The Bias Dynamics Model: Correcting for Meta-Biases in Therapeutic Prediction

Adrian Erasmus

May 26, 2022

aderaasmus@ua.edu
www.adrianerasmus.com

Abstract

Inferences from clinical research results to estimates of therapeutic effectiveness suffer due to various biases. I argue that predictions of medical effectiveness are prone to failure because current medical research overlooks the impact of a particularly detrimental set of biases: meta-biases. Meta-biases are linked to higher-level characteristics of medical research and their effects are only observed when comparing sets of studies that share certain meta-level properties. I offer a model for correcting research results based on meta-research evidence, the bias dynamics model, which employs regularly updated empirical bias coefficients to attenuate estimates of therapeutic effectiveness.

Keywords: medical research, meta-bias, meta-research, medical effectiveness
1. Therapeutic Prediction. It’s commonly assumed that clinical research results can be used to make predictions about the effectiveness of medical interventions. This is done by inferring from measurements of therapeutic efficacy—an intervention’s capacity to cause its intended outcome in a study population—to estimations of therapeutic effectiveness—the general capacity of the intervention to bring about its desired effects in populations outside the study setting. In this paper, I argue that predictions of medical effectiveness are prone to failure because current policies and guidelines don’t account for the impacts of a particularly nefarious set of biases: meta-biases. To help remedy this, I offer a novel model for correcting the results of clinical studies based on evidence about these meta-biases. I refer to this framework as the bias dynamics model.

Researchers quantify therapeutic efficacy as an effect size using some outcome measure. This is a measure of the net difference that exposure to an intervention makes to a particular outcome in a population. Effect sizes are calculated by performing statistical tests on data gathered in clinical trials, such as cohort studies or randomized control trials (RCTs). If a trial meets certain standards of internal validity, then the measured effect size is thought to accurately quantify the relationship between exposure to the treatment and the outcome for the study population. The results from a clinical trial are then used to predict the therapeutic effectiveness of the intervention.

One major challenge facing clinical trials is that they may be internally valid and nevertheless fail to be externally valid. That is, a therapeutic efficacy measure may be accurate
for a study population but inaccurate for a target population. Several factors contribute to the problem of external validity, including problems related to extrapolation and a disregard of mechanistic evidence. Another particularly important consideration is the impact of biases on therapeutic prediction.

Typically, discussions about biases in medical research focus on how they manifest in the methods of a clinical study, threatening the internal validity of medical research. Such biases include confounding, confirmation bias, and reporting bias. However, another form of biases that affect medical research at the meta-level indirectly impact the reliability of research results. I call these meta-biases. Widespread meta-biases present a serious challenge to the reliability of medical evidence, particularly in cases where that evidence is used to estimate therapeutic effectiveness. Some argue that the ubiquity of research biases contributes to a large proportion of published scientific conclusions being false (Ioannidis 2005). More recently, others have suggested that biases, including meta-biases, can, and typically do, lead to systematically exaggerated claims of therapeutic effectiveness (Stegenga 2018; Fuller 2021).

My first aim expands on this line of reasoning, arguing that, despite a growing body of evidence about their impacts, we often fail to account for the effects of meta-biases on therapeutic prediction. I draw a key distinction between what I refer to as methodological biases and the higher-level concept of meta-biases. Methodological biases are those biases that are directly linked to the methodological features of clinical research. Meta-biases, on the other hand, are connected to meta-level properties of a scientific discipline. This distinction helps
demonstrate that disciplinary guidance typically focuses on methodological biases and fails to account for the effects of meta-biases. This failure partially explains why meta-biases, such as publication bias and sponsorship bias, lead to systematic overestimations of therapeutic effectiveness.

My second aim is to offer a strategy to help remedy this problem. I outline a framework for correcting measures of efficacy in line with empirical evidence about the prevalence and effects of meta-biases prior to inferring claims about effectiveness. The model I propose is takes inspiration from the methods used in dynamics, a branch of classical mechanics, whereby calculations of the forces impinging of objects’ motions are adjusted to account for friction. Just as such calculations employ empirical friction coefficients to modulate estimates of the forces acting on objects, I propose the use of empirical bias coefficients to attenuate estimates of therapeutic effectiveness. Such bias coefficients, I argue, should be regularly updated to reflect the best current evidence about meta-biases—they thus change over time. Because of its relation to dynamics calculations, and because I propose that bias coefficients be amended frequently based on empirical findings, I refer to my proposed framework as the bias dynamics model for estimating the effectiveness of medical interventions.

