Why People Overestimate Medicine's Value: The File Drawer Problem in Lay Health Communication

Mícheál de Barra

Author affiliation

Institute of Applied Health Sciences, University of Aberdeen

Author contact

Address: 2nd Floor Health Sciences Building, Foresterhill Campus, Ashgrove Road, Aberdeen,
AB25 2ZD Scotland.

Email: mdebarra@abdn.ac.uk

Tel: 0044 7591052617

Keywords

health informatics; eHealth; medical overuse; word of mouth; cultural evolution.

Abstract

Objectives. People often hold unduly positive expectations about the outcomes of medical treatment. Here I test the following explanation: people who have a positive outcome tend to tell more people about their disease/treatment than people with poor or average outcomes. Akin to the file drawer problem in science, this systematically and positively distorts the information available to others.

Method. If people with good outcomes are more inclined to tell other people, then they should also be more inclined to write online medical product reviews. Therefore, average outcomes in these reviews should be more positive than those found in RCTs. Data on duration of treatment and outcome (i.e., weight / cholesterol change) was extracted from user-generated health product reviews on Amazon.com and compared to RCT data for the same treatments using ANOVA. The sample included 1,675 reviews of cholesterol reduction (Benecol, CholestOff) and weight loss (Orlistat) treatments and the primary outcome was cholesterol change (Bencol and CholestOff) or weight change (Orlistat).

Results. In three independent tests, average outcomes reported in reviews were substantially more positive than outcomes reported in the medical literature ($\eta^2 = 0.01$ to 0.06 ; $p = .04$ to $.001$). For example, average cholesterol change following use of Benecol is -14mg/dl in RCTs and -45mg/dl in online reviews.

Conclusions. People with good outcomes are more inclined to share information and this distorts the information available to others. Patients who select treatments based on word of mouth, electronic or real life, are likely to have unduly positive expectations.

Introduction

People often use medical treatments that are unlikely to have a direct therapeutic benefit (Ernst and Singh, 2006; Evans et al., 2010; Wootton, 2006). Within conventional medicine,

this is described as medical overuse or over-treatment and it is common in both prescribed medications and procedures and, importantly for the present work, in over-the-counter medicine use (Busfield, 2015). Moreover, medical systems like herbal, alternative, complementary, Aruvedic, and Chinese medicine remain popular despite offering few treatments with demonstrable therapeutic benefits (though it is likely that some users experience benefits besides those typically emphasised in biomedical science). For example, over 100 million Europeans currently use traditional and complimentary medicines (WHO, 2013), few of which are supported by evidence. For the purposes of this paper, the term ‘medical overuse’ is used to designate the use of a conventional or alternative therapy that would not be used had the patient had full knowledge of outcome probabilities. The primary focus is on the causes of medical overuse and in particular non-prescription medicines; for a review of the economic, health and environmental consequences of medical overuse, see Thomas and Depledge (2015).

The financial interest of the pharmaceutical and other healthcare industries is undoubtedly an important driver of medical overuse (Busfield, 2015; Li et al., 2012; Thomas and Depledge, 2015). However, an additional or interacting cause is that patients often hold preferences for treatments that are likely to have little therapeutic effect. These preferences are particularly important in the context of over-the-counter treatments, but much recent work suggests that patients’ expectations are important drivers of professional behaviour (Coenen et al., 2006; Covvey et al., 2014) and policy decisions (Taylor and Bury, 2007). There is strong evidence that patients hold unduly positive expectations about treatment outcomes. A recent systematic review found that estimations of benefit were unduly high in at least 63% of the samples studied (Hoffmann and Del Mar, 2015). Benefits were underestimated for just 3% of outcomes. For example, most people overestimate the mortality reduction from breast mammography (Gigerenzer et al., 2009) and weight reduction from common diets (Foster et al., 1997).

These positive expectations about treatment outcomes are likely to drive medical overuse. Many basic psychological theories like subjective expected utility theory (Edwards, 1954)

and social learning theory (Bandura and Walters, 1977) emphasise that positive expectations about the outcome of a given behavior make that behavior more likely. Moreover, there is empirical evidence that people who believe that a medicine will have positive effects are more likely to use that medicine (see e.g., the systematic review presented in Horne et al., 2013). Thus, medical overuse is common and it appears to be driven, in part, by unduly positive expectations regarding treatment outcome. Hence the psychological and social processes that influence people’s inaccurate treatment-related beliefs are an important topic of scientific enquiry.

A recent review of medical reasoning by Lilienfeld et al. (2014) catalogues the processes that predispose people to incorrectly attribute positive change to a treatment. For instance, health often improves due to regression-to-the-mean, because the disease is self-limiting or because of placebo effects, and these positive changes can be incorrectly attributed to the medical treatment. These processes may explain why people come to believe that a treatment works *after one has tried it*. But as Hoffmann and Del Mar (2015) demonstrate, people often hold false beliefs about likely treatment outcomes even when the treatment has never been used. Lilienfeld et al. (2014) say less about how health beliefs spread, but it is plausible that people who try a treatment communicate their positive beliefs to friends, family and others.

An alternative explanation of unduly positive outcome expectations was proposed by de Barra et al. (2014). In contrast with the mechanisms listed in Lilienfeld et al. (2014) this theory does not derive its explanatory power from peoples biased reasoning or faulty logic, but instead from features of the health communication process. Inaccurate health beliefs emerge, they argue, because a non-representative subset of treatment outcomes is communicated from person to person. In the remainder of this paper, I will explore the assumptions of and predictions derived from this theory.

