

Critical Evidence for Proving Causality to Non-statisticians in Court

Palmer Morrel-Samuels
Employee Motivation and Performance Assessment
(palmer@umich.edu)

Peter D. Jacobson
School of Public Health, University of Michigan
(pdj@umich.edu)

Cristina G. Banks
Walter A. Haas School of Business, University of California
(banks@haas.berkeley.edu)

Citation: Morrel-Samuels, P., & Jacobson, P. J., Banks, Cristina, G. (2008, September). Critical Evidence for Proving Causality to Non-statisticians in Court. Paper submitted for presentation at the Third Annual Conference on Empirical Legal Studies, Ithaca, NY. SSRN manuscript ID1133516.

ABSTRACT:

No contemporary guide exists for using statistics to prove causality in court. We outline a new theory explaining comprehension of causal graphs, and claim four hallmarks of causality are critical: Association, Prediction, Exclusion of Alternative Explanations, and Dose Dependence. We test our theory in 63 smoking lawsuits, finding that movants who use all four hallmarks are significantly more likely to prevail ($p < .05$); moreover, number of hallmarks predicts likelihood of prevailing. Results also suggest courts are especially swayed by evidence excluding alternative explanations and/or demonstrating dose dependence ($p < .00001$). We close with guidelines for using causal graphs in court.

As Justice Breyer points out in the introduction to the *Reference Manual on Scientific Evidence*[1], statistical analysis is becoming increasingly important during litigation today. In several cases he said, the Supreme Court justices “placed great weight on a statistical analysis that offered a plausible alternative interpretation” of disputed material facts. Accordingly, guides for attorneys and judges are proliferating, usually at a relatively advanced skill level consistent with the demands of accommodating the Federal Rules of Evidence, *Daubert v. Merrell Dow Pharmaceuticals*[2], and related litigation. However, despite earlier work[3], no comprehensive contemporary guide exists for attorneys who want to use statistical data to create effective demonstrative evidence for non-statisticians.

At its best, statistical analysis is a uniquely effective tool for convincing the jury about a causal linkage. However, persuading non-statisticians about causality poses some distinct challenges, because lay audiences are prone to characteristic errors of judgment and inference wherever statistics is involved, as Kahneman’s research shows[4].

For example, in the conjunction fallacy the probability of a rare but salient conjunction is judged to be higher than the probability of either element alone: People consistently and mistakenly believe that the likelihood of meeting a bank teller who is a feminist is impossibly higher than the likelihood of meeting either a bank teller or a feminist[5].

In this paper we argue that attorneys need to accommodate research on the cognitive processes that non-statisticians use to make causal inferences based on statistical evidence. Note that this task is substantially different from (though it does not supplant) the need to follow the most rigorous best practices for running statistical analyses, as well as creating non-statistical demonstrative exhibits for the courtroom[6]. In a sense, we are suggesting an additional expectation for attorneys who introduce statistics in the courtroom: Statistical analyses have to be more than able to withstand the criticism of an opposing expert; they have to be persuasive in the jury box, where attention span is limited, training rudimentary, and inferential processes imperfect[7]. This brings us to a legitimate and important question: What exactly does it take to convince an ordinary juror that the quantitative evidence at hand provides sufficient proof of a causal linkage?

Although terminology differs, many aspects of causality are similar, albeit not identical, in law, science, philosophy, the behavioral sciences, and statistics. For example, the notion of a proximate (or proximal) cause – the mechanism which directly creates the end effect and is close to it in space and/or time – is important and roughly equivalent in law[8]¹, medicine, logic, and the behavioral sciences. However, some enduring differences seem to resist resolution. For example, epidemiological notions of causality are difficult to reconcile with legal notions of causality, in part because the former offers population-based evidence of a general nature, and the latter requires individual-based evidence of a specific nature. Indeed, one recent decision included the sweeping claim that “Epidemiological studies do not provide direct evidence that a particular plaintiff was injured by exposure to a substance...” a curious contention that, when carried to its logical extreme, makes an entire body of published research virtually inadmissible[9].

In general, courts have held that there are two types of causal relations—proximate (or legal) cause, and cause in fact. To determine whether the defendant’s conduct caused the plaintiff’s harm, courts use one of two tests. The first is known as the “but for” test: but for the defendant’s conduct, no harm would have occurred. The other is the “substantial factor” test: was the defendant’s conduct a substantial factor in causing the accident? As many have pointed out, however, everything is related to everything else, leading to an endless series of causal connections. As a result, courts limit on policy grounds the extent to which causation will be applied. To recover damages, the harm must be foreseeable: would a reasonable person have foreseen or anticipated that the defendant’s behavior

¹ A commonly-encountered modern citation for the definition of the term comes from *Pawsey v. Scottish Union and National Insurance Co* (1908): "Proximate cause means the active, efficient cause that sets in motion a train of events which brings about a result, without the intervention of a force started and working actively from a new and independent source."

would place others at risk of harm? If not, the defendant is not liable. For example, if I am driving an automobile, it is foreseeable that a pedestrian may cross at a crosswalk. It may not be foreseeable that a bicyclist will run a stoplight and ride into traffic. In the former, the driver may be liable for hitting the pedestrian, but may not be liable for hitting the bicyclist. Even in this simple example, a discussion of liability necessarily entails notions of causality, and those notions can be critically non-identical in different disciplines and fields of endeavor.

