an expert's opinion also allows courts to better evaluate motions to exclude proffered opinion testimony.

In pharmaceutical litigation, as in the products liability field generally, cases involving complex questions of medical causation often turn on the "battle of the experts." As this battle takes on heightened importance, more and more litigants - citing *Daubert's* focus on the expert's methodology and procedural rules requiring disclosure of the expert's reliance materials - have successfully sought to review the raw data underlying the opinions proffered by opposing experts. In some cases, the testifying expert relies upon his own published studies and actually possesses the data underlying them. More often, the expert relies upon scientific studies published by others. In these latter cases, the testifying experts likely have no access to the data; the courts must arbiter subpoenas duces tecum and motions to quash involving the production of sensitive data from scientists who have nothing to do with the case.

Whether the testifying expert witness is a "primary" expert (*i.e.*, published or participated in the study upon which his opinions rely) or a "secondary" expert (*i.e.*, is relying on a published paper describing a study in which he played no role), requests to produce the raw data underlying published scientific studies typically are countered with several arguments:

- The published scientific study alone is sufficient, because it has been peer-reviewed;
- The data contain confidential and sensitive patient information, disclosure of which would infringe on patients' privacy rights;
- Production of raw data would be costly and time-consuming, and this burden should not be imposed on the expert;
- Production of raw data would have a "chilling effect" on further scientific research;
- Because the lawyer's job is to "spin" or "twist" facts to support his case, the lawyer would misuse the data to the potential detriment of a legitimate researcher; and
- It would be unfair to ask the trier of fact to decide the causation question based on information available at a single point in time, when science requires repeated replication of results before causal conclusions may be reached.[8]

Courts and commentators alike have examined and responded to these concerns.

In this article, we begin in Part II by discussing why access to raw data is critical in litigation today. Next, in Part III we summarize the arguments against disclosure and review the solutions that have been proposed or adopted to balance the competing concerns of scientists and litigants. Based on these *ad hoc* approaches and the reasoning underlying them, in Part IV we propose a set of cross-disciplinary guidelines designed to streamline the process of creating, disclosing, and protecting raw data sought in litigation. We hope these proposals generate meaningful discussions among scientists, lawyers, and judges facing these issues and, ultimately, contribute to a consistent, well-reasoned approach going forward.

***If the cross-examination is not meaningful, then the right becomes empty.***

**II. THE ARGUMENTS IN FAVOR OF DISCLOSING RAW DATA IN LITIGATION**

*A. The Disclosure Of Raw Data Levels The Playing Field Between The Parties*

The right to cross-examine an opposing witness is guaranteed by the civil rules and due process and is premised on the fundamental principles of fair play and balanced justice. "In almost every setting where important decisions turn on questions of fact, due process requires an opportunity to confront and cross-examine adverse witnesses."[9] This right, however, is not limited to a mere question-and-answer session with a witness testifying under oath. If the cross-examination is not meaningful, then the right becomes empty. "Due process mandates that a judicial proceeding give all parties an opportunity to be heard on the *critical and decisive allegations* which go to the *core* of the parties' claim or defense and to present evidence on the contested facts."[10]

In pharmaceutical lawsuits and other personal injury cases where liability turns on the often complex issue of medical causation, one cannot overstate the importance of *meaningful* cross-examination of the opposing expert witness. A single epidemiological study purporting to find an association between a product and an adverse health effect can spawn hundreds, if not thousands, of individual multi-million dollar lawsuits. Absent a meaningful opportunity to cross-examine the methodology of that study, or its applicability to the facts of a particular lawsuit, the manufacturer faces a considerable disadvantage in defending against potentially crushing liabilities. The mass distribution of products, coupled with increasing use, if not outright sponsorship,[11] of epidemiologic research, compels courts and scientists alike to devise procedures that rigorously test science introduced into the courtroom. Thus, to launch an effective challenge against an opposing party's expert, counsel should be able to challenge either the conclusion he has reached or the basis for it.

If an expert cites a published scientific study as the basis for his opinion that Drug "X" caused the plaintiff's injury, it is necessary but not sufficient to know what the study's authors reported in the publication. While the published summary of the study may support the expert's causation opinion, a closer look at the underlying data may demonstrate that the study was improperly designed or administered or that the data were improperly collected or analyzed. Such methodological improprieties, which violate the *Daubert* standard and require exclusion of the proffered testimony, often cannot be discerned from the publication alone. For example, one court found, on review of the data underlying the expert's opinion, that certain information had been collected but was excluded arbitrarily from the study sample. As a result, the opinion testimony was excluded.[12] Had the data been left unexamined, the unreliable methodology would have gone undetected.

In short, "[n]othing causes greater prejudice than to have to guess how and why an adversarial expert reached his or her conclusion."[13] Without the raw data, parties have no choice but to guess at what the

***If an expert cites a published scientific study as the basis for his opinon that Drug "X" caused the plaintiff's injury, it is necessary but not sufficient to know what the study's authors reported in the publication.***

undisclosed information might reveal. Similarly, nondisclosure of the underlying data hinders courts from performing their gatekeeping role; if the data are not produced, the court cannot confirm that appropriate, reliable methodologies have been used. "The value of the conclusions turns on the quality of the dat[a] and the methods used by the researcher in his analysis of that data …. [I]f the conclusions or end product of a research effort is to be fairly tested, the underlying data must be available to others equally skilled and perceptive."[14]

*B. The Disclosure Of Raw Data Provides A Practical, Cost-Effective Alternative to Requiring Experts to Generate Independent Research*

Neither lawyers nor their clients would be likely to characterize litigation as "fast-paced." However, the wheels of justice turn at breakneck speed in contrast to the necessarily deliberate pace of scientific research. Research progresses via the continued replication of previous results. Often, researchers gather and analyze data over a matter of years or even decades.

