Social Science & Medicine xxx (2017) 1–8

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Reporting bias inflates the reputation of medical treatments: A comparison of outcomes in clinical trials and online product reviews☆

Mícheál de Barra

Health Psychology Group, Institute of Applied Health Sciences, University of Aberdeen, 2nd Floor Health Sciences Building, Foresterhill Campus, Ashgrove Road, Aberdeen, AB25 2ZD Scotland, United Kingdom

ARTICLE INFO

Article history:
Received 24 September 2016
Received in revised form
10 January 2017
Accepted 20 January 2017
Available online xxx

Keywords:
Health informatics
eHealth
Medical overuse
Word of mouth
Cultural evolution

ABSTRACT

Objectives: People often hold unduly positive expectations about the outcomes of medicines and other healthcare products. Here the following explanation is tested: 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 treatment outcomes are more inclined to tell others, then they should also be more inclined to write online medical product reviews. Therefore, average treatment outcomes in these reviews should be more positive than those found in randomised controlled trials (RCTs). Data on duration of treatment and outcome (i.e., weight/cholesterol change) were extracted from user-generated health product reviews on Amazon.com and compared to RCT data for the same treatments using ANOVA. The sample included 1675 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 the reviews were substantially more positive than the outcomes reported in the medical literature ($\eta^2 = 0.01$ to 0.06; $p = 0.04$ to 0.001). For example, average cholesterol change following use of Benecol is −14 mg/dl in RCTs and −45 mg/dl in online reviews.

Conclusions: People with good treatment outcomes are more inclined to share information about their treatment, which distorts the information available to others. People who rely on word of mouth reputation, electronic or real life, are likely to develop unduly positive expectations.

1. Introduction

People often use medical treatments that are unlikely to have a direct therapeutic benefit (Ernst and Singh, 2006; Evans et al., 2010). 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 such as herbal, alternative, complementary, Aruvedic, and Chinese medicine remain popular despite offering few treatments with demonstrable therapeutic benefits (though it is possible that some users experience benefits besides those typically emphasised in biomedical science). For example, over 100 million Europeans currently use traditional and complimentary medicine treatments (WHO, 2013), few of which are supported by scientific 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 have been used had the patient had full knowledge of outcome probabilities. The primary focus of this work is the causes of medical overuse in regards to 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 pharmaceutical and other healthcare industries is undoubtedly an important driver of medical overuse (Busfield, 2015; Thomas and Depledge, 2015). However, an additional or interacting cause is that patients often hold preferences

☆ This research was supported in part by the Prof Roy Weir Career Development Fellowship.

E-mail address: mdebarra@abdn.ac.uk.

http://dx.doi.org/10.1016/j.socscimed.2017.01.033
0277-9536/

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. 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 cognitive theory (Bandura, 1986) emphasise that positive expectations about the outcome of a given behaviour make that behaviour 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 (Horne et al., 2013).

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 prior to use.

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. The remainder of this paper explores 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 individuals developing overly positive or 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. 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 appear to 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* (Anderson, 1998; Hu et al., 2006). However, the under-reporting of negative outcomes might equivalently be described as an over-reporting of positive outcomes. Thus, the term *reporting bias* will be used here, where 'bias' is meant in the statistical sense (i.e., a biased sample) rather than 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 beliefs and 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 not precisely 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 behavioural 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 (Fryling et al., 2011) and several empirical studies (Betsch et al., 2011; Gregory et al., 2011; 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 *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 offline and online contexts (King et al., 2014), suggesting that findings from the online domain may generalise to the offline domain.

1.1. Research questions

de Barra et al. (2014) demonstrated that the reputation of several alternative medicines, measured by averaging the outcomes in multiple Amazon reviews, is more positive than one might expect based on clinical trials. The present work replicates and extends this finding in several ways. First, the 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 so, 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 of this study, research question one (RQ1), is to test the generality of the 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

Please cite this article in press as: de Barra, M., Reporting bias inflates the reputation of medical treatments: A comparison of outcomes in clinical trials and online product reviews, Social Science & Medicine (2017), http://dx.doi.org/10.1016/j.socscimed.2017.01.033

unduly positive.

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 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. 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 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 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 effect and the reporting bias will contribute to its reputation. The third research question concerns the relative size of these factors. Even if the 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 reporting bias to the consequence of the placebo effect, regression to the mean, and other factors discussed by Lilienfeld et al. (2014), it is not possible in the present analysis because none of the trials include a no-treatment control and, therefore, we do not know what would have happened to the patients had they been left untreated. The third research question is how large is the consequence of the reporting bias relative to the size of the treatment's therapeutic benefit (RQ3).