When people use a medical treatment there is generally a broad range of outcomes, with some people improving and others deteriorating. A subset of the people who use a given treatment

will communicate their experience to other people. For example, they might tell friends and colleagues that they lost weight after using a weight loss drug or that their cholesterol count has unexpectedly increased since they started using statins. Exposure to this kind of health information is likely to influence the recipient’s health beliefs and health behaviour (see below). Health beliefs based on the outcome experience of small samples of people are likely to be error prone, with some individuals developing overly positive and others developing overly negative treatment expectations. However, what de Barra et al. (2014) additionally suggest is that the subset of individuals who actively communicate information about their treatment/outcome is not representative of the total outcome distribution. Rather, they propose that people who have positive outcomes are *more* inclined to share information than people with negative or neutral outcomes. This has important implications. If there is a positive correlation between outcome positivity and probability of information sharing, then the information circulating about the treatment will be systematically and positively biased because people with poorer outcomes will be relatively rare. Such a communication pattern could account for the unduly positive treatment outcome expectations discussed above. Note that a similar under-representation of poor to middling reviews has been documented in the marketing literature where it is termed the *under-reporting bias*. The same term will be used here, but note ‘bias’ here refers to bias in the statistical sense (i.e., a biased sample) rather than bias in the psychological sense (i.e., a deviation from some normative standard of reasoning, as in Tversky and Kahneman, 1974).

The notion that exposure to other people’s health outcomes might influence subsequent health behaviour is consistent with a range of models of health behaviour. Social cognition models, like the health belief model (Rosenstock, 1966), the theory of planned behaviour (Ajzen, 1985) and descendent theories, assume that people choose to adopt a particular behaviour, in part, because of a belief that that behaviour will lead to a preferred outcome with an acceptably high probability. The crucial issue of how these beliefs are formed is less well described by these theories, but it seems safe to assume that observation of another

person’s outcome is an influential event. Imagine for example, that we encounter a person who speaks highly of a cholesterol reduction drug they have used. This encounter might influence key health behaviour determinants like (a) the subjective probability that using this cholesterol reduction drug will have the desired outcome, (b) how much this outcome – cholesterol reduction – would positively improve our well-being, (c) the degree to which we see the health behaviour, taking a pill twice daily in perpetuity, as achievable for us – our perceived behavioral control, or (d) it might influence our perception of the social norm, that is the degree to which we see ignoring high cholesterol as socially acceptable. That we are tuned to learn from other’s outcomes is also consistent with observational learning theory (Bandura, 1965; Fryling et al., 2011) and several empirical studies (Betsch et al., 2011; Gregory et al., 2011; Takahashi et al., 2003; Winterbottom et al., 2008) as well as the sampling framework within cognitive psychology (Fiedler and Juslin, 2006).

A central prediction from de Barra et al.’s *under-reporting bias* theory is that the reputation of the treatment (i.e., the average outcome among people who choose to describe their treatment to others) should be substantially more positive than the average outcome as measured in a clinical trial of representative patients. Online medical product reviews provide a good arena in which to test this prediction because this form of information sharing leaves a lasting digital trace that can be quantified and analysed. Moreover, the psychological and contextual factors that lead people to share information about medical products appears to be similar in the offline and online contexts (King et al., 2014) suggesting that findings from the online domain may generalise to the offline domain.

Research questions

De Barra et al. (2014) demonstrated that the reputation of several alternative medicines, measured by averaging the outcomes in multiple Amazon.com reviews, is more positive than one might expect based on clinical trials. This present work replicates and extends this finding

in a number of ways. Firstly, this previous work focused on alternative/unorthodox treatments (the Atkins diet, herbal fertility treatments). One possible explanation for this finding is that people who have average/negative outcomes after using an unorthodox treatment may be unlikely to share their experiences because they are ashamed to have made a poor medical decision or because their experiences provide little new information to a broadly sceptical audience. If this is the case, then the explanatory scope of the theory presented here is quite narrow and it cannot explain why outcome expectancies for conventional, commonly used medicines are unduly positive. The first and main objective is, therefore, to test the generality of the *under-reporting bias* by examining if the reputation of conventional treatments (i.e., treatments widely supported by national health agencies and commonly prescribed by medical doctors) is also unduly positive (Research Question 1, RQ1).

The second objective is to assess whether the magnitude of this distortion is similar across pairs of treatments targeting similar health problems (RQ2). If the under-reporting bias has a consistent effect – that is, it biases the reputation of all treatments by a similar degree – then it will have little effect on the average rank order of treatments. That is to say, if treatment A is more effective than treatment B, then treatment A will also have a better reputation than treatment B, though both will have a positively distorted reputation. On the other hand, if the under-reporting bias acts more strongly in some circumstances than in others, the reputation of treatment B might be more positive than the reputation of treatment A. If, as one might expect, people apply the rule *choose the treatment with the best reputation*, then subtle differences in the extent to which under-reporting biases operate may have a substantial effect on people’s treatment choice. Hence a sub-goal of this paper is to examine if the reputational distortion is similar across treatments.

Both the treatment’s true beneficial treatment effect and also the under-reporting bias will contribute to its reputation. The third research question concerns the relative size of these factors. After all, even if the under-reporting bias is present, it might be relatively minor and the reputation might largely be a function of its true therapeutic effect. Although it would be

interesting to compare the consequences of the under-reporting bias to the consequence of the placebo effect, regression to the mean and other factors discussed by Lilienfeld et al. (2014), this is not possible in the present analysis because none of the trials include a no-treatment control and, hence, we do not know what would have happened to the patients had they been left untreated. The third research question (RQ3) is how large is the consequence of the under-reporting bias relative to the size of the treatment’s therapeutic benefit.

In many online reviews, people report telling friends, family or colleagues about their medical treatment. For example, reviewers might report “As I’ve been telling everyone at work, this drug just does not work - a total waste of money!!”. An additional novel prediction derived from the underreporting bias theory is that people who have average or negative outcomes should be less likely to report this kind of information sharing than people who have a positive outcome. This fourth and final research question (RQ4) is important because it bridges offline and online information sharing and would – if supported – lend weight to the argument that the underreporting bias is a generally important process in shaping the reputation of medical treatments.

Note that the present work does not directly address the question of why people with positive outcomes are more inclined to share information. In the discussion section we review relevant findings from other fields and speculate on the implication of these results for this question.

