The MIT Press

RESEARCH ARTICLE

Incorporating costs and benefits to the evaluation of uncertain research results: Applications to cancer research funding

Han Zhuang[1] iD and Daniel E. Acuna[2] iD

[1]School of Information Studies, Syracuse University, Syracuse, NY, USA
[2]Department of Computer Science, The University of Colorado at Boulder, Boulder, CO, USA

ABSTRACT

Correctness is a key aspiration of the scientific process, yet recent studies suggest that many high-profile findings may be difficult to replicate or require considerable evidence for verification. Proposals to fix these issues typically ask for tighter statistical controls (e.g., stricter *p*-value thresholds or higher statistical power). However, these approaches often overlook the importance of contemplating research outcomes' potential costs and benefits. Here, we develop a framework grounded in Bayesian decision theory that seamlessly integrates cost-benefit analysis into evaluating research programs with potentially uncertain results. We derive minimally acceptable prestudy odds and positive predictive values for cost and benefit levels. We show that tolerance to inaccurate results changes dramatically due to uncertainties posed by research. We also show that reducing uncertainties (e.g., by recruiting more subjects) may have limited effects on the expected benefit of continuing specific research programs. We apply our framework to several types of cancer research and their funding. Our analysis shows that highly exploratory research designs are easily justifiable due to their potential benefits, even when probabilistic models suggest otherwise. We discuss how the cost and benefit of research could and should always be part of the toolkit used by scientists, institutions, or funding agencies.

1. INTRODUCTION

Accurate results should be one of the primary outcomes of science. One way to probe the accuracy of a research finding is to have independent research teams attempt to *replicate* the claims of a previous study. This probing has shown less than ideal results for correctness. For example, replicability of high-profile scientific findings in social science and medical research is relatively low (Camerer, Dreber et al., 2018; Ioannidis, 2005a; Open Science Collaboration, 2015) and the reproducibility of cutting-edge machine learning is also questionable (Raff, 2019). Although failures to replicate a study do not imply inaccurate results (Goodman, Fanelli, & Ioannidis, 2016), the trend is worrying and should be further examined.

Mathematical models such as the one proposed by Ioannidis (2005b) partially explain why published research findings might be inaccurate: low prestudy odds, low statistical power, and high reward for publication priority. Researchers have suggested solving these limitations

using a number of more stringent research protocols, such as preregistration or smaller *p*-value threshold to accept a hypothesis (Benjamin, Berger et al., 2018; Nosek, Ebersole et al., 2018). Most of these new suggestions, however, only consider the *chance* of inaccurate results. The reality of research is significantly more complex and not so binary. For example, exploratory research implicitly tolerates high chances of inaccurate results because of the high potential benefits. Little is formally known, however, about how research uncertainty, cost, and benefits should be combined to produce optimal research decisions. Without this understanding, we risk discarding seemingly unpromising research paths that may have large benefits in the long run. We believe decision theory can be used to combine cost-benefit and probabilistic considerations within a unified framework. In this article, we demonstrate how this framework can help us understand decisions more pragmatically in real research scenarios, such as cancer research.

Scientific correctness has been substantially explored before. Theoretically, we can derive simple models that illuminate how prestudy odds and incentives behind scientific publications are related to scientific correctness. For example, Ioannidis (2005b) examined three types of common practices that affect correctness: low statistical power research design, bias in research, and multiple research groups studying the same question. Another way of inquiring about science's correctness is to investigate whether a published result is reproducible by other research groups. For example, drug companies attempt to reproduce results from the scientific literature as a first step in their research, but this step is surprisingly unsuccessful (Begley & Ellis, 2012). Other large-scale studies have shown that nonreproducibility occurs beyond drug-related research (e.g., Gilbert, King et al., 2016; Open Science Collaboration, 2015) and machine learning (Raff, 2019). Because theoretically the probabilities of true results and replicability are not necessarily related (Goodman et al., 2016), they can form two complementary probabilistic approaches to quantify correctness in science. Several proposals have been put forward to try to improve correctness by making statistical thresholds more stringent (Benjamin et al., 2018) and requiring preregistration (Nosek et al., 2018). Thus, the question of correctness in science has been substantially studied before from a probabilistic perspective.

We can incorporate additional realistic factors of scientific correctness. For example, public decision-making always requires the consideration of the expected value of different actions. These values, however, are typically not used when proposing new *p*-values thresholds or suggesting to increase replicability. If we formulate the *doing* of science as a decision-making process, we can factor in the costs and benefits of our research decisions. For example, we can estimate how many inaccurate results we can afford for a research project with a certain cost and benefit ratio (Djulbegovic & Hozo, 2007). This trade-off can be especially relevant for resource-strapped institutes and funding agencies. Importantly, we can investigate this aspect of research decision problems using standard cost-benefit analysis (CBA) (Layard & Glaister, 1994). We can blend both the probabilistic aspects of the correctness of scientific results with the benefits and costs of those results. This blending is particularly coherent in the framework of Bayesian decision-making (e.g., Gelman, Carlin et al., 2013) where prestudy odds, statistical significance, statistical power, and correctness of outcomes can be integrated with costs and benefits.

In the present study, we combine the ideas of the correctness of scientific findings with their expected costs and benefits. We use Bayesian decision-making theory to study how cost-benefit ratios might affect our tolerance to inaccuracy. We model research decisions in cancer research. We further explore the effectiveness of different types of research strategies as a

function of prestudy odds, statistical power, and funding information (e.g., for NIH). The highlights of our study are as follows:

- Bayesian decision theory can seamlessly incorporate costs and benefits into the analysis of potentially false research results.
- Potentially false research results should not stop researchers from pursuing a research program when benefits of other future results outweigh costs—but effects vary dramatically depending on the quality of research design.
- Marginal improvements in costs brought by increments in the quality of research study designs (e.g., higher prestudy odds or higher-powered experiments) rapidly diminish, revealing the limited effect of "better science" on costs.
- Cost and benefit analysis of cancer research funded by NIH prescribes different research designs for various cancers (e.g., suggesting highly exploratory studies for prostate cancer and meta-analysis of existing literature for brain cancer).

1.1. Rigor in Scientific Publications

Checking the correctness of scientific results is an essential part of modern science. Statistical training at all levels of education prepares students at least in part to verify results (Bond, Perkins, & Ramirez, 2012). Graduate students, especially, study the prevalence of a research issue, current techniques for solving it, and techniques for evaluating the quality of the evidence (Remler & Ryzin, 2014). In clinical science and public health, clinicians need to understand quantities such as prevalence of a disease, and sensitivity and precision of treatments (Bush, 2011). Although these quantities might not be used in practice constantly, they signal that evidence-based medicine has become the dominant approach during the application of scientific advances (Sackett, Rosenberg et al., 1996).

Metaquestions about the correctness of science itself are relatively new. Data availability to draw quantitative conclusions about the state of the issue has been a major problem, with data and large-scale analyses only happening relatively recently (Fortunato, Bergstrom et al., 2018; Open Science Collaboration, 2015). Very little funding and few centers have been devoted to metalevel questions of scientific correctness because they depend on these data sets. Also, the level of funding is comparatively small compared with other branches of science. For example, the Science of Science: Discovery, Communication, and Impact (SoS:DCI) program (formerly SciSIP) at NSF is a relatively new compared to other fields (Directorate for Social, Behavioral, and Economic Sciences, 2019).

The correctness of scientific research findings is always an important topic in different communities of science. There is a long line of thought elevating scientific correctness and its refutation as an integral part of knowledge production. Popper argued that research findings from an inductive logic perspective cannot be verified to be true, but can only be falsified (Popper, 2005 [1934]). Similarly, Kuhn argued that revolutionary science brings paradigm shifts to refute previous research findings (Kuhn, 2012 [1962]). Beyond philosophy of science and history of science, researchers from different scientific disciplines have explored scientific correctness in published work. Lack of replication has been used to spearhead a push to study scientific correctness further. As examples, Ioannidis (2005a) and Open Science Collaboration (2015) showed that subsequent studies in clinical science replicating highly cited work failed to replicate results if better research design was used (e.g., large sample size, better controls). Even though it is hard to conclude how many research findings are inaccurate from these

replication studies, some researchers argue that subsequent research studies are necessary to improve the correctness of science (Nissen, Magidson et al., 2016).

2. CBA OF SCIENCE

We now briefly review background literature on CBA of science and proposals to estimate it.