It would be impossible for a party to generate its own independent research to counter expert testimony premised on the results of a decade-long epidemiological study involving hundreds or thousands of subjects. In such a setting, the only meaningful way to challenge the expert's conclusions is to analyze the existing data and methodology of the study on which those conclusions are based.

The Court of Appeals for the Second Circuit recognized this practicality in a case that turned on studies performed by Dr. Irving Selikoff. Dr. Selikoff reportedly had found a synergistic effect between asbestos and tobacco, increasing the risks of cancer to smokers with histories of asbestos exposure. The defendant tobacco companies issued a subpoena to Dr. Selikoff seeking the data underlying his study, on which the plaintiffs' experts relied. The Court upheld the subpoena, observing that, "[t]hough the tobacco companies could conduct their own studies in an effort to controvert the findings of Dr. Selikoff, it seems that an inordinate amount of time would be required in order to duplicate the Selikoff study, and it is clear that scrutiny of the Selikoff data would provide a logically permissible manner in which to attack the findings."[15]

***By providing a practical way of leveling the playing field between the parties in litigation, disclosure of raw data advances fair play and equal justice.***

*C. The Disclosure Of Raw Data Furthers The Public Interest By Making More Information Publicly Available*

It goes without saying that a fair and just result in litigation is in the public interest. As one court has noted: "We believe that it is human to monetize welfare losses associated with grief, pain and suffering, humiliation, mental anguish, and other intangible injuries so that we can make plaintiffs whole. What we do not do, again for reasons grounded in humanity, is force a defendant to compensate a plaintiff if the plaintiff does not show that the defendant has probably done something to him."[16] By providing a practical way of leveling the playing field between the parties in litigation, disclosure of raw data advances fair play and equal justice.

But this is not the only benefit of disclosure of raw data in litigation. First, millions of dollars in government grants are infused every year into scientific research. The results of this scientific research often form the basis for federal policies, rules, and regulations.[17] Thus, it stands to reason that this research should be made available to those whose lives may one day be governed by it. Since April 2000, federal law has required such disclosure.[18] The law provides that any member of the public may access data from federally funded, scientific research projects via Freedom of Information Act ("FOIA") requests, if:

(a) the data relate to *published* research findings produced under an award that the Federal Government

used in developing an agency action that has the force and effect of law;

(b) the data are from a final, "published" report - that is, (i) published in a peer-reviewed scientific journal, or (ii) a federal agency has publicly and officially cited the research findings in support of an agency action that has the force and effect of law; and

(c) the data are not exempt from disclosure under the nine exceptions to the Freedom of Information Act.[19]

Second, the knowledge that data may one day be disclosed and analyzed in litigation may serve to reinforce the scientific process -- *i.e.*, the cautious, incremental progression requiring testing and re-testing of an hypothesis before causal conclusions are drawn or rejected.[20] As has been noted, "[f]raud in science is a problem that has been around for a long time, and legitimate concerns about it remain."[21] While outright fraud is probably (and thankfully) rare, the grant system provides incentives to finding "positive" correlations between, for example, the studied toxin or drug and an injury. A "positive" correlation generally prompts more funding for additional research. Livelihoods, career advancements, publications, and public recognition turn on research findings and continued funding. The "publish or perish" reality of many academic and research organizations exerts palpable pressure on researchers to find something in the data worthy of publication or funding for further study.

A "negative" finding, on the other hand, may mean finding a new topic to study:

> When the stakes are high - as is true in cutting edge research with millions of research dollars at stake, in drug development efforts where a negative study might stop development of a potential blockbuster medicine, or when millions of dollars in damages may be in play - the incentives are commensurately high for researchers to push the limits of accuracy in providing data.[22]

Negative findings often are not published, and they seldom make the headlines on CNN. There is therefore some incentive to find correlations in the data that may not, on further review, mean anything. Disclosure of data in litigation may effectively counter this perverted incentive.

Finally, as explained below, the raw data underlying epidemiology studies are not disclosed as part of the peer review process. Disclosure during litigation, then, may provide a greater base of information upon which the scientific community can comment.

## III. THE ARGUMENTS AGAINST DISCLOSING RAW DATA

### *A. Published Articles Have Already Been Subject To Peer Review*

Researchers facing requests for raw data underlying their published studies frequently claim that the peer review process obviates any need for inquiry into the data. However, the fact that a published summary study has been peer-reviewed does not mean that the study employed sound methodologies or appropriately analyzed the data gathered. The Supreme Court has recognized that "[p]ublication (which is but one element of peer review) is not a *sine qua non* of admissibility; it does not necessarily correlate with reliability."[23]

> ***In fact, peer review is an editorial process, developed long ago to help editors identify those articles among the many submitted that were suitable for publication. It was never intended to be, and it is not, a rigorous evaluation of the data underlying the author's conclusions.***

Journal peer review and publication are relevant to, but not dispositive of, the admissibility of testimony: "[I]f peer review alone was dispositive, then the *Frye* standard of general acceptability in the scientific

community would have remained adequate."[24] Thus, courts recognize the difference between peer review and review of the data, and generally hold that "peer review" is no substitute for a review of the data: "[A]lthough [the researcher] regularly summarizes his cases and comments on his findings, no one has yet publicly reviewed the actual core data indicating the basis and evidence upon which he classifies a patient's 'exposure' or 'nonexposure' to" the drug at issue.[25]

In fact, peer review is an editorial process, developed long ago to help editors identify those articles among the many submitted that were suitable for publication. It was never intended to be, and it is not, a rigorous evaluation of the data underlying the author's conclusions.[26] Although peer review is a standard practice among biomedical journals, the process is inconsistently implemented and is not intended to identify all the problems or shortcomings of a study. Generally speaking,

