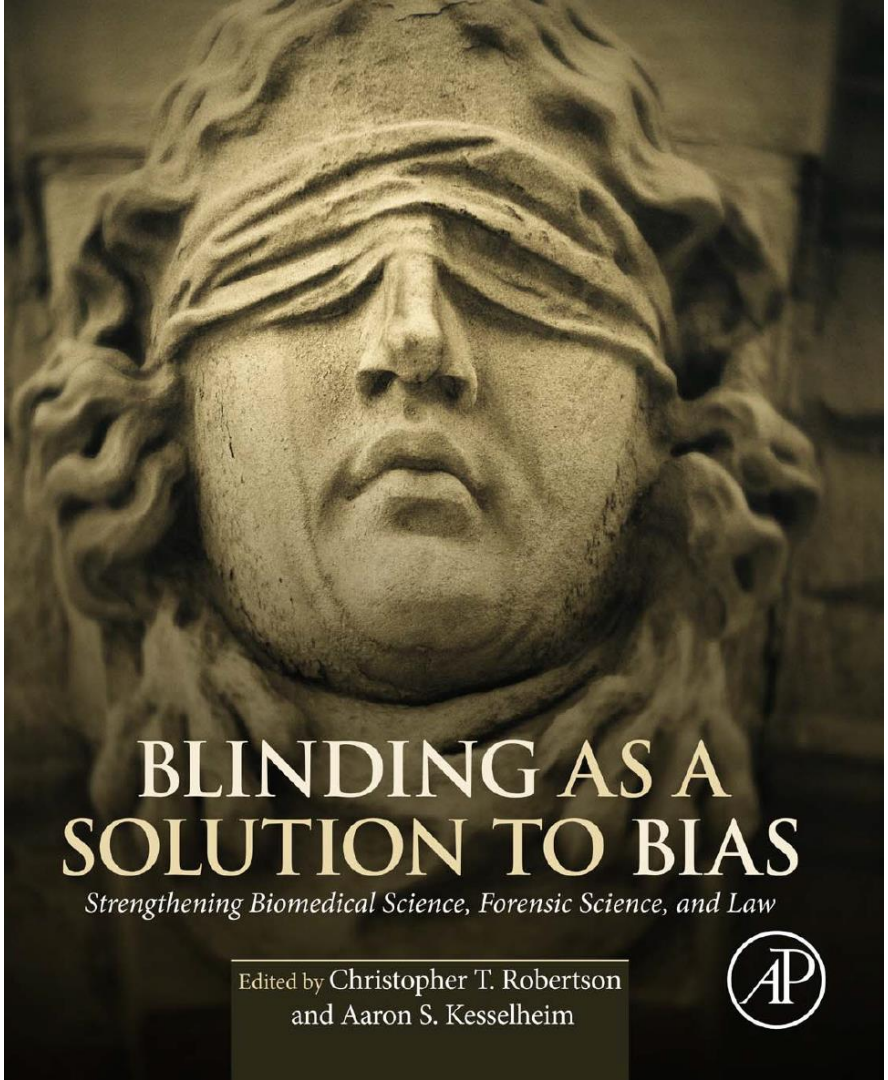

BLINDING AS A SOLUTION TO BIAS
Strengthening Biomedical Science, Forensic Science, and Law
Edited by Christopher T. Robertson and Aaron S. Kesselheim
AP

