Causal Analysis of Absolute and Relative Risk Reductions

Björn Meder[1], Charley M. Wu[2,3,4], and Felix G. Rebitschek[3,5]
[1]Institute for Mind, Brain and Behavior, Health and Medical University, Potsdam, DE
[2]Human and Machine Cognition Lab, University of Tübingen, Tübingen, DE
[3]Max Planck Institute for Human Development, Berlin, DE
[4]Department of Computational Neuroscience, Max Planck Institute for Biological Cybernetics, Tübingen, DE
[5]Harding Center for Risk Literacy, Faculty of Health Sciences Brandenburg, University of Potsdam, Potsdam, DE

Abstract

Any new medical innovation must first prove its benefits with reliable evidence from clinical trials. Evidence is commonly expressed using two metrics, summarizing treatment benefits based on either absolute risk reductions (ARRs) or relative risk reductions (RRRs). While both metrics are derived from the same data (e.g., observed frequencies of a disease in a treatment and control group), they implement conceptually and causally distinct ideas. Here, we analyze these risk reductions measures from a causal modeling perspective, revealing an isomorphic relationship to central measure of causal strength and causal Bayes nets. First, we show that ARR is equivalent to ΔP, while RRR is equivalent to causal power, thus clarifying the implicit causal assumptions. Second, we show how this formal equivalence establishes a relationship with causal Bayes nets theory, offering a basis for incorporating risk reduction metrics into a computational causal modeling framework. Drawing on these analyses, we demonstrate that under dynamically varying baseline risks, ARRs are inadequate when generalizing treatment effects to novel contexts. Specifically, the inherent assumption of a linear parameterization of the underlying causal graph leads to incorrect conclusions when generalizing to baseline risks differing from those from which the original effect was obtained. For instance, generalizing the effect of a vaccine to new contexts with different baseline risks (e.g., from one population to another). Our analyses highlight the shared principles underlying risk reduction metrics and measures of causal strength, emphasizing the potential for explicating causal structure and inference in medical research.

Keywords: causal models, relative risk reduction, absolute risk reduction, causal power, causal Bayes net, causal modeling, risk communication, counterfactuals

During the Covid-19 pandemic, the quantification of risk posed an important challenge. How dangerous was it to go to the grocery store in person vs. paying extra to have them delivered to your doorstep? How much safer was it to go to crowded places once you had one or two vaccines? We all struggled with these dilemmas, while medical professions fought to communicate the most accurate information to the public. Assessing and communicating the effectiveness of vaccinations and treatments against Covid-19 made this challenge even more complex, further complicating the public's understanding of risk and benefits.

This struggle highlights a broader issue in public health: the quantification and communication of treatment effects. Quantifying the causal impact of treatments and communicating medical information in an effective manner remains a key challenge for public policy and informed decision making (Bonner et al., 2021). Two widely used metrics for quantifying treatment benefits are relative and absolute risk reductions, which convey information about risk changes, such as the reduced risk of disease for vaccinated individuals or the reduced risk of recurrence for patients undergoing chemotherapy. These measures play a crucial role in summarizing study outcomes, in the communication of treatment benefits to healthcare professionals and the public, as well as tools for generalizing medical effects to new contexts and populations.

Mathematically, absolute and relative risk reductions are calculated from the same information, such as the frequency of an undesirable outcome in a randomized control trial (RCT). *Absolute risk reductions* (ARRs) correspond to the arithmetic difference in event rates between a treatment and a control group, quantifying the size of the effect and providing an estimate of how likely an individual will benefit from a treatment. *Relative risk reductions* (RRRs) correspond to a normalized difference in event rates, where the reduced risk in the treatment group is expressed relative to the event rate in the control group. While derived from the same data, psychologically the two formats can yield diverging evaluations in both lay people and health care professionals (Akl et al., 2011; Bobbio et al., 1994; Covey, 2007; Marcatto et al., 2013; Perneger & Agoritsas, 2011). These discrepancies have sparked debates on the relative merits of each metric, leading to recommendations for reporting both quantities when communicating treatment benefits and harms (Higgins et al., 2023; Moher et al., 2010).

Here, we highlight central connections between the risk reduction measures used in medical research on the one hand, and causal learning theory and the framework of causal Bayes nets theory on the other hand. First, we demonstrate that absolute and relative risk reductions are mathematically equivalent to two prominent measures of causal strength, namely the ΔP model (Cheng & Novick, 1992; Waldmann & Holyoak, 1992) and the

---

Björn Meder https://orcid.org/0000-0002-9326-400X
Charley M. Wu https://orcid.org/0000-0002-2215-572X
Felix G. Rebitschek https://orcid.org/0000-0001-7653-8278
CMW is supported by the German Federal Ministry of Education and Research (BMBF): Tübingen AI Center, FKZ: 01IS18039A and funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC2064/1–390727645.
Correspondence concerning this article should be addressed to Björn Meder, Institute for Mind, Brain and Behavior, Department of Health, Health and Medical University Potsdam, Olympischer Weg 1, 14471 Potsdam, Germany. E-mail: bjoern.meder@hmu-potsdam.de

causal power model (Cheng, 1997; Novick & Cheng, 2004). Thus, measures from the causal modeling literature are incidentally used under different labels in medical research and risk communication, yet these connections are not well known (but see Sprenger, 2018). Second, we show how this equivalence establishes connections with the formalism of causal Bayes nets (Glymour, 2003; Pearl, 2000; Pearl & Mackenzie, 2018; Spirtes et al., 2000), offering a basis for incorporating risk reductions into a causal modeling framework. This integration reveals how the choice of a particular risk reduction measure inevitably corresponds to different assumptions about how causes probabilistically influence an undesirable outcome. Third, we highlight the diverging consequences of employing either absolute or relative risk reductions to measure treatment effects when generalizing study results to scenarios with different baseline risks, demonstrating the limitations of absolute risk measures for counterfactual causal inference.

From Risk Reductions to Measures of Causal Strength

The purpose of risk reductions metrics is to quantitatively summarize treatment effects, such as the impact of a vaccine on disease cases. This problem is both conceptually and mathematically related to the problem of causal induction: How can we induce unobservable causal relations from observed data and quantify their strength? The nature of the inference mechanisms that support the development of both intuitive and scientific theories about the world is a key issue in philosophy (Cartwright, 2007), psychology (Waldmann, 2017), statistics (Pearl et al., 2016), and machine learning (Schölkopf, 2022). Because medical interventions naturally refer to cause-effect relations, researchers have also emphasized the importance of causal inference in the medical domain (Bareinboim & Pearl, 2016; Etminan et al., 2020; Greenland et al., 1999; Prosperi et al., 2020; Sanchez et al., 2022; Stovitz & Shrier, 2019). At the same time, explicit specification and consideration of causal structure is rare, despite the potential of recent advancements for clinical research. For instance, causal modeling techniques provide a calculus for expressing the key difference between observational and interventional probabilities, thereby supporting the development of interventional clinical predictive models and the calculation of counterfactuals (Prosperi et al., 2020; Sanchez et al., 2022).

Here, we offer a causal analysis of risk reduction metrics, showing that measures from the causal modeling literature are commonly applied under different labels in both medical research and risk communication. Our analyses clarify the implicit assumptions arising from the alignment between risk reduction formats and metrics of causal strength. They also reveal the limitations of using ARRs to predict a treatment's effect in scenarios where the baseline risk differs from that used for the initial estimation.

Measuring treatment effects: Absolute and relative risk reductions

Absolute and relative risk reductions are measures for quantifying changes in the risk associated with the occurrence of an adverse binary health outcome (e.g., disease present or absent). Applied to an experimental design with two groups (treatment vs. control), the ARR of a treatment corresponds to the arithmetic difference of the probability of the adverse outcome under the treatment (e.g., vaccine) compared to the control group (e.g., placebo):

$$ARR = P(\text{event}|\text{control}) - P(\text{event}|\text{treatment}) \tag{1}$$

In applied settings, these probabilities and their difference are typically expressed as percentages. Consider the Covid-19 pandemic, where different vaccines were evaluated in randomized controlled trials (e.g., Baden et al., 2021; Polack et al., 2020; Voysey et al., 2021). For instance, in the BNT162b2 (Pfizer-BioNTech) phase 3 trial, 162 out of 18,325 people (0.88%) in the placebo group developed symptomatic Covid-19, compared to 8 out of 18,198 people (0.04%) in the vaccine group (Polack et al., 2020). The absolute risk reduction is the difference between these event rates, which is $0.88\% - 0.04\% = 0.84$ percentage points (Table 1). This measure provides a quantitative estimate of the vaccine's ability to reduce disease cases.

Another way of expressing treatment benefits is the *relative risk reduction* (RRR). The RRR normalizes the arithmetic difference (i.e., ARR) on the probability of the disease in the control group:

$$RRR = \frac{P(\text{event}|\text{control}) - P(\text{event}|\text{treatment})}{P(\text{event}|\text{control})} \tag{2}$$

Given the observed frequencies in the BNT162b2 trial, the RRR is 95%, as it reduces the number of infections from 162 cases in the placebo group to 8 cases in the vaccine group. Risk reductions and treatment benefits are frequently expressed in such relative terms. For instance, in vaccine epidemiology the *efficacy* of a vaccine is commonly defined as the RRR estimated from a randomized placebo-controlled study (Greenwood & Yule, 1915; Tentori et al., 2021; Weinberg & Szilagyi, 2010).

Table 1

Data from the BNT162b2 vaccine trial against Covid-19 (Polack et al., 2020). Shown are registered (symptomatic) Covid-19 cases in the vaccine and placebo group. Computed are the probability of the event (Covid-19) in each group, absolute risk reduction (ARR), and relative risk reduction (RRR).

	Covid-19	No Covid-19	P(Event)	ARR	RRR
Control: Placebo	162	18,163	0.88%	0.84%	95%
Treatment: Vaccine	8	18,190	0.04%		

While ARRs and RRRs are calculated from the same data, their values can vary greatly. For instance, a 50% RRR could correspond to an ARR of 30% if the event rate decreases from 60% to 30%. But, it could also correspond to an absolute decrease of 1%, if the event rate decreases from 2% to 1%. The former would indicate a substantial effect, while the latter might be negligible in practice because the absolute risk of experiencing the undesirable event is low even without the treatment. Thus, a given relative reduction can

correspond to widely varying absolute reductions, with these differences crucially depending on the baseline risk.

