Sponsorship bias in clinical trials: growing menace or dawning realisation?

Tom Jefferson
Senior Associate Tutor, University of Oxford; Researcher, Nordic Cochrane Centre; Visiting Professor, Newcastle University, UK
Corresponding author: Tom Jefferson. Email: tj@cochrane.dk

Introduction

If therapeutic decisions in healthcare are to be informed by the results of clinical research, patients and prescribers must be able to trust the research evidence presented to them. In recent decades, the credibility of much of the evidence base for some of the most popular therapeutic and preventive interventions has been undermined by the identification of sponsorship bias.

Sponsorship bias is the distortion of design and reporting of clinical experiments to favour the sponsor’s aims. By using the word ‘sponsor’, I am not implying here that the origins of bias are solely or principally commercial. Sponsors are the funders and stakeholders active within the design, setting-up, running and reporting of clinical trials, including the members of research teams.

Until recently, the distortions introduced by sponsorship bias were recognised as important but difficult to identify with certainty because of the secrecy surrounding pharmaceutical trials. Recent developments, such as relaxation of access to regulatory material,1 have led to relatively successful efforts to identify and describe sponsorship bias.

The problem of sponsorship bias was recognised a century ago. In 1917, Torald Sollmann warned of the effects of secrecy and closeness between those who manufacture drugs, those who test them and those who publicise them. As a member of the Council on Pharmacy and Chemistry of the American Medical Association (forerunner of the U.S. Food and Drug Administration), Sollmann was able to identify ‘poor quality’ and secrecy (publication bias) as major threats to the credibility of the reports submitted to the Council.

The problem of sponsorship bias was recognised a century ago. In 1917, Torald Sollmann warned of the effects of secrecy and closeness between those who manufacture drugs, those who test them and those who publicise them. As a member of the Council on Pharmacy and Chemistry of the American Medical Association (forerunner of the U.S. Food and Drug Administration), Sollmann was able to identify ‘poor quality’ and secrecy (publication bias) as major threats to the credibility of the reports submitted to the Council.

Some of the papers masquerade as “clinical reports,” sometimes with a splendid disregard for all details that could enable one to judge of their value and bearing, sometimes with the most tedious presentation of all sorts of routine observations that have no relation to the problem...

...when commercial firms claim to base their conclusions on clinical reports, the profession has a right to expect that these reports should be submitted to competent and independent review. When such reports are kept secret, it is impossible for anyone to decide what proportion of them are trustworthy, and what proportion thoughtless, incompetent or accommodating...

...Those who collaborate should realize frankly that under present conditions they are collaborating, not so much in determining the scientific value, but rather in establishing the commercial value of the article.2

In Sollmann’s simple observations lie the nub of the problem. When the clinical trial is designed to advance public understanding, its design and reporting are less likely to be distorted by bias. However, when trials are run for reasons other than advancing knowledge – such as licensing, profit or career enhancement – bias is likely to creep in, even if it has not been present from the start.

It was not until six decades after Sollmann’s article that empirical investigations of sponsor bias began. In 1980, Elina Hemminki reported her pioneering study of 566 reports of clinical trials of psychotropic drugs submitted in support of licensing applications in Sweden and Finland over a period of four non-consecutive years in the 1960s and 1970s.3 Hemminki’s aims were threefold: to define the number of trials accompanying applications; to determine the quality and publication fate of these trials; and to assess whether data contained in the reports could be used to identify harms. Her research may have been the first use of regulatory documents to explore reporting bias and its association with the contents of reports. The list of problems she identified might well have been compiled 30 years later: secrecy; cherry picking of results; publication bias; ghost authorship; distorted design in favour of testing short-term effectiveness; and the inverse association between the reporting of harms and the likelihood of publication.4
To investigate the specific nature of sponsorship bias in this article, I review subsequent applications of Hemminki’s conceptually simple expedient of comparing the reports of clinical trials of medicinal products submitted to regulators with subsequently published reports of the same trials. Although the available evidence relates principally to clinical trials of commercial interest, this should not be taken to imply that academia has higher standards of reporting – available evidence suggests that it does not — but rather that sponsorship bias is more difficult to study in non-commercial trials.

One manifestation of sponsor bias is choosing comparators to give the impression that new drugs are more effective or safer than existing alternatives. Further on, in this essay, I present examples showing how design distortions can alter the conclusion of a clinical trial or mislead readers.