2.1. Decision-Making in Scientific Research

Accurate research findings should be the primary goal of science, but there are many other factors involved in the process of science. For example, problematic research designs (e.g., low-powered design) might lead to positive research outcomes by chance, and research implementation can be imperfect (e.g., nonrepresentative sample from the population). In addition, peer review might not prevent all errors in research articles because of knowledge limitations in reviewers and the nontransparency of resources (e.g., computer code or raw data). Therefore, inaccurate research findings can still be published because of the limitations of research designs, research implementations, peer review, or other reasons, such as responses to incentives (Kornfeld, 2012; Weingart, 2002). Researchers and funding agencies know that inaccurate research findings might be published, but the self-correcting nature of science and the potential benefit of research programs still motivate to continue them. For example, The United States' National Institutes of Health (NIH) routinely takes into account costs and benefits when allocating funding to competing needs (Gillum, Gouveia et al., 2011; Goodman, 2004), acknowledging that the accuracy of research should not be the most important factor under certain circumstances.

Most research that studies the correctness of science takes a probabilistic perspective without much regard to costs and benefits. For example, the original article that popularized the idea that most published research has inaccurate results (Ioannidis, 2005b) mentions a scenario with multiple research groups competing to find the same research finding. The author translated this into probabilistic measures of prestudy odds and biases but it would also make sense if translated into costs and benefits: The benefit of gaining priority—the reason behind competition—is much larger than the cost of publishing an erroneous result. There are many other similar examples where inaccuracy of science is less important than its expected value (Nissen et al., 2016; Ulrich & Miller, 2020).

Research decision-making is important in science policy and funding policy and needs to be accountable and reasonable. Thus, it is important to include the costs and benefits of research and the uncertainty of the accuracy of research findings into a trade-off to balance multiple factors when performing and publishing research. A first step is to explore the multiple trade-offs that scientists and institutions consider in research to incorporate CBA into their decision-making. Some research concerning the correctness of science tends to overlook these costs and trade-offs for simplicity's sake. For example, Miller and Ulrich (2016) examined the trade-offs between false-positive rate, statistical power, and other parameters within research processes, but the costs and benefits of research as another dimension of research is unstudied. Assigning costs and benefits to scientific activities might be unintuitive for researchers, but whether we like it or not, values are often associated with life, disease, and health. For example, government agencies release their estimates of the value of life and the value of the burden of diseases. Additionally, there are human resources (e.g., professors, students, administrators) as well as "monitoring" resources (university's infrastructure, city investments, and local, state, or federal funding) devoted to performing experiments (Partha

& David, 1994). Thus, science has many trade-offs that are sometimes better captured by costs and benefits in conjunction with probabilities.

2.2. Estimating the Costs and Benefits of Research

Researchers and governments use a variety of estimates of the costs and benefits of research, depending on the field and data availability. For example, in some research fields, such as biomedical research, the costs and benefits of research have been established by health economics researchers. Health economics research considers the benefit of research as the increment in lifetime of patients (Cutler, 2007) or the decrease in mortality rate or in disability rate (Cutler, 2005; Cutler & Meara, 2001). In health economics, the cost is commonly thought of as the cost of funding and the cost of implementation of the technologies or research infrastructure (Cutler & McClellan, 2001). However, these proposed methods to measure the costs and benefits of research are not perfect. For example, the time invested by researchers and other human resources is commonly not considered (Drummond, Davies, & Ferris, 1992). In other fields, the sole notion of measuring costs and benefits is controversial. In environmental research, there are debates about what constitutes a benefit to the environment (Atkinson & Mourato, 2008), where the life of human beings is often frowned upon (Ackerman & Heinzerling, 2002; Bockstael, Freeman et al., 2000; Stirling, 1997).

Estimates are generally starting points to guide public policy, and methods for measuring costs and benefits tend to improve over the years (Fuchs, 2000). Also, funding agencies attempt to measure the benefits and costs of their research decisions to justify their previous funding decisions. However, the viewpoints of the benefits and costs of research to researchers and funding agencies might be different. Therefore, existing measurement methods from health economics might exactly fit the requirements of biomedical research funding agencies, and some existing measures of the benefits and costs of research might be used by decision-makers in science policy from the viewpoint of the benefit to the public (e.g., extending life expectancy of human beings).

2.3. Models for CBA Applied to Research

Compared to health care, there are few studies of CBA on academic research. The main gap is the trade-off framework of the uncertainty of the correctness of research and their corresponding benefits and costs. In health care, CBA is a much more common tool including actions (e.g., prescribe a drug) applied by decision-makers (e.g., doctors) to users of a system (e.g., patients). At a metaresearch level, the question is more abstract—that is, whether to do the studies in the first place or not while considering the uncertainty of the research outcome. Even though, some previous studies have identified the costs and benefits of a group of research problems, such as the benefits of research against diseases (Gillum et al., 2011; Sampat, Buterbaugh, & Perl, 2013), and how to make decisions for research with uncertain accuracy?

There is some research on attempts to make decisions with the benefits and costs of actions and their uncertainties in the medical setting (e.g., drug prescriptions). For example, the work of Djulbegovic and Hozo (2007) extends the analysis of Ioannidis (2005b) to include the "benefit/harm" ratios of health-care research. They derive rules for continuing or stopping research based on minimization of regret—ideas previously explored for the clinical setting in Djulbegovic, Hozo et al. (1999). Importantly, regret minimization is a process that looks *back* (e.g., regret) at what was done and then evaluates past plans. However, this model might not be applicable to research decision-making for the future because these research decisions

can be considered as a problem of value maximization, such as evaluation of a path of action *forward* (Sutton & Barto, 2018). As examples of this approach, Wang, Middleton et al. (2003) studied the costs and benefits of implementing an electronic medical records system, using possible future paths, and Nichol (2001) studied the costs and benefits of vaccination with its uncertain efficiency. These two examples are applications of decision-making for medical settings with their expected benefits and expected costs from CBA. However, how to conduct a trade-off between uncertain variables in research (e.g., statistical power and false-positive rate) and the benefits and costs of research has not been studied.

2.3.1. Other models

There have been some attempts to directly estimate benefits and costs using statistical models. For example, linear or interactive models were proposed by Lundvall (2004). In a linear model, the benefits and costs of research are directly produced by the research. For example, biomedical research might directly produce a valuable innovation to reduce the cost of treatments. The cost of biomedical research can be the cost of the experiment or the research project. However, research builds on other research, leading to an interactive model. In an interactive model, the benefits and costs of research can be produced by other research or activities (Cowan, David, & Foray, 2000). For example, some research might not produce direct value to patients, but provide fundamental knowledge for vaccine research.

2.4. Summary

CBA of science is a crucial part of research decisions. Most research on potentially problematic factors in the veracity of science takes a probabilistic perspective. However, science obeys many real demands, such as the benefits of its outcomes and the costs involved in creating those outcomes. The estimates of such factors are complex and vary from field to field. Models that balance costs and benefits for scientific research at a metalevel can guide the *doing* of science for individuals and institutions. This CBA guidance adds shades of meaning to the claim that most published research is false. The key is to realize that even if inaccurate results are published, these exploratory works can bring value for future research. The degree to which costs and benefits ratios change this dynamic has been understudied even though they form part, albeit informally, of decisions made by scientists, institutions, and foundations.

3. METHODS FOR CBA FOR A RESEARCH PROGRAM

In this section, we present a general framework for CBA of research. First, we build this framework by extending Ioannidis (2005b) with Bayesian decision theory (Duda, Hart, & Stork, 2012). Second, we describe methods for estimating benefits and costs of research in health economics. Third, we present how the return on research investment changes with different levels of estimated research correctness.

3.1. A Framework for CBA of Research

We will use terminology related to uncertainty, costs, and benefits. In Table 1, we list all the terminology used throughout the article.

3.1.1. A mathematical model of correctness of research findings

Research studies in medical sciences can be modeled as a binary outcome problem, examining whether a relationship exists between two or more groups. To establish such a relationship,

Table 1. Notation used throughout the article. True positive is when there is a positive research finding if there is true relationship. A false positive is when there is a positive research finding if there is a false relationship. True negative and false negative are similarly defined.

Symbol or term	Description	Explanation
α	Significance level or Type I error	False positive rate
β	Type II error	False negative rate
$1 - \beta$	Statistical power	True positive rate
PPV	Positive predicted value	The ratio between the number of true positive relationships and positive relationships
RF_P	Positive research finding	Research finding claims that there is a relationship
RF_N	Negative research finding	Research finding claims that there is no relationship
TR_T	True relationship	This relationship exists
TR_F	False relationship	This relationship does not exists
R	Prior true-to-false relationships ratio	The ratio between the number of true relationships and false relationships
λ_{TP_C}	Cost of continuing research under a true positive	Cost of continuing research with a true positive relationship
λ_{TP_S}	Cost of stopping research under a true positive	Cost of stopping research under a true positive relationship
λ_{FP_C}	Cost of continuing research under a false positive	Cost of continuing research under a false positive relationship
λ_{FP_S}	Cost of stopping research under a true positive	Cost of stopping research under a true positive relationship

scientists commonly use likelihood tests to accept or reject scientific hypotheses. Statistical tests often rely on establishing significance test thresholds such as α (for Type I error—the rejection of a true null hypothesis, false positive rate) and β (for Type II error—failing to reject a false null hypothesis, false negative rate).