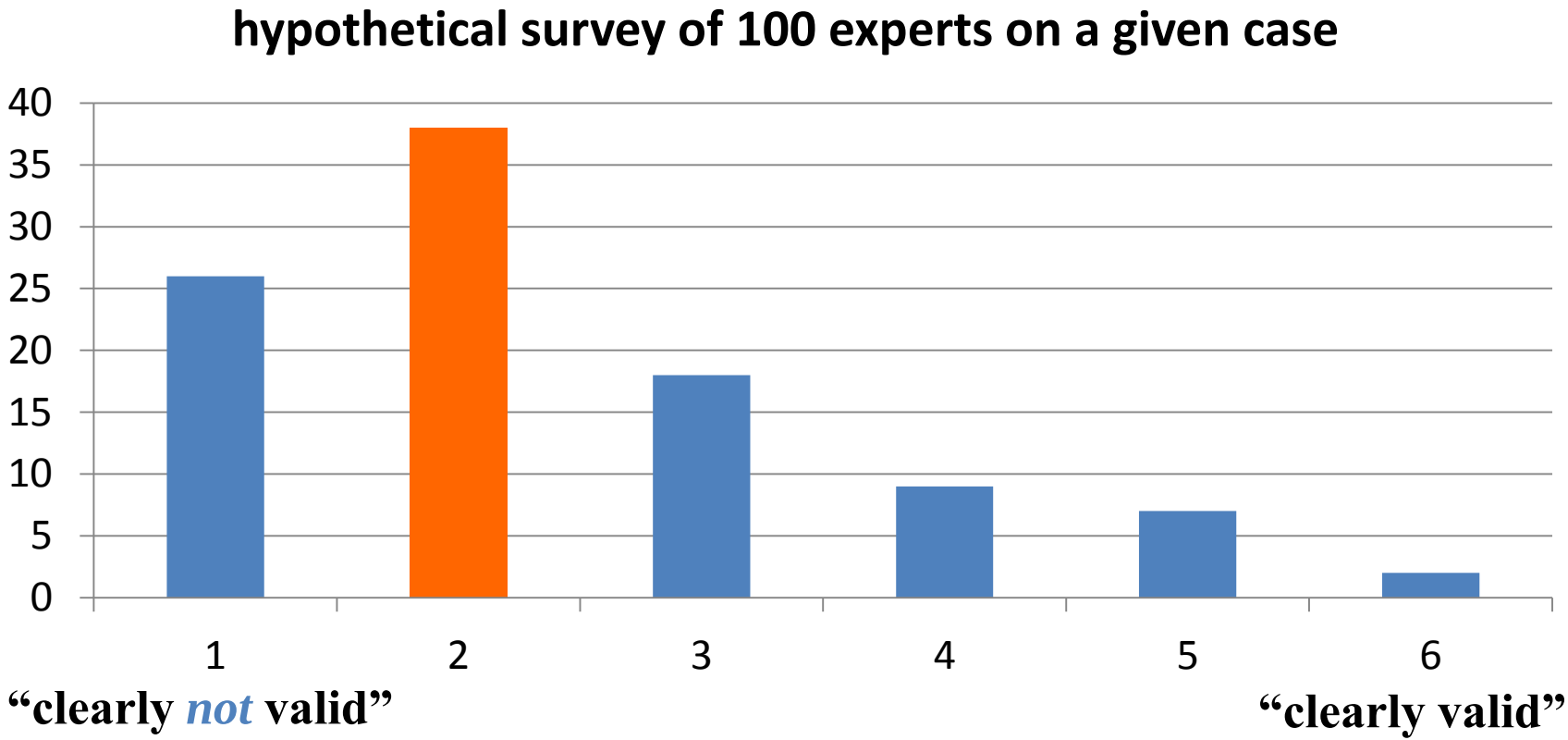
Bias in

LITIGATION SCIENCE

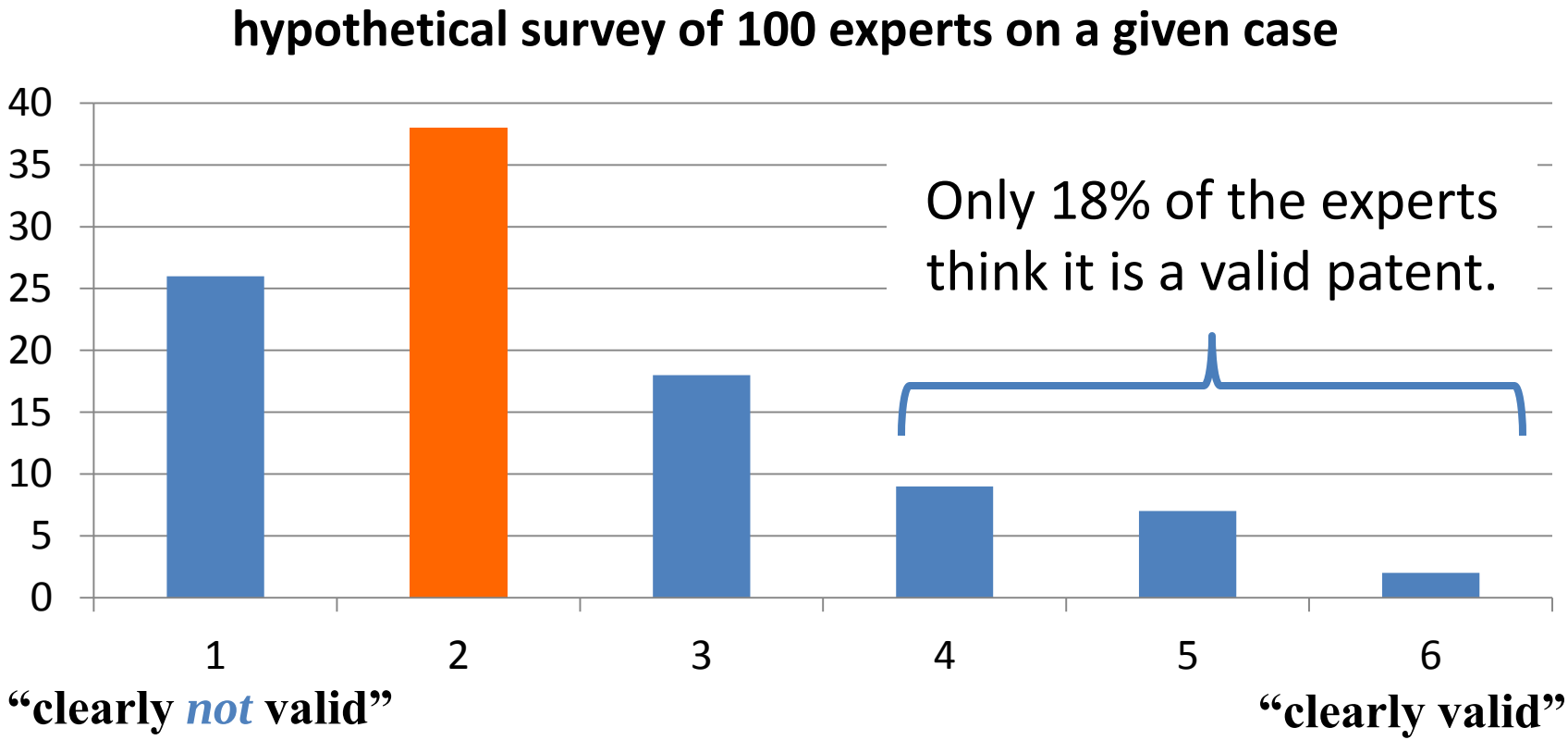
Expert Biases

- **Selection**
- **Affiliation**
- **Compensation**
- **Hindsight**

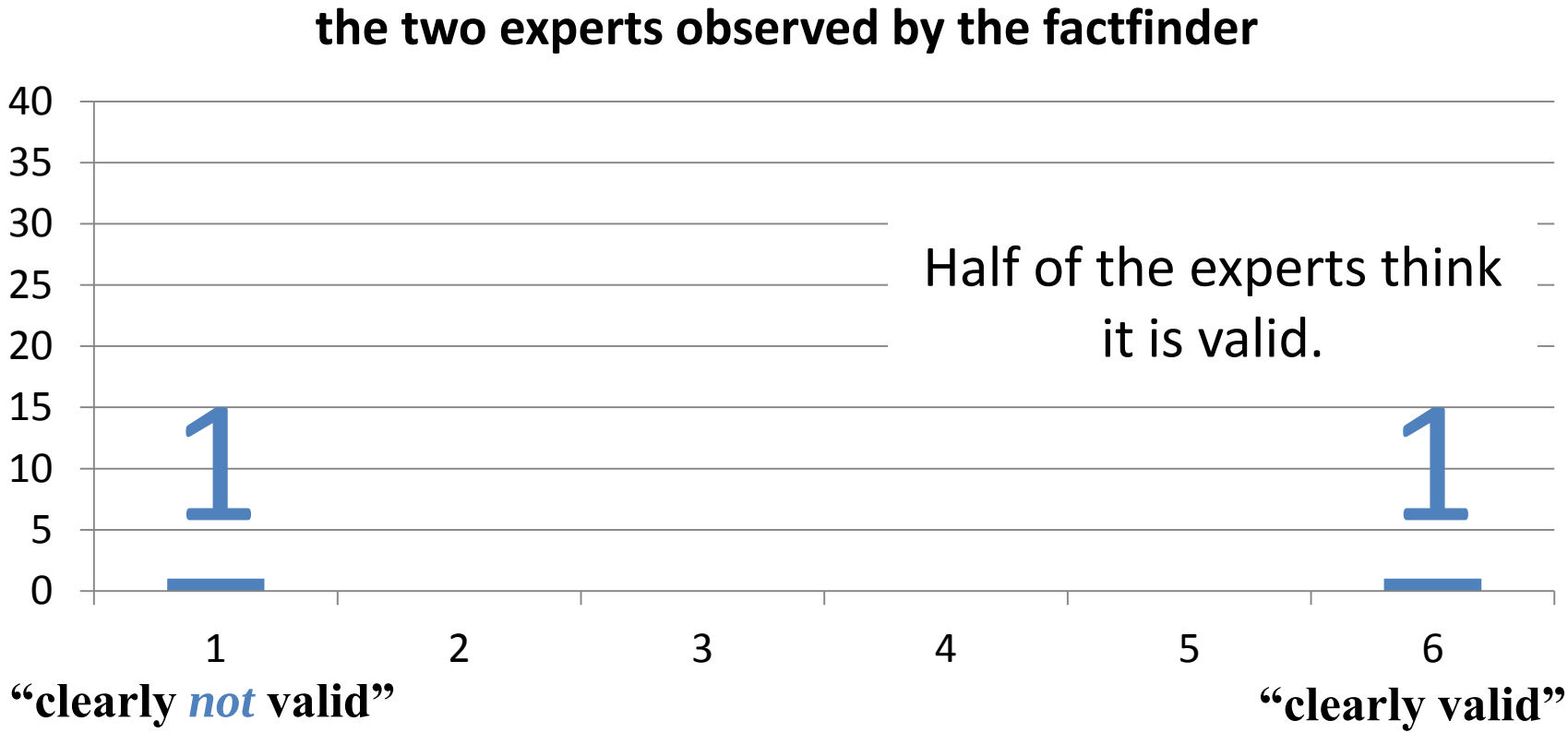
Selection


hypothetical survey of 100 experts on a given case
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
“clearly *not* valid”
“clearly valid”

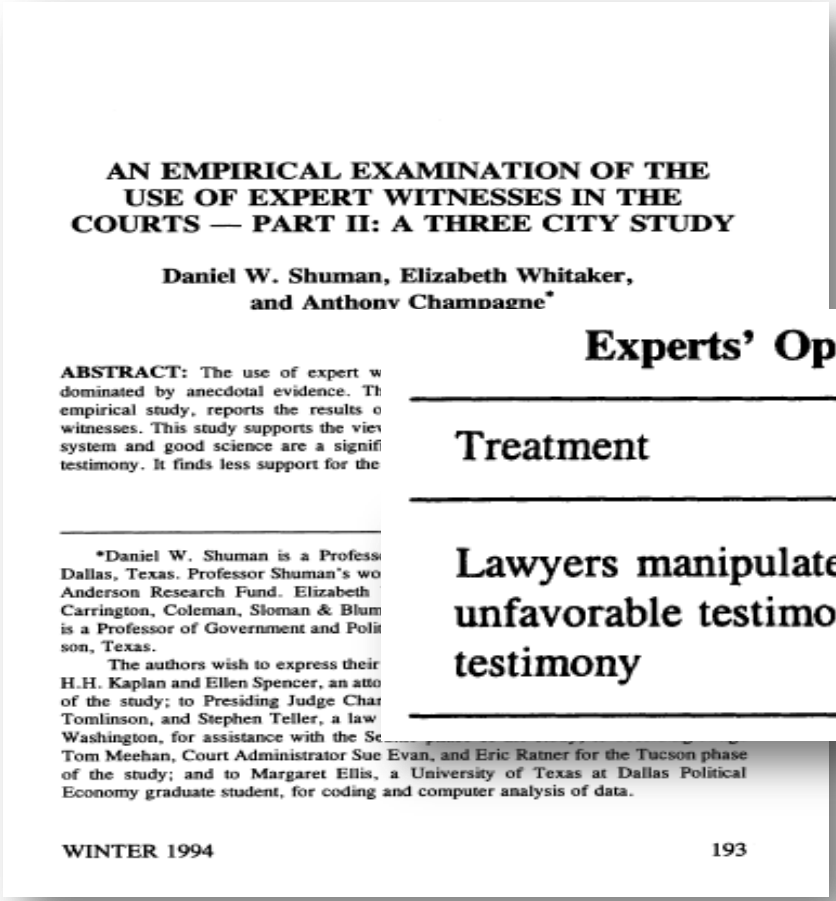
Selection


hypothetical survey of 100 experts on a given case
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
Only 18% of the experts think it is a valid patent.
“clearly *not* valid”
“clearly valid”

Selection


the two experts observed by the factfinder
40
35
30
25
20
15
10
5
0
1
1
Half of the experts think it is valid.
1
2
3
4
5
6
“clearly not valid”
“clearly valid”

Affiliation Bias



AN EMPIRICAL EXAMINATION OF THE USE OF EXPERT WITNESSES IN THE COURTS — PART II: A THREE CITY STUDY

Daniel W. Shuman, Elizabeth Whitaker, and Anthony Champagne*

ABSTRACT: The use of expert w
dominated by anecdotal evidence. Th
empirical study, reports the results o
witnesses. This study supports the viev
system and good science are a signifi
testimony. It finds less support for the

*Daniel W. Shuman is a Profess
Dallas, Texas. Professor Shuman's wo
Anderson Research Fund. Elizabeth
Carrington, Coleman, Sloman & Blum
is a Professor of Government and Polit
son, Texas.

The authors wish to express their
H.H. Kaplan and Ellen Spencer, an atto
of the study; to Presiding Judge Char
Tomlinson, and Stephen Teller, a law
Washington, for assistance with the Se
Tom Meehan, Court Administrator Sue Evan, and Eric Ratner for the Tucson phase of the study; and to Margaret Ellis, a University of Texas at Dallas Political Economy graduate student, for coding and computer analysis of data.

WINTER 1994 193

Experts' Opinions on Their Treatment by Lawyers

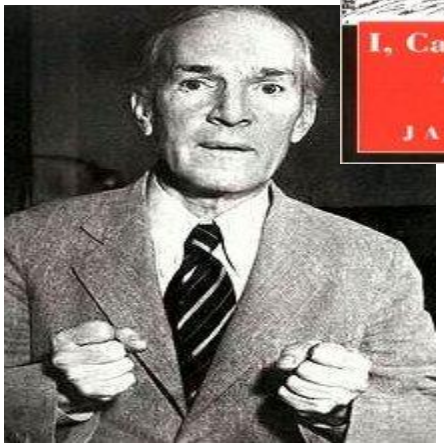
Treatment	Number
Lawyers manipulate their experts to weaken unfavorable testimony and strengthen favorable testimony	63 (77%)

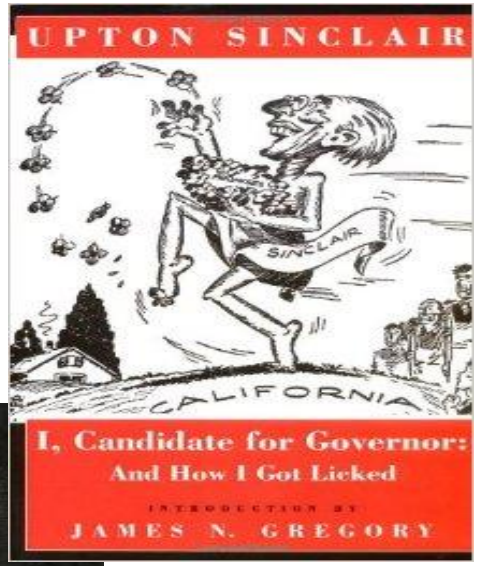
Shuman (1994). An empirical examination of the use of expert witnesses. *Jurimetrics, 193: 201*.

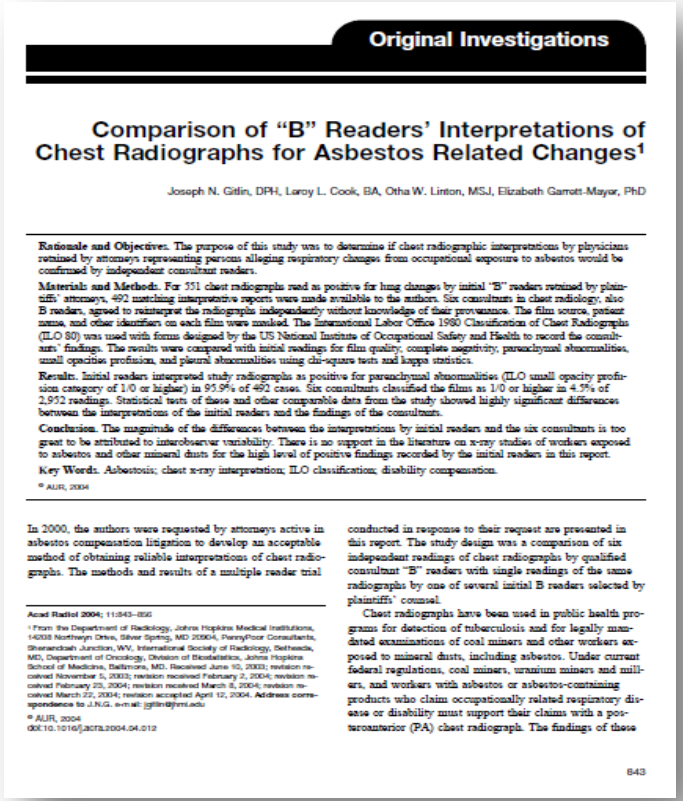
Compensation Bias

"It is difficult to get a man to understand something when his salary depends upon him not understanding it!"

-Upton Sinclair







Original Investigations

Comparison of "B" Readers' Interpretations of Chest Radiographs for Asbestos Related Changes[1]

Joseph N. Gitlin, DPH, Leroy L. Cook, BA, Otha W. Linton, MSJ, Elizabeth Garrott-Mayer, PhD

Rationale and Objectives. The purpose of this study was to determine if chest radiographic interpretations by physicians retained by attorneys representing persons alleging respiratory changes from occupational exposure to asbestos would be confirmed by independent consultant readers.

Materials and Methods. For 351 chest radiographs read as positive for lung changes by initial "B" readers retained by plaintiffs' attorneys, 492 matching interpretative reports were made available to the authors. Six consultants in chest radiology, also B readers, agreed to reinterpret the radiographs independently without knowledge of their provenance. The film source, patient name, and other identifiers on each film were masked. The International Labor Office 1980 Classification of Chest Radiographs (ILO 80) was used with forms designed by the US National Institute of Occupational Safety and Health to record the consultants' findings. The results were compared with initial readings for film quality, complete negativity, parenchymal abnormalities, small opacities profusion, and pleural abnormalities using chi-square tests and kappa statistics.

Results. Initial readers interpreted study radiographs as positive for parenchymal abnormalities (ILO small opacity profusion category of 1/0 or higher) in 95.9% of 492 cases. Six consultants classified the films as 1/0 or higher in 4.5% of 2,952 readings. Statistical tests of these and other comparable data from the study showed highly significant differences between the interpretations of the initial readers and the findings of the consultants.