Methods

Design overview

Many online retailers allow people to write a review of products they have purchased. If people who have better outcomes are more likely to tell other people about their experiences, we might also expect them to be more likely to write an online review. The approach of this paper is to test whether medical outcomes reported in Amazon (a large international online

retailer) reviewers are representative of the outcomes of the same products as reported in clinical trials. If no reporting bias exists, the average outcomes reported in online reviews should be broadly similar to the average outcomes reported in the scientific literature.

Data collection

To test this hypothesis, I searched for three medical products that met the following criteria. (1) The medical product should have more than 300 online reviews. Assuming that one in four reviews have analysable data and that online review averages are approximately 0.5 standard deviations (SDs) more positive, this will give a greater than 80% power to detect a difference at the $p < .05$ threshold of statistical significance. (2) Reviews of the product should contain specific quantitative information about their health. Pain medication, for example, would fail to meet this criteria because changes in pain are generally expressed in qualitative rather than quantitative terms. (3) High quality scientifically collected data on the true effect of the treatments should be available. Randomised controlled trials (RCTs) and longitudinal studies allow us to accurately estimate the average outcome when someone begins treatment. (4) Products should be orthodox medical treatments.

Three products met these criteria: Benecol Smart Chews Caramels (made by Raisio) and Nature Made CholestOff (PharPharmavite), two cholesterol reduction treatments, and Alli Orlistat (GlaxoSmithKline), a fat absorption inhibitor for people seeking to lose weight. It is likely that many other treatments also meet these criteria, but these are the three that were discovered first while searching a range of medical products on the Amazon.com. Data on the Atkins diet from de Barra et al. (2014) were also reanalysed for comparative purposes (see de Barra, 2014 for data access). All data came from Amazon.com, the US version of the on-line retailer. No other medical products were assessed.

Estimates of “true effects” were derived from a number of clinical trials, the details of which can be found below.

Treatments 1 and 2: Benecol and CholestOff.

Nine hundred and eight reviews of cholesterol reduction products written on or before March 18th, 2015 were included (Benecol $N = 526$, CholestOff $N = 382$). Information about blood lipid levels and duration of drug use was extracted from each review. For example, from a review “I have been using this product for 2 years. Within the first 3-4 months my cholesterol was down 30 points. Sometimes I forget to take my pill, but usually I remember. Just got cholesterol tested last week: down from 245 to 196.”, the total cholesterol change was -49 mg/dL and the duration was two years. If, as in the case above, change over two time periods was presented, only the longer time period was used. If the review reported lipid changes for two or more individuals (“my wife and I started using this product...”), only the author’s change was recorded. In 161 of the reviews, the reviewer mentioned either a change in total cholesterol or a pre- and post- treatment cholesterol level. Low-density lipoprotein (LDL) change was mentioned in 50 of the reviews.

Lipid changes in Amazon reviews were compared with lipid changes reported in a systematic review (Wu et al., 2009). I extracted the average baseline lipid level and the average follow-up lipid level from each relevant study arm (Devaraj et al., 2006, 2004; Doornbos et al., 2006; Goldberg et al., 2006; Jauhiainen et al., 2006; Korpela et al., 2006; Maki et al., 2001; Matvienko et al., 2002; Mensink et al., 2002; Miettinen et al., 1995; Polagruto et al., 2006; Quílez et al., 2003; Seki et al., 2003; Woodgate et al., 2006). Four of the 20 studies reported in Wu et al. (2009) were excluded: in two cases the original study report was unavailable and in two cases the intervention involved substantial dietary changes, over and above stanols/sterols (Hallikainen and Uusitupa, 1999; Jones et al., 1999).

Treatment 3: Orlistat

In the 767 reviews of the weight-loss drug Orlistat written up to 28 February, 2015, a specific weight change could be calculated in 250 cases. In some reviews, duration of treatment was

expressed in terms of bottles purchased rather than time. Given that Alli guidelines suggest one pill with each fatty meal, I assumed that people used two pills per day. As in the blood lipid drugs above, the longest time period was used when two were available and if data from two or more people were presented in a single review, only information from the author was extracted.

Comparison trials were selected from a recent systematic review of drug treatments for weight loss (Yanovski and Yanovski, 2014). Two of the included trials examined the relevant drug (Orlistat) at the relevant dosage (60mg) (Hauptman et al., 2000; Rössner et al., 2000). Where necessary, results were extracted from figures using a online graph digitiser (<http://arohatgi.info/WebPlotDigitizer/app/>). Percentage changes were converted to absolute changes by multiplying by average weight at time zero. Weight at time of diet onset rather than weight during run in period was taken as baseline.

Data analysis

Research Questions 1 and 2.

RQ1 (Is there a mean difference between Amazon review outcomes and RCT outcomes) and RQ2 (Is that difference consistent across medical treatments) were addressed using ANOVA models. An estimate of “true” treatment effect was created by collating RCT results from different trials using a meta-analysis. Owing to well-recognised problems in the biomedical literature (under-reporting of negative results, outcome switching, testing of products on samples of people where maximum effect is expected rather than testing on representative users, see Charlson and Horwitz, 1984; Dwan et al., 2013; Gurwitz et al., 1992), these estimates are likely to be an overestimate of the true effect and, hence, they make for a conservative test of the hypothesis. Note that secular trends or regression to the mean result in general improvement in patient condition, independent of treatment. It is appropriate, therefore, to compare the Amazon review weight and cholesterol changes to the baseline

versus endpoint clinical changes rather than to the control versus intervention difference. The comparison of RCT outcome versus Online outcome answers RQ1 while the interaction effect (CholestOff v Benecol or Orlistat v Atkins) answers RQ2.

Research Question 3

The reputation of a treatment is a function of both the under-reporting bias and the therapeutic effect, but what is the relative importance of these two factors? To answer this question, I first calculated the treatment's true benefit by taking the difference between the treatment group and control group at endpoint and expressing this difference in standard deviations of baseline variability. Where multiple estimates were available, an average weighted by sample size was used. To calculate the reputational distortion due to under-reporting, I measured the difference between intervention group at endpoint and the average review outcome at the same time point: this was also expressed in standard deviations of baseline RCT variability. To estimate the relative importance of therapeutic benefit and under-reporting bias, these two effect sizes were compared.

