The Art of Evidence and the Morality of Medical Decisions

Brian Dolan

For this Grand Rounds, Dr. Brian Dolan looks at epistemological challenges to the tenets of Evidence Based Medicine by focusing on some of the ways that statistical data is presented as evidence. Dr. Dolan seeks more generally to explore the contours of quantification and digitization in biomedicine in search of uses of the graphic method and the status of producing pictures of numbers. One area implicated in the relationship between statistical representation and EBM is the way risk factors and decision making techniques are communicated in the physician-patient relationship, hence the question of the morality that underlies the art of evidence. Background reading by Thomas Newman is attached.

Dr. Thomas Newman is a professor of clinical epidemiology and pediatrics at the University of California, San Francisco, who specializes in the study of neonatal jaundice. He is an expert in a disease called kernicterus, which means ‘yellow kern’, referring to the discolored appearance of parts of the brain revealed at autopsy. It is caused by hyperbilirubinemia, an excess of a chemical compound that also causes the yellowish halo to form around bruises. That the name of the condition was derived from morbid anatomy suggests the severity of its neurotoxicity.

The frequency of the disease seems to have hit a trough after a peak in the 1950s and 1960s, only to reemerge in the 1990s. However it is at this unfortunate time that Dr. Newman and his colleagues had been suggesting that jaundice in newborns was being over treated. In a commentary published in the British Medical Journal, Dr. Newman expressed his concern that his opinion might have led to clinical oversight in such cases. However, a statistical correlation is unlikely. The disease is extremely rare. A registry of kernicterus cases in the US that was maintained between 1984 and 2002 recorded 125 babies diagnosed with the condition, five of which died from it. Newman states that a Kaiser Permanente database shows no cases out of 239,000 births during the 1990s.

However in 2000, seven mothers whose children had been diagnosed as having kernicterus formed PICK (Parents of Infants and Children with Kernicterus), an advocacy group that promotes awareness, prevention, and treatment of the disorder. Since kernicterus was linked to high levels of bilirubin, they wanted the American Academy of Pediatrics to recommend that all newborns have a blood test to screen for risk before they were discharged from the hospital. Newman, a physician, a specialist in the field, and an expert witness in malpractice cases relating to this issue, wrote about how much he admired those women who were “heroically fighting to prevent an awful disease”, but he could not get on their side. “I am a proponent of evidence based medicine”, he explained, “and am reluctant to endorse a new screening recommendation that is not based on good evidence.”

The problem, suggestive of a moral dilemma in the practice of modern medicine, was that the experiences these women had did not amount to enough (or the right kind of) evidence to take medical action.

The power of narrative, as Dr. Newman as well as most media pundits, political strategists and novelists recognized, is beguiling. However tragic and heartbreaking individual’s experiences are—and an emotional chronicle of them is archived on PICK’s web site—they do not meet the rigors of EBM. As one ethicist wryly stated, “the plural of anecdote is not data.” Stories are

---

1 Department of Anthropology, History & Social Medicine, UCSF, 3333 California Street, San Francisco, CA 94143-0850. dolanb@dahsm.ucsf.edu


emotive, and they often distort the way people estimate probabilities. As Newman wrote, “if we are trying to estimate the risk of kernicterus, one method we use is to base the estimate on how readily we can recall or imagine a case, and in what level of detail. This technique, called the availability heuristic, leads us to overestimate probabilities of events that we can easily and vividly imagine.” 4 In essence, clarity of expression correlates with persuasiveness.

It is no small irony (and subject of rhetorical analysis) that Western science has relied on enumeration instead of narration to gain credibility. There is further irony in the fact that a key technique for representing the sum of the data that is taken to underwrite evidential claims is itself subject to the same strictures as story telling. Precisely because of the complexity of statistical investigation, the “graphic method” is employed to produce succinct snapshots of its results. As stated in a recent article in the Journal of Clinical Epidemiology that examined the uses of graphs in medical publications, “Often, a graph is a more attractive presentation format than text or tables. … There is evidence of a substantial advantage in recall for graphical over textual information.” 5 For this reason, as I will elaborate on below, physicians are increasingly encouraged to use the graphic method to communicate risk assessment to their patients about life and death decisions. The beauty of the graphs is the allure of their simplicity. Logically, the same critical lens should be used to scrutinize the construction of graphs as is used to examine language games. In short, the art of evidence that is deployed in the process of making moral decisions cannot be taken for granted.