> [T]he reviewers do not themselves perform the experiments or primary research to confirm its results, nor do they ordinarily access the underlying data - the data presented are rarely the raw data; they have long since been analyzed. They instead accept the data as presented by the authors and that the methods used matched the methods described, then decide whether the conclusions reached are appropriate and interesting in the relevant discipline, and whether the methods described were appropriate.[27]

Commentators in the medical research field believe that this method of peer review is flawed.[28] In 2004, the Pharmaceutical Research and Manufacturers of America "establish[ed] an online database of clinical trial results of drugs marketed in the United States."[29] That database (available at www.clinicalstudyresults.org) contains summaries of controlled clinical trials completed since October 2002. The database summarizes both published and unpublished studies, regardless of whether the findings were positive or negative. While the database is not an exhaustive compilation of all clinical trials - participation in the program is voluntary, and the website does not include raw data or information on early-stage clinical trials[30] - its creation acknowledged the "valuable function of making clinical trial results for many marketed pharmaceuticals more transparent. More importantly, it [was] designed as a key tool to provide information to practicing physicians and their patients."[31]

In other scientific fields, raw data typically are shared upon request prior to approval and publication of a study. The *American Economic Review*, for example, has adopted a policy "to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Authors of accepted papers that contain empirical work, simulations, or experimental work *must provide to the Review, prior to publication, the data, programs, and other details* of the computations *sufficient to permit replication."*[32]

Among medical journals, authors also are expected to provide their data upon request. For example, *Nature's* policy states that, because "[a]n inherent principle of publication is that others should be able to replicate and build upon the authors' published claims[,] … authors are required to make materials, data and associated protocols available in a publicly accessible database … or, where one does not exist, to readers promptly on request."[33] Likewise, within the field of psychology, there is a general assumption that raw data will be provided upon request. The code of ethics adopted by the American Psychological Association (APA) mandates that data from published studies be disclosed to anyone who asks to review it, so that the conclusions drawn may be verified.[34]

Widespread recognition of the variability in the quality and scope of peer review and concerns about the limitations of the peer review process have prompted many in the biomedical community to critically investigate the peer review process. Despite an upsurge in research on the topic, Dr. Drummond Rennie, editor of JAMA, still sees many of the serious shortcomings of the peer review process that he highlighted over 20 years ago:

> One trouble is that despite this [peer review] system, anyone who reads journals widely and critically is forced to realize that there are scarcely any bars to eventual publication. There

> seems to be no study too fragmented, no hypothesis too trivial, no literature citation too biased or egotistical, no design too warped, no methodology too bungled, no presentation of results too inaccurate, too obscure, and too contradictory, no analysis too self-serving, no argument too circular, no conclusion too trifling or too unjustified, and no grammar and syntax too offensive for a paper to end up in print.[35]

Thus, access to raw data in litigation would help to allay the concerns expressed by many, such as long-time medical science-writer for the *New York Times*, Dr. Lawrence Altman, "that because editors and reviewers only examine what authors summarize, not raw data, the system can provide false reassurances that what is published is scientifically sound."[36]

Analysis of raw data during litigation can and has led to the discovery of improprieties in landmark studies published in major journals by leading scientists in the field. In the early 1990s, a case filed under the Superfund Act involving lead pollution in Utah led to the disclosure and examination of datasets compiled by Dr. Herbert L. Needleman, who had been retained to testify on behalf of the government. The defense experts' reanalysis of the datasets - which were the bases of ground-breaking studies published in leading journals - so troubled them that they contacted the National Institutes of Health's office of scientific misconduct. A hearing was held at the University of Pittsburgh, where Dr. Needleman was then conducting his research. The Hearing Board concluded:

> [T]here was a deliberate misrepresentation of the procedures actually used in the conduct of [Dr. Needleman's study] as reported in 1979 and thereafter. This was particularly true with respect to the procedures used in subject selection, exclusion, and classification for the analyses that included 158 children. Although we find no evidence that these deviations from the published procedures were done in order to bias the results …. [t]he Board unanimously believes that Dr. Needleman was deliberately misleading in the published accounts of the procedures used in the 1979 study.[37]

The federal Office of Research Integrity agreed that the methodologies used in Dr. Needleman's study "were inaccurately and incompletely reported in the original publication and thereafter."[38] While it found insufficient evidence to support a finding of scientific misconduct under federal regulations, it adopted the Hearing Board's conclusion that the "sequence of statements about these procedures [for excluding subjects] also reveals a 'pattern of errors, omissions, contradictions, and incomplete information from the original publication to the present.'"[39]

The need for access to the data underlying a research study on which an expert relies is a simple extension of the importance, recognized by courts and litigators, of reviewing an

opposing expert's data. In *In re: Silica Products Liability Litigation,*[40] for example, a multidistrict case involving some 10,000 plaintiffs seeking to recover for alleged injuries due to silica inhalation, defendants discovered:

> [T]he over 9,000 Plaintiffs who submitted Fact Sheets [some, in violation of the court's order, did not] were diagnosed with silicosis by only 12 doctors. In virtually every case, these doctors were not the Plaintiffs' treating physicians, did not work in the same city or even state as the Plaintiffs, and did not otherwise have any obvious connection to the Plaintiffs. Rather than being connected to the Plaintiffs, these doctors instead were affiliated with a handful of law firms and mobile x-ray screening companies.[41]