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!!”. An additional novel prediction derived from the reporting bias theory is that people who have average or negative outcomes should be less likely to report 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 reporting bias is a generally important process in shaping the reputation of medical treatments.

2. Methods

2.1. Design overview

Some 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.com (a large international online retailer) reviews are representative of the outcomes of the same products 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. Amazon.com posts all reviews that meet basic criteria (e.g., relate to the product, does not contain personal information).

2.2. Data collection

Three medical products that met the following criteria were sought. First, the products had 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. Second, the reviews contained specific quantitative information about the reviewer's health. Pain medication, for example, would fail to meet this criterion because changes in pain are generally expressed in qualitative terms. Third, high quality, scientifically collected data on the true effect of the treatments were available. Randomised controlled trials (RCTs) and longitudinal studies allow us to accurately estimate the average outcome when someone begins treatment. Finally, products were 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 Amazon.com. Data on the Atkins diet from de Barra et al. (2014) were also reanalysed for comparative purposes (for data access, see de Barra, 2014). 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 several clinical trials, the details of which can be found below. Baseline-versus-endpoint differences rather than control-versus-baseline differences were extracted from these trials. This method enables a fair comparison between the Amazon.com data, which is also derived from baseline-endpoint differences. Thus, change scores from the Amazon.com data and clinical data reflect regression to the mean, placebo effects, and other factors, as well as the true effect of the treatment.

2.2.1. 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. Take, for example, this review:

> I have been using this product for 2 years. Within the first 3–4 months my cholesterol was down 30 points. 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 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). The average baseline lipid level and the average follow-up lipid level were extracted from each relevant study arm (Devaraj et al., 2006; Devaraj et al., 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

Please cite this article in press as: de Barra, M., Reporting bias inflates the reputation of medical treatments: A comparison of outcomes in clinical trials and online product reviews, Social Science & Medicine (2017), http://dx.doi.org/10.1016/j.socscimed.2017.01.033

(Hallikainen and Uusitupa, 1999; Jones et al., 1999).

2.2.2. Treatment 3: Orlistat

In the 767 reviews of the weight-loss drug Orlistat written on or before 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, people were assumed to use two pills per day. As with the blood lipid drugs above, the longest time period was used when two were presented in the review, 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 (60 mg) (Hauptman et al., 2000; Rössner et al., 2000). Where necessary, results were extracted from figures using an 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.

2.3. Data analysis

2.3.1. 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 answered 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 (e.g., 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), 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 difference rather than to the control versus intervention difference. The comparison of RCT outcome versus Online outcome answers RQ1 while the interaction effect (CholestOff vs. Benecol or Orlistat vs. Atkins) answers RQ2.

2.3.2. Research question 3

The reputation of a treatment is a function of both the reporting bias and the therapeutic effect, but what is the relative importance of these two factors? To answer this question, the treatment's true benefit was calculated 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 reporting bias, the difference between intervention group at endpoint and the average review outcome at endpoint was calculated and expressed in standard deviations of baseline RCT variability. To estimate the relative importance of therapeutic benefit and reporting bias, these two effect sizes were compared.

2.3.3. Research question 4

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) was developed to identify instances of information sharing (Rinker, 2013). For example, "advi" identified sentences containing the words "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 at the reader, was included. Buying the product for someone was counted as information sharing, as was recommending that others to buy/use or not to buy/use it. 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.

3. Results

The supplementary materials include 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 vs. fraud) and (2) a propitiatory "fake review" identification algorithm suggests that between 90% and 100% of reviews are reliable. All datasets have been placed in a repository (de Barra, 2017).

3.1. Research questions 1 and 2

3.1.1. Cholesterol reduction

The average cholesterol change reported by Benecol users was −45.32 mg/dl ($SD = 33.08$). As Fig. 1 indicates, this cholesterol change is substantially larger than that reported in any of the nine comparable trials (range: −24.00, −9.28 mg/dl). Average cholesterol change listed by CholestOff reviewers was −30.37 mg/dl ($SD = 41.09$), again larger than any comparator trial (range: −17.40, −3.48 mg/dl).