The probability of a true relationship existing depends also on the prior probability of a relationship existing in the first place. We define the prestudy odds of true relationships, R, as the ratio between the number of true relationships and the number of false relationships. Therefore, $P(TR_T)$ (the prestudy probability) is $R/(R + 1)$. The probability that a researcher obtains a positive finding given a true relationship is $P(RF_P|TR_T)$. This likelihood is known as statistical power and denoted as $1 - \beta$. Using all these quantities, we can now derive the probability that our research findings translate into true relationships.

The simplest case is one in which there are no biases that will mislead the research findings other than the intrinsic randomness in the system. In this regime, the probability that researchers get a positive research finding and a true relationship is a joint distribution of the prestudy probability and the likelihood of the result (or the power of the study, $1 - \beta$): $P(TR_T, RF_P) = (1 - \beta)R/(R + 1)$. Similarly, another joint probability to represent getting a positive finding with a false relationship is $P(TR_N, RF_P) = \alpha/(R + 1)$. From these joint distributions, we

can obtain $P(\text{RF}_P) = ((1 - \beta)R + \alpha)/(R + 1)$. With these expressions, we can now derive the positive predictive value (PPV)

$$\text{PPV} = P(\text{TR}_T|\text{RF}_P) = P(\text{TR}_T, \text{RF}_P)/P(\text{RF}_P), \tag{1}$$

$$= \frac{(1 - \beta)R}{(R - \beta \times R + \alpha)}. \tag{2}$$

We can derive the surprising result that most published studies are more likely to be false than true if we combine values used in common scientific practice for R, α, and β. For example, a study might start with an R of 1 in 100, low power $(1 - \beta = 2/5)$, and significance of $1/20$ $(\alpha = 0.05)$, leading us to a PPV of approximately 8%: The probability of an existing true relationship is only 8% given a positive research finding. In general, because scientists tend to study highly unlikely effects (i.e., low prestudy odds), PPV tends to stay below 0.5, which suggests that most research under these assumptions is likely to be false.

The problem is even more pronounced when we introduce other factors, such as bias. If a research study contains bias, such as selective reporting, then the negative relationships might end up presented as positive due to problematic research practices (Ioannidis, 2005b). Bias also reflects the quality of research design in a field. After bias (denoted as μ here) is involved, PPV with bias is

$$\text{PPV}_\mu = P(\text{TR}_T|\text{RF}_P)_\mu = \frac{R\mu\beta + R(1 - \beta)}{R + \alpha - R\beta + \mu - \mu\alpha + R\mu\beta} \tag{3}$$

Based on Eq. 3, bias μ, will reduce the value of PPV if all other variables are the same. For example, continuing with the same parameters as in the example above, with a bias of 1/10, the PPV drops from 8% to only 3%.

3.1.2. Two-category cost estimation with Bayesian decision theory

Here, we take a Bayesian decision theory point of view to evaluate discrete decisions with probability and costs associated with them (Duda et al., 2012). We will restrict our analysis to settings where there are two possible (unknown) states of nature and two possible actions. To apply Bayesian decision theory to Section 3.1.1, we define two states of nature w_1 and w_2, which we wish to infer. We have priors about their existence $P(w_1)$ and $P(w_2) = 1 - P(w_1)$. We have two conditional probabilities $P(x|w_1)$ and $P(x|w_2)$ that define observation probabilities for a given state of nature. Although we cannot observe this state, we can use Bayes' rule to update our belief about it using observations, $P(w_1|x) = P(x|w_1)P(w_1)/P(x)$ and $P(w_2|x) = P(x|w_2)P(w_2)/P(x)$.

Decision theory goes further by defining actions that we can take based on the costs associated with choosing one state of the world versus the other. We will define two actions a_1 and a_2, which denote the action of choosing the state of world 1 and 2, respectively. We will define the costs by λ_{ij} as the cost of taking action a_i if the state w_j is true. Using Bayesian inference, we can estimate the expected cost of each action as follows

$$\text{Cost}(a_1|x) = \lambda_{11}P(w_1|x) + \lambda_{12}P(w_2|x) \tag{4}$$

$$\text{Cost}(a_2|x) = \lambda_{21}P(w_1|x) + \lambda_{22}P(w_2|x) \tag{5}$$

The best decision is given by

$$a_* = \arg \min_i \text{Cost}(a_i|x).$$ (6)

For example, using these costs, an agent should choose action a_1 if $\text{Cost}(a_1|x) < \text{Cost}(a_2|x)$ or, equivalently, if the following inequality holds

$$(\lambda_{21} - \lambda_{11})P(x|w_1)P(w_1) > (\lambda_{12} - \lambda_{22})P(x|w_2)P(w_2)$$ (7)

With the reasonable assumption that $\lambda_{21} > \lambda_{11}$ and $\lambda_{12} > \lambda_{22}$ (a wrong prediction is costlier than an accurate prediction), we can more compactly decide to take a_1 if the following condition is true

$$\frac{P(x|w_1)}{P(x|w_2)} > \frac{(\lambda_{12} - \lambda_{22})}{(\lambda_{21} - \lambda_{11})} \frac{P(w_2)}{P(w_1)}$$ (8)

From Eq. 8 we know that if the ratio between the cost and the benefit of actions is significantly large, then the decision point will be influenced dramatically.

3.1.3. Research decision making and correctness of research findings

We now apply Eq. 8 to guide decisions about research studies. Actions a_1 and a_2 will be related to whether we continue pursuing a research path (a_1) or not (a_2). The observation x will be the outcome of a research finding ($x = \text{RF}_P$, research finding is positive; $x = \text{RF}_N$, research finding is negative). The states of nature are whether there is a true relationship or not ($w_1 = \text{TR}_T$ and $w_2 = \text{TR}_F$).

Translated into the research process, the conditional probabilities in Eq. 8 are

$$P(x|w_1) = P(\text{RF}_P|\text{TR}_T) = 1 - \beta,$$ (9)

and

$$P(x|w_2) = P(\text{RF}_P|\text{TR}_F) = \alpha,$$ (10)

true positive and false positive probabilities, respectively. In this context, $P(w_1|x)$ equals PPV and $P(w_2|x)$ equals $(1 - \text{PPV})$. We rename λ_{11} as λ_{TP_C} denoting the cost of continuing research when there is a true relationship—there is a true positive—and we continue research, TP_C. Conversely, we rename λ_{21} to λ_{TP_S}, denoting the cost of stopping research when there is a true relationship—there is a true positive and we stop research, TP_S. Similarly, we rename λ_{12} to λ_{FP_C}, and λ_{22} to λ_{FP_S} (see Table 1 for details). Finally, $P(\text{TR}_T)/P(\text{TR}_F) = R$.

Based on these quantities, we should continue doing research if the following condition holds:

$$\text{PPV} > \frac{\lambda_{\text{FP}_C} - \lambda_{\text{FP}_S}}{(\lambda_{\text{TP}_S} - \lambda_{\text{TP}_C}) - (\lambda_{\text{FP}_S} - \lambda_{\text{FP}_C})}.$$ (11)

We further denote the cost incurred by the problem continuing as C (such as the cost of diseases on society) and the research investment to solve the problem as I (such as funding by NIH). We further simplify the problem by representing λ_{FP_C}, λ_{FP_S}, λ_{TP_C}, and λ_{TP_S} with only C and I using the following logic. Here, we assume that the cost of the problem (C) and the investment to solve the problem (I) can be observed first, because some costs of the problems can be hard to quantify. To make this analysis feasible, we applied the methods from health economic to estimate C (see Section 5.3) and assume I as the cost of research on the problem and the expenditure on the problem. If a research finding is a false positive, then the cost of the action

of continuing this research (λ_{FP_C}) is the sum of the cost of the problem (C) and the research investment (I) (see Eq. 12). The cost of the action of stopping this research (λ_{FP_s}) is only the cost of the problem, C (see Eq. 13). If a research finding is true positive, then the problem is solved by research, and the cost of the action of continuing this research (λ_{TP_c}) is only the cost of research investment, I (see Eq. 15). The cost of the action of stopping this research (λ_{TP_s}) is the cost of the problem, C (see Eq. 14). Therefore, we have the following equations:

$$\lambda_{FP_C} = C + I \tag{12}$$

$$\lambda_{FP_s} = C \tag{13}$$

$$\lambda_{TP_s} = C \tag{14}$$

$$\lambda_{TP_C} = I \tag{15}$$

In Eq. 12, we estimate that the total cost of continuing false positive research is the sum of the cost of the problem and the investment in the problem. In Eqs. 13 and 14, we estimate the total cost of stopping research with either a true positive or a false positive is the cost of the problem, because we can not use the research finding to solve the problem and assume the cost of the problem will not be changed. In Eq. 17, we estimate the total cost of continuing true positive research as the investment in the problem, because we assume that when the problem is solved the cost can be diminished to zero.