Research Question 4

Instances of information sharing were identified as follows. Using the qdap (Rinker, 2013) package, I developed an R script that scanned every review for sentences containing words/strings relating to the act of sharing (told, tell, inform, advi, know) or the likely target of sharing (wife, husband, friend, colleague, family, brother, sister, everyone, no one, everybody, nobody, anyone). If the string appeared in any form, the sentence was extracted. For example, "advi" identified sentences with advice, advised and so on. Each of these sentences was then examined manually to confirm that it described information sharing. Any past or planned information sharing (e.g., "I can't wait to tell my dr what I did INSTEAD OF Lipitor") was coded as a share. Only information directed towards a third party, rather than directed

to the reader, was included. Buying the product for someone was counted as information sharing, as was telling them to buy or not to buy it. Recommending that other people use or do not use the product was also categorised as information sharing. A randomly selected 10% of the reviews were manually examined to ensure that the algorithm was reliably identifying information sharing. In these 164 reviews, there were two discrepancies; in both cases the manual coder missed an instance of sharing that the algorithm identified.

Results

In the supplementary materials I present an overview of the data and describe analyses that suggest that fraudulent reviews are rare/absent in this dataset. Briefly, (1) the outcome distributions are very similar in both RCTs and online reviews suggesting similar data generation processes (e.g., weight loss) rather than distinct data generation processes (e.g., weight loss v fraud) and (2) a propitiatory “fake review” identification algorithm suggests that between 90% and 100% of reviews are reliable.

Research Questions 1 and 2.

Cholesterol reduction

The average cholesterol change reported by Benecol users was -45.3232323 mg/dl (SD = 33.0833296). As Figure 1 indicates, this cholesterol change is substantially larger than that reported in any of the nine comparable trials (range: -24, -9.2807424 mg/dl). Average cholesterol changes listed by CholestOff reviewers was -30.3709677 mg/dl (SD = 41.0877645), again larger than any comparator trial (range: -17.401392, -3.4802784 mg/dl).

A meta-analysis of studies in Figure 1 indicates that cholesterol change while using Benecol is -13.8333444 mg/dl (95% CI: -17.7468858, -9.919803) and the change for CholestOff users

Figure 1: Mean change in blood lipids as reported in Amazon.com reviews in clinical trials. Dot size is proportional to sample size. LDL = Low-density lipoprotein. PSE = phytosterol sterol ester. PS = phytosterol, plant sterol or plant stanol.

is -12.5194175 mg/dl (95% CI: -16.2303425, -8.8084926). These estimates are derived from before and after comparisons in the intervention groups using a fixed effects model. Using a two-way ANOVA, average cholesterol change as reported in Amazon reviews was compared to change estimated from the meta-analysis, see Table 1. Results indicate a strong overall effect of data source: that is, cholesterol reduction was substantially larger in online reviews (see RQ1). There is a statistically significant interaction effect between data source (RCT v Amazon review) and treatment (Benecol v CholestOff). This is consistent with the hypothesis that the under-reporting bias distorts the reputation of Benecol more strongly than it does the reputation of CholestOff (see RQ2).

Table 1: ANOVA comparing cholesterol changes reported in online reviews versus RCTs.

	SS	df	MS	F	<i>p</i>	eta2
Online v RCT	77024	1	77024	54.92	0	0.057
CholesterOff v Benecol	8371	1	8371	5.97	0.01	0.007
Interaction	5885	1	5885	4.2	0.04	0.005
Within	1267754	904	1402			
Total	1359034	907				

Average LDL changes show a similar pattern. The change reported in Benecol reviews, -30.9666667 mg/dl, SD = 17.2436562, was larger than that reported in any trial (range: -26, -8.4153132 mg/dl). In sterol/stanol trials, LDL change range (-23.201856, -2.7068832) mg/dl, did not include the average loss reported in the online reviews: -27.4 mg/dl, SD = 34.7144745. However, because few reviewers recorded specific LDL scores (Benecol N = 30, CholestOff N = 20), no further analysis was performed.

Weight changes

I tested whether average weight loss in online reviews was statistically different to weight loss in RCTs at two to three months and at five to seven months. (Atkins diet data is included for comparison, see de Barra et al., 2014) These periods were selected because all of the relevant clinical trials reported participant outcomes during one (Hauptman et al., 2000) or both (Foster et al., 2003; Gardner et al., 2007; Rössner et al., 2000; Truby et al., 2006) of these periods. Where more than one clinical trial was available, the changes in weight were combined using a meta-analysis. Online review scores were averaged over the same period.

Figure 2: Weight change following use of Orlistat as reported in online reviews (data points and smoothened line with standard error shaded) and randomised controlled trials (broken lines).

Results of the two ANOVA analyses (two to three months and five to seven months) are shown in Table 2. Weight loss was substantially larger among the online reviewers than in the clinical trial participants. Weight loss was larger in the Atkins diet than in the Orlistat medication, but this difference appears to be accounted by true differences in treatment effect

342 rather than by differential distortion (i.e., there is no interaction effect, see RQ2).

Table 2: ANOVA comparing weight change in amazon reviews versus weight change in RCTs for two medical products over two time periods.