This discrepancy can have profound implications for communicating medical information to policy makers, healthcare professionals, and laypeople. For instance, RRRs tend to yield more positive assessments of treatment benefits through healthcare professionals and laypeople than when the same effects are expressed as ARRs (Akl et al., 2011; Covey, 2007; Forrow et al., 1992; Malenka et al., 1993). Several researchers have therefore argued that only reporting RRRs can be misleading when communicating health information to practitioners and the general public (Ancker et al., 2006; Gigerenzer & Edwards, 2003; Gigerenzer et al., 2007; Hembroff et al., 2004; Sprenger & Stegenga, 2017). For instance, during the Covid-19 pandemic some researchers criticized the communication of vaccine efficacy as a RRR (e.g., "the vaccine has a 95% efficacy", reflecting a 95% relative reduction in disease cases compared to the placebo group). They argue that ARRs are more sensitive to the baseline risk and can more readily illustrate how many people need to be vaccinated to prevent one additional case of the disease, and that RRRS are more likely to lead to misinterpretations because people would confuse the underlying reference groups and to whom the risk reduction applies (Brown, 2022; Marabotti, 2022; Olliaro et al., 2021). Similar arguments have been proposed in other domains, such as consumer psychology and advertising (Madrigal et al., 2024) and philosophy of science (Sprenger & Stegenga, 2017; Stegenga, 2015). To offer a balanced view, guidelines for reporting treatment effects in RCTs and systematic reviews recommend to communicate both ARRs and RRRs (Higgins et al., 2023; Moher et al., 2010; Working Group GPGI, 2016).

Measuring causal strength: ΔP and causal power

Two common metrics for quantifying the strength of a causal relation are ΔP and *causal power*. In psychology, the ΔP model was originally proposed in the context of evaluating the statistical contingency between two events (Jenkins & Ward, 1965), such as the dependency between responses and outcomes. Variants of the model have been proposed in causal learning theory (Cheng & Novick, 1990, 1992; Waldmann & Holyoak, 1992), psychophysics (Allan et al., 2008), philosophy of science (Carnap, 1962; Crupi & Tentori, 2014), and decision science (Nelson et al., 2022; Wu et al., 2017). The ΔP rule is also closely related to the influential Rescorla-Wagner model of classical conditioning (Danks, 2003; Rescorla & Wagner, 1972).

Applied to assessing the strength of a causal relation between a candidate cause C and an effect E, ΔP formalizes causal strength as the difference of the likelihood of the effect given the presence and absence of the cause, respectively:

$$\Delta P = P(e|c) - P(e|\neg c) \tag{3}$$

where $P(e|c)$ is the probability of the effect given that the cause is present, and $P(e|\neg c)$ is the probability of the effect given that the cause is absent. Typically, these probabilities are estimated from the relative frequency of events where effect E occurs in the presence and absence of C. If the cause increases the likelihood of the effect, ΔP is positive (e.g., smoking increases the likelihood of lung cancer). Conversely, if the cause decreases the likelihood of the effect, ΔP is negative (e.g., a vaccination decreases the probability of disease). If the

presence of the putative cause does not change the likelihood of the effect, ΔP is zero (i.e., an ineffective treatment).

Causal power (Cheng, 1997; Glymour, 2003; Novick & Cheng, 2004), denoted as q_c, provides an alternative measure of causal strength corresponding to the probability that C generates or prevents E in the absence of alternative influences on the effect. Typically, this quantity does not coincide with the probability of the effect in the presence of C, because this empirical probability also includes the influence of other (unobserved) causes. For preventive causes ($\Delta P < 0$), causal power is defined as

$$q_c = \frac{P(e|c) - P(e|\neg c)}{P(e|\neg c)} = \frac{-\Delta P}{P(e|\neg c)} \tag{4}$$

Because causal power denotes a theoretical probability — the probability of generating or preventing the effect if no other causes were present — it is always strictly non-negative and ranges between 0 and 1. Like ΔP, causal power is zero if the cause neither reduces nor raises the probability of the effect. For other scenarios, the two measures usually give different values (see section A1 in the Appendix). In preventive scenarios where the cause reduces the probability of the effect (negative ΔP), a power of 1 results if the effect never occurs in the presence of the cause, regardless of the level of the baseline probability.

Equivalence of measures of causal strength and risk reductions

Measures of risk reduction and measures of causal strength have been proposed in different fields, with distinct problems and applications in mind. In the medical literature, risk reductions are used to measure treatment benefits. In the cognitive science and causal modeling literature, measures of causal strength quantify the magnitude of causal relations and provide normative benchmarks for human causal learning (Buehner et al., 2003; Griffiths & Tenenbaum, 2005; Lober & Shanks, 2000; Meder et al., 2014; Waldmann & Holyoak, 1992). The terminology differs across fields, even though the employed measures are formally equivalent. In medical research, common terms include "treatment and control group" "event rates", "baseline risk", and "efficacy" of treatments. Conversely, the causal modeling literature uses the generic terms "cause" and "effect" to refer to different conditions and explain how people infer causal relations from observed frequencies, with recurring debates on how the "strength" of the relations can be expressed formally.

Despite these differences in nomenclature, the mapping of concepts is straightforward, especially in experimental designs. The treatment group corresponds to instances in which the candidate cause C is present (e.g., vaccine), whereas the control group corresponds to instances in which the candidate cause is absent (e.g., placebo). The effect E is the undesirable event (e.g., disease) that the treatment is supposed to prevent. Accordingly, the definition of ARR aligns with the ΔP measure (Equations 1 and 3):

$$\begin{aligned} ARR &= P(\text{event}|\text{control}) - P(\text{event}|\text{treatment}) \\ &= P(e|\neg c) - P(e|c) \\ &= |\Delta P| \end{aligned} \tag{5}$$

Notably, in the causal literature, ΔP is consistently defined as $P(e|c) - P(e|\neg c)$, yielding negative values for ΔP in case of a preventive cause that reduces the likelihood of the effect, and positive values for generative causes that increase the likelihood of the effect. By contrast, in the medical literature, both reductions and increases are commonly expressed as non-negative numbers (often percentages), with the direction indicated by the verbal label: $P(e|c) < P(e|\neg c)$ indicates a risk reduction, and $P(e|c) > P(e|\neg c)$ signifies a risk increase. Otherwise, however, the metrics are defined identically.

Analogously, RRR is mathematically equivalent to causal power in preventive scenarios (Eq. 2 and 4):

$$\begin{aligned} RRR &= \frac{P(\text{event}|\text{control}) - P(\text{event}|\text{treatment})}{P(\text{event}|\text{control})} \\ &= \frac{P(e|\neg c) - P(e|c)}{P(e|\neg c)} \\ &= \frac{|\Delta P|}{P(e|\neg c)} \\ &= q_c \end{aligned} \tag{6}$$

Since causal power represents a theoretical probability – the likelihood of generating or preventing the effect in the absence of other causes – it is always strictly non-negative, ranging from 0 to 1.

Thus, while each field employs these measures for distinct purposes, using unique labels and terminology, they are mathematically identical. Of course, formal equivalence does not imply that all risk reductions are causal. Causal interpretations of risk reductions are often justified, for example when risk reductions are used to quantify treatment effects in RCTs (e.g. vaccine efficacy). However, since the probabilities used to compute risk reductions can be calculated over arbitrary sets of events, additional criteria must be met to justify a causal interpretation, such as the independence of the candidate cause from other factors influencing the effect (Cheng, 1997). In experimental studies, randomization establishes this independence. Accordingly, risk reductions derived from experimental data justify a causal interpretation, consistent with the normative constraints imposed in the causal learning literature to infer causation from statistical regularities (Cheng, 1997; Glymour, 2003; Pearl, 2000). The situation is more complicated with non-experimental data, since confounding variables can introduce additional causal influences. Thus, while risk reductions and measures of causal strength can be computed over arbitrary sets of events, whether a causal interpretation is warranted depends on additional criteria.[1]

[1]Formally, to mark the distinction between observational and interventional probabilities, one can use Pearl's (2000) *do*-operator (Lagnado & Sloman, 2004; Meder et al., 2008; Waldmann & Hagmayer, 2005). In this case, the observational probability $P(e|c)$ denotes that the presence of candidate cause C has merely been observed, whereas $P(e|\text{do } c)$ represents explicitly that the presence of the candidate cause was set by means of intervention, rendering it independent of other events. Throughout this paper, our focus lies on interventional scenarios, hence the quantities entering the computations are interventional probabilities (as for instance resulting from experimental designs like randomized control trials). For the sake of simplicity, we omit the do-operator notation, but it is important to keep this constraint in mind as it is a common

Counterfactual inference through causal modeling

What are the implications of the equivalence between measures of causal strength on the one hand, and measures of risk reductions on the other hand? Besides highlighting fundamental parallels in the metrics used in different disciplines, our integration provide pathways for applying causal modeling techniques to make predictions for novel contexts. For instance, against the backdrop of the Covid-19 pandemic, healthy volunteers were deliberately exposed to a strand of the SARS-CoV-2 virus (Killingley et al., 2022), a so-called human challenge trial (Adams-Phipps et al., 2023). While ethically controversial, such studies are scientifically valuable because they assess the impact of a treatment under the most extreme condition: maximal exposure to the pathogen. This is a very different situation from standard clinical trials, where participants are naturally exposed to the risk, which typically corresponds to much smaller event rates (e.g., Table 1). Conversely, there are situations where the baseline risk is effectively zero, such as smallpox, which was declared eradicated by the World Health Organization (WHO) in 1980 following global vaccination efforts (Smith & McFadden, 2002). These two situations represent the endpoints of a spectrum where the baseline risk for an undesirable event ranges from 0 to 100%. Some medical problems are characterized by fairly stable baseline risks (e.g., hypertension or type 2 diabetes), whereas other health issues like the Covid pandemic were characterized by high volatility, with baseline risk strongly varying across time, populations, geography, and personal behavior.

