# How to measure effect sizes for rational decision-making[1]

Ina Jäntgen

Faculty of Philosophy, University of Cambridge, United Kingdom

E-mail address: ij271@cam.ac.uk

*This is a working paper. Comments are welcome. Please do not cite draft without permission.*

**Abstract**

Absolute and relative outcome measures measure a treatment's effect size, purporting to inform treatment choices. I argue that absolute measures are at least as good as, if not better than, relative ones for informing rational decisions across choice scenarios. Specifically, this dominance of absolute measures holds for choices between a treatment and a control group treatment from a trial and for ones between treatments tested in different trials. This distinction has hitherto been neglected, just like the role of absolute and baseline risks in decision-making that my analysis reveals. Recognizing both aspects advances the discussion on reporting outcome measures.

[1] For commentary and discussion, I am grateful to Alexander Bird, Jacob Stegenga, Arif Ahmed, Neil Dewar, Nicholas Makins, Cristian Larroulet Philippi, Adria Segarra, Sophia Crüwell, Adrian Erasmus, Oliver Holdsworth, Hamed Tabatabaei Ghomi, Charlotte Zemmel and Zinhle Mncube.

Ina Jäntgen
Working paper.

# How to measure effect sizes for rational decision-making

## 1. Introduction

In biomedical research, scientists often perform trials to test how effective treatments are. In such research, the collected data is analyzed using outcome measures, which describe how the tested treatment and the outcome of interest relate. Such outcome measures are usually interpreted as measuring the effect size of the treatment. They provide information for policymakers, patients and others aiming to decide between treatments.

Not all outcome measures provide the same information though. In this paper, I focus on outcome measures for binary variables. Here, two classes of measures, absolute and relative ones, differ in how they describe a treatment's effect size. Consider the Heart Protection Study which tested the effectiveness of a cholesterol-lowering drug called 'simvastatin' to prevent heart attacks and deaths amongst men with or at risk of heart disease (Heart Protection Study Collaborative Group 2002). The study found a so-called relative risk reduction of 18 % of coronary death. The so-called risk difference was 1.2 %.[2] Only the former effect size was reported. Yet, the difference in the described effect size is striking. Aiming to decide on taking simvastatin, which effect size is informative for a decision-maker? The relative? The absolute? Or perhaps both? More generally, how should we measure effect sizes to inform rational decision-making?

In this paper, I argue that absolute measures are at least as good as, if not better than, relative ones for informing rational decisions across choice scenarios. More precisely, absolute but not relative measures always provide sufficient probabilistic information to choose between a treatment and a control group treatment from a trial. For choices between treatments tested in distinct trials, we instead need information about the difference in the probabilities of the outcome of interest given the treatments, i.e. the difference in the absolute risks. Absent any knowledge about the probabilities of the outcome given control group treatments, i.e. the baseline risks, outcome measures do not

[2] I explain these outcome measures in section 2.

provide this information. If we as deciding agents instead know the baseline risks, then the absolute risks can be derived from both classes of outcome measures. If the baseline risks are known to be equal across trials but are themselves unknown, then absolute measures but not relative ones always provide sufficient information to choose between treatments from distinct trials. Overall, for informing rational decision-making, absolute measures dominate relative ones.

My analysis exposes the conditions under which both absolute and relative measures carry the probabilistic information a rational decision-maker needs, and when only absolute ones do so. Moreover, it identifies the role of absolute and baseline risks in rational choices. Recognizing both aspects advances the discussion on how to report effect sizes to inform treatment choices. In particular, Jacob Stegenga and his co-authors argue that only absolute measures but not relative ones are suited to inform rational decisions between treatments (Stegenga 2015; Sprenger and Stegenga 2017; Stegenga and Kenna 2017; Stegenga 2018; see also Worrall 2010). By contrast, I show when relative measures are just as good as absolute ones for this purpose. Still, I demonstrate that relative measures do not provide decision-relevant information that cannot be provided by absolute measures, including in choice scenarios Stegenga's work fails to consider. This finding strengthens the case against the need for relative measures, questioning suggestions to report both absolute and relative measures (Hoefer and Krauss 2021). Moreover, in biomedical research, most studies report only effect sizes measured in relative terms like the Heart Protection Study (Elliott et al. 2021). My results suggest that this practice could fail to inform treatment choices. Finally, I show that one could also report absolute and baseline risks to guide decisions, providing an alternative to reporting outcome measures.

I proceed as follows: In section 2, I introduce absolute and relative outcome measures. In section 3, I model two choice scenarios using expected utility theory, one involving outcome measures from a single trial and another involving outcome measures from distinct trials. Moreover, I identify absolute and baseline risks as an alternative to outcome measures for informing decisions. In section 4, I use the decision models to identify the conditions under which absolute or relative measures provide decision-relevant

information. As established by Sprenger & Stegenga (2017), absolute measures but not relative ones always do so for choices between a treatment and a control group treatment. I show that this argument does not hold for choices between treatments tested in different trials, a distinction Sprenger & Stegenga (2017) neglect. I then show that absolute measures are still at least as good as, if not better than, relative ones for informing such choices. Overall, absolute measures dominate relative ones. In section 5, I conclude with three options for reporting outcome measures supported by my analysis.