	SS	df	MS	F	p	eta2
5 TO 7 MONTHS						
RCT v Online	7430.074	1	7430.074	19.214	0	0.028
Atkins v Orlistat	455.868	1	455.868	1.179	0.28	0.002
Interaction	952.48	1	952.48	2.463	0.12	0.004
Within	258313.208	668	386.696			
Total	267151.629	671				
3 TO 5 MONTHS						
RCT v Online1	1364.129	1	1364.129	4.318	0.04	0.008
Atkins v Orlistat1	2840.358	1	2840.358	8.991	0	0.016
Interaction1	154.446	1	154.446	0.489	0.48	0.001
Within1	177543.953	562	315.915			
Total1	181902.886	565				

Research question 3

Treatments effects (i.e., difference between control and intervention group changes) for Orlistat, Benecol and CholestOff were -.14, -.41 and -.52 SDs of baseline variance, respectively. The reputational distortion for the same three treatments (i.e., the online average minus the RCT intervention group endpoint average) were -.59, -2.58, and -1.55. Thus, for Orlistat, positive reputation derived from the under-reporting bias is about four times larger than positive reputation derived from the treatment's pharmacological effect. For CholestOff and Benecol, the under-reporting enhancement to reputation is three and six times larger than enhancement to reputation that stems directly from the drug action, respectively.

Research Question 4

In 79 of the 1596 Benecol, CholestOff, and Orlistat reviews, an instance of information sharing was identified. Because information sharing was reported rarely, I examined the relationship with number of stars awarded to the product rather than weight / cholesterol change. Stars are a grading system in which users give more stars to products they are satisfied with (possible range: one to five). This increases the sample size available because all reviewers include a star rating but only a subset include a weight or cholesterol change. The data were analysed using an ordered logistic regression in which treatment kind, duration, and sharing (binary, share vs no share) were used to predict number of stars. As Table 3 shows, reviews that included information sharing were 0.55 stars more positive than reviews without a mention of information sharing (95% CI: 0.07; 1.07).

Table 3: An ordered logistic regression analysis predicting number of stars awarded. Note: Sharing refers to reviews that mention the sharing of information concerning the treatment.

	Value	Std. Error	t value	95% CI
No sharing (ref)				
Sharing	0.55	0.25	2.17	0.07; 1.07
Benecol (ref)				
CholestOff	0.19	0.14	1.32	-0.09; 0.46
Orlisat	-0.46	0.11	-4.14	-0.68; -0.24

Discussion

People with positive outcomes are more heavily represented in online reviews than in RCTs. This effect is large, and it is consistent across the three new medical products assessed here. This result is consistent with the hypothesis that people with more positive outcomes are more inclined to write reviews of medical products than people with average or poorer-than-average outcomes and it suggests that the under-reporting bias documented in de Barra et al. (2014) generalises to conventional medicines.

At a time where online user-generated health information is becoming influential in people’s health decision making (O’Neill et al., 2014) this over-representation of good outcomes (or, equivalently, under-representation of average or poor outcomes) has important implications: people who form beliefs and/or make decisions based on these outcomes are likely to engage in medical overuse.

In this study, the reputation of the treatment was benchmarked against RCT results. It

should be noted however, that the population of people who buy and use the medical product from Amazon.com may be different from the population of people who participant in trials. A priori, we would expect this sampling difference to work against our hypothesis. RCTs of commercial health products are themselves subject to a publication bias with “positive” outcomes more likely to be published (Dwan et al., 2013). Furthermore, restrictive eligibility criteria mean that the treatment is tested on the sample where it is most likely to have an effect, rather than on the sample most representative of future users (see e.g., Charlson and Horwitz, 1984; Gurwitz et al., 1992). For these and a range of other reasons (Goldacre, 2012; Ioannidis, 2005; Jones et al., 2013), results of RCTs are an optimistic estimate of what will actually happen in the real world. Thus, while our ignorance about the differences between the online reviewers and the RCT participants is a limitation of this study, it is more likely that this difference inflates the possibility of a false negative rather than a false positive.

These results tentatively suggest that the difference between the real benefit and the reputed benefit is not consistent across treatments: Benecol’s reputation is more distorted than CholestOff’s. This variability in reputational distortion may be important. For the majority of human history – and in some cases today – choices between medical treatments are mainly based on ad hoc observations and word-of-mouth information rather than on the results of carefully controlled trials or meta-analyses (Evans et al., 2010). The data presented here indicate that some treatments will appear better than others for reasons besides true differences in effectiveness. During the long-term evolution of folk-medicine and other non-scientific medical cultures, medical innovations which benefited health may have been less important than medical innovations that effectively milked this distortion. Treatments good at appearing good will spread at the expense of treatments good at doing good.

The reputation of a treatment is a function of three factors: it’s pharmacological effect, the under-reporting bias, and extra-treatment effects like secular trends, regression to the mean, and the placebo effect. This third factor can only be assessed in trials with a “no treatment control arm” as well as a placebo arm because these different factors might cancel each other

out (e.g., substantial regression to the mean and equally large negative secular trends results in no change) and, hence, they cannot be measured. But the pharmacological effect and the under-reporting bias can be tentatively measured. In the three treatments discussed here the under-reporting bias results in three to six times larger improvements in average outcome than the actual pharmacological content of the treatment. This suggests that the under-reporting bias has significant consequences for the reputation of medical treatments.

Under what conditions is the reputation of a medical treatment maximally distorted? The answer to this question depends on the psychological, biological and social processes that cause people with positive outcomes to tell more people about their treatment. The data presented here sheds little light on this issue, but it is worth speculating. One straightforward possibility is that when people continue to have poor health, they remain in a depressed state, with low mood and little energy or motivation to engage in information sharing. However, high cholesterol does not generally result in ill health in and of itself. Another possibility is that people prefer not to dwell on past periods of sustained ill health. A period of recovery, in contrast, is a more positive experience and, hence, people may be more motivated to discuss it – including treatment choices – with others. Alternatively, telling people that your treatment failed involves telling them that you are still sick, and this is something people may want to avoid. A positive outcome, on the other hand, conveys the message that one is now free from the disease. Some people may believe that choosing an ineffective treatment may reflect poorly on their own decision making capacity and hence may be less inclined to share “failures” than “successes”. It is notable that this distortion of reputation is not limited to unconventional and alternative medicines: Orlistat and plant stanols/sterols (Benecol/CholestOff active ingredient) are indicated for weight loss and dyslipidemia, and are promoted by NHS trusts and the British Dietetic Association. Thus, it would not appear to be the case that only counter-intuitive or surprising positive outcomes, such as those that follow the use of less conventional medicines, are shared more frequently. A final possibility is that people with positive and negative outcomes tell others at the same rate *while using the*

treatment, but that people with negative outcomes tend to cease treatment more quickly. In future research, it may be interesting to evaluate if similar mechanisms drive overly positive medical reviews, film reviews, and other non-medical product reviews (c.f., Anderson, 1998; Hu et al., 2006)