A meta-analysis of studies in Fig. 1 indicates that cholesterol change while using Benecol is −13.83 mg/dl (95% $CI = -17.75, -9.92$) and the change for CholestOff users is −12.52 mg/dl (95% $CI = -16.23, -8.81$). These estimates are derived from before and after comparisons in the intervention groups using a fixed effects model. Heterogeneity was low in both analyses (CholestOff $I^2 = 0.00\%$, 95% $CI = 0.00\%, 50.33\%$, Benecol $I^2 = 21.52\%$, 95% $CI = 0.00\%, 62.53\%$). Using a two-way ANOVA, average cholesterol change as reported in Amazon reviews was compared to change estimated from the meta-analysis. Results indicate a strong overall effect of data source, $F(1,904) = 54.92$, $p < 0.01$, $\eta^2 = 0.057$, that is, cholesterol reduction was substantially larger in online reviews (see RQ1). There is a statistically significant interaction effect between data source (RCT vs. Amazon review) and treatment (Benecol vs. CholestOff), $F(1,904) = 4.2$, $p = 0.04$, $\eta^2 = 0.005$. These results are consistent with the hypothesis that the reporting bias distorts the reputation of Benecol more strongly than it does the reputation of CholestOff (see RQ2).

Average LDL changes show a similar pattern. The change reported in Benecol reviews, −30.97 mg/dl ($SD = 17.24$), was larger than that reported in any trial (range: −26.00, −8.42 mg/dl). In sterol/stanol trials, LDL change range (−23.20, −2.71) mg/dl, did not include the average loss reported in the online reviews: −27.40 mg/dl ($SD = 34.71$). However, because few reviewers recorded specific LDL scores (Benecol $N = 30$, CholestOff

Please cite this article in press as: de Barra, M., Reporting bias inflates the reputation of medical treatments: A comparison of outcomes in clinical trials and online product reviews, Social Science & Medicine (2017), http://dx.doi.org/10.1016/j.socscimed.2017.01.033

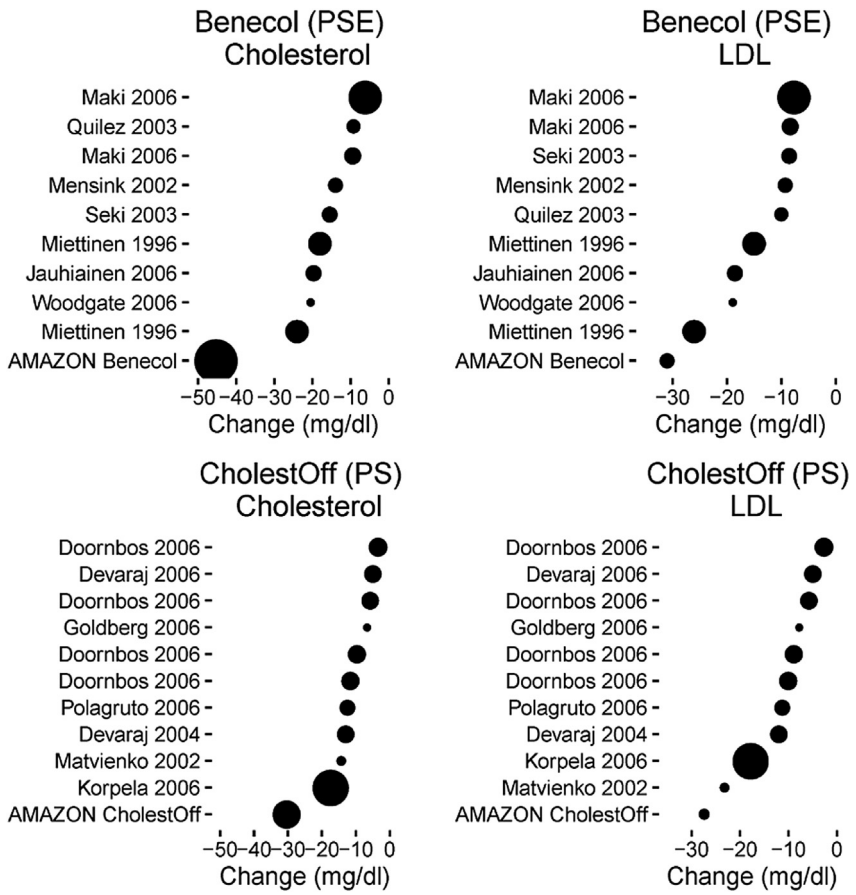

Fig. 1. Cholesterol change reported in Clinical Trials and Amazon.com Reviews. Mean change (i.e. baseline v endpoint) in blood lipids as reported in Amazon.com reviews and in clinical trials. Dot size is proportional to sample size. LDL = Low-density lipoprotein. PSE = phytosterol sterol ester. PS = phytosterol, plant sterol or plant stanol.

$N = 20$), no further analysis was performed.