Conclusion. The magnitude of the differences between the interpretations by initial readers and the six consultants is too great to be attributed to interobserver variability. There is no support in the literature on x-ray studies of workers exposed to asbestos and other mineral dusts for the high level of positive findings recorded by the initial readers in this report.

Key Words. Asbestosis; chest x-ray interpretation; ILO classification; disability compensation.

© AUR, 2004

In 2000, the authors were requested by attorneys active in asbestos compensation litigation to develop an acceptable method of obtaining reliable interpretations of chest radiographs. The methods and results of a multiple reader trial conducted in response to their request are presented in this report. The study design was a comparison of six independent readings of chest radiographs by qualified consultant "B" readers with single readings of the same radiographs by one of several initial B readers selected by plaintiffs' counsel.

Chest radiographs have been used in public health programs for detection of tuberculosis and for legally mandated examinations of coal miners and other workers exposed to mineral dusts, including asbestos. Under current federal regulations, coal miners, uranium miners and millers, and workers with asbestos or asbestos-containing products who claim occupationally related respiratory disease or disability must support their claims with a posteroanterior (PA) chest radiograph. The findings of these

Acad Radiol 2004; 11:843–856

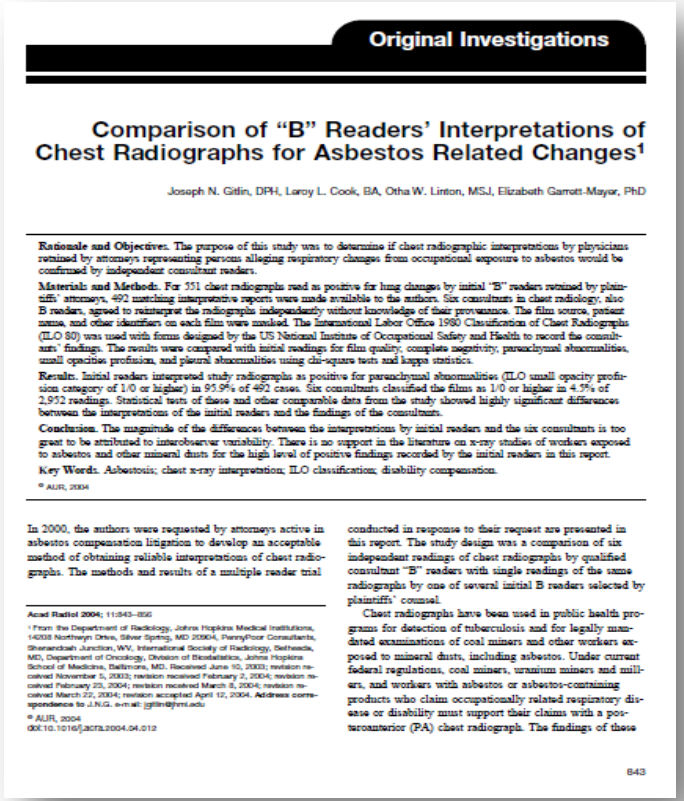
[1] From the Department of Radiology, Johns Hopkins Medical Institutions, 14208 Northwyn Drive, Silver Spring, MD 20904, PennyPoor Consultants, Shenandoah Junction, WV, International Society of Radiology, Bethesda, MD, Department of Oncology, Division of Biostatistics, Johns Hopkins School of Medicine, Baltimore, MD. Received June 10, 2003; revision received November 5, 2003; revision received February 2, 2004; revision received February 23, 2004; revision received March 8, 2004; revision received March 22, 2004; revision accepted April 12, 2004. Address correspondence to J.N.G. e-mail: jgitlin@jhmi.edu

© AUR, 2004
DOI:10.1016/j.acra.2004.04.012

Real Life Experiment:

Have plaintiff's expert witnesses review about 600 x-rays to determine whether they had abnormalities (asbestosis), then have "independent" experts review the same files.

Gitlin et al. (2004). Comparison of "B" readers' interpretations of chest radiographs. *Acad. Radiol.*, *11*(8): 843-856.



Original Investigations

Comparison of "B" Readers' Interpretations of Chest Radiographs for Asbestos Related Changes[1]

Joseph N. Gitlin, DPH, Leroy L. Cook, BA, Otha W. Linton, MSJ, Elizabeth Garrott-Mayer, PhD

Rationale and Objectives. The purpose of this study was to determine if chest radiographic interpretations by physicians retained by attorneys representing persons alleging respiratory changes from occupational exposure to asbestos would be confirmed by independent consultant readers.

Materials and Methods. For 351 chest radiographs read as positive for lung changes by initial "B" readers retained by plaintiffs' attorneys, 492 matching interpretative reports were made available to the authors. Six consultants in chest radiology, also B readers, agreed to reinterpret the radiographs independently without knowledge of their provenance. The film source, patient name, and other identifiers on each film were masked. The International Labor Office 1980 Classification of Chest Radiographs (ILO 80) was used with forms designed by the US National Institute of Occupational Safety and Health to record the consultants' findings. The results were compared with initial readings for film quality, complete negativity, parenchymal abnormalities, small opacities profusion, and pleural abnormalities using chi-square tests and kappa statistics.

Results. Initial readers interpreted study radiographs as positive for parenchymal abnormalities (ILO small opacity profusion category of 1/0 or higher) in 95.9% of 492 cases. Six consultants classified the films as 1/0 or higher in 4.5% of 2,932 readings. Statistical tests of these and other comparable data from the study showed highly significant differences between the interpretations of the initial readers and the findings of the consultants.

Conclusion. The magnitude of the differences between the interpretations by initial readers and the six consultants is too great to be attributed to interobserver variability. There is no support in the literature on x-ray studies of workers exposed to asbestos and other mineral dusts for the high level of positive findings recorded by the initial readers in this report.

Key Words. Asbestosis; chest x-ray interpretation; ILO classification; disability compensation.

© AUR, 2004

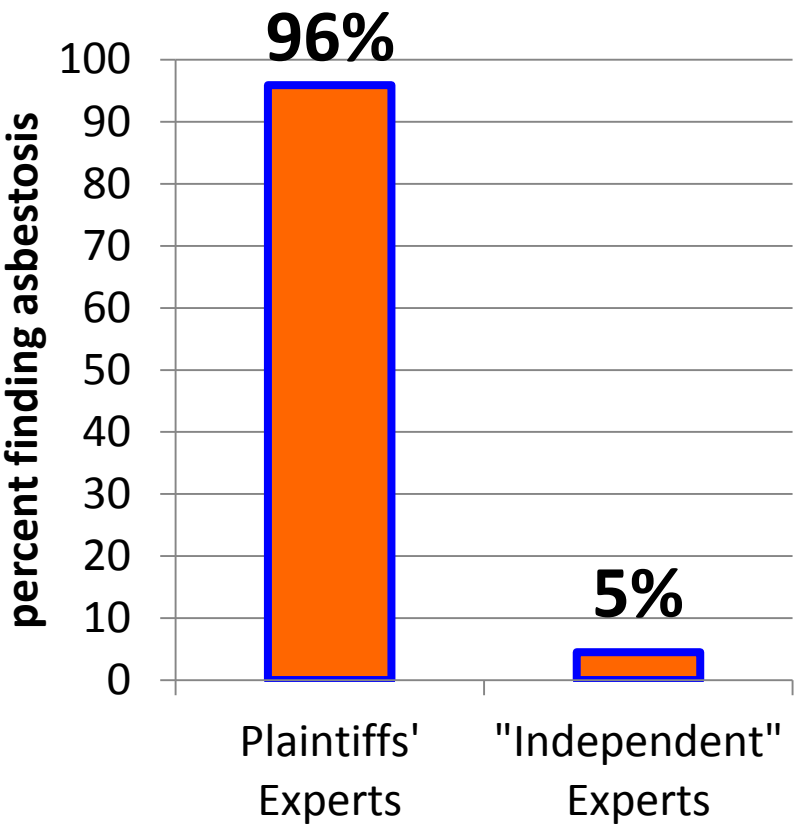
In 2000, the authors were requested by attorneys active in asbestos compensation litigation to develop an acceptable method of obtaining reliable interpretations of chest radiographs. The methods and results of a multiple reader trial conducted in response to their request are presented in this report. The study design was a comparison of six independent readings of chest radiographs by qualified consultant "B" readers with single readings of the same radiographs by one of several initial B readers selected by plaintiffs' counsel.

Chest radiographs have been used in public health programs for detection of tuberculosis and for legally mandated examinations of coal miners and other workers exposed to mineral dusts, including asbestos. Under current federal regulations, coal miners, uranium miners and millers, and workers with asbestos or asbestos-containing products who claim occupationally related respiratory disease or disability must support their claims with a posteroanterior (PA) chest radiograph. The findings of these

Acad Radiol 2004; 11:843–856

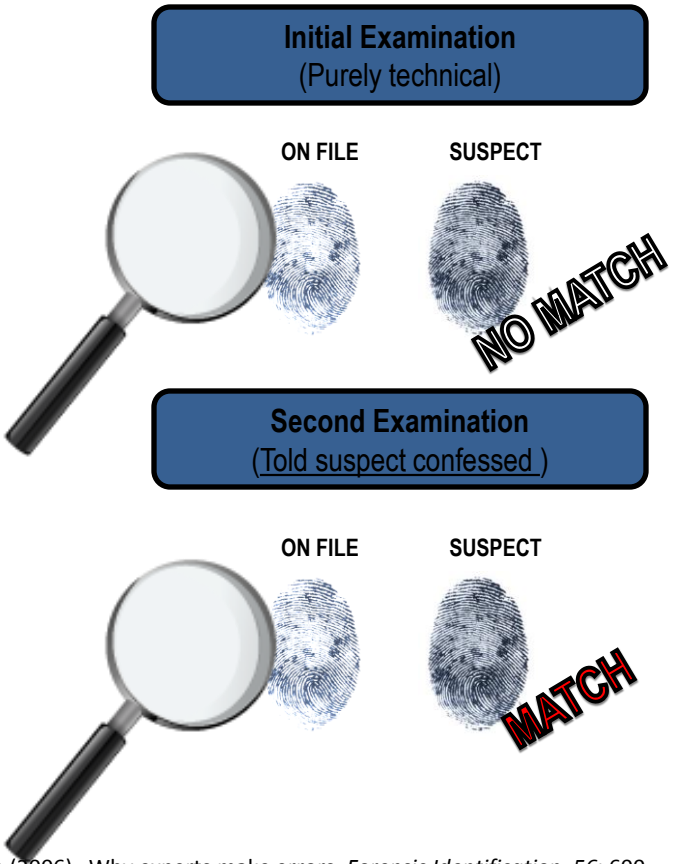
[1] From the Department of Radiology, Johns Hopkins Medical Institutions, 14208 Northwyn Drive, Silver Spring, MD 20904, PennyPoor Consultants, Shenandoah Junction, WV, International Society of Radiology, Bethesda, MD, Department of Oncology, Division of Biostatistics, Johns Hopkins School of Medicine, Baltimore, MD. Received June 10, 2003; revision received November 5, 2003; revision received February 2, 2004; revision received February 23, 2004; revision received March 8, 2004; revision received March 22, 2004; revision accepted April 12, 2004. Address correspondence to J.N.G. e-mail: jgitlin@jhmi.edu