This study adds to a body of literature that implicates *sampling biases* rather than *cognitive biases* as the source of erroneous belief. Beliefs are formed in a way that are analogous to the way pollsters assess support for electoral candidates; by sampling a subset of the total population of relevant events (Galesic et al., 2016, 2012). As in electoral and scientific sampling however, the observed sample is for various reasons non representative of the total population. For instance, Fiedler (2000) suggests that many supposed “cognitive biases” are better understood as a consequence of non-representative samples. In a more recent example, Galesic et al. (2012) finds that people’s tendency to overestimate how much better (or worse) off they are relative to others is a consequence of “convenience” sampling the set people within one’s own social milieu, people who are not representative of the population. It is likely that such biased sampling – enabled by processes like the under-reporting bias described here – contribute to many of the misbeliefs in medicine’s chequered history.

In summary, our data replicates and extends de Barra et al. (2014) by showing that people with positive outcomes are more likely to write reviews of weight loss and cholesterol reduction drug treatments. This differential tendency to write reviews results in a large distortion of the reputation of medical treatments. This finding has important implications for health-care: when people rely on word-of-mouth information to evaluate and choose between health products, they are likely to get an unduly positive impression of the curative value of that treatment. A distorted perception of health outcomes may lead to the selection of health treatments that would, were it not for this distortion, be avoided. This communication pattern may be one explanation for the widespread overestimation of treatment benefits (Hoffmann and Del Mar, 2015).

The R analysis syntax, along with the relevant datasets, are available for download (de Barra, 2016)

Acknowledgements

Marie Johnston, Gertrude (Turu) Stadler, Alberto Acerbi, Pontus Strimling, Kimmo Eriksson and Marijn de Bruin helped with different elements of this project. This project was funded in part by the Prof Roy Weir Career Development Fellowship.

References

- Ajzen, I., 1985. From intention to actions: A theory of planned behavior, in: Kuhl, J., Beckman, J. (Eds.), *Action-Control: From Cognition to Behavior*. Springer, Heidelberg, pp. 11–39.
- Anderson, E.W., 1998. Customer satisfaction and word of mouth. *Journal of Service Research* 1, 5–17.
- Bandura, A., 1965. Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of Personality and Social Psychology* 1, 589.
- Bandura, A., Walters, R.H., 1977. *Social learning theory*.
- Betsch, C., Ulshöfer, C., Renkewitz, F., Betsch, T., 2011. The influence of narrative v. statistical information on perceiving vaccination risks. *Medical Decision Making* 31, 742–753.
- Busfield, J., 2015. Assessing the overuse of medicines. *Social Science & Medicine* 131, 199–206.
- Charlson, M.E., Horwitz, R.I., 1984. Applying results of randomised trials to clinical practice: Impact of losses before randomisation. *Br Med J (Clin Res Ed)* 289, 1281–1284.

477 Coenen, S., Michiels, B., Renard, D., Denekens, J., Van Royen, P., 2006. Antibiotic prescribing
478 for acute cough: The effect of perceived patient demand. *Br J Gen Pract* 56, 183–190.

479 Covvey, J.R., Johnson, B.F., Elliott, V., Malcolm, W., Mullen, A.B., 2014. An association
480 between socioeconomic deprivation and primary care antibiotic prescribing in scotland.
481 *Journal of Antimicrobial Chemotherapy* 69, 835–841.

482 de Barra, M., 2016. Why people overestimate medicine’s value: The file drawer problem in
483 lay health communication. *SocArXiv* <https://osf.io/htkn6/>.

484 de Barra, M., 2014. Reputation of medical treatments: Dataset and analysis files.

485 de Barra, M., Eriksson, K., Strimling, P., 2014. How feedback biases give ineffective medical
486 treatments a good reputation. *Journal of Medical Internet Research* 16.

487 Devaraj, S., Autret, B.C., Jialal, I., 2006. Reduced-calorie orange juice beverage with plant
488 sterols lowers c-reactive protein concentrations and improves the lipid profile in human
489 volunteers. *The American Journal of Clinical Nutrition* 84, 756–761.

490 Devaraj, S., Jialal, I., Vega-López, S., 2004. Plant sterol-fortified orange juice effectively
491 lowers cholesterol levels in mildly hypercholesterolemic healthy individuals. *Arteriosclerosis,
492 Thrombosis, and Vascular Biology* 24, e25–e28.

493 Doornbos, A., Meynen, E., Duchateau, G., Van Der Knaap, H., Trautwein, E., 2006. Intake
494 occasion affects the serum cholesterol lowering of a plant sterol-enriched single-dose yoghurt
495 drink in mildly hypercholesterolaemic subjects. *European Journal of Clinical Nutrition* 60,
496 325–333.

497 Dwan, K., Gamble, C., Williamson, P.R., Kirkham, J.J., 2013. Systematic review of the
498 empirical evidence of study publication bias and outcome reporting bias-an updated review.
499 *PloS One* 8, e66844.

500 Edwards, W., 1954. The theory of decision making. *Psychological bulletin* 51, 380.

501 Ernst, E., Singh, S., 2006. Trick or treatment: The undeniable facts about alternative

502 medicine. W.W. Norton & Co.

503 Evans, I., Thornton, H., Chalmers, I., 2010. Testing treatments: Better research for better
504 healthcare. Pinter & Martin Ltd.

505 Fiedler, K., 2000. Beware of samples! A cognitive-ecological sampling approach to judgment
506 biases. *Psychological Review* 107, 659.