3.1.2. Weight changes

Is average weight loss in online reviews statistically different to weight loss in RCTs at two to three months and at five to seven months? These periods were selected because all 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: Study effects were homogeneous for Orlistat trials at five to seven months ($I^2 = 0.00\%$, 95% $CI = 0.00\%$, 0.00%) and Atkins trials at five to seven months (I^2: 0.00%, 95% $CI = 0.00\%$, 0.00%); heterogeneity was present for Atkins trials at two to three months ($I^2 = 70.97\%$, 95% $CI = 1.25\%$, 91.47%). Online review scores were averaged over the same period. Only one trial assessed Orlistat weight change at two to three months.

Consistent with the data graphed in Fig. 2, results of the two ANOVA analyses suggest that weight loss was substantially larger among the online reviewers than in the clinical trial participants (two to three months: $F(1,562) = 4.32$, $p = 0.04$, $\eta^2 = 0.008$; five to seven months: $F(1,668) = 19.21$, $p < 0.01$, $\eta^2 = 0.028$). The average weight loss was larger in the Atkins diet reviews than in the Orlistat medication reviews, but this difference appears to be accounted by true differences in treatment effect rather than by differential distortion (i.e., there were no interaction effects, see RQ2: two to three months: $F(1,562) = 0.49$, $p = 0.48$, $\eta^2 = 0.001$; five to seven months: $F(1,668) = 2.46$, $p = 0.12$, $\eta^2 = 0.004$).

3.2. Research question 3

Treatments effects (i.e., difference between control and intervention group changes) for Orlistat, Benecol, and CholestOff were −0.14, −0.41 and −0.52 *SD*s 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 −0.59, −2.58, and −1.55. Thus, for Orlistat, positive reputation derived from the reporting bias is about four times larger than positive reputation derived from the treatment's pharmacological effect. For CholestOff and Bencol, the reporting bias enhancement to reputation is three and six times larger than enhancement to reputation that stems directly from the drug action, respectively.

3.3. 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, its relationship to *number of stars* rather than to *weight/cholesterol change* was examined. Stars are a grading system in which product reviewers give more stars to products they are satisfied with (possible range: one to five). This increases the sample size 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. Reviews that included

Please cite this article in press as: de Barra, M., Reporting bias inflates the reputation of medical treatments: A comparison of outcomes in clinical trials and online product reviews, Social Science & Medicine (2017), http://dx.doi.org/10.1016/j.socscimed.2017.01.033

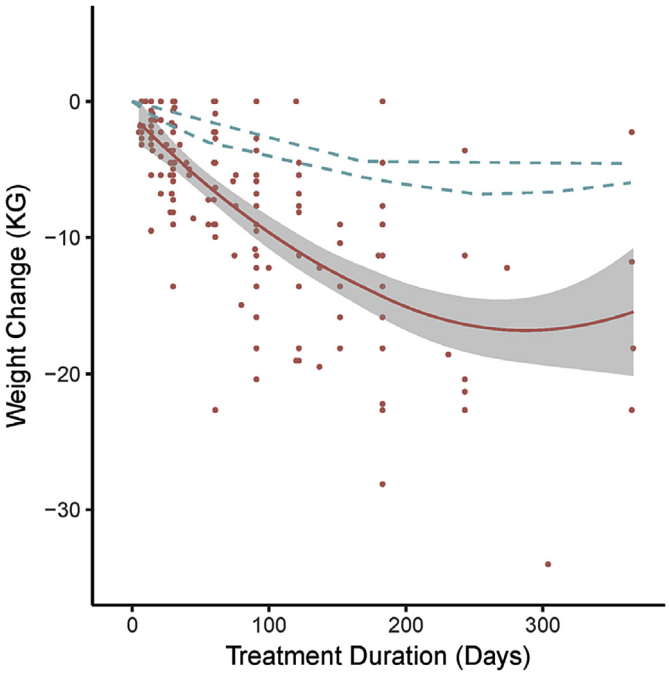

Fig. 2. Weight change reported in Clinical Trials and Amazon.com Reviews. Orlistat Amazon.com review weight change is depicted by the data points (individual reviews) and by the loess smoothened unbroken line and shaded 95% CI area around this estimate. Broken lines depict two comparison randomised controlled trials (baseline versus endpoint changes).

information sharing were 0.55 stars more positive than reviews without a mention of information sharing (95% *CI* = 0.07, 1.07).

4. Discussion

People with positive treatment outcomes are more heavily represented in online reviews than in RCTs. This effect is large and it is present in all 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 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 has important implications: People who form beliefs and make decisions based on these outcomes are likely to engage in medical overuse.