We can manipulate these equations to obtain the minimal PPV:

$$\text{PPV} > \frac{I}{C}, \tag{16}$$

and minimal prestudy odds (R) values and minimal statistical power $1 - \beta$ without bias:

$$R > \frac{I}{C - I} \frac{\alpha}{1 - \beta}, \tag{17}$$

$$1 - \beta > \frac{I}{C - I} \frac{\alpha}{R} \tag{18}$$

and minimal R and $1 - \beta$ with bias:

$$R > \frac{I}{C - I} \frac{(1 - \mu) \times \alpha + \mu}{(1 - \beta) + \beta\mu}, \tag{19}$$

$$1 - \beta > \frac{I}{C - I} \frac{(1 - \mu) \times \alpha + \mu}{R \times (1 - \mu)} - \frac{\mu}{1 - \mu}. \tag{20}$$

Eq. 16 is surprisingly simple: The minimally acceptable PPV is the ratio of investment to cost. This means that the more we invest, the greater statistical power (i.e., large randomized trials) and higher prestudy odds we need. On the other hand, if we cannot afford to invest large amounts of money on a costly research question, then we are happy to settle for lower PPVs. Eqs. 17 and 18 are the thresholds for decisions with no bias, and Eqs. 19 and 20 have bias included. According to Eq. 17 (a prestudy odds threshold with no bias) and Eq. 19 (a prestudy odds threshold with bias), a research problem with a big cost can tolerate low prestudy odds of research, compared with another research problem with the same research investment. In Eq. 19, with bias μ included, it will require higher minimal prestudy odds: Comparing Eq. 19 to Eq. 17 suggests that bias is equivalent to increasing the false positive rate and reducing

statistical power. According to Eq. 18 (a statistical power threshold with no bias) and Eq. 20 (a statistical power threshold with bias), a research problem with a big cost can tolerate a low statistical power, compared with another research problem with the same research investment. In Eq. 20, with bias μ included, a higher minimal statistical power will be required. This is because if the prestudy odds is greater than one, then bias can lead to more true positive research findings. But when the prestudy odds is smaller than one, bias brings more false positive research findings, and to make sure we have enough true positive research findings for a positive return from decisions, the minimal statistical power must be bigger than the scenario of Eq. 18, without bias.

3.2. Improving Statistical Power and Prestudy Odds on Research Outcomes

Another important question is to investigate how changes in PPV (e.g., brought by increased statistical power) change research outcomes, such as the expected costs. More specifically, we will show that constant increments in PPV bring increasingly smaller reductions in expected costs. To see this, we use Eq. 4 and the equivalences described in Section 3.1.3 (i.e., $\lambda_{11}P(w_1|x) + \lambda_{12}P(w_2|x)$ $\lambda_{11} = I$, $P(w_1|x) = PPV$, $\lambda_{12} = I + C$, and $P(w_2|x) = 1 - PPV$) to obtain the following expected cost of doing research:

$$\text{Cost(continue|RF}_P) = \text{Total Cost} = I \times PPV + (I + C) \times (1 - PPV). \tag{21}$$

If we change the PPV by changing power $(1 - \beta)$ or prestudy odds (R), this would produce marginal changes in the cost above:

$$\frac{d\text{Total Cost}}{d(1 - \beta)} = -C \times \frac{\alpha}{((1 - \beta) + \alpha)^2} \tag{22}$$

$$\frac{d\text{Total Cost}}{dR} = -C \times \frac{1 - \beta}{((1 - \beta)R + \alpha)^2} \tag{23}$$

Following Eq. 22 (derivative of the cost of research problem as a function of statistical power) and Eq. 23 (derivative of the cost of research problem as a function of prestudy odds), more statistical power or higher prestudy odds always reduces the expected costs (both derivatives are always negative). However, because the statistical power and preodds interact with each other in the denominator in both derivatives, this marginal decrease in expected cost *decreases* with higher statistical power or prestudy odds (see denominators in Eqs. 22 and 23). Taken together, the benefits of increasing power and prestudy odds have their limits.

3.3. Estimation of the Cost of Research Problem and Research Investment from Health Economics

Public decision-makers and funding agencies have to estimate the benefits and costs of different actions to make decisions accountable and transparent (Layard & Glaister, 1994). Because different funding agencies might have different missions, their measures of benefit and cost of research might be different. To keep our CBA and our terms consistent with Sections 3.1.2 and 3.1.3, we estimate the cost of a research problem as the negative benefit to patients (e.g., how much are patients willing to pay for the cure to a disease). Similarly, we estimate the research investment as the cost of research.

The cost for a disease, PV(Cost), is measured as the present value (PV) of lost life expectancy of patients times the value of a year-life (Cutler & McClellan, 2001). Different age groups

have different life expectancies Mathematically, total cost can be expressed as

$$\text{Cost} = \left(\sum_i \left(\text{LE}_{\text{np}_i} - \text{LE}_{\text{patients}_i} \right) \times \text{ID}_i \right) \times \text{VAYL}, \tag{24}$$

where LE_{np_i} is the life expectancy of the normal population at age i, $\text{LE}_{\text{patients}_i}$ is the life expectancy of the patient population at age i, ID_i is the number of disease incidences at age i, and VAYL is the value of a year-life.

We take as an example disease, cancer. For cancer, the total research investment will be the combination of cancer medical care cost (CMC, e.g., cancer care in the United States) and the cost of research into cancer (CRC):

$$\text{Investment} = \text{CMC} + \text{CRC}. \tag{25}$$

Because we are considering cost as the total discounted cost into PV, we will do the same for investment. If a research program is designed to finish in N years, we consider the effect of time on the present cost and investment as follows

$$\text{PV(Cost)} = \frac{\text{Cost}}{\text{IF}} - \sum_{i=0}^{N} \frac{\text{Cost}}{(1 + \text{IF})^i} \tag{26}$$

$$\text{PV(Investment)} = \sum_{i=0}^{N} \frac{\text{Investment}}{(1 + \text{IF})^i}, \tag{27}$$

where IF is the inflation rate. Notice that in the PV of cost, we have subtracted the cost that the research cannot save.

There are other possible measures of the benefits and costs of biomedical research and research in general, such as the cost of disability of patients of a disease. Our method is a demonstration of how our framework generates different research action recommendations from the single consideration of the correctness of research.

4. DATA ON BIOMEDICAL RESEARCH

In this work, we use data from cancer research to illustrate how cost and investment influence research decision-making. Cancer is one of the leading causes of death in the world (National Cancer Institute, 2019a). The United States has comprehensive data about cancer patients, cancer care expenditure, and cancer research investment. We focus on cancer research with better data availability. In particular, we analyze colon and rectum cancer, brain cancer, lung cancer, female breast cancer, prostate cancer, lymphoma cancer, and ovarian cancer.

For all our analyses, we use the year 2010 as a source of data, extending when necessary. The cancer research investment is estimated as the amount of funding from NIH in that year (National Institutes of Health, 2019). Another investment in cancer is the cancer care expenditure. We use the total of cancer care expenditure in 2010 provided by the National Cancer Institute (2019b).

For our estimate of the cost of cancer for patients, we use the loss in a person's life expectancy as a base. More specifically, we collected survival time data from SEER-18 (Surveillance, Epidemiology, and End Results Program with 18 registry sites) to estimate life expectancy (Cutler, 2007; Luce, Mauskopf et al., 2006). These data have 18 registry sites (San Francisco-Oakland, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, Atlanta, San Jose-Monterey, Los Angeles, Alaska Natives, Rural Georgia, California, Kentucky, Louisiana, New Jersey, and

Greater Georgia) with a follow-up in 2016. Also, we collected the life expectancy of the normal population in 2010 from the Social Security Administration (2017). In addition, we collected incidence count data in 2010 from CDC for patients at different age groups for the previously mentioned kinds of cancer (Centers for Disease Control and Prevention, 2019).