Despite these differences between disciplines, (though we know of no reference to support our contention in its entirety[10]²) it seems fair to say that lay notions of causality, at least among thoughtful members of the population at large, depend upon a core of four logical cornerstones which are common in all disciplines where causality is at issue: Sufficiency (“X sufficient to cause Y”), Necessity (“X is necessary to cause Y”), Proximity (“X is close enough in the chain of events to be considered a cause, or the cause, of this Y”), and Plausibility (“X causes Y by an argument that is specific, likely, and accessible to common sense.”)³. If it’s reasonable to claim that ordinary jurors make causal attributions based on **Sufficiency, Necessity, Proximity, and Plausibility** then we should be able to see how those four logical requirements are manifest in litigation, particularly litigation where statistical evidence is prominent.

Sufficiency: The court has a clear precedent laying out guidelines for what constitutes a sufficient cause, at least from the perspective of likelihood. Specifically, in *Hazelwood* 1977 the court required that evidence of a causal linkage should be “two or three standard deviations” beyond what one would expect to see if there were nothing operating other than random variation[11]. It’s important to note that this standard does not address magnitude, only statistical significance. For example, one aspirin a day helps lower the risk of death by heart disease in women by about 38% over the course of many years, but it does not reduce the likelihood to zero percent; the magnitude of that decrement reflects the effect size of the treatment[12]. This brings up the important distinction between statistical significance (which measures the likelihood of seeing the observed pattern of results if nothing more than random variation were responsible) and effect size (which quantifies the magnitude of an impact.)

Disagreement about Sufficiency: One occasionally encounters examples of disagreement between disciplines about the sufficiency of evidence for causality. For example, some respected philosophers (e.g., Carl Cranor) currently argue that all tests of statistical significance are misleading and lack evenhandedness[13]. However, such arguments usually rest on a remarkably narrow construal of what statistical significance means⁴, and overlook substantial court precedent that in some litigation (e.g., torts in epidemiology) non-statistical evidence, even when very compelling, is inadmissible and no substitute for carefully designed studies that generate actual data[14] and tests of statistical significance[15]. Other unusual standards for determining sufficiency are also encountered with some regularity. For example, litigation involving the EEOC occasionally invokes the Rule of 4/5^{ths} [16] which becomes patently unworkable for very small and very large datasets. Similarly, as several experts on evidence have pointed out[17], some courts follow a legal precedent[18] unknown in the scientific literature that requires a Risk Ratio of 2 (where an RR of 1 indicates that the risk of the exposed group equals the risk of an unexposed control group) to support any arguments of a causal linkage.

Necessity: The idea of a necessary cause is both venerable and straightforward. The notion of necessity allows us to distinguish between the incidental and the essential elements in a causal chain of events. But, although it is not often acknowledged as such, determinations of necessity require a degree of precision that is, in itself, conducive to a clearheaded analysis of cause and effect. For this

2 We can only offer substantiation for the independent elements in our claim by citing research programs by Kahneman, Carrol, and others; e.g., see work by Kahneman above, and Carroll, Hatakenaka and Rudolph (2006)

3 Plausibility here is intended to include parsimony in a sense that is consistent with Occam’s Razor.

4 For example, Cranor, 2006, *ibid*, pg. 231 argues that statistical significance necessarily requires a “statistical anomaly.”

reason it's entirely sensible that courts place a high premium on precise specificity while discussing causal linkages, as *NYC Transit v. Beazer* 1979 shows[19].

Proximity: It is necessary to control for less-proximal causes in the chain of events, so auxiliary factors that are removed in time and/or space will not become confounded with more proximal causes. For example, in *EEOC v. Joe's Stone Crab* 2000 the court determined that it was necessary to run statistical analyses that partial out (i.e., control for) the impact of factors associated with society at large[20]. Moreover, as *People Who Care v. Rockford* 1997 shows, covariates must "correct for salient explanatory variables" or the probative value of the entire analysis is lost[21]. Nor is it permissible, the court found in *Shehan* 1997, to simply control for one prominent factor, such as age in an employment discrimination case, and consider that the requirement for proximity has been met[22].

Plausibility: The courts are consistent and essentially unanimous in their determination that, wherever statistical evidence is used to establish claims of a causal linkage, alternate explanations have to be ruled out, and the resulting claims that survive that process of elimination must be both coherent and specific. For example, *Watson* 1988 stipulates that the elements in a causal linkage have to be explicitly specified[23]. And, in those cases where the database is small and no statistically significant results emerge, (as was the case in *Ambrosini* 1996) that a power analysis[24] be conducted to determine whether or not the results merely reflect the fact that the analysis was based on an overly limited number of observations[25]. Moreover, it is not sensible, where claims of causality are involved, to focus all the attention on refuting alternative explanations, as *Mapes Casino* 1968 [26] shows, because doing so can leave the main argument insufficiently supported. Note that plausible coherence necessarily entails the important distinction between statistical significance and effect size described above because if the magnitude of a putative impact is too minor to have caused the effect at bar, then the credibility of the entire argument is forfeited, regardless of the fact that (for example) random variation alone would not be expected to create the observed data more than one time in 10,000 (i.e., $p < .0001$). This critical distinction between statistical significance and effect size has unfortunately been missed in some litigation (e.g., *Craik*, 731 F. 2d at 479 [27]).

In our view, causality is best proven to non-statisticians by showing four key pieces of evidence. Three of these are positive evidence (which the moving party strives to claim have the stature of material facts in the case) and one is negative evidence where alternative explanations are ruled out. For the sake of convenience we'll call these the four hallmarks of causality. Hasty jurors will jump to a conclusion about the presence of a causal link after seeing just one of these four hallmarks; however, stronger arguments of causality require more than one of these four, and Quine, in his classic text on logic, claims that all four are necessary before anyone can be certain of a causal linkage[28].