How can we formally model different scenarios and derive empirically testable predictions if the baseline risk in the novel context is different from the one in which the treatment effect was assessed? We address this question using causal Bayes nets theory, a general modeling framework that combines graphical causal models and probability calculus (Pearl, 2000; Pearl & Mackenzie, 2018; Spirtes et al., 2000). An important advantage of representing causal relations such as treatment effects as generative causal models is that the formalism supports counterfactual generalizations in a systematic way, an important issue in many applied domains including healthcare (Hernán et al., 2019; Prosperi et al., 2020). Here, we utilize the framework to reinterpret absolute and relative risk reductions from a causal modeling perspective, illustrating how these metrics can lead to very different conclusions when generalizing treatment effects to novel contexts.

Causal Bayes nets: Strength and structure

Causal Bayes nets theory provides a formal approach for representing causal relations and modeling probabilistic inferences across causal networks (Danks, 2014; Pearl, 2000; Pearl & Mackenzie, 2018; Pearl et al., 2016; Spirtes et al., 2000). This approach combines qualitative assumptions about the causal structure of the domain with quantitative estimates specifying the magnitude of the causal relations. By explicating the latent causal dependencies and providing a calculus for distinguishing between observational and interventional probabilities, the framework supports the learning of complex causal structures from data and estimation of causal effects from observational data (Hernán et al., 2019; Meder et al., 2008, 2009; Pearl, 2000; Prosperi et al., 2020). It also provides mechanisms to evaluate the generalizability of causal effects identified in experimental studies to other

---

prerequisite for a causal interpretation of risk reductions.

populations where only observational studies can be conducted (i.e., the transportability problem; Pearl & Bareinboim, 2014).

We leverage this framework to highlight how the choice of a particular risk reduction metric leads to very different conclusions when generalizing the causal impact of a treatment to new contexts, such as making predictions for the efficacy of a vaccine in a population characterized by a different baseline risk of infection. We first provide an analysis of risk reduction metrics from the perspective of causal Bayesian networks and how they instantiate distinct assumptions about how causes interact in producing or preventing an effect. We then illustrate how these considerations are not purely technical issues but, in fact, have significant practical implications in applied domains such as health care, where outcomes are often influenced by multiple factors working in combination, and generalization of treatment effects is paramount.

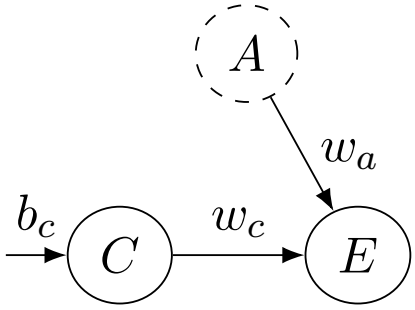

Figure 1

Basic causal model with a candidate cause C, an effect E, and the composite of unmeasured background causes A. Nodes represent domain variables and directed edges denote causal relations. Parameters b_c, w_a, and w_c quantify the strength of the relations.

To illustrate the applicability of the causal Bayes nets framework to medical research, we use the vaccine trial data (Table 1). Figure 1 shows the basic causal model for representing situations involving a single binary cause C and a single binary effect E, both of which can be present or absent (Griffiths & Tenenbaum, 2005; Meder et al., 2014). The nodes within the graph represent the modeled variables, and the directed edges represent the causal relations connecting the variables. This model is appropriate for the case considered here, which focuses on the relation between a vaccine for SARS-CoV-2 (cause C) and Covid-19 (effect E), where randomization renders A and C independent. The arrow connecting C and E represents the presumed causal dependency – the ability of the vaccine to reduce the probability of disease. In addition, the model contains a node A representing the (unobserved) background causes that generate the effect, that is, different ways of contracting Covid-19. (For mathematical convenience, A is assumed to be constantly present; Griffiths and Tenenbaum, 2005.)

The graphical model is complemented by a set of parameters, b_c, w_c, and w_a, typically estimated from data. Parameter b_c denotes the marginal probability of cause C. In an

RCT with two conditions, b_c will usually be about 0.5, as half of the people are randomly assigned to each condition. In other settings, like non-experimental epidemiological studies, this parameter could vary strongly and would typically be estimated from data. Parameter w_a represents the strength of the background cause A, which captures exogenous influences not explicitly represented in the model. An estimate for this influence is provided by the likelihood of the effect in the absence of the cause, $P(e|\neg c)$ — the baseline risk. For instance, in the BNT162b2 trial, 162 out of 18,325 participants in the placebo group developed Covid, hence $w_a = 0.0088$ (0.88%; Table 1). This value reflects the comparably low rate of Covid cases in this particular study. But, of course, over the course of a pandemic, this probability typically fluctuates significantly across time and populations.

Particularly relevant for the present analysis is the parameter w_c, which denotes the strength of cause C — the ability of the vaccine to reduce disease cases. Formally, both ΔP and causal power provide maximum-likelihood estimates for w_c in the causal graph in Figure 1, but under distinct assumptions about how C and A influence the probability of E (Glymour, 2003; Griffiths & Tenenbaum, 2005). Because ARR and RRR are equivalent to ΔP and causal power, respectively, it follows that they also provide maximimum-likelihood estimates for the causal influence of C.

Using ΔP and, equivalently, ARR as an estimate for w_c corresponds to the assumption that the probability of E is a linear combination of C and A, such that the cause decreases the baseline risk by a constant amount. Accordingly, the probability of E in the presence of C (e.g., probability of disease in vaccinated people) is given by

$$\begin{aligned} P(e|c, a; w_a, w_c) &= a \cdot w_a + c \cdot w_c \\ &= a \cdot w_a + c \cdot \Delta P \\ &= a \cdot w_a + c \cdot ARR \end{aligned} \tag{7}$$

where $a, c \in \{0, 1\}$ denote the presence and absence of candidate cause C and background cause A, while w_a and w_c denote their causal influence. According to this linear parameterization, if C is present the probability of effect E occurring is an additive function of w_a and w_c, and reduces to w_a if C is absent. In other words, the probability of the effect (e.g., Covid-19) is determined by subtracting a fixed amount from the baseline risk. However, to ensure a valid probability distribution, it is necessary to externally constrain $w_a + w_c$ to the range of 0 to 1, thereby introducing a dependency of parameters. As we will demonstrate later, the functional form and implied interdependence in the combination of causal influences has profound implications for counterfactual inferences that go beyond a mere summary of the observed data (e.g., treatment effect in a given study).

In contrast, using causal power q_c and, equivalently, RRR as an estimate for w_c offers a more expressive parameterization for graphical models. This approach to causal inference is enabled by a probabilistic generalization of logical functions(Pearl, 1988; Yuille & Lu, 2007). This parameterization specifies how probabilistic influences combine, assuming that causes operate independently in their likelihood of generating or preventing the effect, where causal power q_c serves as estimate for w_c. Importantly, as we demonstrate below, this parameterization enables sound causal counterfactual inferences about novel situations, in contrast to the linear integration assumed under ARR.

In the preventive case, this approach generalizes the logical rule that effect E will occur if A occurs and C does not, such that the causal influence of these factors can be probabilistic (noisy-AND-NOT). Under this assumption, the conditional probability of the effect given the cause is present is given by

$$\begin{aligned} P(e|c,a;w_a,w_c) &= a \cdot w_a(1 - c \cdot w_c) \\ &= a \cdot w_a(1 - c \cdot q_c) \\ &= a \cdot w_a(1 - c \cdot RRR) \end{aligned} \tag{8}$$

which reduces to w_a for the probability of E when candidate cause C is absent. Importantly, this parameterization ensures a valid probability distribution without the necessity to externally constrain the parameter values of w_a and w_c.

Regardless of whether a linear or noisy-logical parameterization of the graph is chosen, the causal-based factorization of the joint probability distribution $P(C, E)$ recovers the empirical probabilities. For instance, in the BNT162b2 trial, 8 out of 18,198 people in the vaccine group developed Covid-19 (Table 1), thus $P(\text{disease}|\text{vaccine}) = P(e|c) = 0.0004$. This probability remains consistent whether computed directly from the empirical data or derived from the parameterized causal model. Specifically, it remains unchanged when using ΔP (absolute risk reduction) or causal power (relative risk reduction) to estimate w_c in Equations 7 and 8, respectively (see Appendix A2 for a numerical example).

The implications of choosing between a linear or noisy-logical parameterization may not be immediately obvious, but in fact the practical implications and downstream advantages are significant. In particular, while both parameterizations accurately capture empirical probabilities, they imply strongly diverging predictions for counterfactual inferences, where the goal is to infer what would happen in scenarios that differ from the one under which the treatment effects were estimated. We investigate these implications in the next section.

To summarize, causal Bayes nets are a modeling framework for causal inference, combining graphical causal models with parameters that quantify the strength of the relations. Different measures of causal strength can be used as estimates for the causal influence of a candidate cause, each reflecting different assumptions about how the causes interact to bring about an effect. Given the mathematical equivalence of ARR to ΔP on one hand, and RRR to causal power on the other (Eqs. 5 and 6), this implies that both approaches are maximum-likelihood estimates for the causal strength parameter of the graph show in Figure 1, although under different parameterizations. Thus, when viewed through a causal modeling lens, selecting either of these metrics to quantify treatment effects incidentally entails the adoption of a specific measure of causal strength and particular assumptions on how causes interact to generate or prevent effects, with looming consequences for generalizing treatment effects to novel situations.

Generalizing causal strength and treatment effects to new contexts

The parameterized graph provides a generative model that supports causal inferences in a principled way, including predictions about the effects of interventions, that is, reasoning about scenarios for which no data is (yet) available. Formally, such counterfactuals can be calculated by adjusting the model's parameters to align with a specific context of

interest and deriving the desired probability distribution. For example, the parameterized causal model in Figure 1 supports modeling a vaccine's impact in contexts characterized by baseline risks different from the ones under which the treatment effect was estimated. The scenario of interest can be simulated by adjusting the strength of the background cause, w_a, to the desired level and deriving the conditional probability distribution of effect E given the presence and absence of the cause (i.e., predicted disease cases with and without vaccine). For instance, a human challenge trial (Adams-Phipps et al., 2023) where all subjects are exposed to the pathogen could be modeled by setting $w_a = 1$. Conversely, a context with a virus-free population, akin to the conditions following smallpox eradication in the 1970s would be simulated by letting $w_a = 0$. These scenarios represent the two extremes of a spectrum ranging from a baseline risk of 0% to 100%.