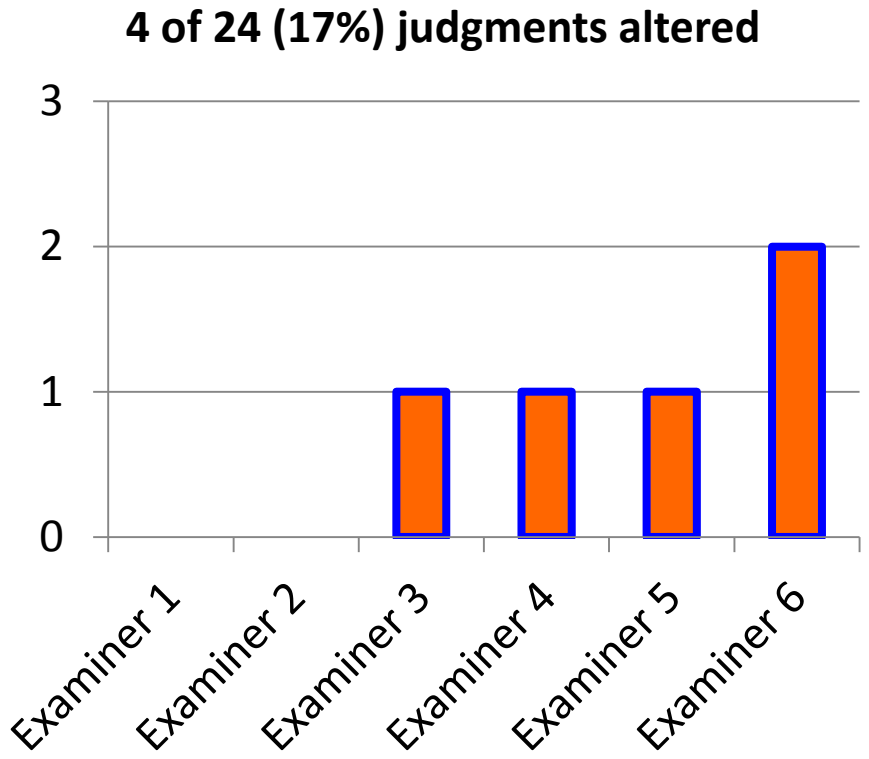
DOI:10.1016/j.acra.2004.04.012



Gitlin et al. (2004). Comparison of "B" readers' interpretations of chest radiographs. *Acad. Radiol.*, *11*(8): 843-856.

Affiliation & Hindsight





Dror & Charlton (2006). Why experts make errors. *Forensic Identification*, *56*: 600.

Law and Human Behavior, Vol. 19, No. 1, 1995

Ex Post ≠ Ex Ante

Determining Liability in Hindsight*

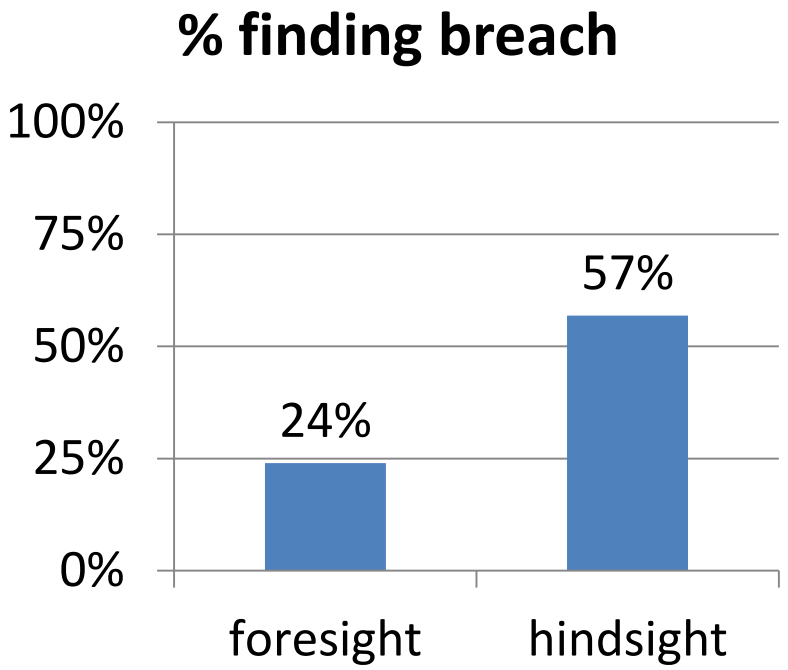
Kim A. Kamin† and Jeffrey J. Rachlinski‡

Participants in three conditions (foresight, hindsight, and a modified hindsight condition designed to ameliorate the hindsight effect) assessed whether a municipality should take, or have taken, precautions to protect a riparian property owner from flood damage. In the foresight condition, participants reviewed evidence in the context of an administrative hearing. Hindsight participants reviewed parallel materials in the context of a trial. Three quarters of the participants in foresight concluded that a flood was too unlikely to justify further precautions—a decision that a majority of the participants in hindsight found to be negligent. Participants in hindsight also gave higher estimates for the probability of the disaster occurring. The debiasing procedure failed to produce any significant differences from the regular hindsight condition. The results suggest that absent an effective debiasing technique, risk assessments made in foresight will be judged harshly in hindsight.

Life involves risk and danger. The potential for accidental harm looms in every environment and situation. When careless conduct causes an accident, injuring people or damaging property, the American tort system obliges a party who has negligently caused damage to pay for it. The tort system recognizes that not every accident is the product of negligence. To obtain compensation, a plaintiff suing for negligence must prove four things: (1) The defendant owed a duty of care to the plaintiff; (2) the duty was breached; (3) the breach caused (4) damage to the plaintiff (American Law Institute [ALI], 1965, p. 4). Negligence law requires that

* The authors gratefully acknowledge the support and advice of David L. Rosenhan and Barbara Tversky. Comments by Derek Koehler and three anonymous reviewers greatly improved earlier drafts. The assistance of Steve Cole, Sonja Lyubomirsky, Phoebe Garfield, and Garner Weng was appreciated. Correspondence and requests for reprints should be addressed to Jeffrey Rachlinski, Cornell Law School, Myron Taylor Hall, Ithaca, NY 14853-4901.
† Stanford University.
‡ Cornell Law School.

Hindsight Bias



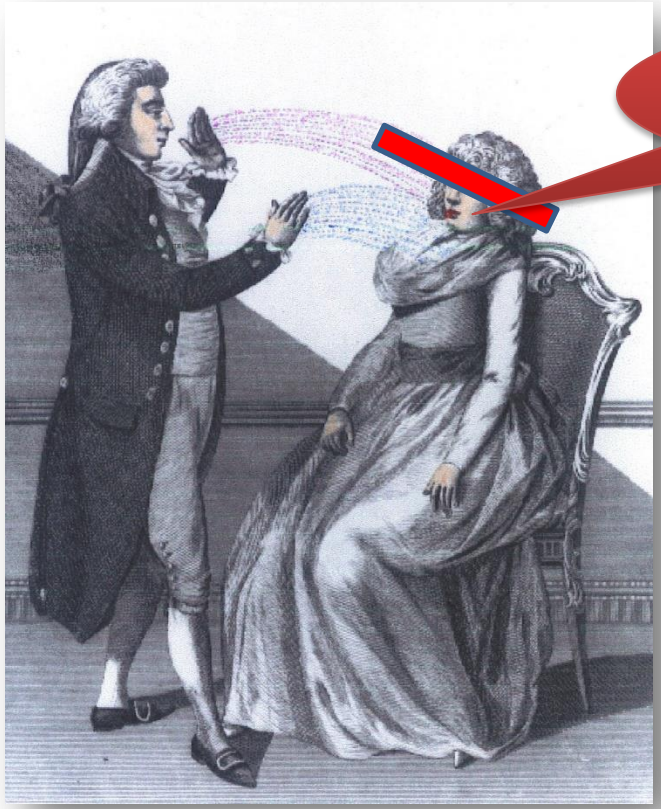
Blinding in

MEDICINE

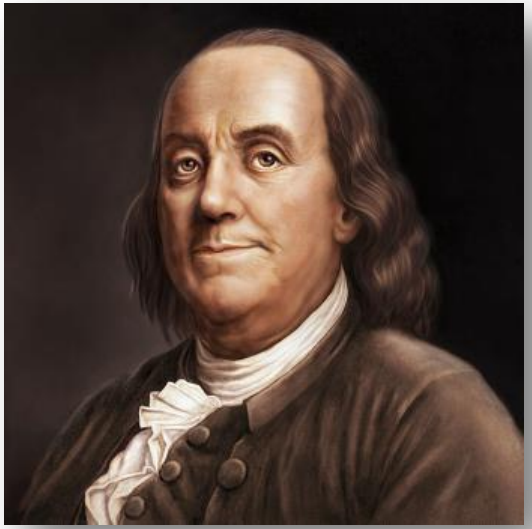
“Animal Magnetism”

Yes! I’m healed!

“Animal Magnetism”



Huh?



Original Investigation

Clinical Trial Evidence Supporting FDA Approval of Novel Therapeutic Agents, 2005-2012

Nicholas S. Downing, AB; Jenerius A. Aminawung, MD, MPH; Nilay D. Shah, PhD; Harlan M. Krumholz, MD, SM; Joseph S. Ross, MD, MHS

IMPORTANCE Many patients and physicians assume that the safety and effectiveness of newly approved therapeutic agents is well understood; however, the strength of the clinical trial evidence supporting approval decisions by the US Food and Drug Administration (FDA) has not been evaluated.

OBJECTIVES To characterize pivotal efficacy trials (clinical trials that serve as the basis of FDA approval) for newly approved novel therapeutic agents.

DESIGN AND SETTING Cross-sectional analysis using publicly available FDA documents for all novel therapeutic agents approved between 2005 and 2012.

MAIN OUTCOMES AND MEASURES Pivotal efficacy trials were classified according to the following design features: randomization, blinding, comparator, and trial end point. Surrogate outcomes were defined as any end point using a biomarker expected to predict clinical benefit. The number of patients, trial duration, and trial completion rates were also determined.

RESULTS Between 2005 and 2012, the FDA approved 188 novel therapeutic agents for 206 indications on the basis of 448 pivotal efficacy trials. The median number of pivotal trials per indication was 2 (interquartile range, 1-2.5), although 74 indications (36.8%) were approved on the basis of a single pivotal trial. Nearly all trials were randomized (89.3% [95% CI, 86.4%-92.2%]), double-blinded (79.5% [95% CI, 75.7%-83.2%]), and used either an active or placebo comparator (87.1% [95% CI, 83.9%-90.2%]). The median number of patients enrolled per indication among all pivotal trials was 760 (interquartile range, 270-1550). At least 1 pivotal trial with a duration of 6 months or greater supported the approval of 68 indications (33.8% [95% CI, 27.2%-40.4%]). Pivotal trials using surrogate end points as their primary outcome formed the exclusive basis of approval for 91 indications (45.3% [95% CI, 38.3%-52.2%]); clinical outcomes for 67 (33.3% [95% CI, 26.8%-39.9%]), and clinical scales for 26 (17.9% [95% CI, 12.6%-23.3%]). Trial features differed by therapeutic and indication characteristics, such as therapeutic area, expected length of treatment, orphan status, and accelerated approval.

CONCLUSIONS AND RELEVANCE The quality of clinical trial evidence used by the FDA as the basis for recent approvals of novel therapeutic agents varied widely across indications. This variation has important implications for patients and physicians as they make decisions about the use of newly approved therapeutic agents.