The graphic method is a multifaceted approach to producing pictures of numbers. It can be like the line graph that Newman reproduced which illustrated his own statistical projection of mortality from car crashes (FIGURE) or something like the smiley-face portrayal of risks for certain therapies (FIGURE). The magnificence of graphs is that they appear so simple, and, as Bruno Latour pointed out, that they are conveniently flat. A graph represents information that can be moved around easily. It can be printed in the space of a few square inches. And from the time the late eighteenth-century Scottish political economist William Playfair first used them, statistical “information may be obtained in five minutes as would require whole days to imprint on the memory, in a lasting manner, by a table of figures.” 6 The immediate impression they make on the mind and memory make them powerful. But as the Yale professor Edward Tufte showed in his work on the aesthetics of statistics, if you sit down and closely examine the uses of graphs as a mode of communication, they often do not assist anyone in “envisioning information.” Inspired by the poignancy of Tufte’s analysis, Ian Hacking wondered “whether the point of the representations is to convey information at all, or rather to convince us that this is solid stuff, not to be challenged, not challengeable.” 7

Graphs are in-your-face statements of (ostensible) fact. As such, they often feature in professional disputes about said facts. When alleged to be worth no more than the paper they are printed on, graphs become points of departure for scientists who venture into laboratory spaces in order to examine the process by which such inscriptions are made. Sociologists of science such as Latour, Steve Woolgar, among others, have shown us how we can “follow” scientists into such arenas where claims about the signified are reasserted through new means, where instruments rather than their inscriptions are scrutinized. In statistical representation there is a certain amount of black-boxing that occurs which is difficult to unpack. But putting aside the issue of how numbers can be manipulated and how impractical in certain circumstances it would be to revisit hundreds of informatics labs to reexamine the work—that is to say the means of production that give value to the graph—I wish to focus on the problems of interpretation.

---

5 Puhun XXX
The construction of facts impacts the “morality” of decision making in ways separate from judgments about the veracity of the spokesperson in the way that Steven Shapin writes about. We need to find a way of examining what is taken as evidence by users of data whose decisions are informed by their interpretations.

Beyond rearticulating the message of the psychologist-sociologist Robert M. Young that “science is social”, meaning both its process and product, drawing attention to the way the “epistemic system” works reminds us of the important point that theories as well as practices are value-laden.  

---

Even an apparently “closed system” of logic such as mathematics, the language of science, is value-laden. “Of” science—with that genitive mathematics becomes an instrument, and as an instrument—of analysis and representation—its uses need to be thoughtfully considered. The first moral decision with regard to evidence (mathematical or otherwise) is how to present it.

***

The UCLA professor of history of science Theodore Porter informed us some time ago that “The credibility of numbers … is a social and moral problem.” Things seem to matter when there are numbers behind them. Not much is thought of the value of zero. Yet zero—a sign that represents nothing—is immensely important. It is where

---

8 Young; Rheinberger.
9 Porter, Trust in Numbers, 1.
counting necessarily begins. It is the foundation of statistics. It would mean the world to someone to be told they had zero chance of getting sick. But zero does not correspond to anything: by that I mean it does not (cannot) represent nothing. Semioticians such as Brian Rotman have enjoyed demonstrating that never-the-less meaning is made of the value of nothing, the sign that stands in place of the invisible.\textsuperscript{10} It is a fundamental point in philosophy, and a concept that mattered to Nietzsche. Truth, he says, is what happens when the illusory nature of illusions is forgotten. To leave this stream of consciousness I wish to make a summary point about the semiotics of numbers: it gives us a way into the conceptual world where symbols that are assigned value take on literal currency in moral, political, and fiscal economies.