After seeking discovery from the "diagnosing" physicians and screening companies, defendants learned that the plaintiffs' "exposure histories" were taken "by people with no medical training, who had significant financial incentives to find someone positive for exposure to silica." In addition, although "outside of the small cadre of doctors who diagnose for screening companies, even a single case of a dual diagnosis of silicosis and asbestosis is extremely rare," the defendants learned that nearly half of these plaintiffs previously had been diagnosed with asbestosis - and filed cases seeking money damages for that illness - by the very same doctors who had now "diagnosed" their silicosis.[42]

After a hearing on the defendants' *Daubert* motions, the court held that the silicosis diagnoses were inadmissible. It also remarked upon the damage done by these diagnoses:

> Limited judicial resources [were] consumed weeding out meritless claims, costing the judiciary, costing other litigants whose suits are delayed, and ultimately costing the public ….
>
> Defendant companies pay significant costs litigating meritless claims. And what harms these companies also harms the companies' shareholders, current employees, and ability to create jobs in the future.
>
> And, potentially, every meritless claim that is settled takes money away from Plaintiffs whose claims have merit. And not only are those with meritorious claims denied just compensation, they are potentially denied full and meaningful access to the courts.[43]

This litigation story is little different from a research study reporting collectively on hundreds of individuals. Because the peer review process does not filter out studies that rely upon data that have been improperly collected or analyzed, courts should not (and under *Daubert*, cannot) excuse the researcher from producing his data simply because the published study was peer-reviewed.[44]

*B. The Time And Cost Involved In Producing Raw Data Impose An Undue Burden On The Researcher*

Some scientists facing a subpoena or motion to compel raw data have argued that compliance with the request would be so costly and time-consuming as to create an unfair burden. Moreover, responding to the request would take them away from their socially valuable research. As an initial matter, datasets are now maintained electronically and transmitted via e-mail or disk. Standardized programs to shield the identity of study participants, which are discussed below, further minimize time, cost and disruption.

Additionally, courts can and often do fashion orders designed to make disclosure of raw data less burdensome.[45] For example, courts have frequently ordered the party requesting the information to compensate the scientist (or, more

***Some scientists facing a subpoena or motion to compel raw data have argued that compliance with the request would be so costly and time-consuming as to create an unfair burden.***

likely, his research assistant) for the time and cost associated with production of the data.[46] This solution also serves to avoid the "shell game" problem of choosing who will testify; there is no "cost penalty" to being the scientist willing to step into the witness box.

A different burden may be presented by data other than computerized datasets, and in these cases, a different solution may be appropriate. One New York court, presented with this argument in *In re R.J. Reynolds*, a products liability suit involving tobacco, found it compelling.[47] There, Reynolds had requested the "production of data, tapes, documentation relating to interviews, questionnaires, medical records, death certificates, x-rays, autopsies and computer tapes, as well as previous or follow-up studies using a subset or superset of data covering the medical investigations." Reynolds also had conceded that the materials requested were "not needed … to evaluate the studies' conclusions." The court held that, because the subpoena was "sweeping and indiscriminate" and "overbroad," the defendant's request for data was "so oppressive as to hinder the normal functioning" of the researcher and granted the motion to quash.[48]

Such collateral materials, which are less likely to have been reduced to electronic form during the course of the study, do present as the *Reynolds* court observed, different burden issues. Undoubtedly, *Reynolds'* concession that these materials were not needed to evaluate the studies was a material point.

Thus, it appears that some heightened burden may be appropriate for non-electronic, particularly voluminous documents. For example, if a review of the electronic dataset were to reveal data anomalies that could be explained only by checking the interview forms, then a case might be made for a secondary production from that limited source. This type of measured, sequential production could produce relevant information while minimizing the burden to the scientist.

At the end of the day, lay witnesses have long been subjected to the subpoena power of the courts; often, although they may have no interest in the outcome of a lawsuit, they are required to produce documents or testify about issues on which they have personal knowledge. Likewise, expert witnesses who have specialized knowledge about matters at the heart of litigation should be required to produce the raw data on which the just outcome of the case will turn, particularly where the testifying expert - who has agreed to appear on behalf of a party and will be compensated for his time and testimony - has control of the data.

*C. Disclosure Of The Raw Data Would Invade The Privacy Concerns Of The Individuals Participating In The Study And Violate The Law*

Another frequent objection to disclosure of data is that the data contain information of a sensitive personal nature. Research, scientists argue, is possible only because individuals trust the researcher with

confidential information. If raw data were disclosed in litigation, not only would the researcher risk violating health privacy laws, but individuals would avoid sharing information in the future. Ultimately, they say, this would adversely affect the quality of research.

Courts have recognized the validity of this concern.[49] *In Deitchman v. E.R. Squibb & Sons, Inc.*, for example, the defendant sought to obtain raw data (the "Registry") compiled on victims of genital carcinoma to evaluate the plaintiffs' claims that their cancer had been caused by a drug ingested while they were *in utero*. The court recognized the researcher's concerns, noting that if the confidentiality of the Registry participants were compromised, "all society will be the poorer … [and] a unique and vital resource for learning about the incidence, causes, and treatment of adenocarcinoma will be lost."[50]

Nevertheless, the data in that case ultimately were disclosed: "We leave it to the district judge ... to fashion as inventive a [protective order] as the necessities of this unique case dictate, one which allows Squibb the least necessary amount of information to avoid a miscarriage of justice without doing needless harm to [the researcher] or his Registry."[51]

In most cases, litigants seeking raw data do not want or need information that identifies the study participants. They seek to review other characteristics recorded by the researchers about the study participants, and determine whether the inclusion or exclusion of these characteristics in the analysis affects the study's conclusions. Likewise, the electronic compiling of data lends itself to easy redaction of confidential information. Study participants' names and other identifying information often can be removed from the database produced to the litigants with just a few strokes on the computer keyboard.