Editorial page 361

Author Video Interview at jama.com

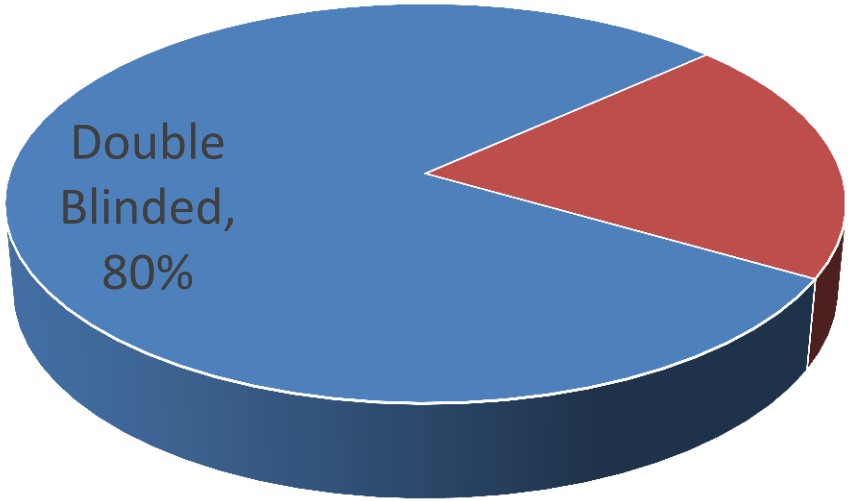
Related articles pages 378 and 385

Supplemental content at jama.com

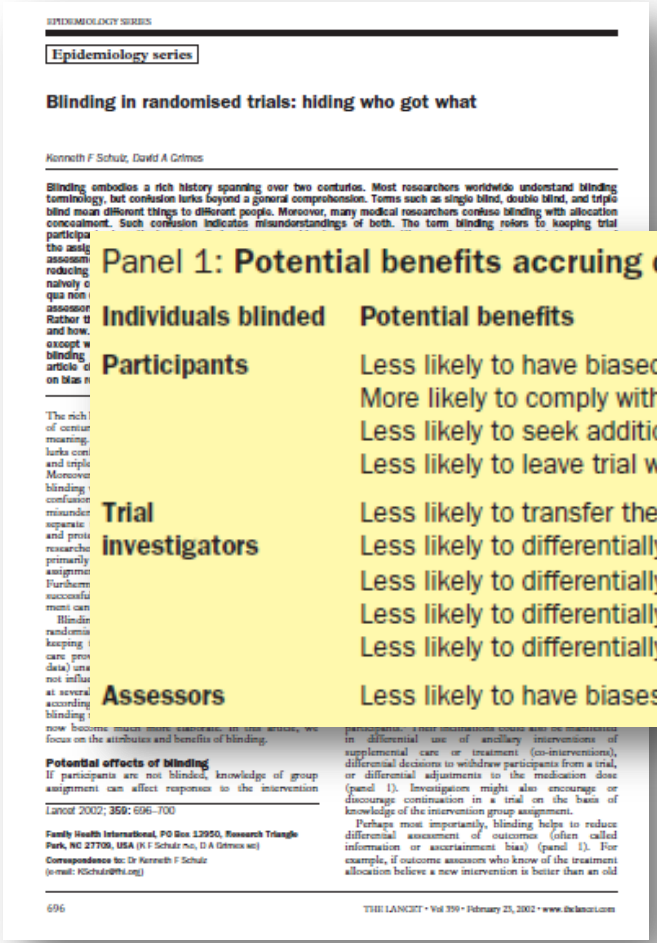
Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Joseph S. Ross, MD, MHS, Section of General Internal Medicine, Yale University School of Medicine, PO Box 208093, New Haven, CT 06520 (joseph.ross@yale.edu).

JAMA. 2014;311(4):368-377. doi:10.1001/jama.2013.282034



What biases does blinding prevent?



EPIDEMIOLOGY SERIES

Epidemiology series

Blinding in randomised trials: hiding who got what

Kenneth F Schulz, David A Grimes

Blinding embodies a rich history spanning over two centuries. Most researchers worldwide understand blinding terminology, but confusion lurks beyond a general comprehension. Terms such as single blind, double blind, and triple blind mean different things to different people. Moreover, many medical researchers confuse blinding with allocation concealment. Such confusion indicates misunderstandings of both. The term blinding refers to keeping trial

Potential effects of blinding

If participants are not blinded, knowledge of group assignment can affect responses to the intervention

Lancet 2002; 359: 696–700

Family Health International, PO Box 12950, Research Triangle Park, NC 27709, USA (K F Schulz PhD, D A Grimes MD)

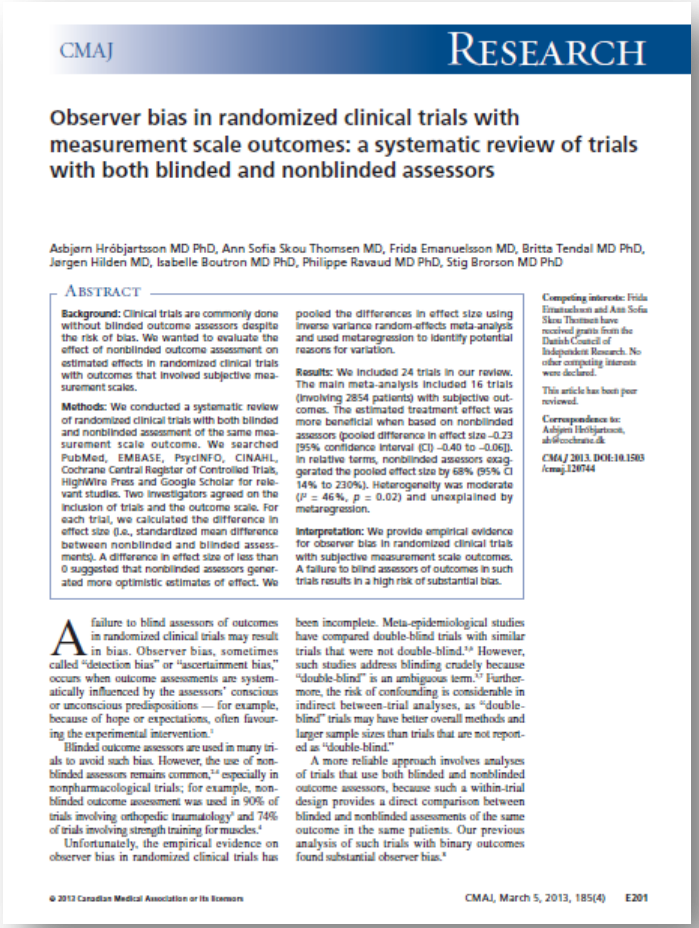
Correspondence to: Dr Kenneth F Schulz (e-mail: KSchulz@fhi.org)

differential use of ancillary interventions of supplemental care or treatment (co-interventions), differential decisions to withdraw participants from a trial, or differential adjustments to the medication dose (panel 1). Investigators might also encourage or discourage continuation in a trial on the basis of knowledge of the intervention group assignment.

Perhaps most importantly, blinding helps to reduce differential assessment of outcomes (often called information or ascertainment bias) (panel 1). For example, if outcome assessors who know of the treatment allocation believe a new intervention is better than an old

Panel 1: **Potential benefits accruing dependent on those individuals successfully blinded**

Individuals blinded	Potential benefits
Participants	Less likely to have biased psychological or physical responses to intervention More likely to comply with trial regimens Less likely to seek additional adjunct interventions Less likely to leave trial without providing outcome data, leading to lost to follow-up
Trial investigators	Less likely to transfer their inclinations or attitudes to participants Less likely to differentially administer co-interventions Less likely to differentially adjust dose Less likely to differentially withdraw participants Less likely to differentially encourage or discourage participants to continue trial
Assessors	Less likely to have biases affect their outcome assessments, especially with subjective outcomes of interest



CMAJ RESEARCH

Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors

Asbjørn Hróbjartsson MD PhD, Ann Sofia Skou Thomsen MD, Frida Emanuelsson MD, Britta Tendal MD PhD, Jørgen Hilden MD, Isabelle Boutron MD PhD, Philippe Ravaud MD PhD, Stig Brorson MD PhD

ABSTRACT

Background: Clinical trials are commonly done without blinded outcome assessors despite the risk of bias. We wanted to evaluate the effect of nonblinded outcome assessment on estimated effects in randomized clinical trials with outcomes that involved subjective measurement scales.

Methods: We conducted a systematic review of randomized clinical trials with both blinded and nonblinded assessment of the same measurement scale outcome. We searched PubMed, EMBASE, PsycINFO, CINAHL, Cochrane Central Register of Controlled Trials, HighWire Press and Google Scholar for relevant studies. Two investigators agreed on the inclusion of trials and the outcome scale. For each trial, we calculated the difference in effect size (i.e., standardized mean difference between nonblinded and blinded assessments). A difference in effect size of less than 0 suggested that nonblinded assessors generated more optimistic estimates of effect. We pooled the differences in effect size using inverse variance random-effects meta-analysis and used metaregression to identify potential reasons for variation.

Results: We included 24 trials in our review. The main meta-analysis included 16 trials (involving 2854 patients) with subjective outcomes. The estimated treatment effect was more beneficial when based on nonblinded assessors (pooled difference in effect size −0.23 [95% confidence interval (CI) −0.40 to −0.06]). In relative terms, nonblinded assessors exaggerated the pooled effect size by 68% (95% CI 14% to 230%). Heterogeneity was moderate (I^2 = 46%, p = 0.02) and unexplained by metaregression.

Interpretation: We provide empirical evidence for observer bias in randomized clinical trials with subjective measurement scale outcomes. A failure to blind assessors of outcomes in such trials results in a high risk of substantial bias.

Competing interests: Frida Emanuelsson and Ann Sofia Skou Thomsen have received grants from the Danish Council of Independent Research. No other competing interests were declared.

This article has been peer reviewed.

Correspondence to: Asbjørn Hróbjartsson, ah@cochrane.dk

CMAJ 2013. DOI:10.1503/cmaj.120744

A failure to blind assessors of outcomes in randomized clinical trials may result in bias. Observer bias, sometimes called "detection bias" or "ascertainment bias," occurs when outcome assessments are systematically influenced by the assessors' conscious or unconscious predispositions — for example, because of hope or expectations, often favouring the experimental intervention.[1]

Blinded outcome assessors are used in many trials to avoid such bias. However, the use of nonblinded assessors remains common,[2-4] especially in nonpharmacological trials; for example, nonblinded outcome assessment was used in 90% of trials involving orthopedic traumatology[3] and 74% of trials involving strength training for muscles.[4]

Unfortunately, the empirical evidence on observer bias in randomized clinical trials has been incomplete. Meta-epidemiological studies have compared double-blind trials with similar trials that were not double-blind.[5,6] However, such studies address blinding crudely because "double-blind" is an ambiguous term.[7,8] Furthermore, the risk of confounding is considerable in indirect between-trial analyses, as "double-blind" trials may have better overall methods and larger sample sizes than trials that are not reported as "double-blind."

A more reliable approach involves analyses of trials that use both blinded and nonblinded outcome assessors, because such a within-trial design provides a direct comparison between blinded and nonblinded assessments of the same outcome in the same patients. Our previous analysis of such trials with binary outcomes found substantial observer bias.[9]

Hróbjartsson 2013

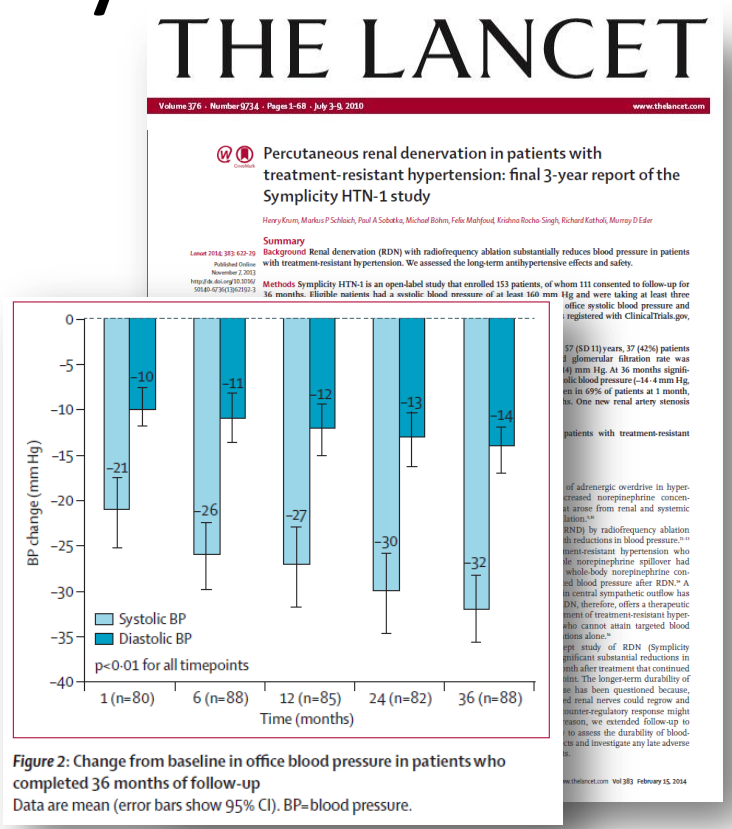
- Systematic review of 24 studies
- "nonblinded assessors exaggerated the … effect size by 68%."