**Sponsorship bias reflected in reporting biases**

In the two decades that followed Hemminki’s observations, the problems of sponsorship bias and its most important consequences were investigated and defined by a handful of investigators who assessed projects submitted to regulators or research ethics committees. Easterbrook and colleagues found that only 48% of trials submitted to an ethics committee between 1984 and 1987 were published, and that publication was more likely if statistically significant differences had been found.

In 1992, Dickersin and colleagues added another dimension. Using a similar sample, they documented publication bias associated with external sponsorship. Their questionnaire survey found that the investigators were reluctant to submit reports of disappointing research for publication, although NIH-funded studies had a higher publication rate than industry-backed studies.

Melander and colleagues assessed 42 placebo-controlled trials of selective serotonin reuptake inhibitors submitted to the Swedish regulators between 1983 and 1999. They found that analysis methods set out in the protocols were ignored and instead the analyses most favourable to new products were presented. Multiple publications of the same ‘positive’ trials were also frequent.

Chan and colleagues followed up a cohort of 274 trial protocols submitted to a Danish research ethics committee in the 1990s and compared their content with subsequent reports. This comparison revealed systematic differences in outcome reporting associated with statistical significance of the comparisons. Higher rates of significant harms were associated with a lower likelihood that these would be reported publicly. The investigators denied suppressing data despite clear evidence to the contrary.

An important innovation in methods used to study sponsorship bias was reported in two papers published in 2008. These compared information in freely available Food and Drug Administration regulators’ reports on the strengths and weaknesses of the population of clinical trials submitted by sponsors to support applications for licensing with information in publications. This investigative innovation was important: these Food and Drug Administration reports are often very detailed and list all the trials relevant to the indications specified in the applications for licensing. Although those investigating possible sponsorship bias did not see the actual submission, they had access to the views of regulatory reviewers. In some instances, they had re-run their original analyses with data provided by the sponsor with each submission. Both these assessments revealed major discrepancies between Food and Drug Administration assessments and conclusions in subsequently published reports of the same trials.

In a further important analysis, Psaty and Kronmal compared internal sponsor documents with mortality data presented to the Food and Drug Administration with two published trials of the drug rofecoxib for Alzheimer’s disease and cognitive impairment. The sponsor concealed from the Food and Drug Administration their own internal analysis showing excess mortality in the intervention arms, while claiming no excess mortality by the simple expedient of ignoring the deaths in the two off-treatment weeks of follow-up. Neither publication contained any statistical analysis and the reporting of the time between exposure and death was unclear.

**Understanding sponsorship bias**

Thanks to the work of groups who accessed regulatory material by statutory means, litigation or media pressure, the decade beginning in 2009 witnessed major advances in understanding sponsorship bias and its effects. All three of the examples which follow have used Hemminki’s comparison method.

The group working in the German Institute for Quality and Efficiency in Health Care, empowered by legislation to access regulatory submissions, has produced cogent evidence of systematic selective reporting of data, both to regulators and in publications, with overemphasis on benefits and under-reporting or non-reporting of harms.

By highlighting the discrepancies among different sources of data for the same trials, primarily publications (when available) and regulatory documents,
other assessments have led to a fundamental reappraisal of the properties of drugs and biologics.\textsuperscript{18–30} More detail of each of these assessments is provided in Figure 1 (adapted from Jefferson et al.\textsuperscript{31}).

Two other reviews of comparison studies looked specifically at how harms were reported in publications and their regulatory submission counterparts and came to similar conclusions.\textsuperscript{32,33} At the time of writing, the assessment by Golder et al. is more up-to-date and comprehensive (it includes some of the studies already cited), and concluded that ‘the percentage of adverse events that would have been missed had each analysis relied only on the published versions ranged between 43\% and 100\%, with a median of 64\%’.

In summary, this comprehensive body of evidence documented the presence of distortions by sourcing published records and comparing them in detail with information that is not usually visible. It documented the presence of the bias introduced by sponsors’ attempts at ‘establishing the commercial value of the article’ – to use Sollmann’s words written over a century ago.

Although evidence of sponsor bias is now undeniable, however, it does not detail how distortions are generated, and how they are sometimes hidden or misunderstood. I shall illustrate this with three examples with which I am particularly familiar.

Tamiflu

The influenza antiviral Tamiflu (oseltamivir, Roche) is a drug that has been extensively stockpiled as a precaution against a predicted influenza pandemic. In 2010, Roche published a randomised, double-blind trial of Tamiflu\textsuperscript{34} in three different strengths, daily for 5 days, versus placebo, in 395 volunteers.\textsuperscript{35} The ‘Methods’ section of the publication reported that:

![Figure 1. Synopsis extract of trial WP 16263 showing overall study design with matching placebo.](image-url)

Figure 1 shows the synopsis of the matching 8544-page clinical study report.