4.1. Inconsistent distortion?

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 Cholest-Off's. (There was no difference in distortion between the Atkins diet and the Orlistat medication, but high heterogeneity in weight loss trials precludes firm conclusions from this comparison.) This variability in reputational distortion may be important. For much of human history — and in some cases today — choices between medical products are mainly based on observations and word-of-mouth information rather than on the results of carefully controlled trials or meta-analyses (Evans et al., 2010). The current study indicates that some treatments will appear better than others for reasons besides true differences in effectiveness. Given that the reporting bias has a three to six times larger influence on the drugs' reputation than the medicinal benefit, it is not surprising that the reputation is influenced by differences in the reporting bias. During the long-term evolution of folk-medicine and other non-scientific medical cultures, medical innovations that benefited health may have been less important than medical innovations that effectively exploited this distortion. Treatments good at appearing effective will spread at the expense of treatments that actually are effective.

4.2. Sampling errors versus cognitive biases

This study adds to a body of literature that implicates *sampling biases* rather than *cognitive biases* as the source of erroneous belief (Fiedler, 2000). According to this perspective, beliefs are formed in a way that are analogous to how pollsters assess support for electoral candidates — by sampling a subset of the total population of relevant events (Galesic et al., 2016). But as in electoral sampling, the observed sample is often unrepresentative of the total population. For example, Galesic et al. (2012) found 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 of people within one's own unrepresentative social milieu. This analysis of online reviews suggests that the treatment-outcomes available as a sample are also unrepresentative and thus the findings provide a parsimonious explanation for some cases of health-related *unrealistic optimism*, also known as the *positive illusion* (Shepperd et al., 2015). It is likely that sampling biases — enabled by processes like the reporting bias described here — contribute to many of the misbeliefs in medicine's chequered history (Wootton, 2006).

4.3. Cultural evolution of a medicine's reputation

Echoing the present findings, diffusion chain experiments find that positively valenced information is transmitted more readily than neutral information (Bebbington et al., 2016). Yet, the same and other studies (Fessler et al., 2014; Moussaïd et al., 2015) found that negative/risk information also has a survival advantage. Why then are negative outcomes not also over-represented in the Amazon dataset? Diffusion chain experiments are an imperfect model of the processes described here. A would-be Amazon reviewer is not deciding whether to relay some N^{th} hand information or not, but rather is deciding whether to describe a health experience which he or she has chosen to undergo. The motives for sharing may be different. But these diffusion chain experiments do suggest that reputation of these treatments will undergo further changes if it is relayed down through several generations of people. This negativity bias, coupled with other content biases (Miton and Mercier, 2015), may explain the unduly negative reputation of vaccines among some individuals.

Under what conditions is the information circulating about 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 results presented here shed little light on this issue, but it is worth speculating. Perhaps 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 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

Please cite this article in press as: de Barra, M., Reporting bias inflates the reputation of medical treatments: A comparison of outcomes in clinical trials and online product reviews, Social Science & Medicine (2017), http://dx.doi.org/10.1016/j.socscimed.2017.01.033

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 (the active ingredient in Benecol/CholestOff) are indicated for weight loss and dyslipidemia, and are promoted by NHS trusts and the British Dietetic Association.

Perhaps the explanation for the distorted reputation rests less on features of human psychology and more on features of the disease process. 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 capacity to engage in information sharing. However, high cholesterol does not generally result in ill health in and of itself. Another 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. Conceptually, this is somewhat like Tanaka et al. (2009) notion that harmful treatments that prolong a disease can be more effective in spreading because they are displayed to others for a longer time than more effective treatments. For the chronic diseases examined here, however, it is more likely that people will continue to use rather than abandon treatments with seemingly positive effects (Colombo et al., 2014; Grandy et al., 2013).

4.4. Limitations

One important limitation of this study is that people who buy products online and people who participate in clinical trials may be different for reasons other than the reporting bias. We might expect these differences to work against the 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 (Charlson and Horwitz, 1984).

Does the reporting bias operate in health communication among friends, family, and colleagues as well as in online medical product evaluation? In the reviews presented here, people who were more satisfied with the product were more likely to describe instances where they told other people about the product (see also King et al., 2014). Nevertheless, tests to confirm or disconfirm this reporting bias in other health communication media (e.g., web forums) or in real-world communication would be a welcome addition to the literature.

The medical treatments studied here were not selected in a systematic fashion. Rather, appropriate criteria were devised and then the Amazon health section was scanned to find qualifying treatments. Undoubtedly several other treatments meet these criteria, and, although unlikely, these results could logically be a result of chance or some selection bias. In the future, automated analysis of multiple online medical products should enable a more comprehensive analysis.

4.5. Conclusions

These results replicate and extend de Barra et al. (2014) findings 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 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).