2. What are Meta-Biases? Medical research is plagued by various forms of bias—errors or deviations from the truth in results or inferences (Cochrane Handbook 8.2.1). Biases are typically thought to occur due to some property of either the design or implementation of a research method or the interpretation of the data gathered through experimentation. But this
conception of bias doesn’t adequately capture the full range of systemic distortions that occur in medical research.

2.1. Methodological Bias and Meta-bias. Standard examples of bias include confounding, confirmation bias, and reporting bias. Biases like these are directly linked to methodological features of clinical research. As such they are relatively well-understood. We have a good idea of how they manifest and have invested a great deal into developing strategies to mitigate or prevent them. Take confounding, the underdetermination of the association between a treatment and an outcome due to an imbalance of some factor in the experimental and control groups of a study. Some strategies for lowering the chance of confounding in clinical research are randomization, matching, stratification, and multivariate analysis. Because of their direct relation to the processes within clinical trial methodology, I call these methodological biases.

Perhaps the most damaging form of methodological bias is reporting bias. This bias occurs when results are distorted due to the selective disclosure of the analyses performed and results obtained in a clinical trial. Practices that lead to reporting bias include withholding unfavorable or nonsignificant results, publishing only a subset of the analyzed data, reporting secondary outcomes as primary outcomes when the latter yielded nonsignificant results, and adding entirely new outcomes to a published study. While difficult to detect, there’s well-known evidence of rampant practices that constitute reporting bias (Goldacre et al. 2016).

Because of the scrutiny reporting bias has received, the mechanisms responsible for it are fairly understood and we’ve developed relatively successful strategies to prevent it. For
instance, one study found a reduction in some of these mechanisms, including questionable design and analytic practices, after a requirement that studies be recorded on official trial registries was introduced at the turn of the millennium (Kaplan and Irvin 2015).

Methodological biases like those described above are emphasized in influential evidence-based medicine (EBM) guidelines. For example, the Cochrane Group’s ‘risk of bias’ tool for evaluating clinical studies lists six broad categories of bias: selection bias, performance bias, detection bias, attrition bias, reporting bias, and a final category for other bias (Cochrane Handbook 8.5.a). Researchers are advised to assess the risk of a particular bias by looking at methodological aspects of a clinical trial. Determining the risk of detection bias, for instance, entails assessing the extent to which analysts in a study were blinded when analyzing outcomes. Confounding, confirmation bias, and detection bias (and other methodological biases) fit well under the standard characterization of bias in medicine. They are indeed linked to problematic procedures in data collection, analysis, and interpretation.

Another set of biases, however, cannot be so explicitly connected to features of the experimental process. Instead, these biases are linked to meta-level properties of science, such as a scientific community’s entrenched values or its inveterate research norms. Because of their connection to meta-level properties that are external to experimental methodology, I call these meta-biases. Meta-biases can influence how research is conducted— they may precipitate or manifest via questionable practices such as p-hacking or selective reporting—yet, in contrast to those of methodological biases, their effects cannot be seen by looking at single studies.
Distortions in results due to meta-biases are only seen when comparing sets of studies with particular meta-level properties.

An exemplar of meta-bias is *publication bias*. This is the tendency of trials with negative results, indicating no causal relationship between a treatment and an outcome, to go unpublished. There’s strong evidence for this. Murad et al. (2018) found that studies with positive results are 3.90 times more likely to be published than those with negative results. This leads to a general weighting toward positive results in published medical research. Here, the distortion is observed by comparing the set of all published studies to the set of all completed studies.

