Does Evidence-Based Hearsay Determine the Use of Medical Treatments?

John P.A. Ioannidis MD, DSc

Stanford Prevention Research Center, Department of Medicine, and Department of Health Research and Policy, Stanford University School of Medicine and Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, and Meta-Research Innovation Center at Stanford (METRICS), Stanford, California, USA

Correspondence: John P.A. Ioannidis, MD, DSc, Medical School Office Building, Room X306, 1265 Welch Rd, Stanford, CA 94305, USA. E-mail: jioannid@stanford.edu
Does Evidence-Based Hearsay Determine the Use of Medical Treatments?

What shapes the use of medical treatments? In theory, the answer seems trivial. Medical treatments require evidence generated from randomized trials, which in turn undergo regulatory review that leads to licensing. Additional evidence from other clinical research as well as patient experience may also further support the wider use of a drug or biologic. Systematic reviews, meta-analyses, and guidelines synthesize and interpret all this evidence and further shape the use of specific treatments. Then the evidence gets applied to specific patient-physician encounters while considering the needs of each individual patient, which is what one might call the evidence-based medicine approach to use of medical treatments (Sackett et al., 1996).

In reality, the use of many, perhaps most, medical treatments does not depend so much on these evidence-based medicine principles but on what one could call evidence-based hearsay. Evidence is possible to subvert at all stages of its generation and dissemination, while fierce marketing, rumors, and beliefs become more powerful than data. Evidence-based hearsay does contain some evidence in the form of numerical data. The days when a few prestigious academic opinion makers could promote the use of drugs with no data are largely gone. Experts now are required to show data such as trials, meta-analyses, and guidelines with numbers, in order to be considered reputable in the medical community (Ioannidis, 2016). Still, perhaps the most influential experts nowadays are patients themselves; patients with the same disease who have tried the same medical intervention and who share their opinions with other patients. Their paradigm has the benefit of representing “real life”. These are real patients telling their real stories. They offer evidence, which can be either qualitative or quantitative, stating exactly what happened to them and offering detailed information about their disease and its evolution or
Patients may exert influence on other patients in many ways. Some depend on communication between single patients or small patient social networks. For example, patients may discuss their medications and (in some community settings) even share their medications with their close relatives and friends. Physicians who insist on asking may find out that their patient has started a new antihypertensive pill because his friend liked it and it tasted better than the pill that their physician had prescribed. However, the greatest impact of such patient recommendations occurs in larger social networks. The largest possible social networks exist in public online dissemination of information on the web. Some databases, like the Database of Individual Patients’ Experience of Illness (DIPEx) collect patient experiences that other patients around the world can then peruse (Herxheimer et al, 2000) and may have some structure and quality control applied by knowledgeable researchers who have built them. However, most information deposited online by patients has no substantial quality screening. A potentially highly influential way for patients to exert their influence is by writing online product reviews. The practice of writing product reviews is widespread for diverse products, not just medical treatments. Customers are prompted to use specific drugs much in the same way they are enticed to buy specific books or see specific movies.

Given this framework, de Barra’s (2017) study makes an important and timely contribution. It quantifies the bias that maybe inherent in such online product reviews by comparing what these patient tools convey about the effectiveness of a treatment versus what we know about the effectiveness of the treatment from more rigorous clinical research, such as
randomized controlled trials. de Barra (2017) documents that people with good treatment outcomes are more inclined to write online medical product reviews.