Acknowledgements

Kimmo Eriksson and Pontus Strimling encouraged me develop the ideas in this manuscript. Turu Stadler advised on some of the statistical analyses and Marie Johnston made many useful suggestions.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.socscimed.2017.01.033.

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. J. Serv. Res. 1 (1), 5–17.

Bandura, A., 1986. Social Foundations of Thought and Action: a Social Cognitive Theory. Prentice-Hall, Inc.

de Barra, M., 2014. Reputation of medical treatments: Dataset and analysis files. Figshare. http://dx.doi.org/10.6084/m9.figshare.843626.

de Barra, M., Eriksson, K., Strimling, P., 2014. How feedback biases give ineffective medical treatments a good reputation. J. Med. Internet Res. 16 (8), 1–10.

de Barra, M., 2017. Reporting bias inflates the reputation of medical treatments: Datasets and analysis syntax. Figshare. https://dx.doi.org/10.6084/m9.figshare.2069646.v3.

Bebbington, K., MacLeod, C., Ellison, T.M., Fay, N., 2016. The sky is falling: evidence of a negativity bias in the social transmission of information. Evol. Hum. Behav. 38 (1), 92–101.

Betsch, C., Ulshöfer, C., Renkewitz, F., Betsch, T., 2011. The influence of narrative v. statistical information on perceiving vaccination risks. Med. Decis. Mak. 31 (5), 742–753.

Busfield, J., 2015. Assessing the overuse of medicines. Soc. Sci. Med. 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 (6454), 1281–1284.

Coenen, S., Michiels, B., Renard, D., Denekens, J., Van Royen, P., 2006. Antibiotic prescribing for acute cough: the effect of perceived patient demand. Br. J. General Pract. J. R. Coll. General Pract. 56 (524), 183–190.

Colombo, O., Ferretti, V.V.V., Ferraris, C., Trentani, C., Vinai, P., Villani, S., Tagliabue, A., 2014. Is drop-out from obesity treatment a predictable and preventable event? Nutr. J. 13 (1), 1.

Covvey, J.R., Johnson, B.F., Elliott, V., Malcolm, W., Mullen, A.B., 2014. An association between socioeconomic deprivation and primary care antibiotic prescribing in Scotland. J. Antimicrob. Chemother. 69 (3), 835–841.

Devaraj, S., Autret, B.C., Jialal, I., 2006. Reduced-calorie orange juice beverage with plant sterols lowers c-reactive protein concentrations and improves the lipid profile in human volunteers. Am. J. Clin. Nutr. 84 (4), 756–761.

Devaraj, S., Jialal, I., Vega-López, S., 2004. Plant sterol-fortified orange juice effectively lowers cholesterol levels in mildly hypercholesterolemic healthy individuals. Arteriosclerosis, Thrombosis, Vasc. Biol. 24 (3), e25–e28.

Doornbos, A., Meynen, E., Duchateau, G., Van Der Knaap, H., Trautwein, E., 2006. Intake occasion affects the serum cholesterol lowering of a plant sterol-enriched single-dose yoghurt drink in mildly hypercholesterolaemic subjects. Eur. J. Clin. Nutr. 60 (3), 325–333.

Dwan, K., Gamble, C., Williamson, P.R., Kirkham, J.J., 2013. Systematic review of the empirical evidence of study publication bias and outcome reporting bias-an updated review. PLoS One 8 (7), e66844.

Edwards, W., 1954. The theory of decision making. Psychol. Bull. 51 (4), 380–417.

Ernst, E., Singh, S., 2006. Trick or Treatment: the Undeniable Facts about Alternative Medicine. W.W. Norton & Co.

Evans, I., Thornton, H., Chalmers, I., 2010. Testing Treatments: better Research for Better Healthcare. Pinter & Martin Ltd.

Fessler, D.M., Pisor, A.C., Navarrete, C.D., 2014. Negatively-biased credulity and the cultural evolution of beliefs. PLoS One 9 (4), e95167.

Fiedler, K., 2000. Beware of samples! A cognitive-ecological sampling approach to judgment biases. Psychol. Rev. 107 (4), 659–676.

Fiedler, K., Juslin, P., 2006. Information Sampling and Adaptive Cognition. Cambridge University Press.

Foster, G.D., Wyatt, H.R., Hill, J.O., McGuckin, B.G., Brill, C., Mohammed, B.S., Klein, S., 2003. A randomized trial of a low-carbohydrate diet for obesity. N. Engl. J. Med. 348 (21), 2082–2090.