Attentive readers will notice that the relation between the four cornerstones and the four hallmarks is a many-to-many relationship: Any one hallmark may lead to inferences about the existence of any one cornerstone, or any number of cornerstones.

The theoretical underpinnings of this approach come from several diverse sources. Primary among these is Kant's notion that humans innately and automatically categorize the world in terms of Quality, Quantity, Relation, and Manner[29]. In Kant's system Quantity contains Unity, Plurality, and Totality. Quality contains Reality, Negation, and Limitation. Relation contains Substance and Happenstance, Cause and Effect, Agent and Object. Modality contains Possibility v. Impossibility, Existence v. Nonexistence, Necessity v. Non-necessity. If Kant is correct, then several distinctions will be especially important during the comprehension of graphs: The distinction between unity and plurality, between cause and effect, and between incidental variation and essential variation. We combine this Kantian idea of innate categorization with an approach to cognition that is sometimes called constructivism – a departure from the mechanistic theoretical approach that most authors typically bring to work on graphic design. In the view of this constructivist approach, viewers

actively work to formulate hypotheses about the graph maker's intended meaning. Evidence for this dynamic extraction of meaning comes from a considerable body of research showing that viewers actively interpret the relations in graphs according to the way the elements are clustered[30] and displayed[31]. General support for our interpretation comes from the fact that listeners use active cognitive processes to extract meaning during conversational exchanges as well[32], relying heavily on the context as a reference point during this process.[33] The final component in this (admittedly rudimentary) theory of graph comprehension comes from extensive research by Kahneman and others[34] showing that some automatic cognitive processes lead to judgments under uncertainty that are subject to a host of predictable errors and erroneous inferences; in our view, these errors of inference are largely and automatically driven by the salience (i.e., prominence) of ostensibly incidental elements by virtue of their color, location, and apparent volume. So, in summary, our theory of graph comprehension has three major elements: 1) That viewers make distinctions about the number of data points, the causal relations, and the chance variations in a graph; 2) That those distinctions are part of an active cognitive process where viewers formulate hypotheses about the graph maker's intended meaning, and; 3) That those inferences about the graph's meaning can be distorted by minor perturbations in the salience of specific elements in the graph.

These four hallmarks will each be addressed in detail directly below; they concern the following: 1) Association; 2) Prediction; 3) Dose-dependence, and; 4) The elimination of alternative explanations. In all the explanatory text regarding the hallmarks we'll use the standard abbreviation "X" to mean the causative variable (also called the independent variable or the predictor variable) and "Y" to mean the outcome variable (also called the dependent variable or the response variable.)

Positive Evidence of association is very straightforward. When simple association exists, X covaries with Y. Descriptive alternatives that define this association include (among others) linear vs. non-linear, positive vs. negative (where one increases as the other declines), unidirectional vs. bi-directional. This is the classic case of correlation that is frequently (though not invariably by any means) a central piece of evidence when claims of a causal linkage are put forward. The strongest evidence of association comes from data collected in a "flow" over an extended period of time, as is the case in much discrimination litigation[35]; in such cases the association is manifest as a statistically significant correlation between X and Y at both the beginning and the end of the observation period.

Positive Evidence of Prediction usually requires that explicit prediction be involved in the strict sense of the word. The requirements of prediction are usually not met by a post hoc analysis where arbitrary time periods are selected after the data have been seen. Notwithstanding the foregoing, predictions from one year to the next, (and those based on days, months, minutes or hours) are usually accorded the full stature of a legitimate predictive linkage, even if the distinction was made post hoc, because of the importance and standardization of the periods involved. In the purest example, first a prediction is made that X will affect Y during some future time period, and then that association between X and Y is indeed observed during the predicted period.

Positive Evidence of Dose-dependence is the gold standard of most clinical research in the medical arena. In the strong instance of dose-dependence, the greater the dosage of X, the greater the subsequently observed change in Y. (In the weaker case, the change in X is correlated with Y as a whole; in essence this is a special case of dose-dependence, where a ceiling effect or a similar non-linearity in Y is assumed to be limiting its response range.)

Negative evidence against counter arguments usually entails the presentation of evidence ruling out myriad alternative explanations. This effort sometimes leads attorneys astray because, in their desire for completeness, they dramatically increase the number of rebutted alternatives and consequently also elevate the likelihood that one or more specific counterarguments will be weak in the eyes of a given juror, who then forgets the relatively minor stature of that particular rebuttal in the context of the more important positive evidence of association, prediction, and dose-dependence. Indeed, when bad-faith arguments are made in an effort to obfuscate and confuse, it is these

multifaceted rebuttals that typically become the target because jurors sometimes give more weight to negative evidence than is justified.[36]

We contend that problems ensue if all 4 tokens of causality are not presented as a mutually supporting, unified, set. It seems plausible to imagine that, if any of the 4 hallmarks is shown without the others, claims of causality are much more likely to fail. If association alone is shown, the argument will fail because correlation does not necessarily entail causation; the underlying fallacy is "Cum hoc ergo propter hoc." If prediction alone is shown the argument will fail because precedence does not necessarily entail causation; the underlying fallacy is "Post hoc ergo propter hoc."

If dose-dependence alone is shown the argument will fail because dose-dependence does not necessarily entail causation; the underlying fallacies include the conjunction fallacy (described above, where a salient conjunction impossibly seems more likely than either of its less salient elements), and, more commonly, the Fallacy of the Single Cause (AKA Ignoring a common cause, where an extraneous third variable simultaneously drives both the putative cause and the putative effect); other errors can also be involved: for example, the failure to account for bi-directional causality, feedback loops, or mediating variables.