The evaluation covered three treatments, juxtaposing their online reviews on Amazon.com (n=1675) against the results noted in the respective randomized controlled trials of the same treatments and indications. Previous work (de Barra, Erriksson, & Strimling, 2014) had looked at alternative treatments (diet and fertility/vitamin interventions), where perhaps it is more easily anticipated that people who did not achieve the desired outcome did not dare share their experimentation with interventions of questionable value. The three treatments in the current study pertained to either cholesterol reduction (Benecol, CholestOff) or weight loss (Orlistat). The reported effectiveness seemed more impressive in the online reports for all three treatments. The difference between reported effectiveness and clinically-proven effectiveness was substantial for all three treatments, and it would change their perceptions from medications of no or small benefit of dubious clinical value to ones with clear clinical indication. The average cholesterol decrease from baseline for Benecol was 14 mg/dl in randomized controlled trials versus 45mg/dl in online reviews. The respective numbers were 13 mg/dl versus 31 mg/dl for CholestOff. The distortion is overall very large, and it is larger for Benecol than for Cholestoff. Orlistat is a minimally effective drug in randomized trials, achieving weight loss corresponding to only -0.14 standard deviations. In contrast, based on the online reviews, it is almost achieving a fourfold change in weight (-0.59 standard deviations). Of note, these effect sizes represent differences between the follow-up and baseline value in the treatment arms of each trial. This is an artificial choice for trial data and aims to correspond to the same information that online reviews communicate. Therefore, the changes include not only the true effect but also regression to the mean and whatever placebo effect may be present. All three treatments have small or
negligible true benefits in randomized trials, but in online reviews they all appear to be very effective.

We should caution that most online reviews written by patients do not contain explicit quantitative information. Most patients use qualitative statements to show how much they were satisfied with some drug and what their experience had been. It is difficult to translate qualitative statements into numerical data. Moreover, it is unknown whether patients may be more likely to report numbers when their review is positive or negative. Online reviews still rate the patient experience with a star system and de Barra’s (2017) study found no significant difference in the average number of stars in online reviews that had numbers versus those that did not. That said, this result cannot totally exclude some modest selection bias in the set of reviews with numbers. Regardless, both quantitative and qualitative statements made by patients may represent powerful hearsay evidence for other patients.

There is at least one major counterargument for the presence of selection bias in online reviews with quantitative information. One may argue that online reviews are not biased (or, at least not as biased as de Barra suggested), but that instead randomized controlled trials are biased to give deflated results compared to the true effect. In this interpretation, online reviews are more trustworthy as they reflect real-life experience with these drugs. Occasionally, some epidemiologists may question the gold standard status of randomized trials as the highest level of evidence. These critics come from different walks of life and have different viewpoints, but their common denominator is their shared belief in using observational data for estimating treatment effects. The record of traditional observational epidemiology (e.g., cohort studies) in making causal claims, in particular about treatment effects, has been miserable: Stan Young has recorded 52 claims made in highly cited observational studies, whereas subsequent randomized controlled
trials refuted all 52 of them (Young & Kerr, 2011). Nevertheless, a huge bubble of comparative effectiveness research grew upon the expectation that observational data could replace randomized trials in estimating effectiveness with sufficient accuracy, and perhaps even higher relevance and generalizability, than trials could do. Lower costs, the convenience of collecting observational data routinely, and the availability of newer methods for reducing bias (e.g., propensity methods) enhanced the arguments in favor of this strategy. Despite high hopes, empirical evidence has shown that data from this new generation of routine, large-scale observational data collection still systematically inflate treatment effects, even for mortality, as compared with subsequent randomized trials on the same treatment comparison (Hemkens, Contopoulos-Ioannidis, & Ioannidis, 2016a). There are many reasons why the observational path may eventually be more expensive, less generalizable, and more misleading (Hemkens, Contopoulos-Ioannidis, & Ioannidis, 2016b). The cost of many observational databases and their infrastructure (e.g., electronic medical records) can be as high or higher than that of few targeted randomized trials. Many of these data are highly selected (e.g., patients referred to specialists in a health care system) and, in the case of online patient reviews, the selection bias may be even larger. In fact, the selection bias is largely of unpredictable magnitude and there is no method to effectively and reliably correct for it. Although some methods exist for bias modeling in observational data (Thompson et al., 2011), none of these currently would be applicable with any reliability to online survey data.

With many powerful academicians, lobbyists, professional societies, funding agencies, and perhaps even regulators shifting away from trials to observational data, even for licensing purposes, clinical medicine may be marching headlong to a massive suicide of its scientific evidence basis. We may experience a return to the 18th century, before the first controlled trial on
scurvy. Yet, there is also a major difference compared with the 18th century: now we have more observational data, which means mostly that we can have many more misleading results.