“Symplicity”

“No clinical advancement has excited the hypertension community … as much as renal nerve ablation via a percutaneous technique.” (Luft 2014)

“Symplicity”

- Open-Label Experiment, 92% report benefit
- Open-Label Experiment, 84% report benefit
- Is blinding possible?
- FDA: “try it.”



THE LANCET

Volume 376 · Number 9734 · Pages 1–68 · July 3–9, 2010 www.thelancet.com

Percutaneous renal denervation in patients with treatment-resistant hypertension: final 3-year report of the Symplicity HTN-1 study

Henry Krum, Markus P Schlaich, Paul A Sobotka, Michael Böhm, Felix Mahfoud, Krishna Rocha-Singh, Richard Katholi, Murray D Esler

Summary

Background Renal denervation (RDN) with radiofrequency ablation substantially reduces blood pressure in patients with treatment-resistant hypertension. We assessed the long-term antihypertensive effects and safety.

Methods Symplicity HTN-1 is an open-label study that enrolled 153 patients, of whom 111 consented to follow-up for 36 months. Eligible patients had a systolic blood pressure of at least 160 mm Hg and were taking at least three

Figure 2: Change from baseline in office blood pressure in patients who completed 36 months of follow-up
Data are mean (error bars show 95% CI). BP=blood pressure.

A Blinded Test

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

A Controlled Trial of Renal Denervation for Resistant Hypertension

Deepak L. Bhatt, M.D., M.P.H., David E. Kandzari, M.D., William W. O'Neill, M.D., Ralph D'Agostino, Ph.D., John M. Flack, M.D., M.P.H., Barry T. Katzen, M.D., Martin B. Leon, M.D., Minglei Liu, Ph.D., Laura Mauri, M.D., Manuela Negoita, M.D., Sidney A. Cohen, M.D., Ph.D., Suzanne Oparil, M.D., Krishna Rocha-Singh, M.D., Raymond R. Townsend, M.D., and George L. Bakris, M.D., for the SYMPLICITY HTN-3 Investigators*

ABSTRACT

BACKGROUND

Prior unblinded studies have suggested that catheter-based renal-artery denervation reduces blood pressure in patients with resistant hypertension.

METHODS

We designed a prospective, single-blind, randomized, sham-controlled trial. Patients with severe resistant hypertension were randomly assigned in a 2:1 ratio to undergo renal denervation or a sham procedure. Before randomization, patients were receiving a stable antihypertensive regimen involving maximally tolerated doses of at least three drugs, including a diuretic. The primary efficacy end point was the change in office systolic blood pressure at 6 months; a secondary efficacy end point was the change in mean 24-hour ambulatory systolic blood pressure. The primary safety end point was a composite of death, end-stage renal disease, embolic events resulting in end-organ damage, renovascular complications, or hypertensive crisis at 1 month or new renal-artery stenosis of more than 70% at 6 months.

RESULTS

A total of 535 patients underwent randomization. The mean (±SD) change in systolic blood pressure at 6 months was −14.13±23.93 mm Hg in the denervation group as compared with −11.74±25.94 mm Hg in the sham-procedure group ($P<0.001$ for both comparisons of the change from baseline), for a difference of −2.39 mm Hg (95% confidence interval [CI], −6.89 to 2.12; $P=0.26$ for superiority with a margin of 5 mm Hg). The change in 24-hour ambulatory systolic blood pressure was −6.75±15.11 mm Hg in the denervation group and −4.79±17.25 mm Hg in the sham-procedure group, for a difference of −1.96 mm Hg (95% CI, −4.97 to 1.06; $P=0.98$ for superiority with a margin of 2 mm Hg). There were no significant differences in safety between the two groups.

CONCLUSIONS

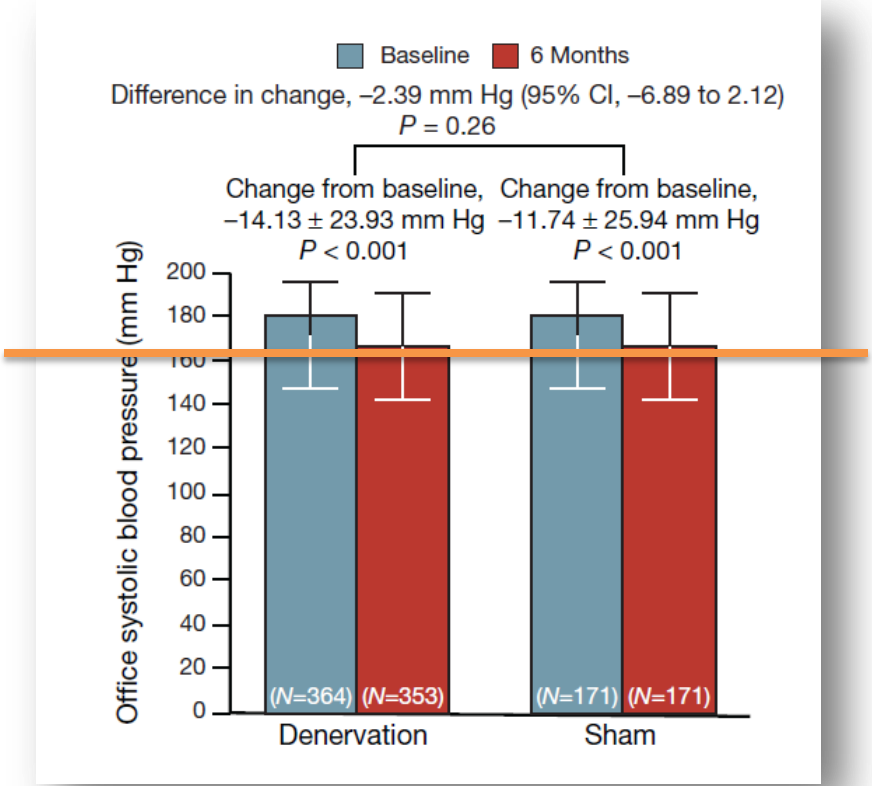
This blinded trial did not show a significant reduction of systolic blood pressure in patients with resistant hypertension 6 months after renal-artery denervation as compared with a sham control. (Funded by Medtronic; SYMPLICITY HTN-3 ClinicalTrials.gov number, NCT01418261.)

From Brigham and Women's Hospital Heart and Vascular Center and Harvard Medical School (D.L.B., L.M.), Boston University School of Public Health (R.D.), and Harvard Clinical Research Institute (R.D., L.M.) — all in Boston; Piedmont Heart Institute, Atlanta (D.E.K.); the Division of Cardiology, Henry Ford Hospital (W.W.O.), and Wayne State University and the Detroit Medical Center (J.M.F.) — all in Detroit; Baptist Cardiac and Vascular Institute, Miami (B.T.K.); New York Presbyterian Hospital, Columbia University Medical Center, and Cardiovascular Research Foundation, New York (M.B.L.); Medtronic CardioVascular, Santa Rosa, CA (M.L., M.N., S.A.C.); University of Alabama at Birmingham, Birmingham (S.O.); Prairie Heart Institute, Springfield, IL (K.R.-S.); Perelman School of Medicine, University of Pennsylvania, Philadelphia (S.A.C., R.R.T.); and University of Chicago Medicine, Chicago (G.L.B.). Address reprint requests to Dr. Bhatt at Brigham and Women's Hospital Heart and Vascular Center, 75 Francis St., Boston, MA 02115, or at dlbhattmd@post.harvard.edu.

*A complete list of investigators in the SYMPLICITY HTN-3 trial is provided in the Supplementary Appendix, available at NEJM.org.

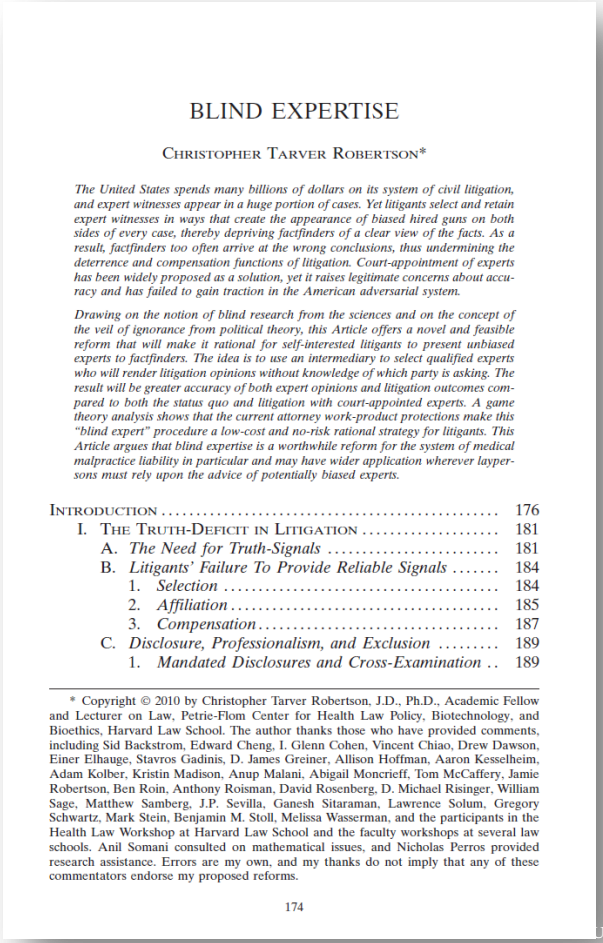
This article was published on March 29, 2014, at NEJM.org.

DOI: 10.1056/NEJMoa1402670



Blinding in

LITIGATION



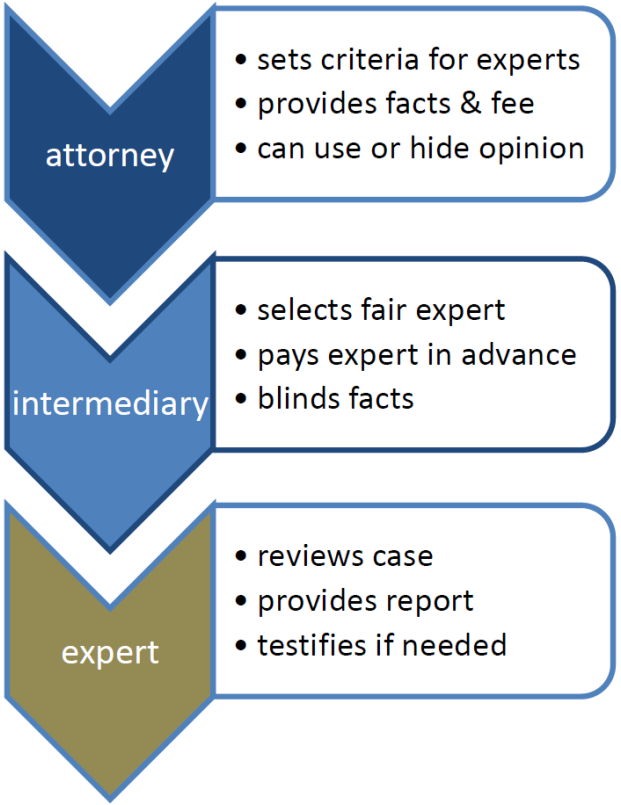
BLIND EXPERTISE

CHRISTOPHER TARVER ROBERTSON*

The United States spends many billions of dollars on its system of civil litigation, and expert witnesses appear in a huge portion of cases. Yet litigants select and retain expert witnesses in ways that create the appearance of biased hired guns on both sides of every case, thereby depriving factfinders of a clear view of the facts. As a result, factfinders too often arrive at the wrong conclusions, thus undermining the deterrence and compensation functions of litigation. Court-appointment of experts has been widely proposed as a solution, yet it raises legitimate concerns about accuracy and has failed to gain traction in the American adversarial system.