To derive counterfactual predictions, such as the probability of disease under different baseline risks (i.e., different values of w_a), the strength w_c of cause C also needs to be quantified. As noted, both ΔP and causal power provide estimates for w_c, but since they instantiate different assumptions of how A and C interact to generate E, they can yield strongly diverging predictions. In the case of the BNT162b2 vaccine trial (Table 1), the RRR (equivalent to causal power) was approximately 95%, while the ARR (corresponding to ΔP) was around 0.84%. These estimates were obtained under a specific baseline risk, which naturally varies under pandemic conditions (e.g., in the BNT162b2 trial the baseline risk, as estimated from the control group cases, was 0.88%; Table 1).

Consider the scenario where the vaccine is administered to another group of people under conditions where the baseline risk is, say, 10% (e.g., healthcare professionals with greater exposure). Building upon the results from the original trial, how many cases of disease should we expect? For instance, if we had 1000 healthcare professionals and a 10% baseline risk of getting infected, we would expect about 100 to get infected without vaccines. But what if they were vaccinated? To answer this question, we need a quantitative estimate of the original treatment effect. If we used the ARR, which assumes that the baseline risk is decreased by a constant amount, the vaccine would be expected to decrease the number of disease cases by only 0.84 percentage points, resulting in around 92 cases (9.16%; Eq. 7).

In stark contrast, when using a RRR we would correctly[2] predict a 95% reduction from the new baseline, yielding an expected 5 disease cases among the vaccinated people (0.5%; Eq. 8). In these scenarios, ARR strongly underestimates the causal impact of the vaccine in terms of its power to prevent disease cases in novel settings.

This divergence between the two measures only becomes larger if the baseline risk further increases. For instance, if the baseline risk in a human challenge study was 100%, using ARR as a measure of vaccine efficacy would predict 99.16% disease cases, whereas a RRR predicts 5% cases. Figure 2 illustrates this divergence across varying baseline risks, showing the predicted probability of disease for three vaccines against Covid-19, using the absolute or relative risk reduction obtained in the corresponding clinical trials (Baden et al., 2021; Polack et al., 2020; Voysey et al., 2021) as causal strength estimates. These considerations illustrate that a linear parameterization, as inherent in ARR and the ΔP model, does not support adequate counterfactual inference. By contrast, RRR and causal

[2]Note that in practice, we wouldn't expect these exact numbers, due to uncertainty about the statistical estimates and other sources of noise. These considerations, while important in practical applications, are omitted here as they do not touch upon the conceptual argument.

power do support an appropriate evaluation of such scenarios.

BNT162b2 (BionTech/Pfizer)

	Covid-19	
Treatment	yes	no
Vaccine	8	18,190
Placebo	162	18,163

$P(\text{Covid}|\text{Vaccine}) = 0.04\%$
$P(\text{Covid}|\text{Placebo}) = 0.88\%$
RRR = 95%
ARR = 0.04%

mRNA-1273 (Moderna)

	Covid-19	
Treatment	yes	no
Vaccine	11	14,123
Placebo	185	13,888

$P(\text{Covid}|\text{Vaccine}) = 0.08\%$
$P(\text{Covid}|\text{Placebo}) = 1.31\%$
RRR = 94.1%
ARR = 1.24%

ChAdOx1 (Astra Zeneca)

	Covid-19	
Treatment	yes	no
Vaccine	30	5,777
Placebo	101	5,728

$P(\text{Covid}|\text{Vaccine}) = 0.52\%$
$P(\text{Covid}|\text{Placebo}) = 1.73\%$
RRR = 70.2%
ARR = 1.21%

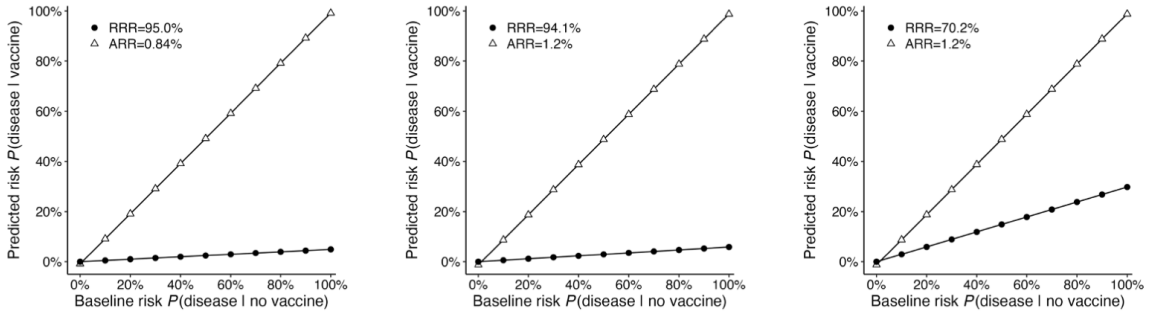

Figure 2

Analysis of counterfactual scenarios where the baseline risk of the disease differs from the risk under which the treatment effect was obtained. The top row shows the observed cases from three randomized control trials conducted to estimate the efficacy of different vaccines against Covid-19 (Baden et al., 2021; Polack et al., 2020; Voysey et al., 2021). The middle row shows the corresponding probabilities of the disease given vaccine and placebo, and the entailed relative risk reduction (RRR) and absolute risk reduction (ARR). The bottom row plots the predicted probability of disease under varying new baseline risks, calculated from a causal model where the impact of the vaccine is quantified using the ARR (corresponding to the ΔP metric) or RRR (corresponding to the causal power metric), as per Equations 7 and 8, respectively.

Another problem with using ARR (or equivalently, ΔP) to quantify treatment efficacy is illustrated by cases in which the baseline risk is extremely low or effectively zero, as seen in scenarios like the eradication of smallpox. In such circumstances, the probability of disease under the linear parameterization inherent in the ΔP model (Eq. 7) erroneously yields a negative probability (Feynman, 1987), although by definition probabilities are limited to the interval [0,1]. Generally, this happens if the baseline risk in the novel context is below the estimate of the absolute reduction. To avoid such cases, it is necessary to constrain the sum of w_c and w_a to the interval [0, 1] to ensure a proper probability distribution. In contrast, the functional form inherent in RRR (and equivalently, in the causal power model) correctly indicates that the probability of the effect is zero. Generally, us-

ing causal power and Equation 8 always yields a valid probability. Thus, causal power and RRRs represent a better-behaved measure than the ΔP model and ARRs when making predictions for scenarios with baseline risks that differ from the original context, thus offering better generalization to counterfactual settings.

Scenario 1

	Effect	
Cause	e	$\neg e$
c	10	90
$\neg c$	15	85

$P(e|c) = 0.1$
$P(e|\neg c) = 0.15$
power $q_c = 0.33$
$\Delta P = -0.05$

Scenario 2

	Effect	
Cause	e	$\neg e$
c	20	80
$\neg c$	50	50

$P(e|c) = 0.20$
$P(e|\neg c) = 0.5$
power $q_c = 0.6$
$\Delta P = -0.3$

Scenario 3

	Effect	
Cause	e	$\neg e$
c	80	20
$\neg c$	90	10

$P(e|c) = 0.8$
$P(e|\neg c) = 0.9$
power $q_c = 0.11$
$\Delta P = -0.10$

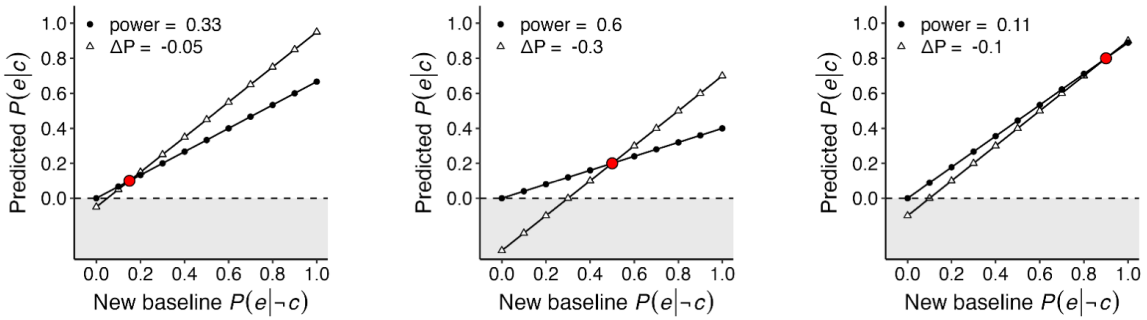

Figure 3

Analysis of three scenarios illustrating the divergence of using ΔP or causal power q_c for generalizing treatment effects to situations with a new baseline risk. The top rows illustrates three different scenarios, where the baseline risk (probability of effect in the absence of the preventive cause) is rather low (left), intermediate (middle), or high (right). The middle row shows the corresponding probabilities and values of ΔP and causal power. The bottom row shows the predicted probability of the effect given the cause under new baseline probabilities, calculated from a causal model where the causal strength of C is quantified using ΔP or causal power q_c (Equations 7 and 8, respectively). The red dot indicates the combination of $P(e|c)$ and $P(e|\neg c)$ from which the strength of C was originally estimated. Shaded area indicate negative predicted probabilities.

But does using ARR and the ΔP model always lead to an underestimation of the treatment effect under a new baseline risk? The answer is no. In fact, we can precisely characterize the circumstances under which ΔP and, equivalently, an ARR, yields lower, higher, or equal probabilities compared to using causal power.

Figure 3 illustrates the diverging implications for three hypothetical scenarios. In each case, the strength of C is estimated under a specific baseline risk (i.e., probability of the effect in the absence of the cause), and then used to make inferences about the likelihood

of the effect in situations with new baseline probabilities. Inspection of the figure highlights several generalizable insights.

First, there is always a cross-over of the predicted $P(e|c)$ based on ΔP vs causal power. Specifically, the predictions intersect (red dot in Fig. 3) at the originally observed $P(e|c)$ and $P(e|\neg c)$. This is necessarily the case because if the new baseline probability is identical to the original one, both metrics should yield the same $P(e|c)$ — namely recover the original probability (see Appendix A2 for a numerical example). Left of this point (i.e., when then new baseline probability is *lower* than the original one), the predicted probability of the effect given the cause based on ΔP is lower than the probability predicted by causal power. Put differently, in this case using ΔP (or equivalently, an ARR) to make predictions *overestimates* the treatment effect. Moreover, negative probabilities will result when the new baseline probability is lower than ΔP. Conversely, right of this point, that is, when the new baseline probability is *higher* than the one under which the strength of C was estimated, the predicted probability of the effect given the cause based on ΔP is higher than the probability predicted by causal power. In this scenario, using ΔP (or equivalently, ARR) to make predictions *underestimates* the treatment effect.