Although some decisions fault movants for failing to rule out alternative explanations, if refutation of alternative explanations alone is emphasized, all claims of causality can evaporate in a fog of what will appear to non-statisticians as nothing more than inconsequential bickering between captious academics; indeed, some judges have voiced unambiguous complaints about this very problem in their decisions [37].

Specifically, this paper tests three hypotheses, the first of which is as follows: In order for an argument of a causal linkage to prevail, the movant must provide evidence of all four hallmarks—Association, Prediction, Exclusion, and Dose Dependence.

The second hypothesis is that movants prevail if and only if they introduce evidence of dose dependence, exclusion of alternative explanations, or both.

The third hypothesis is that the likelihood of prevailing increases with the number of hallmarks in the movant's argument.

METHOD

To determine the minimum number of court cases we would need to test our hypothesis we conducted a power analysis (Cohen & Cohen 1980) so that we could limit the time consuming process of reviewing and coding case summaries. The power analysis indicated that a minimum of approximately 10 cases would provide the desired statistical power given the presumed effect size (Power = .80, predicted effect size = .7, LSN = 10 at alpha = .05). A brief inspection of some appropriate searches in Lexis-Nexis led us to estimate that roughly 1 in 6 case summaries would contain a full description of the movant's causal argument. Accordingly, we planned to inspect 60 case summaries for this study. Given the extreme variability in summaries and indexing terms, we settled on a stepwise three-part method to select a small but representative set of cases, review their summaries, and classify the content of the causal arguments. Similar sequential processes have been advocated in the past where the domain of research poses unusual methodological difficulties [38].

A) Compiling a List of Potentially Relevant Court Cases (The Initial Review): A comprehensive literature search was conducted in Lexis-Nexis using the search terms "[tobacco OR cigarettes] AND causal AND [liability OR tort]." The time period was open, and results came from litigation in State and Federal courts. The first 968 hits from this search were put in a database that was sorted by relevance. A research assistant with experience in law and psychology conducted an initial review to separate cases that probably did contain a causal argument from those that probably did not. (The irrelevant cases lacking causal arguments included pre-trial hearings, motions to remand based on

technicalities, and the like.) This initial review identified 63 case summaries that were potentially relevant. (A full list of the 63 cases passing this initial review is available on request.)

B) Classifying the Elements in Arguments of Causality (The Secondary Review): The 63 cases surviving the initial review were sorted alphabetically to eliminate prior sorting by date and relevance. The primary researcher then conducted a secondary review of these 63 cases, during which the movant's causal arguments were read and the elements of the argument were classified as one of the four hallmarks described above (i.e., evidence of Association and/or Prediction and/or Exclusion and/or Dose Dependence.) During this secondary review, the court's decision was also classified regarding the movant's claim, which either failed or prevailed. If during the secondary review the case was found to contain no argument about a causal linkage then it was deemed irrelevant, and removed from further consideration. For example, some cases that passed the initial review failed the secondary review because they were found to be decided on technicalities in statutes pertaining to time-bound limitations, liability, defects, warnings, and the like. During this secondary review 44 cases were determined to be irrelevant and 19 were deemed relevant. The elements of the movant's causal arguments were then coded and tallied in these 19 cases. (A full list of the 19 relevant cases, as well as case summaries highlighting the text that contains each hallmark, is available upon request.)

C) Compiling a database of the Elements: We built a simple SAS database containing 19 rows (one for each legal case subjected to a secondary review) and six columns containing the following information: Name of the case; Abbreviations of the Hallmarks in the Movant's argument (A, P, E and/or D); Decision (Failed/Prevailed); Determination of Whether or Not the Claim's Outcome Supports the Primary Hypothesis that all four hallmarks are necessary if the movant is to prevail (Yes/No), and; Determination of Whether or Not the Claim's Outcome Supports the Secondary Hypothesis that dose dependence and/or exclusion of alternatives are conducive to a favorable decision (Yes/No);and Total Number of Hallmarks introduced into the movant's argument.

RESULTS

We tested the first and second hypotheses with the Cumulative Binomial Distribution Test [39, 40], a statistical tool that has proven useful for similar analyses since its introduction by Newton in 1676 [41]. The results of the analysis were unambiguous: Of the 19 court summaries examined, 13 cases (68%) confirmed our primary hypothesis, a distribution that cannot be explained by chance alone ($p < 0.05$). When we tested the second hypothesis (that movants prevail if and only if they introduce evidence of dose dependence, exclusion of alternative explanations, or both) we again find statistically significant confirmation: Movants prevailed in 18 out of 19 cases (95%; $p < .00001$) when they introduced evidence of dose dependence, exclusion of alternatives, or both; (in our dataset, dose dependence and exclusion were never introduced without evidence of association, prediction, or both). To test the third hypothesis (namely, that the likelihood of prevailing increases as the number of hallmarks in the argument increases) we used the point biserial correlation between the number of hallmarks introduced and the likelihood of prevailing or failing (coded respectively as 1 or 0). The significant correlation ($r = 0.88$, $n = 19$, $p < .0001$) indicates that the higher the number of hallmarks used in the argument, the higher the likelihood of a favorable decision for the movant.

DISCUSSION

As predicted, our review of litigation revealed that building a causal argument in court requires attention to (what this paper calls) the four hallmarks of causality. Those hallmarks—showing evidence of association, prediction, exclusion of alternative explanations, and dose dependence—seem to be required if arguments of a causal linkage are to prevail. The fact that 68% of the cases confirm our main hypothesis leads us to suspect the hallmarks play an important role in convincing fact finders in court.