Another example of meta-bias involves observed differences between the findings of industry funded studies and those of non-industry funded studies. A vast majority of clinical trials are funded and/or conducted by private organizations that have vested interests in those trials’ outcomes. There’s evidence that industry funded studies tend to generate a higher ratio of results that promote the interests of the funding organization when compared to non-industry funded studies. A recent Cochrane systematic review concluded that industry sponsored trials were 1.27 times more likely to report beneficial outcomes, 1.34 times more likely to show less evidence of harms, and 1.37 times more likely to present more favorable overall conclusions when compared with non-industry funded studies (Lundh et al. 2017).
This tendency of clinical trials to generate results that promote a sponsor’s interests know as
*sponsorship bias.*

Seeing that sponsorship bias is a meta-bias is less straightforward than it is for publication bias. This is because it’s common for methodological biases, such as reporting bias, to occur in efforts to generate findings that benefit a trial sponsor. Nevertheless, there’s clear evidence that the meta-level properties of studies, namely ‘industry funded’ and ‘non-industry funded’, are associated with a systemic distortion towards results that favor trial sponsors’ interests. Indeed, industry sponsorship is not a necessary feature of trials where related methodological biases occur. Nor is it necessarily the case that researchers in industry sponsored trials with positive findings always commit practices that constitute methodological biases. It’s simply that industry sponsored trials tend to generate positive findings in favor of sponsors’ interests. And since most trials are sponsored by industry, and most industry-funded studies report positive findings in favor of their sponsor, industry sponsorship leads to a positive skew in research results.

Still, it may be argued that, on its own, evidence that industry sponsored trials tend to report beneficial outcomes when compared to non-industry funded trials is insufficient for establishing the existence of the funding effect. After all, pharmaceutical companies will often

\[1\] Also called industry funding bias, funding bias, and the funding effect.
halt research programs that run the risk of failure before they reach clinical trials. And non-sponsored research is often slow off the mark in this regard. This provides an ostensible explanation for the higher rates of positive results from sponsored trials. However, data shows that sponsored head-to-head trials tend to generate results that favor the sponsor’s drug over a competitor when compared to non-sponsored head-to-head trials (Flacco et al. 2015). In other words, industry funded trials are more likely to conclude that their drug is superior, or at least not inferior, to competitors than non-industry funded trials. This failure to reach similar conclusions regarding particular treatments further supports the existence of sponsorship bias in medical research.

2.2. *The Neglect of Meta-Biases.* A major issue facing the problem of meta-biases is that they have not been explored to the same extent as methodological biases. As a result, strategies to help prevent them are nascent and somewhat limited. The Cochrane risk of bias tool gestures toward this concern in its ‘other bias’ category, recommending that systematic reviewers simply “State any important concerns about bias not addressed in the other domains in the tool” (Cochrane Handbook 8.5.a.). Sadly, this leaves much open to interpretation.

Failing to commit resources to meta-biases hinders efforts to limit their effects. To prevent publication bias, for instance, some journals have made it policy to solicit and publish research with negative results. Yet, other factors that lead to publication bias, such as the higher likelihood of positive results being cited thus increasing one’s chance of funding and career promotion, go unchecked. And governmental and organizational policy requiring that the
results of certain studies be reported on clinical trial registries within a year of completion are often undercut by non-compliance on behalf of researchers and regulatory loopholes (Goldacre et al. 2018).

Another more recent strategy, the use of funnel plots, aims at detecting publication bias. This formal method, as Holman (2019) notes, doesn’t require access to unpublished studies and is routinely used in medical research. While useful for assessing the risk of publication bias in meta-analytic research, funnel plots don’t provide a way to correct for its effects. Researchers who find a risk of publication bias using funnel plots usually mention it as a caveat in the discussion sections of their studies. Thus, funnel plots don’t directly tackle the challenge of publication bias, but rather reveal the extent of the problem further. Overall, publication bias is a multifaceted problem and preventing it requires immense coordinated efforts.