***Only when replication of experiments over time has yielded consistent, significant results do scientists consider a matter "settled." In contrast, litigation sets two or more parties against one another, each of whom seeks to prove false the other's allegations.***

Accordingly, trial judges typically will "compare the hardship to the party [or person] against whom discovery is sought, if discovery is allowed, with the hardship to the party seeking discovery if discovery is denied."[52] They then can draft protective orders to prevent disclosure of confidential, personal information.

*D. The Scientific Method Is Fundamentally Incompatible With Litigation*

The scientific method seeks truth by repeated testing of the same hypothesis. Only when replication of experiments over time has yielded consistent, significant results do scientists consider a matter "settled." In contrast, litigation sets two or more parties against one another, each of whom seeks to prove false the other's allegations. Therefore, researchers have argued, they should not be forced to disclose their data to those with an interest in demonstrating their conclusions to be false, as this runs counter to the scientific process.

At least one court has found this argument compelling enough to quash a subpoena seeking raw data. There, the court reasoned that "[t]he validity of opinions formed and expressed in the context of disciplines other than the law should be tested by the relevant discipline's requirements for validity or acceptability;" therefore, it would not "substitute the adversarial process of the judicial search for truth for the epistemological standards set by other disciplines."[53]

Few courts, however, have followed this line of reasoning. Nor should they. However awkward the fit between science and law may be, neither field would be improved by the construction of a wall between the two. In litigation, injured parties who

set forth a meritorious causation claim should be compensated. Often, that causation claim cannot be satisfied without expert testimony. Likewise, those who are alleged to have caused harm to others have a right to defend against these allegations. When a plaintiff is permitted to use a scientific study as a "sword" against a defendant (or vice versa), justice requires that the opposing party be permitted to examine the basis for the study. Only then may a jury determine whether the argument has merit.

Similarly, the scientific process itself demands rigorous criticism and skepticism. Science progresses through the questioning of accepted "truth" and debate over novel theories. Every Ph.D. candidate must defend his research against vigorous "cross-examination." There is no reason for excusing researchers from answering legitimate questions about the bases of their conclusions, especially when those conclusions are cited by a litigant against his opponent. Checks on the methodologies employed in collecting and analyzing the raw data underlying scientific studies will encourage more responsible research and more carefully articulated conclusions. This, in turn, will promote more rational public policies. The public benefits by shining light into the scientific laboratory.

*E. The Data Is Protected By A Researcher's "Privilege"*

Some researchers have resisted subpoenas by asserting a researcher's or scholar's "privilege." This argument is often raised when the opposing party seeks data related to a study that is ongoing, continuous, or otherwise incomplete. Generally speaking, however, courts have rejected the notion of a researcher's privilege.[54] Even in those cases where the privilege has been recognized, the courts have acknowledged that a litigant who needs the data from ongoing studies and has no other means of obtaining them should be entitled to review the data, subject to an order protecting the data from disclosure to those not directly involved in the litigation.

Of the federal courts, only the Seventh Circuit has recognized even a limited privilege for ongoing research.[55] Even in these two cases, however, the courts noted that academic freedom was not absolute and that trial judges must balance the researcher's interest in disclosing his data at the time and place of his choosing with the requesting party's right to meaningfully confront the evidence levied against it. Moreover, in both cases, the data requested either would not be used against the requesting party or included information identifying the study's participants.[56]

Likewise, one intermediate appellate state court has held that researchers have a genuine interest in being the "first presenters" of the results of their studies: "While ... medical investigations are still in progress, they should not be subjected to examination and criticism by people whose interests are arguably antithetical to the medical scientists."[57] This is because disclosure prior to publication "would have the effect of denying to these doctors the opportunity of first publication of their studies. It could also have a chilling effect and discourage further scientific endeavors." However, as in the cases examined by the Seventh Circuit, the subpoena at issue in this case was remarkably broad, and the party seeking disclosure had admitted that most of the materials were not necessary to evaluating the studies' conclusions.[58]

In short, courts appear to invoke the so-called "researcher's privilege" to block disclosure of data from ongoing research only when the request is overly burdensome, it seeks materials that are unnecessary to the party's case, or the data include identifying information not needed by the requesting party. In addition, these courts have suggested that a request for data more narrowly tailored to the facts and legal claims at issue would likely be upheld, subject to an appropriate protective order.[59]

On the other hand, at least one court has rejected a scientist's attempt to avoid production of data from an ongoing study.[60] In that case, the plaintiff, who had sued several vinyl chloride manufacturers alleging that workplace exposure to the chemical had caused her husband's death, sought data from studies examining the health effects of vinyl chloride. Defendants objected, arguing (among other things) that because the research was ongoing and incomplete, it would not be admissible at trial and therefore "could not possibly contain any

information reasonably calculated to lead to the discovery of admissible evidence." The court disagreed, noting that "there is not enough information … to make a determination as to whether any of the scientific techniques or theories … will be relied upon by any of the witnesses at trial." Accordingly, the motion to compel the documents was granted even though the research was ongoing and not yet the subject of a published article.[61]

Situations where a researcher adds to a dataset over many years, publishing new articles when additional data are added and the full set reanalyzed, provide yet another angle from which this issue must be considered. Longitudinal epidemiology studies may examine the same individuals over a matter of decades to determine the effects of toxic exposure early in life. The most recent data in these studies likely are the most informative as to the long-term impact, if any, of that exposure. But, if analysis of the most recent data does not reveal any association between the exposure and the hypothesized effects, the researcher may decide not to publish that finding; as noted above, "non-findings" often are not reported. Meanwhile, the lack of association might be critical to a litigant defending against claims involving the same toxic exposure and effects studied. In such instances, where the litigant has no other means of proving its case and a protective order can be fashioned to avoid unwanted disclosure of the data, justice requires that the data be disclosed.