Moreover, in the current environment, even published randomized trials may be more likely to have inflated results (Gluud, 2006). If so, this makes the inflation seen in observational data and in online patient reviews even more worrisome and misleading. Published evidence from randomized trials (and the systematic reviews, meta-analyses and guidelines that they feed into) is already an amalgam of evidence-based medicine and hearsay. There is extensive evidence that even today many trials do not pre-register their protocols, and few pre-register their protocols with full details about the analytical plans and outcomes. About half of randomized trials remain unpublished and among those that do get published, selective outcome and analysis reporting, frank manipulation of outcomes, inappropriate spin, and misleading interpretations are highly prevalent (Ioannidis, 2014). Similar biases get propagated and enhanced in systematic reviews, meta-analyses, and guidelines that use this randomized evidence. Court documents have revealed the intense manipulation of this evidence by sponsoring companies (Vedula et al, 2012).

One may argue that randomized trials, meta-analyses, and guidelines are more likely to be targets of manipulation by eager sponsors than by single patient who express their enthusiastic or disapproving experiences honestly in online reviews. However, the spillover and impact of evidence-based hearsay eventually concentrates on patients. Patients receive drugs that are endorsed by regulators, prescribed by their physicians, and marketed favorably by sponsors. These steps create an aura of positive expectations. Bias at multiple steps may be pushing in the same direction in this whole chain of misinformation.

If we dismantle randomized trials and licensing becomes dependent on observational data that show some preliminary safety, the impact of online patient reviews may become even more
influential. Moreover, whichever piece of the evidence-based hearsay chain is more influential is likely to receive the attention of marketing and data manipulator teams from the sponsors. If subversion and manipulation can happen in randomized trials, online patient reviews can easily become the target of even more pervasive subverting and manipulating efforts. Creating hearsay in the absence of rules should be very easy. Online patient reviews have practically no rules currently. By their very nature, they cannot have rules of the sort that scientific work has, such as protocols, analysis plans, registration, peer-review, and the like. Patients express their views freely, and therefore subverted and manipulated opinions are unavoidable. Marketing and advertising is, alas, an effective science.

Options to decrease the impact of evidence-based hearsay are limited. Well-trained physicians, pharmacists or other health practitioners may help, but countering misleading evidence by health professionals is not easy. Academic detailing methods (Sumera, & Avorn, 1990) have some effectiveness for counter-balancing marketing advertisement, but their effectiveness is usually limited and most of the time resources for their wide implementation are lacking. It is also unknown how effectively we can alert the public about the limitations of observational data and, in particular, of online reviews for making treatment decisions. We need additional studies like de Barra’s (2017) to evaluate the generalizability of the inflation in the treatment effects for different drugs and different settings. A caveat consumer is due in online reviews and we need to inform people about the potential for bias and the anecdotal nature of these assessments. Such information about the potential of bias is also relevant for all the other parts of the evidence-based hearsay chain. Transparent communication of both the evidence and its accompanying caveats may help. It remains to be studied how exactly this communication should happen and who the most effective communicators might be. Positive illusion, impact of
concrete anecdotes, and emotions may be more powerful than rational decision-making. Peers (other patients) may thus be essential communicators to counter bias.

These communication efforts should not fall into the trap of nihilism or unhealthy skepticism about all medical treatments. There are many treatments that we know to be both effective and safe. Some of them are actually underused (Glazsiou et al, 2017). These effective treatments are often avoided for no good reason. For example, many parents are declining childhood vaccinations based on the dissemination of fraudulent research about vaccines and autism or other weird speculations. One can only dread what would happen if online patient reviews became the target of anti-vaccine propagandists. Creating negative hearsay evidence has never been so easy.

In the same vein, monitoring of online patient reviews and other means of wide-impact non-peer reviewed information is useful to perform for multiple other drugs besides the three medications that de Barra (2017) evaluated. Such monitoring could give us more insights about the dynamics of this hearsay. It may help us understand the factors that shape how often this hearsay biases evidence in favor or against specific interventions and whether this bias is stable over time or can change in magnitude or even direction depending on various circumstances. Understanding the science of hearsay may be useful to help us salvage evidence-based medicine.
References