It is interesting that dose dependence and exclusion of alternatives – two hallmarks that are typical of clinical studies in the medical domain and experimental research in the behavioral sciences respectively – seem to have a potentiating effect on evidence of association and prediction – hallmarks that are more typical of lay conversations about causality. The results from the second hypothesis suggest that there are at least two different profiles of hallmarks that sway non-statisticians: One where association and/or prediction is supplemented by evidence of dose dependence, and one where these two common hallmarks are supplemented by negative evidence that rules out attractive alternatives. In the first case, perhaps the non-statisticians in court are being persuaded by dose dependence because they have adopted the gold standard of causality used in pharmaceutical research. In the second case, perhaps fact finders are being persuaded by exclusion of alternatives because they have implicitly accepted Disraeli’s view (“There are lies, damn lies, and statistics...”) and assume that if a statistician cannot find some evidence of an alternative explanation, then it must be a very compelling argument indeed. Settling these questions about the cognitive processes of jurors will have to wait until we have access to a larger dataset or an experimental manipulation; however, it is clear that some combinations of hallmarks are especially powerful in the courtroom – a contention supported by the fact that 95% of the arguments in our study were potentiated by either dose dependence, exclusion of alternatives, or both.

As a final piece of evidence showing that the four hallmarks play an important role in deliberations of causality, it is wise to call attention to the simple but compelling fact that the likelihood of a favorable decision is very highly correlated with the number of hallmarks in the movant’s argument. Clearly, if the movant has any choice in the matter at all, it is wisest to build an argument of causality that contains evidence of all four hallmarks – association, prediction, exclusion of alternative explanations, and dose dependence – so that the chances of prevailing are maximized.

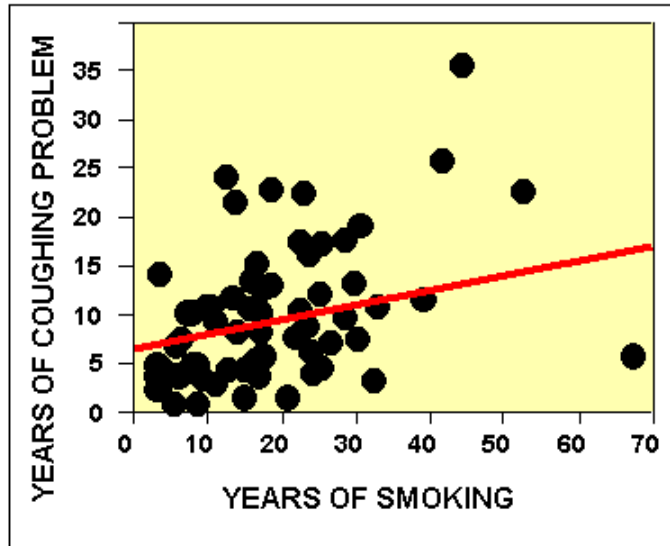
The remainder of this paper addresses the next issue of concern: Given the fact that the four hallmarks are important, how should they be shown graphically in exhibits. Although it is not necessary that all elements in a causal argument be distilled into graphs and plots, we have found (especially where very large datasets are concerned) that it is often sensible to utilize the added facilitation to understanding and memory that plots and graphs can provide.

Here are some samples of how we recommend graphing the four hallmarks of causality. Our data come from two sources: We begin by showing a simple association plot from data collected annually by the Center for Disease Control and Prevention for NHANES—the US government’s National Health and Nutrition Examination Survey.) We then move on to show how graphical evidence from a large statistical database can be used to prove causality even to readers who lack advanced statistical training.

=== remainder of page blank ===

Association without proof of a causal linkage <insert figure here>

Association: Years of Smoking Tracks Years of Coughing



There is a strong and statistically significant correlation between Years of Smoking and Years of Coughing, but this association is far from irrefutable proof of a causal linkage between the two. For example, coughing may have preceded smoking, or a third variable might cause both (e.g., working in an environment where many people smoke and particulates are pervasive), to name just two alternative explanations.

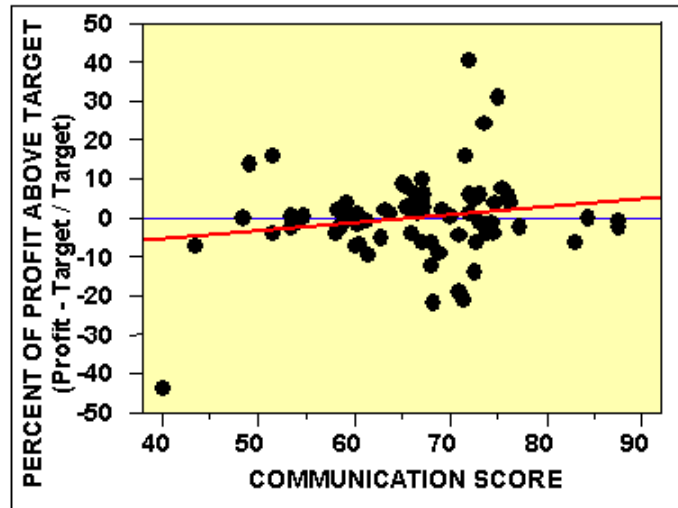
Statistical Note: Data are from NHANES 2003-2004; n = 4696 (86) aggregated by age at time of interview; r = .32; p = 0001

2

==== remainder of page blank ====

Association as part of a proof of a causal linkage <insert figure here>

Association: Communication Tracks Profit in the Clinic



Between 2002 and 2005 Fallon Clinic—a health care provider with approximately 2000 employees and 38 clinics—conducted four annual employee surveys to identify and measure the drivers of good corporate performance within their organization. Results showed that the better the clinic’s Communication Score, as measured by the employee survey, the better the profitability. (Profitability was operationalized as Percent Profit Above Target, which is profit variance as a percent of the budget target.) As the plot above shows, clinics with Communication Scores below 65 were significantly more likely to fall below their revenue target. However, association alone does not prove causality.