Similar to zero, there is an illusory nature to probability—where statisticians run numbers against imaginary complete data sets, for instance. To bolster evidence-based medicine, “meta-analyses are used to aggregate the results of underpowered individual studies that are individually incapable of drawing positive conclusions.” It is an interesting way of speaking on behalf of the unspeakable. The selection of data to be analyzed and the choice of statistical tests are critical to such acts of quantitative synthesis and sometimes the results of “meta” as opposed to individual clinical trials are interpreted wildly differently.\textsuperscript{11} What does this say about the basis of making decisions about administering high-dose interferon alfa-2b in melanoma? (Such was a debate between major US conductors of clinical trials and an Oxford University research group.)

This is not a condemnation of statisticians or their work. I suggest that regardless of the reality behind the representation, we need to pay attention to how users interpret results. It is the epistemological issue—not what we know but how we come to know it. Before we get to action (administering doses, screening for risk, etc), we need to focus on the moment of understanding, because that is what determines decision making. The experimental evidence—the diagram, the picture, etc.—that works to allow facts “to speak for themselves” and establish scientific credibility does not speak for itself. Contrary to the position taken by Enlightenment philosophes that there are truths to be taken as self-evident, by suggesting that nature speaks for itself, nothing is self-evident and moral principles are established through informed action.

Scientific evidence is not merely descriptive—it does not just create a portrait of reality. Scientific evidence is a set of instructions about how to look at reality. Statistics and graphs do not provide answers, they stimulate the very questions to be answered. Mathematical proofs themselves, as Brian Rotman pointed out, are injunctive: “define A, compute B, consider C …” The logic of mathematics allows one to venture into another world. “Mathematics creates imaginary worlds,” wrote Rotman in his book \textit{Mathematics as Sign}, “brought into being and controlled through the agency of specialized signs.” Another virtual reality. But it is in such a space that the future is imagined, where meaning is made of the art of evidence.

\* \* \*

Here is an anecdote from a physician’s autobiography.

One day when I was a junior medical student, a very important Boston surgeon visited the school and delivered a great treatise on a large number of patients who had undergone successful operations for vascular reconstruction. At the end of the lecture, a young student at the back of the room timidly asked, ‘Do you have any controls?’ Well, the great surgeon drew himself up to his full height, hit the desk, and said, ‘Do you mean did I not operate on half of the patients?’ The hall grew very quiet then. ‘Yes, that is what I had in mind.’ Then the visitor’s fist really came down as he thundered, ‘Of course not. That would have doomed half of

\textsuperscript{10} Rotman, \textit{Semiotics of Zero}

\textsuperscript{11} See the debate about the Lens and Dawes meta-analysis that played out in the editorial pages of the \textit{Journal of Clinical Oncology} 2002, pp. 4122, etc.
them to their death.’ God, it was quiet then, and one could scarcely hear the small voice ask, ‘Which half?’

The legalistically labeled concept of “evidence based medicine” has generated a new kind of consciousness in biomedical practice. EBM declares that it is important to think about how decisions are made. The phrase is really phenomenal. Evidence based … on what? Research. Collecting the evidence shows that one is aware that someone else has thought about a problem and reached some conclusions. The phrase means also that “medicine”—as a practice and a product—has evidence of its own development. It is not sui generis, but constitutive of a history of research. Understanding the complexity of that research and the means by which the results of it are represented can help give meaning to EBM, for there is nothing self-evident about the evidence upon which medicine is based, or the research upon which the evidence is based.

Since the publication in 1954 of Darrell Huff’s insightful book How to Lie with Statistics, notice has been made of the different ways that data can be manipulated for various ends. “The secret language of statistics”, he writes, “so appealing in a fact-minded culture, is employed to sensationalize, inflate, confuse, and oversimplify.” Neither the data nor the designer of the results need actually “lie”—an accurate graphic representation of the statistics might still mislead the reader. A classic example is in the use of the visual slope that underlies line graphs.