The same can be said of sponsorship bias. Doucet and Sismondo (2008) outline several proposed solutions to sponsorship bias, including financial disclosures, standardized reporting, and trial registration. Such policies, however, fail to cover all the contributing factors. Financial disclosure has been common practice for many years, yet industry funded trials continue to regularly favor sponsors’ interests. Standardized reporting cannot fully address trial design concerns. And while it may make it more difficult, developing strict reporting guidelines will not fully deter those whose goal is to manipulate data through outright fraud. Furthermore, the complexities of rhetoric in articles are hardly solved by introducing reporting standards. Such standardization may contribute to what Steel (2018) refers to as inferential asymmetries.
in the interpretation of clinical results, whereby some stakeholders are less able to infer true
conclusions than others due to the way in which research results are reported. Trial registries
may help solve sponsorship bias as it relates to publication bias, but one could follow
registration procedures correctly and still introduce industry favoring design features, directly
manipulate data, and present findings in a way that promotes sponsors’ interests.

Ultimately, sponsorship bias cannot be so easily reduced to issues of methodology. The
effect can manifest via multiple first-order methodological biases or through outright fraud. It
is, in this sense, multiply realizable—in individual trials, different constellations of biases may
be responsible for generating results that favor sponsor interests. However, it’s often difficult
to know whether methodological biases have occurred in a trial, and when we do know, it’s
difficult to explicitly connect these to a desire or tendency to favor a sponsor’s interests.
Moreover, existing quality assessment tools (QATs) for evaluating evidence, including the risk
of bias tool, often have poor inter-rater and inter-tool reliability (Stegenga 2018). In other
words, there are often epistemic gaps regarding first-order methodological biases connected
to the sponsorship bias and their effects on the quality of evidence. Here, the higher-order
concept of sponsorship bias is useful since we can have clearer evidence to show how industry
funding is correlated with favorable results.

Sponsorship bias can, in this sense, be thought of as a more distal bias responsible for the
systemic distortions in clinical research in comparison to the more proximal methodological
biases that it precipitates or from which it manifests. Indeed, it seems like the funding effect is a
problem of bad barrels rather than bad apples—it’s the entrenched seating of private sponsors as the majority curators of clinical research, not merely pervasive methodological biases, that’s ultimately responsible for the observed distortion in results. That’s what makes sponsorship bias a meta-bias. Thus, preventing sponsorship bias involves the mammoth task of completely overhauling clinical research, requiring an enormous amount of time and resources, all while the effects of the bias persist.

The systematic distortions in results caused by pervasive meta-biases are a serious problem for medical research (Fuller 2018). A growing body of evidence shows that publication bias leads to a systemic overestimation of effect sizes generated through meta-analyses of clinical research (Murad et al. 2018). Because there’s a higher proportion of positive findings than negative findings in the published literature, the pooled results of the published studies will show greater effectiveness than if all studies were included (Stegenga 2018). Likewise, we can infer from evidence about sponsorship bias that it generally skews results such that therapeutic effectiveness is overestimated. If most clinical trials are industry funded and if industry funded trials tend to generate industry favoring results when compared to non-industry funded trials, then this gives us good reason to believe that (a) any given industry funded trial is more likely than not biased, and (b) amalgamations of evidence from all relevant published trials will bake sponsorship bias into their results. Therefore, there are principled reasons to lower our confidence in the results of industry funded RCTs and meta-analyses that include industry funded trials.
This discussion should make three things clear. First, the higher-order concept of meta-biases is helpful. While it might be the case that a particular meta-bias is constituted by methodological biases, we may not have epistemic access to those first-order biases. Second, current guidelines and efforts aimed at preventing the detrimental effects of biases largely neglect meta-biases. And third, the problem of meta-biases is difficult to solve because such biases are typically linked to high-level properties of a scientific discipline. Preventing meta-biases would entail a near-total overhaul of the medical research system. In the absence of effective strategies for preventing meta-biases, there should be a way to attenuate efficacy measures based on what we know about the influence of meta-biases. In the next section, I describe a model for how this can be done.