These analyses demonstrate the application of causal modeling techniques to formally represent and assess hypothetical scenarios through parameter adjustments in the corresponding causal graph. Moreover, when it comes to generalizing causal effects to novel situations with varying baseline risks (as opposed to simply summarizing the effects observed in a particular study), ARRs are inadequate due to their implicit assumption of a constant linear decrease in risk. This aligns with analyses of real-world data from meta-analyses and individual RCTs which suggest that relative outcome measures (i.e., RRRs) tend to be more stable than absolute reductions across studies and subgroups with differing baseline risks (Deeks, 2002; Furukawa et al., 2002; Schmid et al., 1998). Accordingly, it is usually recommended to avoid using absolute measures in meta-analyses that integrate treatment effects across studies (Higgins et al., 2023).

Conclusions

We have integrated concepts from three distinct fields: i) health and medical research, which seeks to quantify and generalize treatment effects, ii) risk communication, which is concerned with the understanding of health information in patients and health professionals, and iii) cognitive psychology, which investigates how people infer causal relations from covariation data. Conceptually, our analyses provide a theoretical foundation for interpreting risk reduction measures through the lens of causal inference, making causal considerations explicit that are otherwise left implicit. From an applied standpoint, the findings demonstrate how the causal Bayes nets framework can be leveraged to clarify the diverging implications of using ARR and RRR when generalizing treatment effects to contexts with varying baseline risks.

The analyses also hold specific implications for the different fields and central debates that have engaged them. First, our analyses underscore the potential of real-world data for the theoretical and experimental analysis of human causal induction (Bramley et al., 2017; Gerstenberg et al., 2021; Gopnik et al., 2004; Griffiths & Tenenbaum, 2009; Holyoak et al., 2010; Meder et al., 2014; Waldmann, 2017). Psychologists have focused on the inference processes that transform raw data (i.e., described or experienced frequencies)

into estimates of causal strength, with recurring debates about the normative and descriptive validity of different metrics (Cheng, 1997; Griffiths & Tenenbaum, 2005; Holyoak & Cheng, 2011; Lober & Shanks, 2000; Meder et al., 2014). Typically, causal learning experiments present participants with stylized scenarios akin to real-world experiments (e.g., hypothetical study data on the joint occurrences of a virus and disease cases), typically with equal base rates for the presence and absence of the candidate cause (Buehner et al., 2003; Griffiths & Tenenbaum, 2005; Meder et al., 2014). While this design facilitates the assessment of the relevant probabilities assumed to enter the computations, it also incurs a mismatch with real-world environments characterized by low probability events, such as the comparably low exposure rate to a pathogen like the Coronavirus. To enhance ecological validity, empirical studies could use real-world data from medical studies (e.g., vaccine trials) as stimuli, where base rates are typically much lower and therefore ARR (ΔP) and RRR (causal power) strongly diverge. Participants could then be prompted to make inferences about contexts with varying baseline risks, where different models imply distinct predictions. Such experiments would provide additional insights into how people generalize their rich causal knowledge to novel situations to inform their judgment and decision making processes. Notably, there are further relations between probabilistic theories of causality proposed in philosophy and psychology, and metrics routinely applied in medical research and practice (Sprenger, 2018). Variants of the ΔP model have also featured prominently in philosophy of science and confirmation theory (Crupi & Tentori, 2014; Tentori et al., 2007) and models of the value of information (Nelson et al., 2022; Nelson, 2005; Wu et al., 2017). These relationships provide further traction for theory integration across disciplinary boundaries, while also highlighting differences among concepts, such as disparities between the definition of generative causal power and relative risk increases (see Appendix A3 for details).

Second, our analyses contribute to a causal inference perspective on medical problems, leveraging the Bayes nets formalism. Causal modeling techniques support the explicit specification of causal structure, which despite recent advances in formal methodology remains rare in clinical research. Notably, the computational framework offers a means to express the crucial distinction between observational and interventional probabilities, facilitating the development of interventional clinical predictive models and the calculation of counterfactuals (Prosperi et al., 2020). Here, we analyzed common measures of risk reduction from a causal inference perspective, highlighting that in contexts like a pandemic, where base rate probabilities fluctuate, using ARRs to quantify vaccine efficacy is inadequate and leads to error-prone generalizations. This limitation extends to related measures like the number needed to treat (Olliaro et al., 2021), which is inversely related to ARRs. Our analyses thus refute arguments that ARRs provide a more appropriate measure for quantifying vaccine efficacy (Brown, 2021, 2022). In fact, whereas a common criticism of RRRs is that they are not sensitive to the baseline risk, for the purpose of estimating treatment effects this could be considered a desirable property, as it renders the measure independent of the particular baseline risk under which the treatment effect is estimated. Accordingly, one practical implication is that in contexts with dynamic baseline risks, RRRs are better-behaved measures of treatment benefits than ARRs. This aligns with analyses showing that relative outcome measures are more consistent across studies than absolute measures (Deeks, 2002; Furukawa et al., 2002; Schmid et al., 1998).

Third, our analyses contribute to the ongoing debate on the interpretation of ARRs and RRRs in health communication. A central concern in this discourse is how laypeople (Carling et al., 2009; Hembroff et al., 2004) and healthcare professionals (Akl et al., 2011; Bucher et al., 1994; Marcatto et al., 2013) interpret and respond to treatment benefits expressed either as relative or absolute risk reduction. We argue that both absolute or relative risk reductions in isolation are inadequate for communicating medical risks and supporting individual decision-making. For instance, while both measures quantify the causal impact of a treatment by a single number derived from the likelihood of an outcome in one group compared to another, research shows that laypeople actually prefer having two numbers: one indicating how many people are likely to develop a disease with treatment and another without treatment (Carling et al., 2009; Trevena et al., 2006). Moreover, informed health decisions also require consideration of additional elements, including the magnitude of the baseline risk[3] (Bodemer et al., 2014; Marcatto et al., 2013; Natter & Berry, 2005; Sheridan et al., 2003), heterogeneity of treatment effects across subpopulations (Kent et al., 2010), the applicability of summary findings to individual patients (Kent & Hayward, 2007; Rothwell, 1995), as well as the quality of evidence on which estimates and recommendations are based (Guyatt et al., 2011; Higgins et al., 2023). Neither ARR nor RRR can account for these requirements. Instead, standards of evidence-based health communication ensure the support of informed and shared decision-making in the health domain (Bonner et al., 2021; Gigerenzer et al., 2007; Woloshin et al., 2023). Promising pathways include so-called Fact Boxes (McDowell et al., 2016, 2019; Schwartz & Woloshin, 2013) that summarize the best available evidence on the benefits and harms associated with medical interventions, as well as carefully designed visual information (Woloshin et al., 2023). Thus, while our analyses demonstrate that relative reductions provide a more accurate assessment of treatment effects when generalizations to varying levels of baseline risk are desirable or required, effective risk communication requires a more comprehensive approach than providing healthcare professionals and the general public with a single number (or two).

References

Adams-Phipps, J., Toomey, D., Więcek, W., Schmit, V., Wilkinson, J., Scholl, K., Jamrozik, E., Osowicki, J., Roestenberg, M., & Manheim, D. (2023). A systematic review of human challenge trials, designs, and safety. *Clinical Infectious Diseases*, *76*(4), 609–619.

Akl, E. A., Oxman, A. D., Herrin, J., Vist, G. E., Terrenato, I., Sperati, F., Costiniuk, C., Blank, D., & Schünemann, H. (2011). Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database of Systematic Reviews*, (3).

[3]While neither ARRs nor RRRs provide direct information on how likely an adverse outcome is to occur in the first place, the baseline risk (i.e., event rate in the control group) provides an upper limit for the size of the ARR. Thus, one might argue that indirectly absolute reductions provide some information on the baseline risk. In our opinion, this statistical property is too subtle for communicating information about baseline risks in practical contexts. Similarly, an ARR provides information on how likely an individual will actually benefit from a treatment. For instance, if the ARR is 2%, this implies that, on average, 2 out of 100 people would actually benefit from the treatment. In other words, 2 fewer people would be infected by the virus, and these individuals are the ones who in fact benefit from the treatment. However, this conclusion, too, is rather subtle and is only valid if the baseline risk in the new context remains the same.

Allan, L. G., Hannah, S. D., Crump, M. J., & Siegel, S. (2008). The psychophysics of contingency assessment. *Journal of Experimental Psychology: General*, *137*(2), 226–243.

Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association*, *13*(6), 608–618.

Baden, L. R., El Sahly, H. M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S. A., Rouphael, N., Creech, C. B., et al. (2021). Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *New England Journal of Medicine*, *384*(5), 403–416.

Bareinboim, E., & Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, *113*(27), 7345–7352.

Bobbio, M., Demichelis, B., & Giustetto, G. (1994). Completeness of reporting trial results: Effect on physicians' willingness to prescribe. *The Lancet*, *343*(8907), 1209–1211.

Bodemer, N., Meder, B., & Gigerenzer, G. (2014). Communicating relative risk changes with baseline risk: Presentation format and numeracy matter. *Medical Decision Making*, *34*(5), 615–626.

Bonner, C., Trevena, L. J., Gaissmaier, W., Han, P. K., Okan, Y., Ozanne, E., Peters, E., Timmermans, D., & Zikmund-Fisher, B. J. (2021). Current best practice for presenting probabilities in patient decision aids: Fundamental principles. *Medical Decision Making*, *41*(7), 821–833.

Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017). Formalizing neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, *124*(3), 301–338.

Brown, R. B. (2021). Outcome reporting bias in COVID-19 mRNA vaccine clinical trials. *Medicina*, *57*(3), 199.

Brown, R. B. (2022). Relative risk reduction: Misinformative measure in clinical trials and covid-19 vaccine efficacy. *Dialogues in Health*, 100074.

Bucher, H., Weinbacher, M., & Gyr, K. (1994). Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *Bmj*, *309*(6957), 761–764.

Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1119–1140.

Carling, C. L., Kristoffersen, D. T., Montori, V. M., Herrin, J., Schünemann, H. J., Treweek, S., Akl, E. A., & Oxman, A. D. (2009). The effect of alternative summary statistics for communicating risk reduction on decisions about taking statins: A randomized trial. *PLoS Medicine*, *6*(8), e1000134.

Carnap, R. (1962). *Logical foundations of probability* (Vol. 2). The University of Chicago Press.

Cartwright, N. (2007). Hunting causes and using them. *Cambridge Books*.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*(2), 367–405. https://doi.org/10.1037/0033-295X.104.2.367

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, *58*(4), 545–567.

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, *99*(2), 365–382.

Covey, J. (2007). A meta-analysis of the effects of presenting treatment benefits in different formats. *Medical Decision Making*, *27*(5), 638–654.

Crupi, V., & Tentori, K. (2014). State of the field: Measuring information and confirmation. *Studies in History and Philosophy of Science Part A*, *47*, 81–90.

Danks, D. (2003). Equilibria of the rescorla–wagner model. *Journal of Mathematical Psychology*, *47*(2), 109–121.

Danks, D. (2014). *Unifying the mind: Cognitive representations as graphical models*. MIT Press.

Deeks, J. J. (2002). Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Statistics in medicine*, *21*(11), 1575–1600.

Etminan, M., Collins, G. S., & Mansournia, M. A. (2020). Using causal diagrams to improve the design and interpretation of medical research. *Chest*, *158*(1), S21–S28.

Feynman, R. P. (1987). Negative probability. In B. Hiley & F. Peat (Eds.), *Quantum implications: Essays in honour of David Bohm* (pp. 235–248). Routledge.

Forrow, L., Taylor, W. C., & Arnold, R. M. (1992). Absolutely relative: How research results are summarized can affect treatment decisions. *The American Journal of Medicine*, *92*(2), 121–124.

Furukawa, T. A., Guyatt, G. H., & Griffith, L. E. (2002). Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analyses. *International Journal of Epidemiology*, *31*(1), 72–76.

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review*, *128*(5), 936–975.

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: From innumeracy to insight. *BMJ*, *327*(7417), 741–744.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, *8*(2), 53–96.

Glymour, C. (2003). Learning, prediction and causal bayes nets. *Trends in Cognitive Sciences*, *7*(1), 43–48.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and bayes nets. *Psychological Review*, *111*(1), 3–32.

Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, 37–48.

Greenwood, M., & Yule, G. U. (1915). The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general. *Proceedings of the Royal Society of Medicine*, *8*, 113–194.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*(4), 334–384.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, *116*(4), 661–716.

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., DeBeer, H., et al. (2011). Grade guidelines: 1. introduction—grade evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, *64*(4), 383–394.

Hembroff, L. A., Holmes-Rovner, M., & Wills, C. E. (2004). Treatment decision-making and the form of risk communication: Results of a factorial survey. *BMC Medical Informatics and Decision Making*, *4*, 1–9.

Hernán, M. A., Hsu, J., & Healy, B. (2019). A second chance to get causal inference right: A classification of data science tasks. *Chance*, *32*(1), 42–49.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2023). *Cochrane handbook for systematic reviews of interventions version 6.4 (updated august 2023)* (J. P. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch, Eds.). Cochrane. Retrieved May 14, 2024, from https://training.cochrane.org/handbook

Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, *62*, 135–163.

Holyoak, K. J., Lee, H. S., & Lu, H. (2010). Analogical and category-based inference: A theoretical integration with Bayesian causal models. *Journal of Experimental Psychology: General*, *139*(4), 702–727.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, *79*(1), 1–17. https://doi.org/10.1037/h0093874

Kent, D. M., & Hayward, R. A. (2007). Limitations of Applying Summary Results of Clinical Trials to Individual PatientsThe Need for Risk Stratification. *JAMA*, *298*(10), 1209–1212. https://doi.org/10.1001/jama.298.10.1209

Kent, D. M., Rothwell, P. M., Ioannidis, J. P., Altman, D. G., & Hayward, R. A. (2010). Assessing and reporting heterogeneity in treatment effects in clinical trials: A proposal. *Trials*, *11*, 1–11.

Killingley, B., Mann, A. J., Kalinova, M., Boyers, A., Goonawardane, N., Zhou, J., Lindsell, K., Hare, S. S., Brown, J., Frise, R., et al. (2022). Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. *Nature Medicine*, *28*(5), 1031–1041.

Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 856–876.

Lober, K., & Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review*, *107*(1), 195–212.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, *115*(4), 955–984.

Madrigal, R., Armstrong Soule, C. A., & King, J. (2024). Manipulating consumers with the truth: Relative-difference claims in advertising and inferences of manipulative intent. *Journal of Advertising*, 1–18.

Malenka, D. J., Baron, J. A., Johansen, S., Wahrenberger, J. W., & Ross, J. M. (1993). The framing effect of relative and absolute risk. *Journal of General Internal Medicine*, *8*, 543–548.

Marabotti, C. (2022). Efficacy and effectiveness of Covid-19 vaccine-absolute vs. relative risk reduction. *Expert Review of Vaccines*, *21*(7), 873–875.

Marcatto, F., Rolison, J. J., & Ferrante, D. (2013). Communicating clinical trial outcomes: Effects of presentation method on physicians' evaluations of new treatments. *Judgment and Decision Making*, *8*(1), 29–33.

McDowell, M., Gigerenzer, G., Wegwarth, O., & Rebitschek, F. G. (2019). Effect of tabular and icon fact box formats on comprehension of benefits and harms of prostate cancer screening: A randomized trial. *Medical Decision Making*, *39*(1), 41–56.

McDowell, M., Rebitschek, F. G., Gigerenzer, G., & Wegwarth, O. (2016). A simple tool for communicating the benefits and harms of health interventions: A guide for creating a fact box. *MDM Policy & Practice*, *1*(1), 2381468316665365.

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2008). Inferring interventional predictions from observational learning data. *Psychonomic Bulletin & Review*, *15*(1), 75–80. https://doi.org/10.3758/PBR.15.1.75

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2009). The role of learning data in causal reasoning about observations and interventions. *Memory & Cognition*, *37*(3), 249–264. https://doi.org/10.3758/MC.37.3.249

Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review*, *121*(3), 277–301. https://doi.org/10.1037/a0035944

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2010). Consort 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ*, *340*. https://doi.org/10.1136/bmj.c869

Natter, H. M., & Berry, D. C. (2005). Effects of presenting the baseline risk when communicating absolute and relative risk reductions. *Psychology, Health & Medicine*, *10*(4), 326–334.

Nelson, J. D., Rosenauer, C., Crupi, V., Tentori, K., & Meder, B. (2022). The likelihood difference heuristic and binary test selection given situation-specific utilities. *Decision*, *9*, 285–319. https://doi.org/10.1037/dec0000160

Nelson, J. D. (2005). Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review*, *112*(4), 979–999. https://doi.org/10.1037/0033-295X.112.4.979

Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review*, *111*(2), 455–485.

Olliaro, P., Torreele, E., & Vaillant, M. (2021). COVID-19 vaccine efficacy and effectiveness – the elephant (not) in the room. *The Lancet Microbe*, *2*(7), e279–e280.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan kaufmann.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Pearl, J., & Bareinboim, E. (2014). External validity: From do-calculus to transportability across populations. *Statistical Science*, *29*(4), 579–595.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

Pearl, J., & Mackenzie, D. (2018). *The book of why*. Hachette.

Perneger, T. V., & Agoritsas, T. (2011). Doctors and patients' susceptibility to framing bias: A randomized trial. *Journal of General Internal Medicine*, *26*, 1411–1417.

Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J. L., Pérez Marc, G., Moreira, E. D., Zerbini, C., et al. (2020). Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *New England Journal of Medicine*, *383*(27), 2603–2615.

Prosperi, M., Guo, Y., Sperrin, M., Koopman, J. S., Min, J. S., He, X., Rich, S., Wang, M., Buchan, I. E., & Bian, J. (2020). Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence*, *2*(7), 369–375.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations on the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Rothwell, P. M. (1995). Can overall results of clinical trials be applied to all patients? *The Lancet*, *345*(8965), 1616–1619.

Sanchez, P., Voisey, J. P., Xia, T., Watson, H. I., O'Neil, A. Q., & Tsaftaris, S. A. (2022). Causal machine learning for healthcare and precision medicine. *Royal Society Open Science*, *9*(8), 220638.

Schmid, C. H., Lau, J., McIntosh, M. W., & Cappelleri, J. C. (1998). An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Statistics in Medicine*, *17*(17), 1923–1942.

Schölkopf, B. (2022). Causality for machine learning. In H. Geffner, R. Dechter, & J. Y. Halpern (Eds.), *Probabilistic and causal inference: The works of Judea Pearl* (pp. 765–804). Association for Computing Machinery.

Schwartz, L. M., & Woloshin, S. (2013). The drug facts box: Improving the communication of prescription drug information. *Proceedings of the National Academy of Sciences*, *110*(3), 14069–14074.

Sheridan, S. L., Pignone, M. P., & Lewis, C. L. (2003). A randomized comparison of patients' understanding of number needed to treat and other common risk reduction formats. *Journal of General Internal Medicine*, *18*(11), 884–892.

Smith, G. L., & McFadden, G. (2002). Smallpox: Anything to declare? *Nature Reviews Immunology*, *2*(7), 521–527.

Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search*. MIT Press.

Sprenger, J. (2018). Foundations of a probabilistic theory of causal strength. *Philosophical Review*, *127*(3), 371–398.

Sprenger, J., & Stegenga, J. (2017). Three arguments for absolute outcome measures. *Philosophy of Science*, *84*(5), 840–852.

Stegenga, J. (2015). Measuring effectiveness. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *54*, 62–71. https://doi.org/10.1016/j.shpsc.2015.06.003

Stovitz, S. D., & Shrier, I. (2019). Causal inference for clinicians. *BMJ Evidence-Based Medicine*, *24*(3), 109–112. https://doi.org/10.1136/bmjebm-2018-111069

Tentori, K., Passerini, A., Timberlake, B., & Pighin, S. (2021). The misunderstanding of vaccine efficacy. *Social Science & Medicine*, *289*, 114273.