The information presented in each of these figures is the same, as is the “curve.” Nothing is falsified, yet a glance at them could easily mislead the reader, or be easily misinterpreted. While a thoughtful look at the ordinate and abscissa of each graph would help clarify the relationship between the data, graphs are meant to present information quickly, not requiring meditation.

While much attention to has been given to the use of manipulative graphic techniques, especially in advertising, political campaigning, etc., the act of misleading readers might not be intentional. The issue is that graphs are often meant to stimulate action. Risk graphs, for instance, are meant to inform readers of an absolute risk (X amount of radiation causes cancer), or relative risk (Y amount of cigarettes each day has an X chance of causing cancer). Studies show that the location of a risk on a graphic risk ladder was more important than the numbers involved. Graphs also make people more risk averse.

One would not think that graphs published in articles in medical journals would intentionally mislead—and I am not suggesting malicious intent in the following examples—but graphs are nonetheless subject to misinterpretation. Consider a study in 1990 that examined how patients interpreted survival graphs relating to outcomes of disease treatments. Shown projections of the outcomes of two different treatments, the majority of patients (and, independently, their physicians) made a decision to pursue a course of treatment that had worse short-term but better long-term survival rates.

12 Peacock, quoted in Tufte, Beautiful Evidence, p. 145.


14 Lipkus 1999
The authors of this study concluded that “A graphic survival curve appears to provide enough information to assess patient preferences between two alternative treatments.” In other words, graphs might be useful for helping patients make decisions. However, when asked what information in particular they learned from the graph, the authors noted that “Patients appeared to differ from physicians and medical students in their interpretation of the curves.” This observation triggered another investigation.

In 1993 the same authors published the results of a study where physicians and patients were independently asked to decide between two alternative treatments with five-year survival rates projected on three different charts. The only difference between the charts was the amount of space under each curve. Curiously, patients were consistent with their decisions for one treatment when presented with each graph, however physicians tended to change their minds. A number of other studies in subsequent years by different investigators showed that other variations in the presentation of the same data led their patients to make different decisions about treatment.

What was particularly disturbing about these findings was that physicians were throughout this same period being advised to use graphs as an effective way to communicate with their patients about life decisions. Assumptions about the efficacy and utter simplicity of the graphic method were prevailing. In fact, this served two functions: first, it promoted physician-patient communication with the intention of encouraging the patient to make decisions about their future instead of the doctor. This is referred to as “participatory decision making”, and is a feature of modern health care. Second, it was another way that doctors could demonstrate their commitment to EBM, incorporating such reports into their everyday clinical practice. In effect, the clinical encounter would be something like: you have a difficult decision to make about alternative therapies, here is statistical evidence regarding respective outcomes to help you make a decision. Eliminating undocumented clinical judgments based on “intuition”, such an episode captured the spirit of twenty-first century, personalized, evidence-based medical practice. No more ex cathedra statements about the natural order of things derived from incommunicable experience. Physicians now present the evidence upon which they based their recommendations for action in graphic form.

What about the tendency of graphs to be mis- (or variously) interpreted? The decision now for the physician is which graph to

---

15 Mazur & Hickman, 1990

16 Mazur & Hickman, 1993

17 Epstein 2004
present, and exactly how to present it. Another study published in 2005 showed that attention to details such as consistency of graph design might help reduce confusion among physicians and patients alike. Researchers from the University of Michigan found that people misinterpreted survival rates when comparing two graphs of the same physical size where one depicted a five-year survival curve while the other depicted a fifteen-year survival curve.

The study concluded that “people often fail to appropriately adjust the risk perceptions they derive from viewing survival curves to account for different lengths of time displayed.” This is a significant misinterpretation the solution to which is not necessarily a matter of asking readers to be more careful. Psychological data suggests that there is a “cognitive bias” in viewing data that is presented in certain ways. Therefore one somewhat obvious step to take towards fixing the problem is to suggest that publishers standardize graphic design, for instance the spacing used to denote units of time. This would mean that the graph depicting fifteen-years would be three times as long as the five-year graph. Interestingly, there is practically no standardization of the graphic method for publication in medical journals. In fact, in 2003, a look at 120 core clinical journals that are catalogued in PubMed showed that only 6% offered author guidelines to the preparation of graphs. That is to say, 94%, or 113 journals, left it entirely to the author’s discretion. (In 2000, JAMA appointed its first graphs technical editor with the responsibility of examining graphic content.)