3. The Bias Dynamics Model. The model I propose is loosely based on friction dynamics calculations. In these calculations, empirical friction coefficients are used to estimate the effect of friction between two surfaces in calculations of the motions of objects. I propose a similar approach for therapeutic prediction, whereby empirical bias coefficients can be used to estimate the effects that meta-biases have on therapeutic efficacy measures. I refer to my model for correcting for meta-bias as the bias dynamics model.²

² The bias dynamics model draws inspiration from Appendix 5 in Stegenga (2018).
Assume a meta-analysis is conducted to measure the effect of some treatment on some dichotomous outcome. After completing data collection and analysis, the researchers calculate an effect size measured as an absolute risk reduction (ARR). Let $\text{ARR}_S$ refer to the measure of therapeutic efficacy—calculated for a given study population—quantified as an absolute risk reduction. When conducting such a study, we are not just interested in therapeutic efficacy, but also therapeutic effectiveness—estimated for some target population. Since we usually don’t have direct evidence about the extent to which an intervention will work in a target population, $\text{ARR}_S$ is used to infer the general capacity of the treatment in that target population. Let $\text{ARR}_T$ refer to a prediction of therapeutic effectiveness, estimated as an absolute risk reduction. $\text{ARR}_T$ as an *expected frequency* representing the predicted rate of at-risk individuals in the target population who would benefit from the intervention in question.

The aim here is to take what we believe to be a true measure of a treatment’s effect in the study and infer the general capacity of the intervention that’s as close to the truth for a target population. Many in the EBM movement assume that, barring threats to internal validity and

---

3 I do not have space to deal with debates about the generalizability of different outcome measures (cf. Glasziou and Irwig 1995; Stegenga 2018; Fuller 2021). I use ARR for convenience, but the bias dynamics model can be adapted for use with other outcome measures.
relevant differences between study and target populations, we can straightforwardly apply such results. Thus:

\[ \text{ARR}_T = \text{ARR}_S \]

However, as noted, pervasive meta-biases should lead us to attenuate clinical trial results—there’s good reason, besides differences between populations and traditional failures of internal validity, to expect that the effectiveness represented by \( \text{ARR}_T \) is lower than the efficacy represented by \( \text{ARR}_S \).

Lowering the expected effect size can be done by introducing what I refer to as a bias coefficient, \( \delta \), where \( 0 \leq \delta \leq 1 \). Ideally, the bias coefficient represents the effect of all meta-biases on effectiveness claims. By multiplying the therapeutic efficacy measure by the bias coefficient, we lower the expected effectiveness of the treatment in question:

\[ \text{ARR}_T = \delta \text{ARR}_S \]

Since the magnitude of the bias coefficient is between 0 and 1, it will always attenuate the value of \( \text{ARR}_T \). The closer \( \delta \) is to 0, the more meta-biases affect \( \text{ARR}_T \), and likewise, the closer \( \delta \) is to 1, the less meta-biases affect \( \text{ARR}_T \). It’s theoretically possible that the effects of meta-biases are sufficient \( (\delta = 0) \) for us to expect that the intervention will have no effect in the target \( (\text{ARR}_T = 0) \). Conversely, meta-biases could, in principle, have no effect \( (\delta = 1) \), and thus we could conclude (assuming no other problems related to generalization and no other threats to internal validity) that the therapeutic effectiveness will be equal to the therapeutic efficacy \( (\text{ARR}_T = \text{ARR}_T) \). However, given the prevalence of meta-biases, these scenarios are unlikely.
The bias dynamics model provides a way to correct therapeutic effectiveness estimates by taking the effects of meta-biases into account. As mentioned, the bias coefficient ideally represents the effect of all meta-biases on research results, yet, in practice, it’s very unlikely that we could have access to such knowledge. This doesn’t, however, preclude the use of the model. We can determine bias coefficients based on what is known about meta-biases. This is the next step to calls for the use of meta-research evidence in adjusting confidence in clinical findings (Fuller 2018). Furthermore, bias coefficients can fill the gap left in recognized guidelines on biases by helping researchers account for the effects of meta-biases, which are not explicitly listed in QATs. For example, once Cochrane’s risk of bias tool has been used to rate single studies for bias, the bias coefficient can be used to determine an overall rating for a given research program based on evidence about meta-biases. In line with this, an organization like Cochrane should publish domain specific bias coefficients. And, at least in the case of publication bias, bias coefficients can be used in conjunction with funnel plot analyses. Here, researchers may detect publication bias and be more justified in applying the bias dynamics model.