Tentori, K., Crupi, V., Bonini, N., & Osherson, D. (2007). Comparison of confirmation measures. *Cognition*, *103*(1), 107–119.

Trevena, L. J., Davey, H. M., Barratt, A., Butow, P., & Caldwell, P. (2006). A systematic review on communicating with patients about evidence. *Journal of Evaluation in Clinical Practice*, *12*(1), 13–23.

Voysey, M., Clemens, S. A. C., Madhi, S. A., Weckx, L. Y., Folegatti, P. M., Aley, P. K., Angus, B., Baillie, V. L., Barnabas, S. L., Bhorat, Q. E., et al. (2021). Safety and efficacy of the ChAdox1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: An interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *The Lancet*, *397*(10269), 99–111.

Waldmann, M. R. (2017). *The Oxford handbook of causal reasoning*. Oxford University Press.

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 216–227.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, *121*(2), 222–236.

Weinberg, G. A., & Szilagyi, P. G. (2010). Vaccine epidemiology: Efficacy, effectiveness, and the translational research roadmap. *The Journal of Infectious Diseases*, *201*(11), 1607–1610.

Woloshin, S., Yang, Y., & Fischhoff, B. (2023). Communicating health information with visual displays. *Nature Medicine*, *29*(5), 1085–1091.

Working Group GPGI. (2016). Good practice guidelines for health information. *Zeitschrift für Evidenz, Fortbildung und Qualitat im Gesundheitswesen*, *110-111*, 85–92. https://doi.org/10.1016/j.zefq.2015.11.005

Wu, C. M., Meder, B., Filimon, F., & Nelson, J. D. (2017). Asking better questions: How presentation formats influence information search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 1274–1297. https://doi.org/10.1037/xlm0000374

Yuille, A. L., & Lu, H. (2007). The noisy-logical distribution and its application to causal inference. In J. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in Neural Information Processing Systems* (Vol. 20). Curran Associates, Inc.

Appendix

A1. ΔP and causal power: Differences in model behavior

Analogously to absolute and relative risk reductions, ΔP and causal power can yield very different values when applied to the same data (Figure A1). Consider a disease with a baseline risk of $P(e|\neg c) = 0.00002$, i.e. 2 in 100,000 people have the disease. Assume a vaccine reduces this baseline risk to 1 in 100,000, such that $P(e|c) = 0.00001$. In this case, $\Delta P = 0.00001$, but $q_c = 0.5$, as the vaccine reduces the number of cases by 50%.

Generally, the same level of ΔP can entail different levels of causal power, depending on the probability of the effect in the absence of the cause (i.e., $P(e|\neg c)$; or the baseline risk). Conversely, low values of ΔP can imply maximum values of causal power, namely in those cases where the adverse event never occurs in the presence of the cause (i.e., $P(e|c) = 0$). In this case, while the level of ΔP depends on and in fact is equal to the baseline risk $P(e|\neg c)$, causal power is always 1 as the presence of the cause eliminates the presence of the effect (e.g., if a vaccine provides perfect protection against a disease).

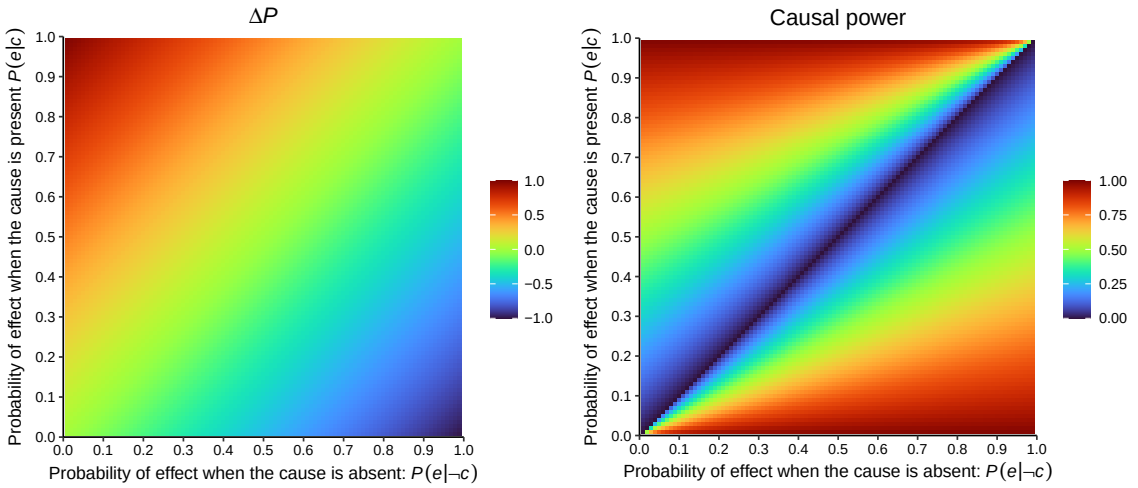

Figure A1

ΔP and causal power as a function of the probability of the effect in the presence and absence of the candidate cause, $P(e|c)$ and $P(e|\neg c)$. ΔP is always in the range $[-1, 1]$, whereas causal power is strictly non-negative in the range $[0, 1]$.

A2. Causal inference with graphical models: Numerical example

To provide a numerical example for modeling causal inference with the graph shown in Figure 1, we use the data from the BioNTech SARS-CoV-2 vaccine trial (Polack et al., 2020, Table 1). First, we show how the parameterized causal model recovers the empirical probabilities. Then, we demonstrate how it can be used to generalize to different baseline risks.

The first step is to estimate parameters w_a and w_c from the observed frequencies (The base rate of the cause, b_c, can be estimated analogously, but is not relevant for the present purposes.) If we use maximum-likelihood point estimates derived from the observed frequencies for estimating w_a and w_c, the conditional probability of the effect given the cause computed via Equations 7 and 8 corresponds exactly to the empirical probability (i.e., proportion of observed disease cases in the vaccine group). We do not consider alternative approaches based on Bayesian methods (for details, see Griffiths & Tenenbaum, 2005; Lu et al., 2008; Meder et al., 2014) here for the sake of simplicity. While Bayesian methods involve more intricate mathematical computations, the conceptual implications do not change.

In the BNT162b2 trial, 8 out of 18,198 people in the vaccine group developed Covid-19 (Table 1). Hence, $P(disease|vaccine) = P(e|c) = 8/181988 = 0.0004$. This probability remains consistent when computing $P(e|c)$ from the parameterized causal model according to Equation 7 and using ΔP (absolute risk reduction) as an estimate for w_c, or when using Equation 8 and using causal power (relative risk reduction) as an estimate for w_c. An estimate for the strength of the background cause, or baseline risk, is provided by the number of disease cases in the placebo group, where 162 out of 18,325 participants developed Covid-19 (Table 1). Thus, $w_a = \frac{162}{18,325} = 0.0088$.

For estimating causal strength, represented by w_c, we can either use ΔP, which is equal to an absolute risk reduction, or causal power, which corresponds to a relative risk reduction. First, we compute ΔP:

$$\begin{aligned} \Delta P &= P(e|c) - P(e|\neg c) \\ &= \frac{8}{18198} - \frac{162}{18325} \\ &= 0.0004 - 0.0088 \\ &= -0.0084 \end{aligned} \tag{A1}$$

Next, we use Equation 7 to compute the conditional probability of the effect given the cause, that is, the probability of Covid-19 given vaccination:

$$\begin{aligned} P(e|c, a; w_a, w_c) &= a \cdot w_a + c \cdot w_c \\ &= a \cdot w_a + c \cdot \Delta P \\ &= a \cdot w_a + c \cdot ARR \\ &= 0.0088 - 0.0084 \\ &= 0.0004 \end{aligned} \tag{A2}$$

This probability is identical to the probability obtained from the observed frequencies as computed above.

Next, we show that the same probability obtains when using causal power q_c instead of ΔP as estimate for w_c (Equation 4), and compute the probability of the effect given the cause according to Equation 8. Using causal power as estimate for w_c corresponds to an alternative parameterization of the graph, instantiating a probabilistic generalization of

logical functions (Pearl, 1988; Yuille & Lu, 2007). Incidentally, this aligns with using a relative risk reduction as a measure of vaccine efficacy. First, we compute the causal power of the vaccine according to the observed data (Table 1):

$$\begin{aligned} q_c &= \frac{P(e|c) - P(e|\neg c)}{P(e|\neg c)} \\ &= \frac{-\Delta P}{P(e|\neg c)} \\ &= \frac{0.0084}{0.0088} \\ &= 0.95 \end{aligned} \tag{A3}$$

Now we can compute the probability of the effect (Covid-19) given the cause (vaccine) in accordance with Equation 8:

$$\begin{aligned} P(e|c, a; w_a, w_c) &= a \cdot w_a(1 - c \cdot w_c) \\ &= a \cdot w_a(1 - c \cdot q_c) \\ &= a \cdot w_a(1 - c \cdot RRR) \\ &= 0.0088(1 - 0.95) \\ &= 0.0004 \end{aligned} \tag{A4}$$

Again, this yields the same probability as calculated directly from the empirical frequencies.

These calculations illustrate how one can perform probabilistic causal inferences with a parameterized causal model. For the present purpose, we have used maximum-likelihood estimates for the graph's parameters that are directly derived from the empirical data. Accordingly, the recalculated probability of effect given cause corresponds exactly to the empirical proportion, both under a linear-additive parameterization based on ΔP and a noisy-logical parameterization based on causal power.

This approach corresponds to a causality-based factorization of the joint probability distribution over C and E, which enables inferences about scenarios that differ from the observed data. In particular, we can perform counterfactual reasoning by making inferences about novel situations for which no data is (yet) available. For instance, we might be interested in making predictions about the expected number of disease cases under a baseline risk different from the one under which the treatment effect was estimated (e.g., for a group of health professionals who have a much higher risk of being exposed to the virus). This inference can be modeled by setting the parameter w_a to the desired baseline risk, where a baseline risk of 1% would correspond to $w_a = 0.01$, a baseline risk of 30% would correspond to $w_a = 0.3$, etc.