Please cite this article in press as: de Barra, M., Reporting bias inflates the reputation of medical treatments: A comparison of outcomes in clinical trials and online product reviews, Social Science & Medicine (2017), http://dx.doi.org/10.1016/j.socscimed.2017.01.033

Fryling, M.J., Johnston, C., Hayes, L.J., 2011. Understanding observational learning: an interbehavioral approach. Analysis Verbal Behav. 27 (1), 191–202.
Galesic, M., Kause, A., Gaissmaier, W., 2016. A sampling framework for uncertainty in individual environmental decisions. Top. Cognitive Sci. 8 (1), 242–258.
Galesic, M., Olsson, H., Rieskamp, J., 2012. Social sampling explains apparent biases in judgments of social environments. Psychol. Sci. 23 (12), 1515–1523.
Gardner, C.D., Kiazand, A., Alhassan, S., Kim, S., Stafford, R.S., Balise, R.R., King, A.C., 2007. Comparison of the Atkins, Zone, Ornish, and LEARN diets for change in weight and related risk factors among overweight premenopausal women: the A TO Z weight loss study: a randomized trial. J. Am. Med. Assoc. 297 (9), 969–977.
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. Am. J. Cardiol. 97 (3), 376–379.
Grandy, S., Fox, K.M., Hardy, E., Group, S.S., et al., 2013. Association of weight loss and medication adherence among adults with type 2 diabetes mellitus: SHIELD (study to help improve early evaluation and management of risk factors leading to diabetes). Curr. Ther. Res. 75, 77–82.
Gregory, T.A., Wilson, C., Duncan, A., Turnbull, D., Cole, S.R., Young, G., 2011. Demographic, social cognitive and social ecological predictors of intention and participation in screening for colorectal cancer. BMC Public Health 11 (1), 38.
Hallikainen, M.A., Uusitupa, M.I., 1999. Effects of 2 low-fat stanol ester–containing margarines on serum cholesterol concentrations as part of a low-fat diet in hypercholesterolemic subjects. Am. J. Clin. Nutr. 69 (3), 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 Fam. Med. 9 (2), 160.
Hoffmann, T.C., Del Mar, C., 2015. Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. JAMA Intern. Med. 175 (2), 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 (12), 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).
Jauhiainen, T., Salo, P., Niittynen, L., Poussa, T., Korpela, R., 2006. Effects of low-fat hard cheese enriched with plant stanol esters on serum lipids and apolipoprotein b in mildly hypercholesterolaemic subjects. Eur. J. Clin. Nutr. 60 (11), 1253–1257.
Jones, P.J., Ntanios, F.Y., Raeini-Sarjaz, M., Vanstone, C.A., 1999. Cholesterol-lowering efficacy of a sitostanol-containing phytosterol mixture with a prudent diet in hyperlipidemic men. Am. J. Clin. Nutr. 69 (6), 1144–1150.
King, R.A., Racherla, P., Bush, V.D., 2014. What we know and don't know about online word-of-mouth: a review and synthesis of the literature. J. Interact. Mark. 28 (3), 167–183.
Korpela, R., Tuomilehto, J., Högström, P., Seppo, L., Piironen, V., Salo-Väänänen, P., et al., 2006. Safety aspects and cholesterol-lowering efficacy of low fat dairy products containing plant sterols. Eur. J. Clin. Nutr. 60 (5), 633–642.
Lilienfeld, S.O., Ritschel, L.A., Lynn, S.J., Cautin, R.L., Latzman, R.D., 2014. Why ineffective psychotherapies appear to work a taxonomy of causes of spurious therapeutic effectiveness. Perspect. Psychol. Sci. 9 (4), 355–387.
Maki, K.C., Davidson, M.H., Umporowicz, D.M., Schaefer, E.J., Dicklin, M.R., Ingram, K.A., et al., 2001. Lipid responses to plant-sterol-enriched reduced-fat spreads incorporated into a national cholesterol education program step i diet. Am. J. Clin. Nutr. 74 (1), 33–43.
Matvienko, O.A., Lewis, D.S., Swanson, M., Arndt, B., Rainwater, D.L., Stewart, J., Alekel, D.L., 2002. A single daily dose of soybean phytosterols in ground beef decreases serum total cholesterol and lDL cholesterol in young, mildly hypercholesterolemic men. Am. J. Clin. Nutr. 76 (1), 57–64.
Mensink, R.P., Ebbing, S., Lindhout, M., Plat, J., Heugten, M. M. van, 2002. Effects of plant stanol esters supplied in low-fat yoghurt on serum lipids and lipoproteins, non-cholesterol sterols and fat soluble antioxidant concentrations. Atherosclerosis 160 (1), 205–213.