## 2. Measuring Effect Sizes for Binary Variables

In empirical research conducting trials, outcome measures are used to state how the measured values of the outcome variable in the control and treatment groups relate (Stegenga 2015). In such a way, outcome measures summarize trial data to form evidence for a causal relationship between treatment and outcome.

In this paper, I focus on outcome measures for binary variables. These are usually defined in terms of the observed frequencies in a trial (see Stewart 2016, Chapter 26), which can be represented as conditional probabilities. Let A denote the tested treatment and A′ the control group treatment. E denotes that the outcome of interest is present and ¬E that it is absent. For ease of exposition, I will throughout focus on two of the most commonly cited outcome measures.[3] Those are:

$$\text{Relative risk: } \mathrm{RR_A} = \frac{\mathrm{P(E|A)}}{\mathrm{P(E|A')}}$$

$$\text{Risk difference: } \mathrm{RD_A} = \mathrm{P(E|A)} - \mathrm{P(E|A')}$$

The relative risk belongs to a class of measures commonly called relative outcome measures. For instance, $\mathrm{RR_A}$=1.25 means that the probability of E given treatment A is 1.25 times the probability of E given the control group treatment. By contrast, the risk difference is usually classified as an absolute outcome measure. To give an example,

[3] Other measures such as relative risk reduction and numbers needed to treat can be derived from these two measures.

$RD_A$=0.05 means that the probability of E is increased by 5 % given the treatment compared to the control group treatment, for instance from 10 % to 15 %.

Binary outcome measures in the first instance measure the strength of a statistical association. If computed numerically, they are often additionally interpreted as measuring the *causal* effect size of the tested treatment (Broadbent 2013; Sprenger and Stegenga 2017). Here, I will not discuss which, if any, causal properties different outcome measures measure, as the following arguments do not depend on an answer to this question. Instead, I will speak loosely about outcome measures as measuring a treatment's effect size.

Quite obviously, relative and absolute outcome measures do not provide the same probabilistic information. Which outcome measures then provide the information we need for rational choices between treatments? In the following, I argue that absolute measures are at least as good as relative ones for informing rational decision-making across choice scenarios.

## 3. Modelling Two Scenarios for Choosing Treatments

Here are two scenarios for choosing treatments:

> ***Single:*** Imagine treatment A was tested in a trial. Treatment A′ is the control group treatment used in the trial testing A, for instance, no treatment, a competitor treatment or a placebo. Based on the reported outcome measures, an agent wants to choose between A and A′, for example, for consuming either treatment or giving it to a patient.
>
> ***Distinct:*** Imagine treatment A was tested in a trial. Moreover, an alternative treatment B was tested in another trial. Based on the reported outcome measures, an agent wants to choose between A and B, for example, for consuming either treatment or giving it to a patient.

The two scenarios involve different outcome measures. In the choice between A and A′, we consider a relative or an absolute outcome measure from the trial testing A (see section 2). By contrast, in the choice between A and B, we consider the outcome measures from

the trial testing A and from the trial testing B. Let B′ denote the control group treatment in the trial testing B. The outcome measures for the trial testing B are:

$$\mathrm{RR_B} = \frac{\mathrm{P(E|B)}}{\mathrm{P(E|B')}}$$

$$\mathrm{RD_B} = \mathrm{P(E|B)} - \mathrm{P(E|B')}$$

To distinguish both scenarios, I assume $\mathrm{A'} \neq \mathrm{B}$ and $\mathrm{B'} \neq \mathrm{A}$. It may be that $\mathrm{A'} = \mathrm{B'}$, a case that I turn to in section 4.2. In this paper, I focus on the two described scenarios since they are the most common. The other possible scenarios are a choice between control group treatments from distinct trials and one between a treatment tested in one trial and a control group treatment from another trial. Both are analogous to a choice between treatments tested in different trials. In analogous ways, the following arguments also hold for these scenarios.

How should an agent decide in the two described scenarios? In line with orthodox decision theory (see Bradley 2017, Part 1), I assume that an agent ought to choose a treatment which maximizes expected utility:

| **Expected utility maximization**: |
| --- |
| Treatment X is better than treatment Y iff $\mathrm{EU(X)} > \mathrm{EU(Y)}$ |

The relevant expected utility calculations are different in the two scenarios. I start with the choice between A and A′, following a decision model provided by Sprenger and Stegenga (Sprenger and Stegenga 2017, 845–48).[4]

Let u(E)=u denote the utility of E and u(¬E)=u' the utility of ¬E. Here, u(X) is a utility function which attaches a value to each possible outcome such as to represent the agent's evaluations of these possible outcomes. Moreover, we assume that consuming treatments