5. RESULTS

In this article, we investigate how to incorporate CBA into decisions about whether or not to continue a program of research. Our work is in the context of recent evidence claiming that a relatively large proportion of research is not replicable and might be problematic. Previous investigations into this problem usually only applied probabilistic approaches, which do not consider the potential benefits of continuing research even if final results are inaccurate. We now apply our framework to several kinds of cancer and show that, depending on the expected benefits, the decision to continue research in spite of potential inaccurate results varies considerably.

5.1. Estimation of the Costs and Benefits of Cancer Research from Health Economics

We estimate the cost of cancer on patients and also cancer research investment as cancer research before making decisions based on them. We start by estimating the costs and investments using the method from Section 3.3. First, we present the total cost of different cancers on patients, cancer care expenditure, and research investment in 2010 (Table 2). In this study, we assume the cost of cancer is consistent across years (even in the future). Then, we incorporate the cumulative time value of these costs and investments with the concept of PV theory. Applied to Table 2, we obtain the PV for the cost of cancer for patients and cancer care expenditure, and for cancer research investment (Table 3).

We should expect to find a correlation between the cost incurred by people and the research investments. Indeed, we found a significant correlation between total costs for cancer patients and cancer care expenditures ($r = 0.771$, $p = 0.0425$, $N = 6$), and a positive but nonsignificant correlation between the total costs on cancer patients and investments in

Table 2. The cost of cancer for patients, cancer care expenditure, and cancer research investment in 2010 (US$ millions). The cost for patients is based on Eq. 24 and with the data of the loss of life time from SEER-18, the incidence count from CDC, and the value of a year of life (US$100,000, common in health economics literature). The cost of cancer care (Medicare payments and patient responsibilities for medical services) is from National Cancer Institute (2019b). The research investment is from National Institutes of Health (2022).

Cancer type	Cost for patients	Cancer care expenditure	Research investment
Lung	132,994	12,120	201
Female breast	107,702	16,499	763
Prostate	104,934	11,848	331
Colon and rectum	73,971	14,140	291
Lymphoma	40,469	12,142	195
Ovarian	11,801	5,116	122
Brain	10,608	4,469	274

Table 3. The cost of cancer for patients, cancer care expenditure, and cancer research investment in PV (US$ millions). In this table, we consider the present value of the above cost for patients as the benefit of cancer research to patients (willingness to pay to avoid loss in life time) and we consider the present value of cancer care expenditure and research investments as the cost of research in Table 2 with an inflation rate of 5%, which is common in literature of health economics.

Cancer type	Cost for patients	Cancer care expenditure	Research investment
Lung	2,659,887	242,414	4,020
Female breast	2,154,049	329,996	15,260
Prostate	2,098,685	236,962	6,620
Colon and rectum	1,479,436	282,810	5,820
Lymphoma	809,387	242,850	3,900
Ovarian	236,026	102,322	2,440
Brain	212,166	89,386	5,480

research ($r = 0.440$, $p = 0.323$, $N = 6$). These results give initial evidence that research is being done in areas that need it most.

5.2. Differential Effect of Cancer Type on Decision-Making

We should not expect to have all cancer types follow the same decision points. Some of them might benefit from exploratory research and others can also benefit from randomized control trials. We can precisely explore these differences by looking at the minimal PPV and R that each of them can afford. These minimal values represent break-even points where the cost of research equals the cost of no research, assuming different research into a certain type of cancer has the same benefit and cost. As an example, we analyzed these break-even points for the relatively likely scenario where we have high statistical power (power = 0.8) and minor bias ($\mu = 0.2$). If we assume cancer research can be completed in 20 years, we found that colon and rectum cancer research can be beneficial with PPV above 32.3% and R above 0.14, but ovarian cancer research needs PPV above 73.3% and R above 0.79. These results suggest that different cancers follow different decision points. Importantly, both PPV are below 50%, suggesting that the importance of research in this area supports continuing research while risking having inaccurate research findings. Finally, our results show different required PPV and R for a positive net gain from research (Table 4), but such results are limited to our data and estimation of costs and benefits of research.

We now want to investigate the effect of bias in the minimal required R. As we see in Section 2.1.3, the minimal required R in research changes as we adjust the factor of bias (μ). We analyze this effect with colon and rectum cancer, and ovarian cancer (Figure 1). Our results suggest that higher bias indeed increases the required prestudy R for the break-even point. For example, this effect is more pronounced for ovarian cancer than colon and rectum cancer.

5.3. Consequence of Improving Statistical Power and Prestudy Odds on Research Outcomes

A common suggestion to improve the condition of scientific research is to make research more rigorous—for example, do a higher statistical power research in a study. However, the

Table 4. Minimal PPV (m-PPV) and R (m-R) for break-even points (between the cost of research problem and research investment) of selected cancer research. By comparing cancer with a relatively higher benefit-cost ratio (such as lung cancer) and cancer with a relatively lower benefit-cost ratio (such as brain cancer), we can see the latter kind of cancer needs higher PPV and R to reach a break-even point.

| | Assume the research will be completed in 20 years | | | |
| | $\alpha = 0.05, \beta = 0.2, \mu = 0.2$ | | $\alpha = 0.05, \beta = 0.4, \mu = 0.5$ | |
Cancer type	m-PPV (%)	m-R	m-PPV(%)	m-R
Lung	15.3	0.051	15.3	0.12
Prostate	19.2	0.068	19.2	0.16
Female breast	26.5	0.10	26.5	0.24
Colon and rectum	32.3	0.14	32.3	0.31
Lymphoma	50.4	0.29	50.4	0.67
Ovarian	73.3	0.79	73.3	1.8
Brain	73.9	0.81	73.9	1.9

effectiveness of this recommendation on the expected cost of the disease is unclear. Therefore, with our cost and benefit estimation method, we now investigate if this suggestion affects diseases differently depending on the cost-benefit ratio.

Indeed, we find that the marginal benefit of improving research rigor decreases. For example, colon cancer sees large benefits with improved statistical power at the beginning but those benefits decrease later on (Figure 2). This result suggests that improving research rigor does indeed bring benefits that are mediated by cost-benefit ratios. Even though the actual
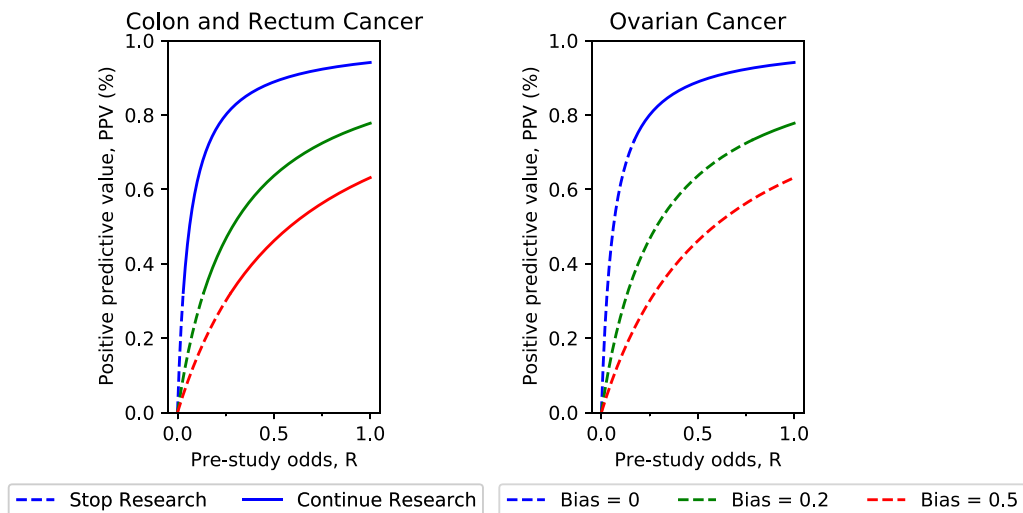