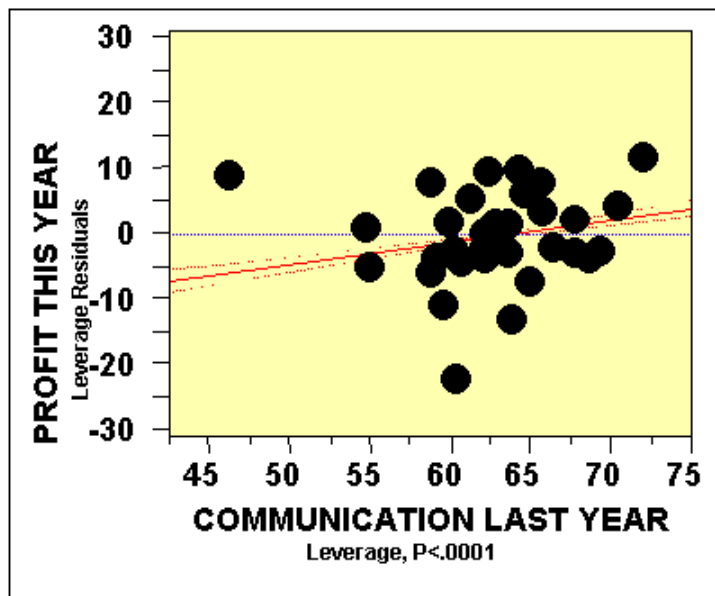
Statistical Note: Data are from Fallon Clinic 2002-2005; n = 4640 (152) aggregated by Year and Clinic; r = .21; p = 0001

3

=== remainder of page blank ===

Prediction as part of a proof of a causal linkage <insert figure here>

Prediction: Communication last Year Predicts Profit this Year



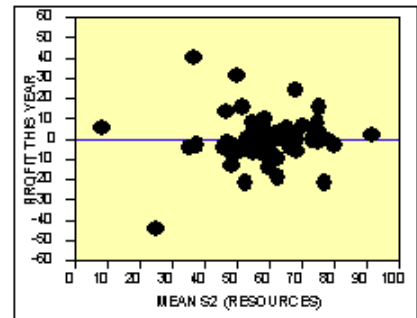
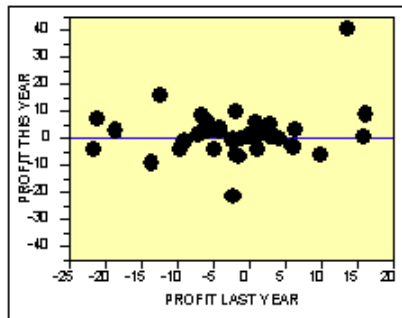
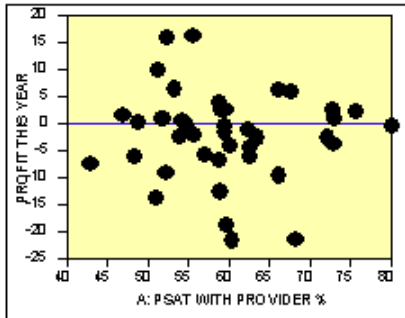
A predictive linkage provides additional support of a causal relationship: There is a statistically significant predictive linkage between last year's Communication Score and this year's Profit. Specifically, the better the clinic's communication score in the previous year, the greater the clinic's Profit Above Target in the subsequent year. (As in the previous slide, Profit This Year is profit variance as a percent of the target.)

Statistical Note: Data are from Fallon Clinic 2002-2005; n = 4640 (152) aggregated by Year and Clinic, controlling for Headcount, Clinic, PSAT with Provider, PSAT with Operations, and Pot of Complaint-free encounters; B = .34 p < .0001 (Accordingly, each increase of 10 points in Communication yields a 3% increase in Profit Above the Target) R-Square = .48. 4

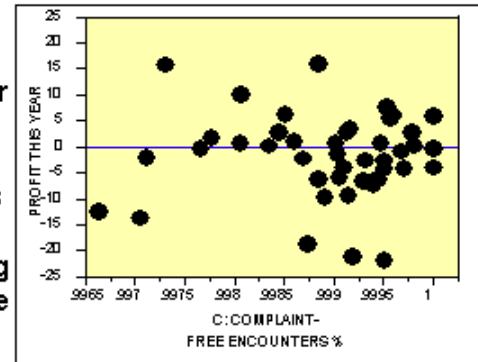
=== remainder of page blank ===

Ruling out alternatives <insert figure here>

Ruling Out Alternatives: Other Metrics Do Not Drive Profit



In virtually all analyses from this study, Profit does not vary consistently or significantly with other variables such as Patient Satisfaction (upper left), Profit Last Year (upper middle), Resources (upper left), or Percent of Complaint-free Encounters (directly to right). The results are important because these and similar metrics were used as covariates to control for confounding effects in the plots shown on the previous and following pages. The fact that the regression lines are flat in these metrics shows that profit in this dataset is not simply driven by these potentially confounding variables.