507 Fiedler, K., Juslin, P., 2006. Information sampling and adaptive cognition. Cambridge
508 University Press.

509 Foster, G.D., Wadden, T.A., Vogt, R.A., Brewer, G., 1997. What is a reasonable weight loss?
510 Patients' expectations and evaluations of obesity treatment outcomes. *Journal of Consulting
511 and Clinical Psychology* 65, 79.

512 Foster, G.D., Wyatt, H.R., Hill, J.O., McGuckin, B.G., Brill, C., Mohammed, B.S., Szapary,
513 P.O., Rader, D.J., Edman, J.S., Klein, S., 2003. A randomized trial of a low-carbohydrate
514 diet for obesity. *New England Journal of Medicine* 348, 2082–2090.

515 Fryling, M.J., Johnston, C., Hayes, L.J., 2011. Understanding observational learning: An
516 interbehavioral approach. *The Analysis of Verbal Behavior* 27, 191.

517 Galesic, M., Kaure, A., Gaissmaier, W., 2016. A sampling framework for uncertainty in
518 individual environmental decisions. *Topics in Cognitive Science* 8, 242–258.

519 Galesic, M., Olsson, H., Rieskamp, J., 2012. Social sampling explains apparent biases in
520 judgments of social environments. *Psychological science* 0956797612445313.

521 Gardner, C.D., Kiazand, A., Alhassan, S., Kim, S., Stafford, R.S., Balise, R.R., Kraemer,
522 H.C., King, A.C., 2007. Comparison of the atkins, zone, ornish, and IEARN diets for change
523 in weight and related risk factors among overweight premenopausal women: The a to z
524 weight loss study: A randomized trial. *JAMA* 297, 969–977.

525 Gigerenzer, G., Mata, J., Frank, R., 2009. Public knowledge of benefits of breast and prostate
526 cancer screening in europe. *Journal of the National Cancer Institute* 101, 1216–1220.

Goldacre, B., 2012. Bad pharma: How medicine is broken, and how we can fix it. Harper-
Collins UK.

Goldberg, A.C., Ostlund, R.E., Bateman, J.H., Schimmoeller, L., McPherson, T.B., Spilburg,
C.A., 2006. Effect of plant stanol tablets on low-density lipoprotein cholesterol lowering in
patients on statin drugs. *The American Journal of Cardiology* 97, 376–379.

Gregory, T.A., Wilson, C., Duncan, A., Turnbull, D., Cole, S.R., Young, G., 2011. Demo-
graphic, social cognitive and social ecological predictors of intention and participation in
screening for colorectal cancer. *BMC Public Health* 11, 38.

Gurwitz, J.H., Col, N.F., Avorn, J., 1992. The exclusion of the elderly and women from
clinical trials in acute myocardial infarction. *Jama* 268, 1417–1422.

Hallikainen, M.A., Uusitupa, M.I., 1999. Effects of 2 low-fat stanol ester-containing mar-
garines on serum cholesterol concentrations as part of a low-fat diet in hypercholesterolemic
subjects. *The American Journal of Clinical Nutrition* 69, 403–410.

Hauptman, J., Lucas, C., Boldrin, M.N., Collins, H., Segal, K.R., 2000. Orlistat in the
long-term treatment of obesity in primary care settings. *Archives of Family Medicine* 9, 160.

Hoffmann, T.C., Del Mar, C., 2015. Patients’ expectations of the benefits and harms of
treatments, screening, and tests: A systematic review. *JAMA Internal Medicine* 175, 274–286.

Horne, R., Chapman, S.C., Parham, R., Freemantle, N., Forbes, A., Cooper, V., 2013.
Understanding patients’ adherence-related beliefs about medicines prescribed for long-term
conditions: A meta-analytic review of the necessity-concerns framework. *PloS One* 8, e80633.

Hu, N., Pavlou, P.A., Zhang, J.J., 2006. Can online word-of-mouth communication reveal
true product quality? Experimental insights, econometric results, and analytical modeling.
Experimental Insights, Econometric Results, and Analytical Modeling (April 1, 2006).

Ioannidis, J.P., 2005. Contradicted and initially stronger effects in highly cited clinical
research. *Jama* 294, 218–228.

552 Jauhiainen, T., Salo, P., Niittynen, L., Poussa, T., Korpela, R., 2006. Effects of low-fat
 553 hard cheese enriched with plant stanol esters on serum lipids and apolipoprotein b in mildly
 554 hypercholesterolaemic subjects. *European Journal of Clinical Nutrition* 60, 1253–1257.

555 Jones, C.W., Handler, L., Crowell, K.E., Keil, L.G., Weaver, M.A., Platts-Mills, T.F., 2013.
 556 Non-publication of large randomized clinical trials: Cross sectional analysis. *BMJ* 347, f6104.

557 Jones, P.J., Ntanios, F.Y., Raeini-Sarjaz, M., Vanstone, C.A., 1999. Cholesterol-lowering
 558 efficacy of a sitostanol-containing phytosterol mixture with a prudent diet in hyperlipidemic
 559 men. *The American Journal of Clinical Nutrition* 69, 1144–1150.

560 King, R.A., Racherla, P., Bush, V.D., 2014. What we know and don't know about online
 561 word-of-mouth: A review and synthesis of the literature. *Journal of Interactive Marketing* 28,
 562 167–183.

563 Korpela, R., Tuomilehto, J., Höglström, P., Seppo, L., Piironen, V., Salo-Väänänen, P.,
 564 Toivo, J., Lamberg-Allardt, C., Kärkkäinen, M., Outila, T., others, 2006. Safety aspects and
 565 cholesterol-lowering efficacy of low fat dairy products containing plant sterols. *European*
 566 *Journal of Clinical Nutrition* 60, 633–642.

567 Li, Y., Xu, J., Wang, F., Wang, B., Liu, L., Hou, W., Fan, H., Tong, Y., Zhang, J., Lu, Z.,
 568 2012. Overprescribing in china, driven by financial incentives, results in very high use of
 569 antibiotics, injections, and corticosteroids. *Health Affairs* 31, 1075–1082.