There may be no intention to mislead readers of graphs, but graphic designers, like the statisticians who crunch numbers, are nevertheless full of intentions with regard to the production of data. What needs to be more carefully examined is the relationship between information and evidence, the latter of which is not only subject to interpretation (as is everything), but is artfully produced.

**The Misinterpretation of Graphs**

There are a number of problems with the need to rely on evidence to inform judgment. First, there may be no evidence. Data may be “suggestive, but not diagnostic” (or sufficient). Second, evidence may contradict the point one wishes to make or the action one desires to stimulate. Third, there is too much evidence to keep on top of. As one medical researcher recently wrote, “to keep abreast with the continuously increasing number of publications in health research, a primary health care professional would need to read an insurmountable number of articles every day covered in more than 13 million references and over 4800 biomedical and health journals in Medline alone.” Information overload relates to crisis. It is what crashes the system. Information overload might be the schizophrenic condition that poses the biggest challenge to the tenets of Evidence Based Medicine. This is where graphs come in. So in 2006 she published a brief “practical guide” to interpreting and understanding meta-analysis graphs.

---

18 Zikmund-Fisher, et al 2005
These instructions on how to read the graphs (or are they tables?) are helpful, but underdetermined. There is more to it. In his 1990 book *Envisioning Information*, Edward Tufte, professor of political economy and statistics at Yale University, already wondered about the ability of members of the information age—“we”—to exist in information-thick worlds. “Doesn’t data”, he asks, “have to be ‘boiled down’ and ‘simplified’?” The answer happens to be yes, but he complains that the question—which he posed apparently on behalf of an imagined bewildered spectator of the times—misses the point that he wants to make. The point is that “the quantity of detail is an issue completely separate from the difficulty of reading. *Clutter and confusion are failures of design, not attributes of information.*”\(^{19}\)

In other words, information overload is merely a condition that a good model of information management can fix. It is all about presentation.

For Tufte, there is nothing to fret about. However complex, the world is representational, and humans turn out to be masters of inscription, capable of selecting and editing out irrelevance to find order. (Tufte seems to think that there is a natural order that is revealed when one looks at it with the right chart, rather than imagining that such an order is a product of the design of the chart.) “Simplified” data is not a manipulation of this reality—the really complex world that is our context—but the right way of looking at the world. An inability to frame beautiful data, the inability to edit out nonsense to get that clear picture, is what renders a sense of information overload. Confusion, and therefore uncertainty, is not a reflection of the world’s complexity, but a pathological state of being. It is a symptom of one’s inability to edit information. His theory, as Lawrence Schehr has pointed out, is reminiscent of the anthropologist-cum-cyberneticist Gregory Bateson’s etiology of schizophrenia. In this world view, the relatively few people who are victims of a “double-bind”, whereby they receive contradictory bits of information or messages that they internalize and grow incapable of sorting out, stuck in a state of confusion and undecidability.

What makes Tufte particularly pertinent to a discussion about evidence-based medicine is that the form of communication that he analyses is so central to what constitutes “evidence” in modern medicine—the presentation of statistically-informed research. Unlike most logical positivists, post-structuralists or deconstructionists who have argued from both sides of the nature/representation divide but who share concerns to analyze the power of narrative, Tufte took on the challenge of examining how the graphic method is used to communicate matters of fact about the world.

Tufte actually says very little about medicine, but his analysis of the graphic method is I think usefully applied to the study of how biomedicine produces evidential statements. What I think is particularly revealing, if also somewhat disturbing, is that it appears the majority of statistical evidence presented in published biomedical research would be quarantined by Tufte for being pathological. This matters not because Tufte’s theory of reality or representation is at issue, but because it raises questions about how evidence is constituted, interpreted and eventually used to make medical decisions.