Figure 1. Prestudy odds for break-even points (between the cost of research problem and investment) in colon and rectum cancer research and ovarian cancer research. In this analysis, we assume these research problems can be solved in 20 years, and α is 0.05 and β is 0.2. Small bias in research refers to μ of 0 and medium bias in research refers to μ of 0.2, and large bias in research refers to μ of 0.5. The comparison between the above two subplots shows that colon and rectum cancer research (research with high benefit-cost ratio) can tolerate lower prestudy, especially when bias in research is relatively large.
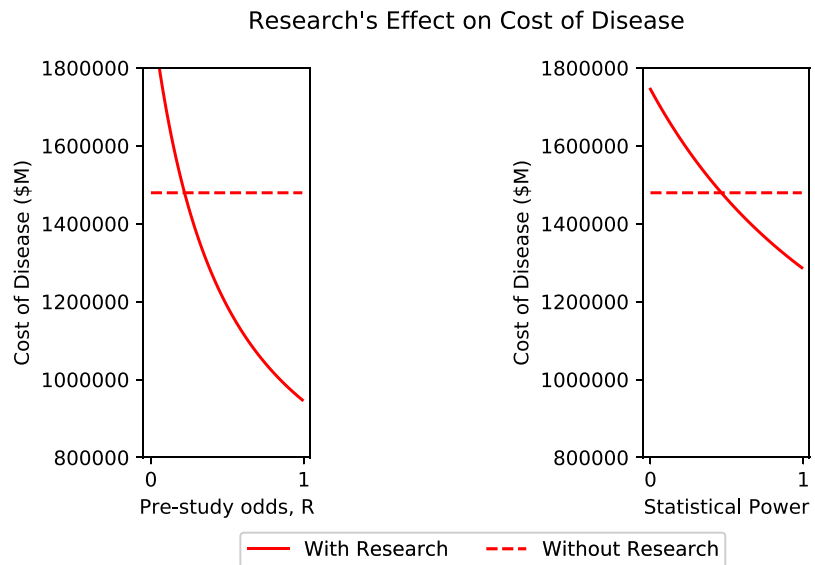
Research's Effect on Cost of Disease

Figure 2. Consequences of improving prestudy odds and power of research on the cost of diseases (the total loss of value of life years from patients, in million dollars). In this analysis, α is 0.05 and μ is 0.2 in both subfigures, power is 0.4 in the left subfigure, and R is 0.2 in the right subfigure. Following Eqs. 22 and 23, the marginal in cost of diseases from prestudy odds and power of research decreases as prestudy odds and power of research increase (this pattern is clearer with research of a lower benefit-cost ratio).

consequences of these research recommendations on the cost of diseases might vary from the missions of funding agencies, this analysis shows that relationship between the statistical power of research and the cost of diseases and the relationship between prestudy odds of research and the cost of diseases might be nonlinear.

5.4. Research Designs Based on Cost and Research Investment

One of our motivations for studying the costs of research problems and research investment is to provide suggestions about the best research designs for different research questions. Previous research has done this but only considering probabilistic factors (i.e., statistical power, prestudy odds ratio, and bias). To demonstrate our extension, we use our measure of the cost of research problem and research investment and apply Ioannidis's estimation of PPV on different kinds of research scenarios considered in Ioannidis (2005b) ("Practical example" column in Table 5). The idea is to have a well-motivated reason to allow research to do more exploratory analysis and for others to do more rigorous (e.g., higher statistical power) research. In this analysis, we made the assumption that different kinds of research in the cancer research field have the same benefits and costs—clearly simplistic.

According to our analysis, RCTs are suitable for all types of cancer under this study, and exploratory epidemiological studies are suitable for lung cancer and prostate cancer research. Even though discovery-oriented exploratory research with massive testing might not be desirable for any type of cancer under this study, research fields with a higher ratio between benefit and cost could afford this kind of study. These suggestions show the difference between research decision-making with and without the costs and benefits of research. Therefore, new possible research actions can be made by funding agencies, according to their own cost and benefit estimation methods.

Table 5. Research suggestions for each field. Naming and threshold taken from Ioannidis (2005b). This table shows that research (assuming it can be completed in 20 years) with relatively higher benefit-cost ratio, such as lung cancer, might be worthy of conducting research that is likely to be false (such as exploratory epidemiological studies).

$1 - \beta$	R	μ	Practical example	PPV	Suggested research fields with cost-benefit analysis (type of cancer)	Suggested research fields without cost-benefit analysis (Ioannidis, 2005b)
0.80	1:1	0.10	Adequately powered randomized controlled trial (RCT) with little bias and 1:1 prestudy odds	0.85	Colon and rectum cancer, brain cancer, lung cancer, female breast cancer, prostate cancer, lymphoma cancer, ovarian cancer	Colon and rectum cancer, brain cancer, lung cancer, female breast cancer, prostate cancer, lymphoma cancer, ovarian cancer
0.95	2:1	0.30	Confirmatory meta-analysis of good-quality RCTs	0.85	Colon and rectum cancer, brain cancer, lung cancer, female breast cancer, prostate cancer, lymphoma cancer, ovarian cancer	Colon and rectum cancer, brain cancer, lung cancer, female breast cancer, prostate cancer, lymphoma cancer, ovarian cancer
0.80	1:3	0.40	Meta-analysis of small inconclusive studies	0.41	Colon and rectum cancer, lung cancer, female breast cancer, prostate cancer	None recommended
0.20	1:5	0.20	Underpowered, but well-performed phase I/II RCT	0.23	Colon and rectum cancer, lung cancer, female breast cancer, prostate cancer	None recommended
0.20	1:5	0.80	Underpowered, poorly performed phase I/II RCT	0.17	Lung cancer	None recommended
0.80	1:10	0.30	Adequately powered exploratory epidemiological study	0.20	Lung cancer, prostate cancer	None recommended
0.20	1:10	0.30	Underpowered exploratory epidemiological study	0.12	None recommended	None recommended
0.20	1:1,000	0.80	Discovery-oriented exploratory research with massive testing	0.0010	None recommended	None recommended
0.20	1:1,000	0.20	As in previous example, but with more limited bias (more standardized)	0.0015	None recommended	None recommended

6. DISCUSSION

In this article, we have investigated how CBA can significantly affect research decisions, extending current metaresearch considerations about the likelihood of published results with low accuracy. We derived decision principles based on Bayesian statistics that incorporate research with potentially inaccurate findings into the scientific process. Our results suggested that indeed costs and benefits justify continuing research even if we risk research with potentially inaccurate findings being produced. CBA, in the end, provides more complex but realistic justifications for different research actions. When applied to several kinds of cancer research, our method provided possible suggestions about research actions based on our estimation of costs and benefits. Considering this framework and the potential significant research recommendations, our proposal could form a useful policy tool to justify potentially

risky research based on the needs of the questions. We now discuss the implications of our method and our contributions.

6.1. Contributions

Our method is an integration of CBA and a Bayesian decision-making framework that is amenable for health economic decisions. Within this framework, we were able to factor in costs and benefits directly into a decision function (Eq. 6). This function only needs the cost structure of the problem (i.e., the λ_{ij} for action i and relationship j in Section 3.1.1). There could be alternative formulations to incorporate CBA, even without needing to resort to Bayesian theory. For example, we could apply the optimal stopping theory developed by Wald (1949), which does not need prior probabilities. An alternative formulation could also add a game-theoretic component, which is closer to the real world but more complicated to apply (Nissen et al., 2016). For example, a regret minimization framework is one way of adding costs in game theory/competition settings (Zinkevich, Johanson et al., 2007). In the work of Djulbegovic and Hozo (2007), regret minimization is used with the goal of establishing the point at which false research is acceptable. Regret minimization, however, is meant to be done *after* decisions are made, while our framework is forward looking. Also, the relatively simplicity of our approach makes it attractive for policy decisions with all the benefits of Bayesian theory—namely, transparency for the researcher and potential users (e.g., funding agencies) to explicitly specify priors and understand the objectives being optimized. Our framework also provides (Bayesian) uncertainty bounds on all estimations and not just single-point ones. In our particular application, we are able to draw and incorporate language and numbers from health economics (e.g., Cutler, 2007). We use the value of life years as the benefit of research which is recommended in health economics as well (McIntosh, Donaldson, & Ryan, 1999). In summary, our framework and applications highlight the potential benefit of research with uncertain accuracy.

A group of researchers have discussed how to guide science policy to react to the concern of the correctness of research. Benjamin et al. (2018) proposed making the current statistical threshold (statistical significance) more stringent. In contrast, McShane, Gal et al. (2019) suggested reconsidering the tendency of publishers to favor research findings with statistical significance. Nissen et al. (2016) suggest that negative research findings of the original research studies should be published, so that false claims can be identified by research communities. These studies bring reasonable suggestions from different perspectives or statistical assumptions, but the dimension of the cost and benefit of research is not included in their reasoning. For example, our result showed that benefits and costs could drive research decision-makers to accept the risk of producing inaccurate findings. The current funding policy of NIH on resource allocation across fields is mainly focused on well-being and the potential value of research (National Institutes of Health, 2015), but it has been unclear how to determine the right amount of risk-taking. Our analysis is able to demonstrate that tolerance of inaccurate research findings changes with the costs and benefits of research. Even though science funding agencies have recognized that some high-risk research might bring a high payoff, there is little formalization of this principle.