In sum, no court has permitted a scientist to shield his data from production based solely on the claim that there is "more to come" - and rightly so. At the end of the day, a party's right to full and fair cross-examination of opposing experts should not be denied because of speculation that someone else may steal the data, particularly where a protective order is in place to keep the data from improper disclosure or use.

## IV. PROPOSED CROSS-DISCIPLINARY PRINCIPLES TO PROMOTE FULL AND FAIR DISCLOSURE OF RAW DATA

Based on the review of the objections raised by researchers responding to requests for data, the solutions that courts have created to balance the concerns of science with litigants' need for data, and our own experience seeking raw data for use in lawsuits, several principles may be distilled. It is hoped that these proposed principles will generate much-needed discussion of these issues among judges, litigants and scientists. Ultimately, we hope that these discussions will generate a well-considered approach going forward.

***At the end of the day, a party's right to full and fair cross-examination of opposing experts should not be denied because of speculation that someone else may steal the data, particularly where a protective order is in place to keep the data from improper disclosure or use.***

Principle No. 1: Raw data from studies on which expert testimony relies should be available to opposing attorneys for use at deposition, trial, and during motion practice. Because researchers expect that their data and methodology will be reviewed in the scientific context, the presumption should be in favor of disclosure in the legal arena as well.

Principle No. 2: Raw data from studies on which expert testimony relies should be available regardless of whether the witness relying upon it is a "primary" or a "secondary" expert.

Principle No. 3: Researchers and attorneys should work together to develop a means by which electronic datasets can be collected and stored to facilitate quick and inexpensive production of datasets in litigation.

Principle No. 4: Where review of the core dataset reveals anomalies that may be relevant to the case at hand and are explained only by additional, non-electronic data, courts should permit the production of these additional materials. A protective order may deal with any appropriate concern for cost or time of production.

Principle No. 5: Except in exceptional circumstances where specific need has been demonstrated, litigants requesting raw data should refrain from seeking disclosure of personally identifying information and courts should require "de-identification" of the data prior to production.

***It has been said that "lawyers use statistics like a drunk uses a lamppost - for support, not illumination."***

Principle No. 6: Where several lawsuits in one or more jurisdictions involve the same causation issue and reliance upon one or more of the same studies, judges, lawyers and scientists should develop a confidential, protected "warehouse" of data so non-testifying researchers need not respond more than once to requests for the same information. Moreover, parties in later cases must show specific need for additional information before issuing subpoenas for more data.

Principle No. 7: If data are part of an ongoing study on which no articles have been published, the data are directly relevant to a litigant's claims or defenses, and the litigant demonstrates specific need for the data, the data should be produced subject to a protective order against further disclosure or use outside the litigation.

Principle No. 8: Confidentiality concerns beyond information identifying study participants should be addressed through protective orders. These protective orders should carry meaningful sanctions for non-compliance.

## V. CONCLUSION

It has been said that "lawyers use statistics like a drunk uses a lamppost - for support, not illumination." It is not surprising, then, that scientists object to producing their data to lawyers who would use the data to support their claims or defenses in litigation. Lawyers, on the other hand, have long been familiar with Mark Twain's "three kinds of lies: lies, damned lies, and statistics." They believe that disclosure of relevant data promotes a full and fair search for the truth, and that reliance on the published studies without reviewing the underlying data is tantamount to admitting liability without seeing a speck of proof.

While this mutual suspicion is not surprising, it is in the best interests of scientists and attorneys to reach an accord on these matters. Science and law are inextricably wed - now so more than ever before - and must find a way to coexist with one another. The disclosure of raw data fosters the shared goals of science and litigation: a search for truth derived from examination of various viewpoints using accurate information. Adherence to and enforcement of protective orders can ensure that raw data produced in litigation will go no further. The case law and commentary provide a road map for a practical, balanced approach to disclosure of researcher's raw data for use in litigation.

***Laura Ellsworth*** is the Partner-In-Charge of the Pittsburgh office of Jones Day. ***Chuck Moellenberg*** coordinates the Trial Practice Group of Jones Day's Pittsburgh office. ***Neelie Simmons*** is an associate in Jones Day's Pittsburgh office.

[1] This article, under the title "Let Data Speak Equally to All: The Increasing Importance of Raw Data in Litigation and a Proposal for Principles Governing the Production, Protection, and Use of Raw Data in the Litigation Context," by Laura E. Ellsworth, Charles H. Moellenberg, Jr., and Neelie S. Simmons, was first published in Mass Torts, Volume 6, No. 2, Winter 2008. 

[2] The authors wish to thank Sarah Moellenberg for her research into various publications' requirements regarding the disclosure of raw data and Joshua Cippel for his research of case law and commentary. The opinions in this

article are those of the authors and do not express the opinions of Jones Day or any of its clients.

[3] *Daubert v. Merrill-Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 589 (1993).

[4] *Kuhmo Tire Co., Ltd. v. Carmichael*, 526 U.S. 137, 152 (1999) (emphasis added).

[5] Federal Rule 26(a)(2)(B) requires disclosure of the identity of any expert witness whose testimony a party plans to introduce at trial, along with a written report "prepared and signed by the witness [which] shall contain a complete statement of all opinions to be expressed and the basis and reasons therefore; [and] *the data or other information considered by the witness in forming the opinions*." FED. R. CIV. P. 26(a)(2)(B) (emphasis added).

[6] *United States v. Procter & Gamble Co.*, 356 U.S. 677, 682 (1958).

[7] *See Thibeault v. Square D Co.*, 960 F.2d 239, 244 (1st Cir. 1992) (citing cases) (internal citations omitted).