Drawing on the notion of blind research from the sciences and on the concept of the veil of ignorance from political theory, this Article offers a novel and feasible reform that will make it rational for self-interested litigants to present unbiased experts to factfinders. The idea is to use an intermediary to select qualified experts who will render litigation opinions without knowledge of which party is asking. The result will be greater accuracy of both expert opinions and litigation outcomes compared to both the status quo and litigation with court-appointed experts. A game theory analysis shows that the current attorney work-product protections make this "blind expert" procedure a low-cost and no-risk rational strategy for litigants. This Article argues that blind expertise is a worthwhile reform for the system of medical malpractice liability in particular and may have wider application wherever laypersons must rely upon the advice of potentially biased experts.

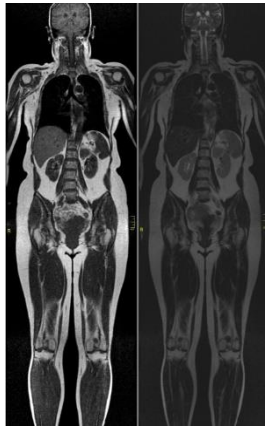
 The author thanks those who have provided comments, including Sid Backstrom, Edward Cheng, I. Glenn Cohen, Vincent Chiao, Drew Dawson, Einer Elhauge, Stavros Gadinis, D. James Greiner, Allison Hoffman, Aaron Kesselheim, Adam Kolber, Kristin Madison, Anup Malani, Abigail Moncrieff, Tom McCaffery, Jamie Robertson, Ben Roin, Anthony Roisman, David Rosenberg, D. Michael Risinger, William Sage, Matthew Samberg, J.P. Sevilla, Ganesh Sitaraman, Lawrence Solum, Gregory Schwartz, Mark Stein, Benjamin M. Stoll, Melissa Wasserman, and the participants in the Health Law Workshop at Harvard Law School and the faculty workshops at several law schools. Anil Somani consulted on mathematical issues, and Nicholas Perros provided research assistance. Errors are my own, and my thanks do not imply that any of these commentators endorse my proposed reforms.

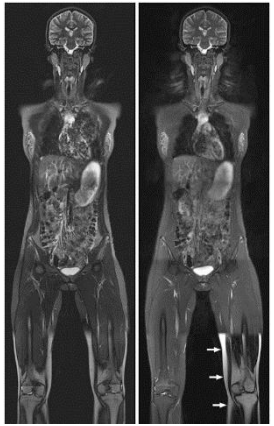


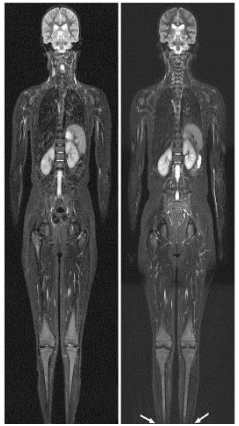
Solves Selection, Affiliation, & Compensation Biases

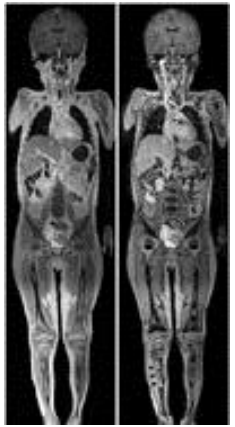
Solving Hindsight Bias

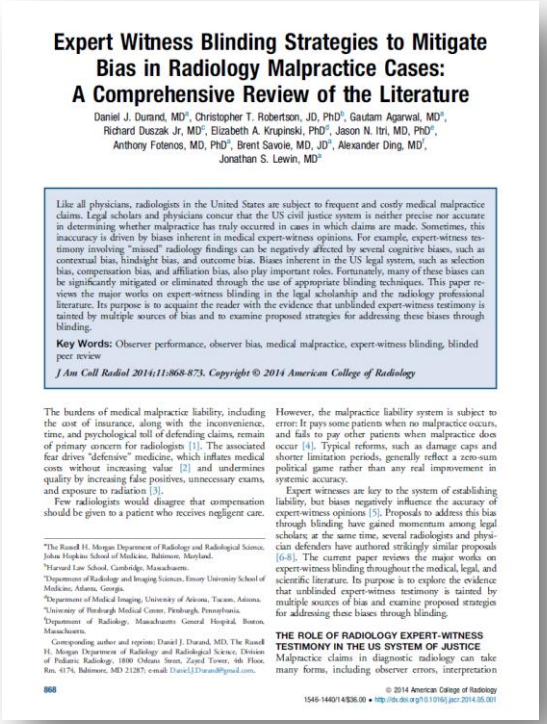
- Remove the outcome data
- Obscure the litigation question











Expert Witness Blinding Strategies to Mitigate Bias in Radiology Malpractice Cases: A Comprehensive Review of the Literature

Daniel J. Durand, MD[a], Christopher T. Robertson, JD, PhD[b], Gautam Agarwal, MD[a], Richard Duszak Jr, MD[c], Elizabeth A. Krupinski, PhD[d], Jason N. Itri, MD, PhD[e], Anthony Fotenos, MD, PhD[f], Brent Savoie, MD, JD[g], Alexander Ding, MD[f], Jonathan S. Lewin, MD[h]

Like all physicians, radiologists in the United States are subject to frequent and costly medical malpractice claims. Legal scholars and physicians concur that the US civil justice system is neither precise nor accurate in determining whether malpractice has truly occurred in cases in which claims are made. Sometimes, this inaccuracy is driven by biases inherent in medical expert-witness opinions. For example, expert-witness testimony involving "missed" radiology findings can be negatively affected by several cognitive biases, such as contextual bias, hindsight bias, and outcome bias. Biases inherent in the US legal system, such as selection bias, compensation bias, and affiliation bias, also play important roles. Fortunately, many of these biases can be significantly mitigated or eliminated through the use of appropriate blinding techniques. This paper reviews the major works on expert-witness blinding in the legal scholarship and the radiology professional literature. Its purpose is to acquaint the reader with the evidence that unblinded expert-witness testimony is tainted by multiple sources of bias and to examine proposed strategies for addressing these biases through blinding.

Key Words: Observer performance, observer bias, medical malpractice, expert-witness blinding, blinded peer review

J Am Coll Radiol 2014;11:868-873.

The burdens of medical malpractice liability, including the cost of insurance, along with the inconvenience, time, and psychological toll of defending claims, remain of primary concern for radiologists [1]. The associated fear drives "defensive" medicine, which inflates medical costs without increasing value [2] and undermines quality by increasing false positives, unnecessary exams, and exposure to radiation [3].

Few radiologists would disagree that compensation should be given to a patient who receives negligent care. However, the malpractice liability system is subject to error: It pays some patients when no malpractice occurs, and fails to pay other patients when malpractice does occur [4]. Typical reforms, such as damage caps and shorter limitation periods, generally reflect a zero-sum political game rather than any real improvement in systemic accuracy.

Expert witnesses are key to the system of establishing liability, but biases negatively influence the accuracy of expert-witness opinions [5]. Proposals to address this bias through blinding have gained momentum among legal scholars; at the same time, several radiologists and physician defenders have authored strikingly similar proposals [6-8]. The current paper reviews the major works on expert-witness blinding throughout the medical, legal, and scientific literature. Its purpose is to explore the evidence that unblinded expert-witness testimony is tainted by multiple sources of bias and examine proposed strategies for addressing these biases through blinding.

[a]The Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins School of Medicine, Baltimore, Maryland.
[b]Harvard Law School, Cambridge, Massachusetts.
[c]Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, Georgia.
[d]Department of Medical Imaging, University of Arizona, Tucson, Arizona.
[e]University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania.
[f]Department of Radiology, Massachusetts General Hospital, Boston, Massachusetts.
Corresponding author and reprints: Daniel J. Durand, MD, The Russell H. Morgan Department of Radiology and Radiological Science, Division of Pediatric Radiology, 1800 Orleans Street, Zayed Tower, 4th Floor, Rm. 4174, Baltimore, MD 21287; e-mail: DanielJDurand@gmail.com.

THE ROLE OF RADIOLOGY EXPERT-WITNESS TESTIMONY IN THE US SYSTEM OF JUSTICE

Malpractice claims in diagnostic radiology can take many forms, including observer errors, interpretation

1546-1440/14/$36.00 ● http://dx.doi.org/10.1016/j.jacr.2014.05.001

Research Questions

- Can blinding actually be implemented in a way that removes bias?
- Can those efforts be successfully communicated to the factfinder?
  - Improve litigation outcome accuracy
  - Create an incentive for litigants to do it

JOURNAL OF EMPIRICAL LEGAL STUDIES

Journal of Empirical Legal Studies
Volume 9, Issue 4, 765–794, December 2012

The Effect of Blinded Experts on Juror Verdicts

*Christopher T. Robertson and David V. Yokum**

"Blind expertise" has been proposed as an institutional solution to the problem of bias in expert witness testimony in litigation (Robertson 2010). At the request of a litigant, an intermediary selects a qualified expert and pays the expert to review a case without knowing which side requested the opinion. This article reports an experiment that tests the hypothesis that, compared to traditional experts, such "blinded experts" will be more persuasive to jurors. A national sample of mock jurors (N= 275) watched an online video of a staged medical malpractice trial, including testimony from two medical experts, one of whom (or neither, in the control condition) was randomly assigned to be a blind expert. We also manipulated whether the judge provided a special jury instruction explaining the blinding concept. Descriptively, the data suggest juror reluctance to impose liability. Despite an experimental design that included negligent medical care, only 46 percent of the jurors found negligence in the control condition, which represents the status quo. Blind experts, testifying on either side, were perceived as significantly more credible, and were more highly persuasive, in that they doubled (or halved) the odds of a favorable verdict, and increased (or decreased) simulated damages awards by over $100,000. The increased damages award appears to be due to jurors hedging their damages awards, which interacted with the blind expert as a driver of certainty. Use of a blind expert may be a rational strategy for litigants, even without judicial intervention in the form of special jury instructions or otherwise.

I. Background

The U.S. legal system tasks judges and jurors—both laypersons as to science—with resolving highly technical questions. These laypersons are asked, for example, to evaluate DNA evidence to determine whether it inculpates a particular defendant, to determine the standard of care for lumbar radiculopathy, to interpret epidemiological data to determine whether a given chemical causes an observed disease, and to ascertain the state of the art in a patent suit for computer software. Thus, in both civil and criminal litigation, expert

*Address correspondence to Christopher T. Robertson, James E. Rogers College of Law, the University of Arizona, PO Box 210176, Tucson, AZ 85721; email: chris.robertson@law.arizona.edu. Yokum is J.D./Ph.D. candidate, James E. Rogers College of Law, Department of Psychology, The University of Arizona.