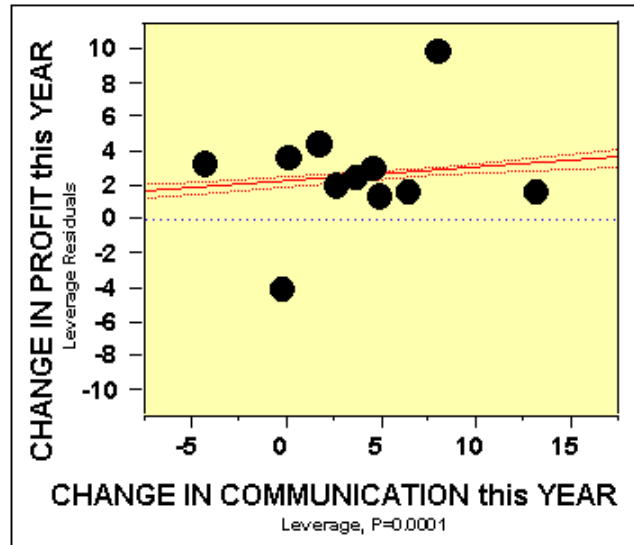


Statistical Note: Data are from Fallon Clinic 2002-2005; n = 152 rows aggregated by Year and Clinic. Plots show correlations between Profit Above Target (i.e., profit variance as a percent of the budget target) and PSAT with Provider ($r = -.06$, NS); Profit Above Target Last Year ($r = .18$, NS); Resources ($r = .16$, NS); and Percent of Complaint-Free Encounters ($r = .01$, NS). 5

=== remainder of page blank ===

Dose-dependence as part of a proof of a causal linkage <insert figure here>

Dose Dependence: The More Communication Improves, the More Profit Improves



In all years of the study, we found a statistically significant linkage between between this year's change in Communication and this year's change in Profit. Specifically, the better the clinic's improvement in communication, the greater their improvement in Profit. With statistically significant evidence of **DOSE DEPENDENCE** and **ASSOCIATION** and **PREDICTION** and very few alternative explanations, the evidence for a causal linkage between Communication and Profitability, at least in this dataset, is consistent and clear: **Good communication facilitated good profitability at Fallon Clinic from 2002 to 2005.**

Statistical Note: Data are from Fallon Clinic 2002-2005; n = 4640 (152) aggregated by Year and Clinic, controlling for Headcount, Clinic, PSAT with Provider, PSAT with Operations, and Pct of Complaint-free encounters; B = .08 p < .0001 (Accordingly, each Annual increase of 10 points in Communication yields a 0.8% annual increase in Profit Above Target) R-Square = .97. 6

Philosophers, researchers, jurors, theologians, statisticians, parents, and legal scholars have been struggling with causality for centuries, so we would surely not presume to have access to any ultimate insights. However, we certainly do know (as that old saying goes) that an example is not a proof, and we are beginning to get a clear idea of the difference. Moreover, we believe that we stand on firm ground to claim that proofs of causality are genuinely possible with methodical and systematic attention to graphic evidence.

References and Cases Cited

1. *Reference Manual on Scientific Evidence*. 2nd ed. 2000, Washington, D.C.: Federal Judicial Center.
2. *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579, 113 S. Ct. 2786, 125 L. Ed. 2d 469, 61 U.S.L.W. 4805, 27 U.S.P.Q.2D (BNA) 1200, and 1993 U.S. LEXIS 4408. (1993).
3. Joseph, G.P., *Modern Visual Evidence*. 1984, New York, NY, USA: ALM Properties, Inc., Law Journal Press.
4. Kahneman, D., *Maps of Bounded Rationality: A perspective on intuitive judgment and choice*, in *Nobel Prize Award Ceremony*. 2002: Stockholm, Sweden.
5. Tversky, A. and D. Kahneman, *Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment*. *Psychological Review*, 1983. 90(4): p. 293-315.
6. Young, R.D. and S. Susser, *Effective Use of Demonstrative Exhibits and Demonstrative Aids*. *Michigan Bar Journal*, 2000. 79(11): p. none given.
7. Kaye, D.H. and G.F.J. Sensabaugh, *Reference Guide on DNA Evidence*, in *Reference Manual on Scientific Evidence*, F.J. Center, Editor. 2000, Federal Judicial Center: Washington, D.C., USA. p. 538 ff.
8. *Pawsey v. Scottish Union and National Insurance Co (1908)*.
9. *DeLuca v. Merrell Dow Pharmaceuticals Inc.*, 911 F.2d 941, 945 & n.6 (3rd Cir. 1990).
10. Carroll, J.S., S. Hatakenaka, and J.W. Rudolph, *Naturalistic Decision Making and Organizational Learning in Nuclear Power Plants: Negotiating Meaning Between Managers and Problem Investigation Teams*. *Organization Studies*, 2006. 27(7): p. 1037 - 1057.
11. *Hazelwood School District v. United States*, 433 U.S. 299, 97 S.Ct. 2736, 53 L.Ed.2d 768 (1977).
12. Chan, A.T., et al., *Long-term Aspirin Use and Mortality in Women*. *Archives of Internal Medicine*, 2007. 167: p. 562-572.
13. Cranor, C.F., *Toxic Torts; Science, Law, and the Possibility of Justice*. 2006, New York, NY, USA: Cambridge University Press.
14. *Haggerty v. Upjohn Co.*, 950 F. Supp. 1160. 1163-64 (S. D. FL. 1996).
15. *Cartwright v. Home Depot USA Inc.*, 936 F. Supp 900, 905 (M. D. FL. 1996).
16. *29 Code of Federal Regulations 1607-3.4(d)*.
17. Imwinkelried, E.J., *The Admissibility and Legal Sufficiency of Testimony About Differential Diagnosis (Etiology): Of Under - and Over - Estimations*. *Baylor Law Review*, 2004. 56, Part 2: p. 391-422.
18. *In re Breast Implant Litigation*, 11 F. Supp. 2d 1217, 1229 (D. Colo. 1998).
19. *New York City Transit Authority v. Beazer*, 440 U.S. 568, 99 S.Ct. 1355, 59 L.Ed.2d 587 (1979).
20. *EEOC v. Joe's Stone Crab, Inc.*, 220 F. 3d 1263 (11 Cir. 2000).
21. *People Who Care v. Rockford Board of Education*, 111 F.3d 528, 537-38 (7th Cir. 1997).
22. *Sheehan v. Daily Racing Form, Inc.*, 104 F. 3d 940, 942 (7th Cir. 1997).
23. *Watson v. Fort Worth Bank and Trust*, 487 U.S. 977, 108 S.Ct. 2777, 101 L.Ed.2d 827 (1988).
24. Cohen, J., et al., *Applied Multiple Regression - Correlation Analysis for the Behavioral Sciences*. 2002, Hillsdale, NJ, USA: Erlbaum Associates.
25. *Ambrosini v Labarraque*, 101 F. 3d. 129, 138 (D.C. Cir. 1996).
26. *Mapes Casino, Inc. v. Maryland Casualty Company*, 290 F.Supp. 186 (Nev., 1968).
27. *Craik v. Minnesota State University Board*, 731 F. 2d, 465,479, (8th Cir, 1984).
28. Quine, W.V.O., *Elementary Logic*. 1941, New York: Harper.