[8] The replication requirement poses unique problems in litigation when one party bases its causation argument (or an expert bases his testimony as to causation) on one or two epidemiology studies. Epidemiology studies, which examine "the incidence, distribution and etiology of disease in human populations," are generally "offered to establish or dispute whether exposure to an agent caused a harmful effect or disease." REF. MAN. ON SCI. EVID. 335 (2000). Importantly, these studies demonstrate *association*, not causation. *Id.* at 336. For that reason, epidemiologists decline to make causal conclusions absent replication of the results obtained, understanding of the biological mechanisms for the association, and similar information. *See id.*; *see also* David C. Bellinger & Kim N. Dietrich, "Low Level Lead Exposure and Cognitive Function in Children," PED. ANNALS, 23:11, 600-605 (November 1994) (repeatedly referring to epidemiology studies suggesting "association"). While this issue is beyond the scope of this article, it is an important topic for future consideration.

[9] *Goldberg v. Kelly*, 397 U.S. 254, 269 (1970) (citing cases).

[10] *Complaint of Bankers Trust*, 752 F.2d 874, 890 (3d Cir. 1985) (emphasis in original).

[11] Though corporate funding of scientific research typically garners the most press, law firms that usually (if not exclusively) represent plaintiffs are now in the business of funding research as well. *See* "Questions Persist Over Lynch's Deal With DuPont," *Prov. J.*, Aug. 9, 2006. Just as there is nothing intrinsically nefarious about corporate funding of medical or epidemiological research, there is nothing necessarily nefarious about research sponsored by plaintiff law firms. But, because no single group has a corner on the research market, the real world of high stakes litigation requires inquiry into scientific evidence.

[12] *See Sheehan v. Daily Racing Form, Inc.*, 104 F.3d 940, 942 (7th Cir. 1997) (rejecting expert testimony because the expert had "fail[ed] to exercise the degree of care that a statistician would use in his scientific work, outside the context of litigation."); *Deitchman v. E.R. Squibb & Sons, Inc.*, 740 F. 2d 556, 563 (7th Cir. 1984) (allowing discovery of data underlying study of persons who had developed genital carcinoma in case involving children exposed in utero to drug DES, because "a study of this sort may have a number of different, but inadvertent, biases present.").

[13] *Reed v. Binder*, 165 F.R.D. 429, 430 (D.N.J. 1996); *see also Deitchman v. E.R. Squibb & Sons, Inc.*, 740 F.2d 556, 563 (7th Cir. 1984) (access to and analysis of epidemiological research data was "absolutely essential" to defendant's ability to prepare proper defense).

[14] *Wright v. Jeep Corp.*, 547 F. Supp. 871, 874 (E.D. Mich. 1982).

[15] *In re American Tobacco Co.*, 880 F. 2d 1520, 1529 (2d Cir. 1989).

[16] *Wright v. Willamette Indus., Inc.*, 91 F.3d 1105, 1108 (8th Cir. 1996).

[17] *See, e.g.*, Sen. Richard Shelby, *Accountability and Transparency: Public Access to Federally Funded Research Data*, 37 Harv. J. Leg. 369, 375 (Summer 2000) ("Americans have a right to know how their tax dollars are spent and whether they are spent wisely, as well as the underlying scientific basis for many of our federal policies and rules.").

[18] *See* 2 C.F.R. Part 215.0 - 215.73.

[19] *See id.* Generally speaking, these nine exceptions involve information related to national security, agency personnel rules and procedures, trade secrets or privileged commercial information, data that would create an invasion of personal privacy if disclosed, and other sensitive agency, law enforcement, and geological or geophysical information. *See* 5 U.S.C. § 552(b).

[20] *See, e.g.*, Richard L. Marcus, *Evidence: Discovery Along the Litigation/Science Interface*, 57 BROOKLYN L. REV. 381, 385-86 & n.13 (Summer 1991) (describing the scientific process as a "quest [that] involves experimental evaluations of hypotheses that incrementally build upon the scientific work of others toward greater insights....").

[21] *See id.* at 388.

[22] *See* William G. Childs, *The Overlapping Magisteria of Law and Science: When Litigation and Science Collide*, 85 NEBRASKA L. REV. 643, 658-59 (2007).

[23] *Daubert*, 509 U.S. at 593.

[24] *Allison v. McGhan Med. Corp.*, 184 F.3d 1300, 1313 (11th Cir. 1999) (rejecting peer-reviewed study as an inadequate basis for expert's opinion). *See also Daubert*, 509 U.S. at 593 ("The fact of publication (or lack thereof) in a peer reviewed journal thus will be a relevant, though not dispositive, consideration."), *Consolidation Coal Co. v. Latusek*, 187 F.3d 628, 1999 WL 592051, at *4 (4th Cir. 1999) ("[T]he fact that the articles were published and subjected to some amount of peer review does not indicate that they were necessarily reliable."); *Black v. Rhone-Poulenc, Inc.*, 19 F. Supp. 2d 592, 600 (S.D. W. Va. 1998) (publication does not ensure reliability of article).

[25] *Deitchman v. E.R. Squibb & Sons, Inc.*, 740 F.2d 556, 563 (7th Cir. 1984).

[26] *See* J.C. Burnham, *The Evolution of Editorial Peer Review*, 263 JAMA 1323-29 (1990).