* * *

Evidence is never an end in itself but rather is suggestive of a course of action. Moral conduct in medicine—the morality of decision making—is not reached by appealing to evidence, it emerges in the acts of interpretation of what evidence means. Graphs that represent research only have value when they are integrated into a system of currency that defines its value.

A challenge to using well-designed statistical data which prevents it from having maximum value is that it is terribly boring. Biased studies produce much more interesting results. Dr. Thomas Chalmers, pioneer of meta-analysis and anointed father of evidence-based medicine, knew this. He coded a number of reports on medical techniques and demonstrated a correlation between how well a study

\(^{19}\) Tufte, *Envisioning Information*, p. 51, emphasis in original.
was designed (RCT) and the “enthusiasm” over its findings. He discovered that in well designed studies, there was zero enthusiasm for the findings.\footnote{Chalmers in Tanur, et al, \textit{Statistics}, p. 135.}

In a world of digital data mining, the graphic method gives the impression of order and simplicity where there is none. Original research is chaotic; it traces the contours of the unknown and presents to the world its latest sketch of what the universe might look like. Saying that one uses evidence to make decisions gives the impression that the decision is itself calculated. But this hides much that is implicit in the act of deliberation.

Physicians and patients are left in what I think is a moral conundrum. Thomas Newman, who I cited at the beginning of the paper, is an example. One evening while watching a video at home he received a call from the hospital regarding an infant with borderline high bilirubin levels but otherwise healthy looking. “I kept asking my wife to stop the film so I could fret about what to do,” wrote Newman. Hospital guidelines said to admit, but statistics (and maybe his own experience) told him to ignore it. And that is what he recommended the nurse do, “and of course,” he continues—throwing off the yoke of moral angst—“the baby did just fine.” But the process of making the decision and the way he worked around what he might even call his “better judgment” was obtrusive. “Who needs this?”, he asked himself. “Next time I’ll just follow the guidelines and admit such kids to the hospital, so I don’t have to worry and can enjoy my movie.”\footnote{Newman 2003, p. 1425.}

Newman’s reasoning and actions not only remind us of the humanity and uncertainty that underlies medical practice, but his insight to the various decisions facing physicians demonstrate what is at stake in working through the evidence. Newman had referred to the role of the “availability heuristic” in shaping perceptions of probability (how readily we can recall a particular case). However there are a whole range of what the legal theorist Cass Sunstein calls “moral heuristics” which are at work all the time in human decision making, which reduce “complex tasks of assessing probabilities and predicting values to simpler judgmental operations.” He is talking about mental short-cuts and rules of thumb that people tacitly use. The use of these heuristics gives rise to intuitions about what is true. Sunstein wants to point out, and I think that Newman would agree, that heuristics also (of course) lead to error. But we are not analyzing human error here, we are examining the art of decision making and the constitution of evidence. It is the \textit{epistemic time} in which decisions are calculated (to use a biased word) that is of interest. As Malcolm Gladwell shows, intuitive judgment can be as fast as a \textit{Blink}.

Even though decisions are often based on incommunicable knowledge, credibility rests on documentation. “Our world requires that decisions be sourced and footnoted, and if we say \textit{how} we feel, we must also be prepared to elaborate on \textit{why} we feel that way.”\footnote{Gladwell, \textit{Blink}, p. 52.} Is it really obvious that the acts of documentation, inscription, representation are more reliable? In the words of a group of physicians who in 2002 wrote about diagnostic tests: “Is this the direction we wish to take?”

The developments in how information is presented offer opportunities to put substance into commonplace healthcare discussions. But does this swing the balance away from the art of medicine? Will it become less of a ‘high touch’ discipline, in which professionals try to support patients through episodes of illness, and more of a ‘high tech’ one, in which reductionist approaches see pathways of illness as a series of dilemmas that can be ‘solved’? There may be intangible, even mysterious, value in the softer art of medicine—is this being endangered?

The practice of medicine will probably continue to be a matter of cherry-picking. But with the help of the social sciences, studying those who study us, it will be a practice of critical reflexivity about how it is that we think we know what we know. It may be difficult to articulate, but therein is the art of evidence.