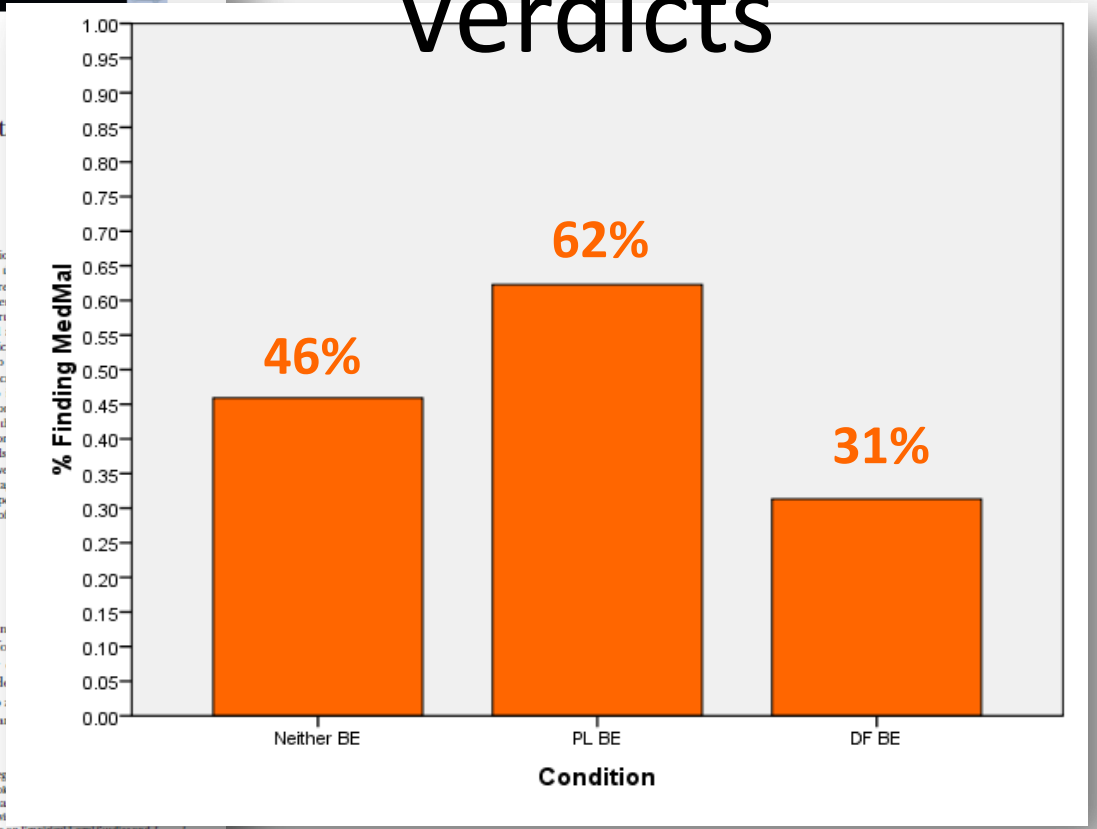
Thanks to Gregory Schwartz, Aaron Kesselheim, and Tom Maues for serving as actors and consultants; to James Greiner, George Kimball Smith, and anonymous reviewers for the Conference on Empirical Legal Studies and *Journal of Empirical Legal Studies* for commenting on drafts, and to Germar Townsend, Tess Gemberling, Carol Ward, Judy Parker, Barbara Lopez, and Ben Sky for excellent research and administrative support.

A Randomized, Controlled, Blinded Experiment

Mock Jurors

- Blind Expert for Plaintiff
- Blind Expert for Defendant
- Control

Verdicts



(*N* = 275, p=.04)

JOURNAL OF EMPIRICAL LEGAL STUDIES

Journal of Empirical Legal Studies
Volume 9, Issue 4, 765–794, December 2012

The Effect of Blinded Expert
Juror Verdicts

*Christopher T. Robertson and David V. Yokum**

"Blind expertise" has been proposed as an institutional solutic
expert witness testimony in litigation (Robertson 2010). At
intermediary selects a qualified expert and pays the expert to re
which side requested the opinion. This article reports an expe
esis that, compared to traditional experts, such "blinded expert
jurors. A national sample of mock jurors (*N*= 275) watched
medical malpractice trial, including testimony from two medic
neither, in the control condition) was randomly assigned to
manipulated whether the judge provided a special jury instruc
concept. Descriptively, the data suggest juror reluctance to
experimental design that included negligent medical care, o
found negligence in the control condition, which represents t
testifying on either side, were perceived as significantly mo
highly persuasive, in that they doubled (or halved) the odds
increased (or decreased) simulated damages awards by ov
damages award appears to be due to jurors hedging their dama
with the blind expert as a driver of certainty. Use of a blind exp
for litigants, even without judicial intervention in the form o
otherwise.

I. Background

The U.S. legal system tasks judges and jurors—both layperson
highly technical questions. These laypersons are asked, fo
evidence to determine whether it inculpates a particular
standard of care for lumbar radiculopathy, to interpret epid
whether a given chemical causes an observed disease, and to
a patent suit for computer software. Thus, in both civil a

*Address correspondence to Christopher T. Robertson, James E. Rogers Colleg
Box 210170, Tucson, AZ 85721; email: chris.robertson@law.arizona.edu. Yo
Rogers College of Law, Department of Psychology, The University of Arizona

Thanks to Gregory Schwartz, Aaron Kesselheim, and Tom Maues for servi
Greiner, George Kimball Smith, and anonymous reviewers for the Conference on Empirical Legal Studies and *Journal of Empirical Legal Studies* for commenting on drafts, and to Germar Townsend, Tess Gemberling, Carol Ward, Judy Parker, Barbara Lopez, and Ben Sky for excellent research and administrative support.

Damages

JOURNAL OF EMPIRICAL LEGAL STUDIES

Journal of Empirical Legal Studies
Volume 9, Issue 4, 765–794, December 2012

The Effect of Blinded Expert Juror Verdicts

*Christopher T. Robertson and David V. Yokum**

"Blind expertise" has been proposed as an institutional solutio expert witness testimony in litigation (Robertson 2010). At intermediary selects a qualified expert and pays the expert to re which side requested the opinion. This article reports an expe esis that, compared to traditional experts, such "blinded expert jurors. A national sample of mock jurors (N= 275) watched medical malpractice trial, including testimony from two medic neither, in the control condition) was randomly assigned to manipulated whether the judge provided a special jury instruc concept. Descriptively, the data suggest juror reluctance to experimental design that included negligent medical care, o found negligence in the control condition, which represents t testifying on either side, were perceived as significantly mo highly persuasive, in that they doubled (or halved) the odds increased (or decreased) simulated damages awards by ov damages award appears to be due to jurors hedging their dama with the blind expert as a driver of certainty. Use of a blind exp for litigants, even without judicial intervention in the form o otherwise.

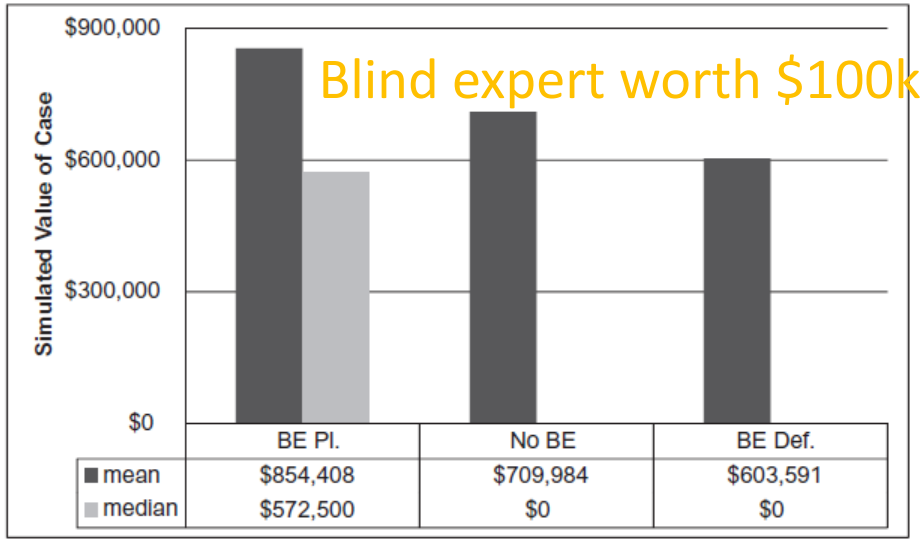
I. Background

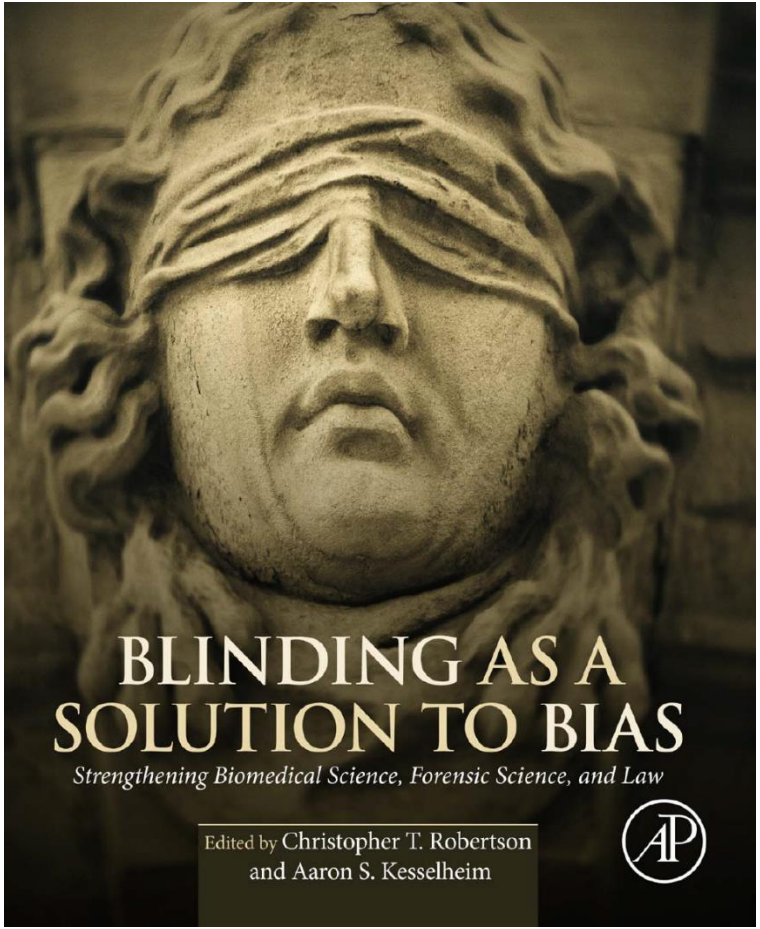
The U.S. legal system tasks judges and jurors—both layperson highly technical questions. These laypersons are asked, fo evidence to determine whether it inculpates a particular standard of care for lumbar radiculopathy, to interpret epid whether a given chemical causes an observed disease, and to a patent suit for computer software. Thus, in both civil a

*Address correspondence to Christopher T. Robertson, James E. Rogers Colleg Box 210170, Tucson, AZ 85721; email: chris.robertson@law.arizona.edu. Yo Rogers College of Law, Department of Psychology, The University of Arizona

Thanks to Gregory Schwartz, Aaron Kesselheim, and Tom Maues for servi Greiner, George Kimball Smith, and anonymous reviewers for the Conference *of Empirical Legal Studies* for commenting on drafts, and to Germar Townsen Parker, Barbara Lopez, and Ben Sky for excellent research and administrativ

Figure 1: Simulation of economic value of case (U.S. dollars) when neither side (no BE), only the plaintiff (BE Pl.), or only the defendant (BE Def.) has a blind expert, including defense verdicts as zeros. Outlier award values were transformed to within two standard deviations, and $500,000 economic damages were assumed. On these assumptions, the tactic of using a blind expert pays over $100,000 on average to the litigant that uses the tactic, conditional on the expert rendering a favorable, usable opinion not rebutted by a blind expert on the other side.





Christopher T. Robertson

ctr5@nyu.edu

THE UNIVERSITY OF ARIZONA®
James E. Rogers College of Law

NYU | LAW