[27] *See* Childs, *supra* note 22, at 655.
[28] *See, e.g.*, Arnold S. Relman and Marcia Angell, *How Good is Peer Review?*, 321 NEW ENGL. J. MED. 828 (1989) ("[P]eer review is not and cannot be an objective scientific process, nor can it be relied on to guarantee the validity or honesty of scientific research").
[29] *See* Rick Whiting, "Drug-Industry Database Will Let Doctors, Consumers Search Trial Results," INFORMATION WEEK (Sept. 7, 2004), available at http://www.informationweek.com.
[30] *See id*.
[31] *See* www.clinicalstudyresults.org/home/.
[32] *See* Submissions, AM. ECON. REV., *at* http://www.aeaweb.org/ear/data_availability_policy).html (Oct. 12, 2007) (emphasis added).
[33] Availability of Data and Materials, NATURE, *at http://www.nature.com/authors/editorial_policies/availability.html* (Oct. 12, 2007).
[34] *See* "Ethical Principles of Psychologists and Code of Conduct," Rule 8.14 (eff. June 1, 2003), *at* http://www.apa.org/ethics/code2002.html ("After research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose, provided that the confidentiality of the participants can be protected and unless legal rights concerning proprietary data preclude their release.").
[35] Drummond Rennie, *Fourth International Congress on Peer Review in Biomedical Publication*, 287 JAMA 2759, 2760 (2002); *see also* Lars Noah, *Sanctifying Peer Review*, 59 U. PITT. L. REV. 677, 698 (1998) ("At best ... editorial peer review manages to filter out obviously sloppy work.").
[36] *See* Lawrence K. Altman, *Peer Review System for Journals Can Get You Into Trouble, at* http://www.lauralee.com/news/peerreview.htm, (Oct. 13, 2007), p. 1 of 3.
[37] Needleman Hearing Board Final Report at i-ii.
[38] Office of Research Integrity Oversight Report, at 5-6.
[39] *See id*.
[40] 398 F. Supp. 2d 563 (S.D. Tex. 2005).
[41] *Id*. at 580.
[42] *Id*. at 622 & 628.
[43] *See id.* at 634 & 636.
[44] Nor should the court excuse the production of raw data simply because the witness relying upon it is a "secondary" expert and, therefore, has not seen the data. While the playing field may be "even" when neither the expert nor the lawyer cross-examining him has reviewed a study's underlying data, such reasoning ignores the ultimate purpose of both science and law. The search for truth is at the heart of what goes on in the courtroom (and in the laboratory). That search cannot be effective — and justice is not served — by rendering both plaintiff and defendant equally ignorant of the data underlying an expert's opinion.
[45] *See Wright v. Jeep Corp.*, 547 F. Supp 871, 877 (E.D. Mich. 1982) ("The solution is not to cover-up [research data] because disclosure is too burdensome but to use the tools available to lessen the burden and to permit the information to become available.").
[46] *See In re American Tobacco Co.*, 880 F.2d 1520, 1529 (2d Cir. 1989); *Wright v. Jeep Corp*., 547 F. Supp 871, 877 (E.D. Mich. 1982) (noting that a reasonable fee "may include not only the professional fee and cost of supplying the documents and remuneration for the inconvenience, but also could include in the appropriate case a charge for a portion of the expenses of the original research").
[47] *See, e.g., In re R.J. Reynolds*, 518 N.Y.S. 2d 729, 733 (Sup. Ct. 1987) ("The loss of the scientists' time in complying with the subpoena would be an unreasonable burden and would unduly interrupt their ongoing medical investigations.").
[48] *See id*. at 731, 733 & 734.
[49] *See, e.g., Farnsworth v. Proctor & Gamble*, 101 F.R.D. 355, 358-59 (N.D. Ga. 1984) (arguing that loss of confidentiality surrounding this information could inhibit future CDC studies).
[50] *See Deitchmann*, 740 F.2d at 560.
[51] *See id*. at 566.
[52] *See id*. at 559.
[53] *In re Snyder*, 115 F.R.D. 211, 215-16 (D. Ariz. 1987).
[54] *See, e.g.*, *Burka v. United States Dep't of Health & Human Servs*., 87 F.3d 508 (D.C. Cir. 1996); *In re American Tobacco Co*., 880 F.2d 1520 (2d Cir. 1989); *Anker v. G.D. Searle & Co*., 126 F.R.D. 515, 519 (M.D.N.C. 1989); *Wright v. Jeep Corp*., 547 F. Supp. 871, 876 (E.D. Mich. 1982).
[55] *See Deitchman v. E.R. Squibb & Sons, Inc*., 750 F.2d 556, 560-61 (7th Cir. 1984); *Dow Chem. Co. v. Allen*, 672 F.2d 1262, 1274-76 (7th Cir. 1982).
[56] *See Deitchman*, 750 F.2d at 562, 564-65 (stating that "[t]he denial of discovery here has effectively precluded Squibb from engaging in any meaningful cross-examination of plaintiffs' experts' opinions" but noting that data including identifying information ought not be produced); *Dow Chem. Co. v. Allen*, 672 F.2d at 1276-77 (noting that "facts could arise sufficient to overcome respondents' academic freedom interests" and suggesting that, if data in question were to be used against party seeking disclosure, balancing test would favor disclosure).
[57] *In re R.J. Reynolds*, 518 N.Y.S.2d 729, 733-74 (Sup. Ct. 1987).
[58] *See id.* at 733, 734 & 786.
[59] *See, e.g.*, *Deitchman*, 740 F.2d at 566 ("We leave it to the district court to fashion as inventive an order as the necessities of this unique case dictate.").
[60] *See Smith v. Dow Chem. Co*., 173 F.R.D. 54 (W.D.N.Y. 1997).
[61] *See id.* at 55, 58 & 59 (citation & internal quotations omitted); *see also Simon v. G.D. Searle & Co.,* 119 F.R.D. 680 (D. Min.. 1987) (ordering production of ongoing research study conducted by defendant's expert and funded by defendant).