570 Lilienfeld, S.O., Ritschel, L.A., Lynn, S.J., Cautin, R.L., Latzman, R.D., 2014. Why ineffective
 571 psychotherapies appear to work a taxonomy of causes of spurious therapeutic effectiveness.
 572 *Perspectives on Psychological Science* 9, 355–387.

573 Maki, K.C., Davidson, M.H., Umporowicz, D.M., Schaefer, E.J., Dicklin, M.R., Ingram,
 574 K.A., Chen, S., McNamara, J.R., Gebhart, B.W., Ribaya-Mercado, J.D., others, 2001. Lipid
 575 responses to plant-sterol-enriched reduced-fat spreads incorporated into a national cholesterol
 576 education program step i diet. *The American Journal of Clinical Nutrition* 74, 33–43.

577 Matvienko, O.A., Lewis, D.S., Swanson, M., Arndt, B., Rainwater, D.L., Stewart, J., Alekel,
 578 D.L., 2002. A single daily dose of soybean phytosterols in ground beef decreases serum total
 579 cholesterol and LDL cholesterol in young, mildly hypercholesterolemic men. *The American*
 580 *Journal of Clinical Nutrition* 76, 57–64.

581 Mensink, R.P., Ebbing, S., Lindhout, M., Plat, J., Heugten, M.M. van, 2002. Effects of plant
 582 stanol esters supplied in low-fat yoghurt on serum lipids and lipoproteins, non-cholesterol
 583 sterols and fat soluble antioxidant concentrations. *Atherosclerosis* 160, 205–213.

584 Miettinen, T.A., Puska, P., Gylling, H., Vanhanen, H., Vartiainen, E., 1995. Reduction of
 585 serum cholesterol with sitostanol-ester margarine in a mildly hypercholesterolemic population.
 586 *New England Journal of Medicine* 333, 1308–1312.

587 O’Neill, B., Ziebland, S., Valderas, J., Lupiáñez-Villanueva, F., 2014. User-generated online
 588 health content: A survey of internet users in the united kingdom. *Journal of Medical Internet*
 589 *Research* 16.

590 Polagruto, J.A., Wang-Polagruto, J.F., Braun, M.M., Lee, L., Kwik-Urbe, C., Keen, C.L.,
 591 2006. Cocoa flavanol-enriched snack bars containing phytosterols effectively lower total and
 592 low-density lipoprotein cholesterol levels. *Journal of the American Dietetic Association* 106,
 593 1804–1813.

594 Quílez, J., Rafecas, M., Brufau, G., García-Lorda, P., Megías, I., Bullo, M., Ruiz, J.A.,
 595 Salas-Salvado, J., 2003. Bakery products enriched with phytosterol esters, α -tocopherol and
 596 β -carotene decrease plasma LDL-cholesterol and maintain plasma β -carotene concentrations
 597 in normocholesterolemic men and women. *The Journal of Nutrition* 133, 3103–3109.

598 Rinker, T.W., 2013. qdap: Quantitative discourse analysis package. University at Buf-
 599 falo/SUNY, Buffalo, New York.

600 Rosenstock, I.M., 1966. Why people use health services. *Millbank Memorial Fund Quarterly*
 601 44, 94–124.

602 Rössner, S., Sjöström, L., Noack, R., Meinders, A., Nosedá, G., 2000. Weight loss, weight
603 maintenance, and improved cardiovascular risk factors after 2 years treatment with orlistat
604 for obesity. *Obesity Research* 8, 49–61.

605 Seki, S., Hidaka, I., Kojima, K., Yoshino, H., Aoyama, T., Okazaki, M., Kondo, K., 2003.
606 Effects of phytosterol ester-enriched vegetable oil on plasma lipoproteins in healthy men.
607 *Asia Pacific Journal of Clinical Nutrition* 12, 282–291.

608 Takahashi, O., Noguchi, Y., Rahman, M., Shimbo, T., Goto, M., Matsui, K., Asai, A., Onishi,
609 M., Koyama, H., Sawada, I., others, 2003. Influence of family on acceptance of influenza
610 vaccination among japanese patients. *Family Practice* 20, 162–166.

611 Taylor, D., Bury, M., 2007. Chronic illness, expert patients and care transition. *Sociology of*
612 *Health & Illness* 29, 27–45.

613 Thomas, F., Depledge, M., 2015. Medicine “misuse”: Implications for health and environ-
614 mental sustainability. *Social Science & Medicine* 143, 81–87.

615 Truby, H., Baic, S., DeLooy, A., Fox, K.R., Livingstone, M.B.E., Logan, C.M., Macdonald,
616 I.A., Morgan, L.M., Taylor, M.A., Millward, D.J., others, 2006. Randomised controlled trial
617 of four commercial weight loss programmes in the uK: Initial findings from the bBC diet
618 trials. *BMJ* 332, 1309–1314.

619 Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases.
620 *Science* 185, 1124–1131.

621 WHO, 2013. WHO traditional medicine strategy 2014-2023.

622 Winterbottom, A., Bekker, H.L., Conner, M., Mooney, A., 2008. Does narrative information
623 bias individual’s decision making? A systematic review. *Social Science & Medicine* 67,
624 2079–2088.

625 Woodgate, D., Chan, C.H., Conquer, J.A., 2006. Cholesterol-lowering ability of a phytostanol
626 softgel supplement in adults with mild to moderate hypercholesterolemia. *Lipids* 41, 127–132.

- 627 Wootton, D., 2006. Bad medicine: Doctors doing harm since hippocrates. OUP Oxford.
- 628 Wu, T., Fu, J., Yang, Y., Zhang, L., Han, J., others, 2009. The effects of phytosterols/stanols
629 on blood lipid profiles: A systematic review with meta-analysis. Asia Pac J Clin Nutr 18,
630 179–86.
- 631 Yanovski, S.Z., Yanovski, J.A., 2014. Long-term drug treatment for obesity: A systematic
632 and clinical review. JAMA 311, 74–86.