Miettinen, T.A., Puska, P., Gylling, H., Vanhanen, H., Vartiainen, E., 1995. Reduction of serum cholesterol with sitostanol-ester margarine in a mildly hypercholesterolemic population. N. Engl. J. Med. 333 (20), 1308–1312.
Miton, H., Mercier, H., 2015. Cognitive obstacles to pro-vaccination beliefs. Trends Cognitive Sci. 19 (11), 633–636.
Moussaïd, M., Brighton, H., Gaissmaier, W., 2015. The amplification of risk in experimental diffusion chains. Proc. Natl. Acad. Sci. 112 (18), 5631–5636.
O'Neill, B., Ziebland, S., Valderas, J., Lupiáñez-Villanueva, F., 2014. User-generated online health content: a survey of internet users in the United Kingdom. J. Med. Internet Res. 16 (4).
Polagruto, J.A., Wang-Polagruto, J.F., Braun, M.M., Lee, L., Kwik-Uribe, C., Keen, C.L., 2006. Cocoa flavanol-enriched snack bars containing phytosterols effectively lower total and low-density lipoprotein cholesterol levels. J. Am. Dietetic Assoc. 106 (11), 1804–1813.
Quílez, J., Rafecas, M., Brufau, G., García-Lorda, P., Megías, I., Bullo, M., Salas-Salvado, J., 2003. Bakery products enriched with phytosterol esters, α-tocopherol and β-carotene decrease plasma lDL-cholesterol and maintain plasma β-carotene concentrations in normocholesterolemic men and women. J. Nutr. 133 (10), 3103–3109.
Rinker, T.W., 2013. Qdap: Quantitative Discourse Analysis Package. Retrieved from. University at Buffalo/SUNY, Buffalo, New York. http://github.com/trinker/qdap.
Rosenstock, I.M., 1966. Why people use health services. Millbank Meml. Fund. Q. 44, 94–124.
Rössner, S., Sjöström, L., Noack, R., Meinders, A., Noseda, G., 2000. Weight loss, weight maintenance, and improved cardiovascular risk factors after 2 years treatment with orlistat for obesity. Obes. Res. 8 (1), 49–61.
Seki, S., Hidaka, I., Kojima, K., Yoshino, H., Aoyama, T., Okazaki, M., Kondo, K., 2003. Effects of phytosterol ester-enriched vegetable oil on plasma lipoproteins in healthy men. Asia Pac. J. Clin. Nutr. 12 (3), 282–291.
Shepperd, J.A., Waters, E.A., Weinstein, N.D., Klein, W.M., 2015. A primer on unrealistic optimism. Curr. Dir. Psychol. Sci. 24 (3), 232–237.
Tanaka, M.M., Kendal, J.R., Laland, K.N., 2009. From traditional medicine to witchcraft: why medical treatments are not always efficacious. PLoS One 4 (4), e5192.
Taylor, D., Bury, M., 2007. Chronic illness, expert patients and care transition. Sociol. Health & Illn. 29 (1), 27–45.
Thomas, F., Depledge, M., 2015. Medicine “misuse”: implications for health and environmental sustainability. Soc. Sci. Med. 143, 81–87.
Truby, H., Baic, S., DeLooy, A., Fox, K.R., Livingstone, M.B.E., Logan, C.M., et al., 2006. Randomised controlled trial of four commercial weight loss programmes in the UK: initial findings from the BBC diet trials. Br. Med. J. 332 (7553), 1309–1314.
Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185 (4167), 1124–1131.
WHO, 2013. WHO Traditional Medicine Strategy 2014–2023.
Winterbottom, A., Bekker, H.L., Conner, M., Mooney, A., 2008. Does narrative information bias individual's decision making? A systematic review. Soc. Sci. Med. 67 (12), 2079–2088.
Woodgate, D., Chan, C.H., Conquer, J.A., 2006. Cholesterol-lowering ability of a phytostanol softgel supplement in adults with mild to moderate hypercholesterolemia. Lipids 41 (2), 127–132.
Wootton, D., 2006. Bad Medicine: Doctors Doing Harm since Hippocrates. OUP Oxford.
Wu, T., Fu, J., Yang, Y., Zhang, L., Han, J., et al., 2009. The effects of phytosterols/stanols on blood lipid profiles: a systematic review with meta-analysis. Asia Pac. J. Clin. Nutr. 18 (2), 179–186.
Yanovski, S.Z., Yanovski, J.A., 2014. Long-term Drug Treat. Obes.: Systematic Clin. Rev. 311 (1), 74–86.

Please cite this article in press as: de Barra, M., Reporting bias inflates the reputation of medical treatments: A comparison of outcomes in clinical trials and online product reviews, Social Science & Medicine (2017), http://dx.doi.org/10.1016/j.socscimed.2017.01.033