Synopses of clinical study reports are usually very useful, exhaustive and highly structured and the content is usually coherent with the rest of the document. However, when we reviewed the certificates of analysis for oseltamivir, which are part of the regulatory statutory requirements in which manufacturers describe the active intervention and its comparator, we discovered that the capsules containing oseltamivir and its placebo, although of the same size, had
Figure 2. Certificates of analysis from Tamiflu trial WP 16263 describing the active capsule and its placebo overall and with details framed and enlarged (in red frames).
Figure 2. Continued.
different coloured caps. This meant that the trial could not have been double-blind, thus raising the likelihood of bias in assessing outcomes. It is difficult to know whether ‘mistakes’ of this kind are simple errors undetected and unremarked upon by regulators. It is clearly not possible for any editor or peer reviewer to identify them either.

Figure 2 displays the certificates of analysis from Tamiflu trial WP 16263 describing the active capsule and its placebo. The relevant passages have been enlarged and bordered in red.

Gardasil

The quadrivalent human papilloma virus vaccine Gardasil made by Merck was launched to provide protection against papilloma viral infection, a precursor of cancer of the cervix. The publication of pivotal trial V501-020 in the New England Journal of Medicine in 2011 reports the comparator to the vaccine as ‘AAHS containing placebo’. Reports of the same item from other sources are notable (Table 1). All contain the words ‘placebo’.

Despite its description as a ‘placebo’, and being mentioned as many as 50 times in the publication, a placebo has by definition no active ingredient. Merck’s adjuvant aluminium hydroxyphosphate sulphate (AAHS) is neither a placebo nor inactive but is actually a very potent adjuvant contained in the vaccine. Its purpose is to stimulate immunity and maintain a high and prolonged immune response. Its use as a control may mask both harms and differences in effectiveness between arms, something that none of the four data sources report. The manuscript does not explain any of this nor are readers warned of the presence of fragments of DNA in the AAHS mesh, of possible recombinant origin.

The practical consequence of using the Gardasil adjuvant as a control is that the clinical difference being estimated is Gardasil plus adjuvant versus adjuvant alone. Mistakes and misreporting of this type seem unlikely to have occurred by chance and leave readers wondering how regulators and journal editors could have missed the facts, considering that all the pivotal trials of Gardasil had the same comparator. What is needed is a trial comparing Gardasil (which contains AAHS adjuvant) with an inactive placebo control.

We have drawn the attention to similar reporting bias in four other major human papillomavirus trials.

Statins

The third and last example of sponsor bias comes from the JUPITER trial of the cholesterol

Table 1. Reporting format for the control used in Gardasil trial V501-020.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Clinical study report</th>
<th>Register</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>V501-020</td>
<td>225 mcg of aluminium as AAHS in normal saline</td>
<td>No ingredients listed; only states ‘placebo’ (NCT00090285)</td>
<td>AAHS containing placebo</td>
</tr>
</tbody>
</table>

AAHS: adjuvant aluminium hydroxyphosphate sulphate.
lowering drug rosuvastatin. JUPITER tested the effects of the statin in preventing primary cardiovascular events in asymptomatic people with elevated C-reactive protein. JUPITER is a significant trial as it was the first trial to test statin use in primary prevention. It is on the basis of the results of trials like JUPITER that the indications for statin use have been expanded to include primary prevention. Although there is little debate about the benefits of statin therapy in those at higher risk of cardiovascular disease, recently attention has shifted to the trade-off between benefits and harms in those prescribed statins for primary prevention. Higher doses of statins are known to be associated with rhabdomyolysis, leading to renal and respiratory failure, but the debate is now on the uncertainty about rates of less serious harms (especially myalgia and low-grade myopathy) in populations at lower risk of cardiovascular disease. Although not immediately life-threatening, the impact of these potential harms on quality of life and remaining mobility, especially in frail elderly people, as statins are extended to wider and older populations, make them potentially very important. The debate swings between the conclusions drawn by the influential Oxford-based Cholesterol Treatment Trialists’ Collaboration on the basis of their series of individual participant’s data meta-analysis on one side, and evidence from observational studies on the other. The Cholesterol Treatment Trialists’ director insists that the benefits of statin use in primary prevention of cardiovascular disease outweigh their harms (the incidence of which they estimate at 1 in 10,000 users). These observations seem to be contradicted by numerous large surveys, and observational studies report that users quit mainly because of harms.