To contribute to the formalization of the trade-off between the costs and benefits of research and the correctness of research, our analysis demonstrates how dramatically different inaccurate research findings can be tolerated with small changes in costs and benefits. For example, because colon and rectum cancer have a higher ratio between the benefit and the cost of research, we could afford inaccurate research to be conducted (see Figure 1). Even though our measure of

the benefits and costs of research is from a health economics perspective and also is not perfect, this increment in the tolerance of inaccurate research findings might motivate researchers and funding agencies to propose a new measure of the benefits and costs of research.

As concern about the replicability of research arises, many researchers attempt interpret this phenomena with statistical variables, such as statistical power and prestudy odds. For example, Ioannidis (2005b) suggested that high statistical power might help the research community to produce more true positive research findings, which should be replicable. Ulrich and Miller (2020), meanwhile, demonstrated that low prestudy odds might be the root cause of a low replicability of previous research findings, according to their statistical model. In this article, we also want to examine how statistical variables, such as statistical power, produce impacts on the return of research. We analyze (see Eqs. 22 and 23) how statistical power, prestudy odds, and bias can affect the tolerance of inaccurate research findings when we assume other variables to have a certain value. The result of this analysis presented the way in which these variables affect the trade-off between the risk of research and the payoff of research (see Figure 2). This result demonstrated that statistical power and prestudy odds reduce cost in a nonlinear manner. If a field expects a significant reduction in the expected cost of research, then that field not only should be more rigorous (e.g., higher prestudy odds and statistical power research), but should also be encouraged to improve its cost to benefit ratio. Of course, there are multiple caveats, such as statistical quantities (statistical power or p-value) that may vary from study to study (e.g., need contexts for interpretations; Betensky, 2019). However, our results can be used to motivate guidelines for proposal review criteria and resource allocation in funding agencies. More specifically, if a research field can fund research with low prestudy odds, as Figure 2 (left) suggests, the magnitude of the reduction of the cost will diminish more slower when the prestudy odds is greater than 0.2. This implies that funding agencies can still fund relatively risky research with a high payoff, but should do so with reasonable statistical power research. If we assume that the number of positive research findings does not change when we control the statistical power of research, then high statistical power research can reduce the cost of research problems, but the effect of this reduction diminishes as the statistical power of research increases.

Our framework can help us embrace publication environments with high uncertainty and tolerance for exploration provided that expected benefits far outweigh costs. CBA can bring principled decisions to current research actions.

6.2. Limitations

Our model and its applications have shortcomings. First, the standard criticisms of Bayesian statistics apply to us too. Although our model is open about its assumptions, we need to assume, without evidence, the very important parameter R, which is related to the true-to-false relationship ratio. While this is usually based on anecdotal evidence about how hard a research question is and how long it has been since someone has advanced a related research question, it is mainly guesswork. Second, there are multiple approaches to cost and benefit estimation for medical research (Grant & Buxton, 2018). Getting an accurate estimate of these values can be highly difficult. Even in major health economics areas, some measurements are not included in the estimations (e.g., psychological costs for patients; Naughton & Weaver, 2014). Researchers have proposed other nontraditional costs and benefits related to scientists careers, such as increased chances of getting tenure after a publication is accepted in a top journal (Gelman, 2018).

In the future, we plan on adding them to our analysis. Also, the inner workings of funding agencies should be incorporated, but these are even harder to estimate because they

contain values that are hard to measure, such as training, taxpayers' mandates, and social good (National Institutes of Health, 2015). Still, the action of forcing us to think about costs and benefits in research is a good first step in contextualizing science and its uses of resources.

When taking our model and results to other settings or areas, we should be cautious about their generalization power. First, in some fields, costs and benefits are controversial or even impossible to define. For example, basic research areas such as mathematics, physics, and their related areas of theoretical computer science do not need to have immediate applications (Glänzel, Schlemmer, & Thijs, 2003; Ke, Ferrara et al., 2015). Also, some fields are not as reliant on research funding but rather on their educational value, which is hard to measure, and perhaps accrues not at a project level but at a multiproject level. The notion of costs and benefits of research should be evaluated with caution.

All in all, our framework and applications could enrich the toolkit that we can use to understand research itself. Many of the discussions on replicability, reproducibility, bias, and transparency are too often framed as questions of possibilities. However, a more realistic scenario is to consider the cost trade-offs between continuing research or stopping it. In our framework, we provide concrete tools and a way of thinking about these issues that we believe are more grounded in how science works—within monetary and time resource budgets.

7. CONCLUSION

The correctness of research findings is a long-term challenge for scientists and science funding agencies. But before we can improve the correctness of research findings concretely, funding agencies might need to consider how much risk they should take and how funding evaluations should be made. In this article, we have expanded a model of the probability that research is true and incorporated the cost-benefit dimension to it. We showed that there are multiple implications for the question of how to manage a program of research from an economics perspective. Our estimated decision points and possible action suggestions for research show many new possibilities for funding agencies to reconsider their funding policy, such as a comprehensive benefit measurement of research.

In the future, we plan to make our framework more realistic by incorporating sequential experiments, where evidence for a relationship is gathered through a series of studies. Similarly, we will expand it to consider that scientists' work and research projects do not happen in a vacuum but rather through competition and collaboration across scientists, funding agencies, and countries. It is likely that these extensions will significantly complicate our model, but we think that these aspects are crucial and warranted in an increasingly team-based and interconnected community of scientists.

In summary, our study provides a nuanced view of the concern that most research findings might be inaccurate. Science is a crucial part of modern societies, and thinking about how it lives within this society is crucial for its government and public support.

AUTHOR CONTRIBUTIONS

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

DATA AVAILABILITY

All the data is self-contained in this article. All results can be derived from the equations and such data.

REFERENCES

Ackerman, F., & Heinzerling, L. (2002). Pricing the priceless: Cost-benefit analysis of environmental protection. *University of Pennsylvania Law Review*, *150*, 1553–1584. https://doi.org/10.2307/3312947

Atkinson, G., & Mourato, S. (2008). Environmental cost-benefit analysis. *Annual Review of Environment and Resources*, *33*, 317–344. https://doi.org/10.1146/annurev.environ.33.020107.112927

Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531–533. https://doi.org/10.1038/483531a, PubMed: 22460880

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A. Wagenmakers, E.-J., ... Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*(1), 6–10. https://doi.org/10.1038/s41562-017-0189-z, PubMed: 30980045

Betensky, R. A. (2019). The *p*-value requires context, not a threshold. *American Statistician*, *73*, 115–117. https://doi.org/10.1080/00031305.2018.1529624

Bockstael, N. E., Freeman, A. M., Kopp, R. J., Portney, P. R., & Smith, V. K. (2000). On measuring economic values for nature. *Environmental Science & Technology*, *34*(8), 1384–1389. https://doi.org/10.1021/es990673l

Bond, M. E., Perkins, S. N., & Ramirez, C. (2012). Students' perceptions of statistics: An exploration of attitudes, conceptualizations, and content knowledge of statistics. *Statistics Education Research Journal*, *11*(2), 6–25. https://doi.org/10.52041/serj.v11i2.325

Bush, H. M. (2011). *Biostatistics: An applied introduction for the public health practitioner*. Nelson Education.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., ... Wu, H. (2018). Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z, PubMed: 31346273

Centers for Disease Control and Prevention. (2019). *United States and Puerto Rico cancer statistics, 1999–2016 incidence archive request*. https://wonder.cdc.gov/cancer-v2016.html

Cowan, R., David, P. A., & Foray, D. (2000). The explicit economics of knowledge codification and tacitness. *Industrial and Corporate Change*, *9*, 211–253. https://doi.org/10.1093/icc/9.2.211

Cutler, D. M. (2005). Intensive medical technology and the reduction in disability. In D. A. Wise (Ed.), *Analyses in the economics of aging* (pp. 161–184). Chicago, IL: University of Chicago Press. https://doi.org/10.7208/chicago/9780226903217.003.0006

Cutler, D. M. (2007). The lifetime costs and benefits of medical technology. *Journal of Health Economics*, *26*(6), 1081–1100. https://doi.org/10.1016/j.jhealeco.2007.09.003, PubMed: 18023900

Cutler, D. M., & McClellan, M. (2001). Is technological change in medicine worth it? *Health Affairs*, *20*(5), 11–29. https://doi.org/10.1377/hlthaff.20.5.11, PubMed: 11558696

Cutler, D. M., & Meara, E. (2001). *Changes in the age distribution of mortality over the 20th century* (Technical report). National Bureau of Economic Research. https://doi.org/10.3386/w8556