29. Kant, I., *Immanuel Kant's Critique of Pure Reason (Kritik der Reinen Vernunft)*. 1929/1781, London: Macmillan.
30. Jarvenpaa, S.L., G.W. Dickson, and G. DeSanctis, *Methodological issues in experimental IS research: Experiences and recommendations*. MIS Quarterly, 1985. 9(1): p. 141-158.
31. Shah, P., R.E. Mayer, and M. Hegarty, *Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension*. Journal of Educational Psychology, 1999. 91(4): p. 690-702.
32. Morrel-Samuels, P. and R.M. Krauss, *Word familiarity predicts temporal asynchrony of hand gestures and speech*. Journal of experimental psychology. Learning, memory, and cognition, 1992. 18(3): p. 615-22.
33. Weick, K.E., K.M. Sutcliffe, and D. Obstfeld, *Organizing and the Process of Sensemaking*. Organization Science, 2005. 16(4): p. 409-425.
34. Tversky, A. and D. Kahneman, *Judgment under uncertainty: Heuristics and biases*. Science, 1974. 185(4157): p. 1124-1131.
35. *Malave v. Potter*, 320 F.3d 321, 2nd Cir. (2003).
36. Mendez, M.A., *California's New Law on Character Evidence: Having Code Section 352 and the Impact of Recent Psychological Studies*. UCLA Law Review, 1984. 1003: p. 1045 - 1048.
37. *Johnson v. Miller*, 864 F. Supp. 1354, 1390 (S.D. GA 1994).
38. Fiske, D.W. and L.V. Jones, *Sequential analysis in psychological research*. Psychological Bulletin, 1954. 51(3): p. 264-275.
39. SAS Institute, I., *JMP Statistics and Graphics Guide*. 2007, Cary, NC: SAS Institute, Inc.
40. *JMP*. 1989-2007, SAS Institute Inc.: Cary, NC.
41. Wallis, J., *De Algebra Tractatus (Treatise of Algebra)*. 1685, Oxford, England: Oxoniae: E Theatro Sheldoniano.

Authors' Note: We wish to thank Dr. Baltej Maini, CEO of Fallon Clinic, and Dr. Diana Palmiere Hunt of EMPA, and Mary Heumann of EMPA, for their indispensable assistance on this project.

SUMMARY:

Many writers claim that statistics have become increasingly important in litigation. However, no comprehensive contemporary guide exists for attorneys who want to use statistical data to create effective demonstrative evidence—an issue that is especially important when non-statisticians use statistics to make inferences about causality. In this paper we outline a new theory explaining the perception, comprehension, and recall of quantitative graphs. As an outgrowth of that theory we propose that four cornerstones are essential for inferences of causality in most disciplines: Sufficiency, Necessity, Proximity, and Plausibility. Consistent with this theory, we contend that four hallmarks of causality are critical. These hallmarks show positive evidence of Association, Prediction, Exclusion of Alternative Explanations, and Dose dependence. We then test our theory by compiling a set of 63 case summaries of smoking injury litigation, isolating the 19 relevant cases where a movant presented a causal argument linking cigarettes to illness, and classifying the elements in the movant's argument. We find that, in 13 out of 19 cases (68%; $p < .05$), if movants use all four hallmarks—showing association, prediction, exclusion, and dose-dependence—their arguments prevail. We also find that exclusion of alternative explanations and dose dependence have an especially powerful impact on non-statisticians in the courtroom: In 18 out of 19 cases (95%; $p < .00001$) movants prevail if they introduce evidence of dose dependence, exclusion of alternatives, or both. A simple additive model also has support: The higher the number of hallmarks introduced, the greater the likelihood of a favorable decision, and the strength of the relationship is exceptionally strong for studies in the behavioral sciences ($r = .88$; $p < .0001$). We close our paper with a discussion examining recent advances from research in experimental psychology, and use those advances to formulate best-practice examples of how these four hallmarks can be shown in quantitative graphs during litigation.