The original Cholesterol Treatment Trialists’ protocol made no mention of harms, concentrating instead only on potential benefits (see Figure 2), while by 2016 the Cholesterol Treatment Trialists was actively seeking harms data from their trials holdings, as they had based their analyses of individual participant’s data mostly on publications. No results are available from this proposed analysis at the time of writing. Bias and distortion lie in originally ignoring harms and in analysing individual data without reference to the source clinical study reports and especially in not checking the definitions, severity scales and case report forms. This is where such harms would be recorded. Examination of the relevant parts of the JUPITER protocol and blank or model case report form would provide the answer. The JUPITER protocol suggests recording adverse events as follows:

4.4.2.2 Recording of adverse events

Adverse events will be identified by means of a standard question such as, “Have you had any health problems since the previous visit?” The question will be asked of each subject at all study visits following randomization. All adverse events must be recorded in the CRF provided.

In addition, non-cardiovascular adverse events are defined as follows:

(a) Adverse Event

An adverse event is the development of an undesirable medical condition or the deterioration of a pre-existing medical condition following or during exposure to a pharmaceutical product, whether or not considered causally related to the product. An undesirable medical condition

Extracts are taken from Astra Zeneca’s protocol for JUPITER.

The publication by Ridker et al. states that ‘there were no significant differences between the two study groups with regard to muscle weakness’, although this refers to serious muscular weakness. It is also of note that all specific definitions regarding possible harms are related to cardiovascular endpoints (i.e. benefits). Absence of common methods and definitions across trials are likely to raise questions about the validity of general statements such as those by Ridker et al.

These examples of distortions in designing and reporting trials of these three blockbuster interventions suggest deeply ingrained sponsorship bias in everyday clinical trial science. Whereas the first two examples could be ascribed to Sollmann’s ‘establishing the commercial value of the article’, the statin meta-studies were organised and run by well-respected academic centres in a network of trialists, which is only partially pharma-funded. The dangers of not referring to clinical study reports, protocols and manuals of operations have not been highlighted here. It is important to recognise that undetected but plausible adverse effects of statins in the hundreds of millions of people who are now being prescribed these drugs could be causing low-grade harms making their lives more difficult.

Tackling sponsorship bias

In any list of priorities for preventing sponsorship bias, disclosure and openness and the avoidance of Sollmann’s ‘secrecy’ ranks as the first and most obvious measure. The example and snapshots of regulatory documents presented in this paper did not fall out of the sky. They were the results of many years of work and effort by my colleagues and me. The degree of scrutiny involved in the review of regulatory documents also points to the necessity of quick access (when the intervention has only just been licensed); a sufficient body of researchers willing and capable of doing this work; and the identification of priority
topics (perhaps on the basis of cost and potential benefit) on which to concentrate scarce reviewing efforts. Legislation would probably be necessary to enable early access by accredited groups worldwide to regulatory submissions and to open up the editorial process completely to the use of regulatory material.

Most of these measures were proposed by Garattini and Chalmers more than a decade ago, with little political engagement. In 1968, at U.S. Congressional Hearings, the statistician Donald Mainland suggested that Congress, ‘take the evaluation of drugs entirely out of the producer’s hands, after the completion of toxicological testing on animals’. Ultimately, separation of those who are keen on establishing the value of healthcare interventions for health and those who promote the commercial worth of interventions will be the only efficient way protecting the interests of the public.

Declarations

Competing Interests: TJ is funded by the Nordic Cochrane Centre and NIHR for his work on statins. Other disclosures are https://restoringtrials.org/competing-interests-tom-jefferson/.

Funding: TJ received a fee from the James Lind Initiative for preparing this article.

Ethics approval: Not applicable.

Guarantor: TJ.

Contributorship: Sole authorship.

Acknowledgements: The author thanks the Dutch Medicines Evaluation Board for providing the JUPITER trial clinical study report.

Provenance: Invited article from the James Lind Library.

References


39. JUPITER (Justification for the Use of statins in Primary prevention: an Intervention Trial Evaluating Rosuvastatin). A randomized, double-blind, placebo controlled, multicenter, Phase III Study of rosuvastatin (CRESTOR) 20 mg in the primary prevention of
cardiovascular events among subjects with low levels of LDL-cholesterol and elevated levels of C-Reactive Protein. JUPITER trial protocol (2003 and 2006 versions).


44. Malhotra A. Saturated fat is not the major issue. *BMJ* 2013; 347: f6340.