Directorate for Social, Behavioral, and Economic Sciences. (2019). *NSF budget request to congress*. https://www.nsf.gov/about/budget/fy2019/pdf/28_fy2019.pdf

Djulbegovic, B., & Hozo, I. (2007). When should potentially false research findings be considered acceptable? *PLOS Medicine*, *4*(2), e26. https://doi.org/10.1371/journal.pmed.0040026, PubMed: 17326703

Djulbegovic, B., Hozo, I., Schwartz, A., & McMasters, K. M. (1999). Acceptable regret in medical decision making. *Medical Hypotheses*, *53*(3), 253–259. https://doi.org/10.1054/mehy.1998.0020, PubMed: 10580533

Drummond, M. F., Davies, L. M., & Ferris, F. L., III. (1992). Assessing the costs and benefits of medical research: The diabetic retinopathy study. *Social Science & Medicine*, *34*(9), 973–981. https://doi.org/10.1016/0277-9536(92)90128-D, PubMed: 1631610

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*. Chichester: John Wiley & Sons.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., ... Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), eaao0185. https://doi.org/10.1126/science.aao0185, PubMed: 29496846

Fuchs, V. R. (2000). The future of health economics. *Journal of Health Economics*, *19*(2), 141–157. https://doi.org/10.1016/S0167-6296(99)00033-8, PubMed: 10947574

Gelman, A. (2018). How to think scientifically about scientists' proposals for fixing science. *Socius*, *4*. https://doi.org/10.1177/2378023118785743

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Boca Raton, FL: CRC Press. https://doi.org/10.1201/b16018

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science". *Science*, *351*(6277), 1037. https://doi.org/10.1126/science.aad7243, PubMed: 26941311

Gillum, L. A., Gouveia, C., Dorsey, E. R., Pletcher, M., Mathers, C. D., ... Johnston, S. C. (2011). NIH disease funding levels and burden of disease. *PLOS ONE*, *6*(2), e16837. https://doi.org/10.1371/journal.pone.0016837, PubMed: 21383981

Glänzel, W., Schlemmer, B., & Thijs, B. (2003). Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon. *Scientometrics*, *58*, 571–586. https://doi.org/10.1023/B:SCIE.0000006881.30700.ea

Goodman, C. S. (2004). *Introduction to health technology assessment*. Falls Church, VA: The Lewin Group.

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, *8*(341), 341ps12. https://doi.org/10.1126/scitranslmed.aaf5027, PubMed: 27252173

Grant, J., & Buxton, M. J. (2018). Economic returns to medical research funding. *BMJ Open*, *8*(9), e022131. https://doi.org/10.1136/bmjopen-2018-022131, PubMed: 30201795

Ioannidis, J. P. A. (2005a). Contradicted and initially stronger effects in highly cited clinical research. *JAMA*, *294*(2), 218–228. https://doi.org/10.1001/jama.294.2.218, PubMed: 16014596

Ioannidis, J. P. A. (2005b). Why most published research findings are false. *PLOS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124, PubMed: 16060722

Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying Sleeping Beauties in science. *Proceedings of the National Academy of Sciences*, *112*(24), 7426–7431. https://doi.org/10.1073/pnas.1424329112, PubMed: 26015563

Kornfeld, D. S. (2012). Perspective: Research misconduct. The search for a remedy. *Academic Medicine*, *87*(7), 877–882. https://doi.org/10.1097/ACM.0b013e318257ee6a, PubMed: 22622208

Kuhn, T. S. (2012 [1962]). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press. https://doi.org/10.7208/chicago/9780226458144.001.0001

Layard, R., & Glaister, S. (1994). *Cost-benefit analysis*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511521942

Luce, B. R., Mauskopf, J., Sloan, F. A., Ostermann, J., & Paramore, L. C. (2006). The return on investment in health care: From 1980 to 2000. *Value in Health*, *9*(3), 146–156. https://doi.org/10.1111/j.1524-4733.2006.00095.x, PubMed: 16689708

Lundvall, B.-Å. (2004). The economics of knowledge and learning. In J. L. Christensen & B.-Å. Lundvall (Eds.), *Product innovation, interactive learning and economic performance*. Leeds: Emerald Group Publishing. https://doi.org/10.1016/S0737-1071(04)08002-3

McIntosh, E., Donaldson, C., & Ryan, M. (1999). Recent advances in the methods of cost-benefit analysis in healthcare: Matching the art to the science. *PharmacoEconomics*, *15*(4), 357–367. https://doi.org/10.2165/00019053-199915040-00003, PubMed: 10537954

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *American Statistician*, *73*, 235–245. https://doi.org/10.1080/00031305.2018.1527253

Miller, J., & Ulrich, R. (2016). Optimizing research payoff. *Perspectives on Psychological Science*, *11*(5), 664–691. https://doi.org/10.1177/1745691616649170, PubMed: 27694463

National Cancer Institute. (2019a). *Cancer statistics*. https://www.cancer.gov/about-cancer/understanding/statistics

National Cancer Institute. (2019b). *Financial burden of cancer care*. https://progressreport.cancer.gov/after/economic_burden

National Institutes of Health. (2015). *NIH-wide strategic plan*. https://www.nih.gov/about-nih/nih-wide-strategic-plan

National Institutes of Health. (2019). *Estimates of funding for various research, condition, and disease categories (RCDC)*. https://report.nih.gov/funding/categorical-spending#/

National Institutes of Health. (2022). *NIH categorical spending*. https://report.nih.gov/funding/categorical-spending#/

Naughton, M. J., & Weaver, K. E. (2014). Physical and mental health among cancer survivors: Considerations for long-term care and quality of life. *North Carolina Medical Journal*, *75*(4), 283–286. https://doi.org/10.18043/ncm.75.4.283, PubMed: 25046097

Nichol, K. L. (2001). Cost-benefit analysis of a strategy to vaccinate healthy working adults against influenza. *Archives of Internal Medicine*, *161*(5), 749–759. https://doi.org/10.1001/archinte.161.5.749, PubMed: 11231710

Nissen, S. B., Magidson, T., Gross, K., & Bergstrom, C. T. (2016). Publication bias and the canonization of false facts. *eLife*, *5*, e21451. https://doi.org/10.7554/eLife.21451, PubMed: 27995896

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114, PubMed: 29531091

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716, PubMed: 26315443

Partha, D., & David, P. A. (1994). Toward a new economics of science. *Research Policy*, *23*(5), 487–521. https://doi.org/10.1016/0048-7333(94)01002-1

Popper, K. (2005 [1934]). *The logic of scientific discovery*. London: Routledge. https://doi.org/10.4324/9780203994627

Raff, E. (2019). A step toward quantifying independently reproducible machine learning research. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (pp. 5485–5495).

Remler, D. K., & Ryzin, G. G. V. (2014). *Research methods in practice: Strategies for description and causation*. Thousand Oaks, CA: Sage.

Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal*, *312*(7023), 71–72. https://doi.org/10.1136/bmj.312.7023.71, PubMed: 8555924

Sampat, B. N., Buterbaugh, K., & Perl, M. (2013). New evidence on the allocation of NIH funds across diseases. *Milbank Quarterly*, *91*(1), 163–185. https://doi.org/10.1111/milq.12005, PubMed: 23488714

Social Security Administration. (2017). *Period life tables*. https://www.ssa.gov/oact/HistEst/PerLifeTables/2017/PerLifeTables2017.html

Stirling, A. (1997). Limits to the value of external costs. *Energy Policy*, *25*(5), 517–540. https://doi.org/10.1016/S0301-4215(97)00041-4

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Ulrich, R., & Miller, J. (2020). Meta-research: Questionable research practices may have little effect on replicability. *eLife*, *9*, e58237. https://doi.org/10.7554/eLife.58237, PubMed: 32930092

Wald, A. (1949). Statistical decision functions. *Annals of Mathematical Statistics*, *20*(2), 165–205. https://doi.org/10.1214/aoms/1177730030

Wang, S. J., Middleton, B., Prosser, L. A., Bardon, C. G., Spurr, C. D., … Bates, D. W. (2003). A cost-benefit analysis of electronic medical records in primary care. *American Journal of Medicine*, *114*(5), 397–403. https://doi.org/10.1016/S0002-9343(03)00057-3, PubMed: 12714130

Weingart, P. (2002). The moment of truth for science: The consequences of the 'knowledge society' for society and science. *EMBO Reports*, *3*(8), 703–706. https://doi.org/10.1093/embo-reports/kvf165, PubMed: 12151322

Zinkevich, M., Johanson, M., Bowling, M., & Piccione, C. (2007). Regret minimization in games with incomplete information. In *Proceedings of the 20th International Conference on Neural Information Processing Systems* (pp. 1729–1736).