Bias coefficients should be categorized by research area and outcome. Meta-biases occur at various rates in different research programs—the prevalence of outcome reporting bias in cancer research differs to that in research on statins (cf. Vera-Badillo et al. 2016; Rezende et al. 2018). Likewise, different outcomes within each research program will have different rates of meta-biases—the rate of publication bias in research on the association between statins and
heart attack may differ from that research on the association between statins and stroke. Naturally, because of this, the values of bias coefficients for given outcomes in particular fields will differ.

Unlike their analog in physics, bias coefficients should be dynamic. Their values should be updated over time with the observation of new evidence about meta-biases. Such an approach is in keeping with the central tenets of EBM and its related organizations, which aim to update guidelines based on the best available evidence. The Cochrane Group has already gestured toward this sort of approach with updates to its research into the effects of industry funding on research outcomes (cf. Lundh et al. 2012; Lundh et al. 2017).

To briefly demonstrate the bias dynamics model, consider a meta-analysis that found statin therapy is associated with a 0.81% decrease in heart attacks over six years (Chou et al. 2016). Thus, $\text{ARR}_S = 0.81\%$. We want to know if using statins will decrease the risk of heart attack in a target population of individuals with high cholesterol, a risk factor for heart attack; in other words, we want to estimate $\text{ARR}_T$. Most would assume that $\text{ARR}_T = 0.81\%$. However, evidence about meta-biases in statin research should be considered.

Bero et al. (2007) found that the odds of statistically significant results in favor of statin therapy in drug-drug comparisons were 16 times greater for industry funded trials than for non-industry funded trials. While these findings don’t indicate the exact effects of meta-biases on statin research, they do warrant a relatively low bias coefficient (indicating a large effect
from meta-biases). In line with this, assume a bias coefficient \( \delta = 0.45 \). This would attenuate the expected frequency of individuals who would benefit from statin therapy; \( \text{ARR}_T = 0.37\% \).

Overall, the bias dynamics model provides a relatively straightforward strategy for researchers correct for the effects of meta-biases. Bias coefficient function as regularly updated attenuating variables representing the known effects of meta-biases. The implementation of such a tool would go some way to producing more reliable predictions of therapeutic effectiveness.

4. Conclusion. Estimations of therapeutic effectiveness suffer greatly due to widespread neglect of the effects of meta-biases. Meta-biases are distinct from methodological biases in that they are not directly linked to the methodological features of clinical research. Rather, they are connected to meta-level properties of a scientific discipline, such as the deep-rooted values of the medical research community or entrenched norms of the system in which medical research is conducted. Current research guidance focuses on methodological biases at the expense of accounting for meta-biases.

The bias dynamics model provides a way to modulate therapeutic predictions. Using evidence on the effects of meta-biases, estimates of therapeutic effectiveness can be attenuated using empirical bias coefficients. The bias dynamics model, and its use of bias coefficients, is appealing for at least two reasons. First, it has the potential to fill a gap in current EBM guidelines, which emphasize the risks of methodological biases, and do not account for meta-bias effects. And second, through the use of empirical bias coefficients, the model provides a
clear way for researchers to use up-to-date evidence to generate more accurate estimations of medical effectiveness.

Further study is, of course, necessary. For instance, there needs to be a reliable program for determining bias coefficients. There’s promising work in this domain. Tabatabaei Ghomi and Stegenga (2021) simulate trial-level data using the higher-order reported results of published trials, such as means and effect sizes. Using simulations allows researchers to not only specify the true effectiveness of the intervention in question prior to analyzing the effects of meta-biases, but also control for methodological biases and researcher idiosyncrasies. Such studies can be conducted using meta-research findings on the prevalence of various meta-biases. In doing so, we can measure their effects on different outcomes in different subdomains of medicine, and from there determine bias coefficients for specific areas of research that can be published on a regular basis to reflect the latest meta-research evidence.
REFERENCES


Flacco, M.E. et al. 2015. “Head-to-head randomized trials are mostly industry sponsored and almost always favor the industry sponsor.” *Journal of Clinical Epidemiology* 68(7): 811–820.


Steel, D. 2018. ‘If the facts were not untruths, their implications were: Sponsorship bias and misleading communication’, Kennedy Institute of Ethics Journal 28(2): 119–144.