Importantly, in this case it does matter whether a linear or noisy-logical parameterization of the graph is adopted, hence whether an absolute or relative risk reduction is used to quantify vaccine efficacy. An absolute risk reduction amounts to a linear-additive

parameterization; accordingly, the predicted probability of the effect given the cause when the baseline risk is 30% (i.e., $w_a = 0.3$) would be

$$\begin{aligned} P(e|c, a; w_a, w_c) &= a \cdot w_a + c \cdot w_c \\ &= a \cdot w_a + c \cdot \Delta P \\ &= a \cdot w_a + c \cdot ARR \\ &= 0.3 - 0.0084 \\ &= 0.2916 \end{aligned} \tag{A5}$$

Thus, when making a generalization using an absolute risk reduction, the inherent linearity assumption yields an (erroneous) estimate of about 29%.

In stark contrast, using a relative risk reduction yields a very different estimate, which aligns with the assumption that the vaccine should prevent the disease in 95% of the cases:

$$\begin{aligned} P(e|c, a; w_a, w_c) &= a \cdot w_a(1 - c \cdot w_c) \\ &= a \cdot w_a(1 - c \cdot q_c) \\ &= a \cdot w_a(1 - c \cdot RRR) \\ &= 0.3(1 - 0.95) \\ &= 0.015 \end{aligned} \tag{A6}$$

For instance, if we had 1000 health care professionals with a 30% base line risk, without vaccination we would expect about 300 to get infected. If they were vaccinated, we would expect that 95% of these cases would be prevented, leaving 15 expected disease cases among the vaccinated people.

A3. Relative risk increases and generative causal power

Analogously to absolute and relative risk reductions, the medical literature uses quantitative estimates to assess *increases* in health-related risks, such as the increased risk of skin cancer from repeated sunlight exposure or the increased risk of lung cancer for smokers. Experimental study designs that regularly examine potentially beneficial interventions for risk reduction also reveal risk increases for adverse events due to the interventions (e.g. allergic responses to medications). Similar to the preventive case there are fundamental mathematical relationships among the concepts used in the medical and causal modeling literature.

The *absolute risk increase* (ARI) is defined analogously to the preventive case as the arithmetic difference between the event rate in a control condition compared to a condition where participants have been exposed to the risk factor or treatment(Equation 1). Thus,

both absolute risk increases and absolute risk reductions align with the ΔP model, quantifying the (beneficial or harmful) change in risk as an absolute difference in the probability of the effect in the presence and absence of the cause.

In contrast, a *relative risk increase* (RRI) as used in the medical literature and generative causal power are not defined identically. The relative increase normalizes the absolute risk increase by the event rate in the control group:

$$RRI = \frac{P(event|treatment) - P(event|control)}{P(event|control)} \tag{A7}$$

However, for generative causes ($\Delta P > 0$), causal power is defined as

$$q_c = \frac{P(e|c) - P(e|\neg c)}{1 - P(e|\neg c)} = \frac{\Delta P}{1 - P(e|\neg c)} \tag{A8}$$

where $1 - P(e|\neg c)$ serves as denominator. This normalization ensures that generative causal power is always in the range [0,1], whereas the calculation of a relative risk increase (Equation A7) does not impose any such constraints, rather making the RRI a multiple of the baseline risk. Generative causal power is zero when the cause does not change the probability of the effect (i.e., $\Delta P = 0$), and obtains the maximum value of 1 if $P(e|c) = 1$, regardless of the value of the baseline risk, $P(e|\neg c)$.

Both relative risk increases and generative causal power quantify increases in the probability of the effect given the cause event. However, normatively generative causal power has several advantages over relative risk increases as defined in the medical literature. First, generative power is constrained to the same range as relative risk reductions, namely [0,1] (or, when expressed as percentages, 0-100%). This symmetry facilitates a causal interpretation analogously to decreases in risk. Second, mathematically causal power is better behaved: because the denominator in the definition of relative risk increases is $P(e|\neg c)$, the measure is undefined if the effect never occurs in the absence of the cause (Equation A8). To illustrate, assume 10 people eat a fish dish (cause C) and get food poisonings afterwards (effect E), whereas 10 other people eat no fish and develop no food poisoning. Accordingly, $P(e|c) = 1$ and $P(e|\neg c) = 0$. Using Equation A7 entails division by zero, leaving the risk increase undefined. By contrast, generative causal power is defined and takes value 1 (Equation A8), suggesting that the fish caused the food poisoning. Third, relative increases can loom large even if the overall risk remains very low. For instance, a factor that increases the risk from 0.001% (1 in 100,000) to 0.002% (2 in 100,000) increases the relative risk by 100%, although the absolute risk is still fairly low. Causal power, being bounded between 0 and 1, does not suffer from this shortcoming:

$$\begin{aligned} q_c &= \frac{P(e|c) - P(e|\neg c)}{1 - P(e|\neg c)} = \frac{\Delta P}{1 - P(e|\neg c)} \\ &= \frac{(0.0002 - 0.0001)}{(1 - 0.00001)} = \frac{0.0001}{(1 - 0.0001)} \\ &\approx 0.00001 \end{aligned} \tag{A9}$$

This low value better aligns with the intuition that the candidate cause has only a weak influence on the likelihood of the effect. Fourth, using generative causal power ensures that the estimate can be incorporated into a causal modeling framework and be used to represent w_c, thereby supporting counterfactual inferences and the evaluation of hypothetical scenarios. Thus, from a normative perspective generative causal power is a better-behaved metric that relative risk increases as defined in the medical literature. Pragmatically though, it seems unlikely to replace an established measure like relative risk increase.

Like in the preventive case, using an absolute risk increase or, equivalently, ΔP, as estimate for w_c in the causal graph (Fig. 1) corresponds to a linear parameterization. In contrast, using generative causal power (but not an RRI) instantiates a so-called noisy-OR parameterization, which extends the logical OR by allowing the cause events A and C to influence E probabilistically (Cheng, 1997; Glymour, 2003; Pearl, 1988; Yuille & Lu, 2007). Given a noisy-OR parameterization, the conditional probability of the effect given the cause (e.g., probability of infection given vaccine) can be computed from the causal model's parameters as follows:

$$P(e|c, a; w_c, w_a) = c \cdot w_c + a \cdot w_a - c \cdot w_c c \cdot a \cdot w_a \tag{A10}$$

where $a, c \in \{0, 1\}$ denote the presence and absence of candidate cause C and background cause A, and w_a and w_c denote their causal strength.

Like in the preventive case, using causal power has several advantages compared to using a linear parameterization where ΔP provides the estimate for w_c. Figure A2 illustrates this using three scenarios analogous to the analysis of preventive causes and risk reductions (Figure 3). In each instance, the causal impact of the cause is assessed under a specific baseline risk $P(e|\neg c)$, and this estimate is then utilized to draw inferences about the likelihood of the effect in scenarios with new baseline probabilities, that is, $P(e|c)$. Analogous to the preventive case, there is always a crossover of the predicted $P(e|c)$ based on ΔP versus causal power. The intersection is always located at values of $P(e|c)$ and $P(e|\neg c)$ that were originally used to estimate causal strength, because for this particular combination both measures recover the original $P(e|c)$. To the left of this point, i.e., when the new baseline probability is *lower* than the original one, the predicted conditional probability of the effect given the cause based on ΔP (or, equivalently, an ARR) is lower than the probability predicted by the causal power model. Conversely, to the right of this point, i.e., when the new baseline probability is higher than the one under which the strength of C was estimated, the predicted probability of the effect given the cause based on ΔP is higher than the probability predicted by causal power. Moreover, as in the preventive scenarios a linear parameterization based on ΔP does not always yield valid probability estimates. Specifically, while in the preventive case negative probabilities result if the new baseline probability is lower than the (absolute) value of ΔP, in generative scenarios the probability estimates exceed unity if the baseline risk is higher than $1 - \Delta P$.

Scenario 1

	Effect	
Cause	e	$\neg e$
c	15	85
$\neg c$	10	90

$P(e|c) = 0.15$
$P(e|\neg c) = 0.10$
power $q_c = 0.06$
$\Delta P = 0.05$

Scenario 2

	Effect	
Cause	e	$\neg e$
c	80	20
$\neg c$	50	50

$P(e|c) = 0.8$
$P(e|\neg c) = 0.5$
power $q_c = 0.6$
$\Delta P = 0.3$

Scenario 3

	Effect	
Cause	e	$\neg e$
c	90	10
$\neg c$	80	20

$P(e|c) = 0.9$
$P(e|\neg c) = 0.8$
power $q_c = 0.5$
$\Delta P = 0.1$

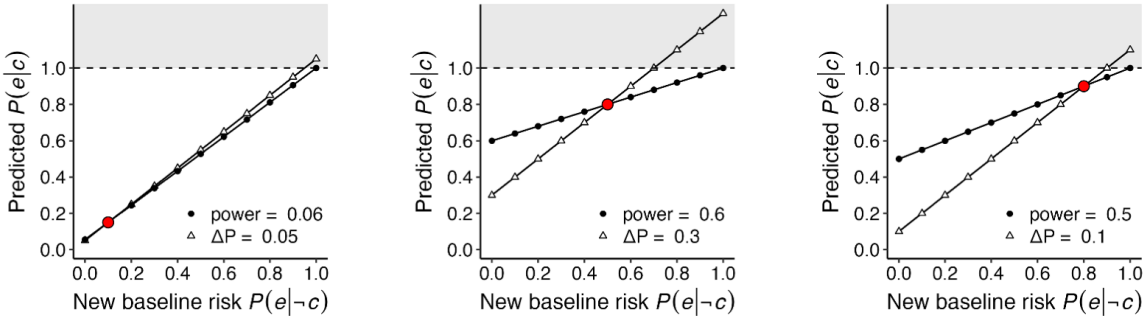

Figure A2

Analysis of three scenarios illustrating the divergence of using ΔP or causal power q_c for generalizing generative causes to situations with a new baseline risk. The top rows illustrates three different scenarios, where the baseline risk (i.e., probability of effect in the absence of the generative cause) is low (left), intermediate (middle), or high (right). The middle row shows the corresponding probabilities and values of ΔP and causal power. The bottom row shows the predicted probability of the effect given the cause under new baseline probabilities, calculated from a causal model where the causal strength of C is quantified using ΔP or causal power q_c (Equations 7 and A10, respectively). The red dot indicates the combination of $P(e|c)$ and $P(e|\neg c)$ from which the strength of C was originally estimated. The shaded area indicates predicted probabilities that exceed unity.