[4] Sprenger & Stegenga (2017) do not distinguish between a choice involving treatments from a single trial and one involving treatments tested in distinct trials.

comes at a cost. Broadly construed, such costs include all harmful effects of consuming the treatment, for instance, negative side effects. Let a denote the cost of consuming A and a′ the cost of consuming A′.[5] The expected utilities of A and A′ are calculated as follows:

$$(1)\ \mathrm{EU(A)} = \mathrm{P(E|A)u} + \mathrm{P(\neg E|A)u'} - \mathrm{a} = \mathrm{P(E|A)}\ (\mathrm{u} - \mathrm{u'}) + \mathrm{u'} - \mathrm{a}$$

$$(2)\ \mathrm{EU(A')} = \mathrm{P(E|A')u} + \mathrm{P(\neg E|A')u'} - \mathrm{a'} = \mathrm{P(E|A')}\ (\mathrm{u} - \mathrm{u'}) + \mathrm{u'} - \mathrm{a'}$$

(1) and (2) jointly with expected utility maximization provide a decision model:

$$(3)\ \mathrm{EU(A)} > \mathrm{EU(A')}\ \text{iff}\ \mathrm{P(E|A)}\ (\mathrm{u} - \mathrm{u'}) - \mathrm{a} > \mathrm{P(E|A')}\ (\mathrm{u} - \mathrm{u'}) - \mathrm{a'}$$

For the choice involving several trials, the relevant expected utility calculations are (1) and with b denoting the cost of consuming B:

$$(4)\ \mathrm{EU(B)} = \mathrm{P(E|B)u} + \mathrm{P(\neg E|B)u'} - \mathrm{b} = \mathrm{P(E|B)}\ (\mathrm{u} - \mathrm{u'}) + \mathrm{u'} - \mathrm{b}$$

Again, we derive a decision model from combining (1) and (4) with expected utility maximization:

$$(5)\ \mathrm{EU(A)} > \mathrm{EU(B)}\ \text{iff}\ \mathrm{P(E|A)}\ (\mathrm{u} - \mathrm{u'}) - \mathrm{a} > \mathrm{P(E|B)}\ (\mathrm{u} - \mathrm{u'}) - \mathrm{b}$$

In section 4, I will rely on both decision models to analyze when absolute and relative measures provide decision-relevant probabilistic information. To do so, I will assume that the observed frequencies used to calculate outcome measures are numerically equivalent to the agent's decision-relevant credences. For instance, in (3), I take $\mathrm{P(E|A)}$ to denote both the agent's credence in E occurring if she would take treatment A and the observed frequency of E given A in the trial testing A. This is a substantial simplification; there are several inferences involved in forming a credence on a treatment's effect size for a treatment choice based on one calculated from trial frequencies (Fuller and Flores 2015).

[5] In separating costs and utilities, I follow the model in Sprenger & Stegenga (2017). I will assume here that costs and utilities are commensurable such that we can combine both to calculate expected utility. I also assume the costs are certain. Both assumptions are implicit in the model presented in Sprenger & Stegenga (2017).

Indeed, analyzing how outcome measures figure in such inferences is an important task I bracket in this paper. Nevertheless, this omission poses no threat to my argumentation. If reported outcome measures are to inform rational decision-making at all, then they must in principle provide the needed probabilistic information. The mentioned simplification allows us to analyze when absolute or relative measures succeed in doing so.

The decision models (3) and (5) already show an alternative to reporting outcome measures for informing treatment choices: report absolute and baseline risks. As we can see in (3), in a choice involving treatments from a single trial, we can decide between treatments if we know $P(E|A)$ and $P(E|A')$ in addition to costs and utilities.[6] Moreover, as we can see in (5), in a choice involving treatments from distinct trials, we can decide between treatments if we know $P(E|A)$ and $P(E|B)$ in addition to costs and utilities. These conditional probabilities are usually called absolute risks when referring to the treatment group, that is $P(E|A)$ and $P(E|B)$ (Stewart 2016, Chapter 26). When referring to the control group, that is $P(E|A')$ and $P(E|B')$, I will call them baseline risks. Reporting absolute and baseline risks provides sufficient probabilistic information for rational decision-making. Note that it is feasible to report these risks; anyone calculating any outcome measure has access to them.[7]

Still, in biomedical research, most studies assessing the effectiveness of treatments report effect sizes to inform treatment choices, rather than solely the underlying absolute and baseline risks. Hence, it is important to analyze when this practice successfully informs decisions. Moreover, this practice could be warranted. Perhaps effect sizes extrapolate better to target populations than absolute and baseline risks. Or perhaps laypeople reason

[6] This verdict holds even though reporting $P(E|A)$ and $P(E|A')$ provides some unnecessary information; reporting the mere difference between these probabilities, i.e. $RD_A$, is always already sufficient for deciding between A and A′ (see section 4.1).

[7] It is worth emphasizing that even if researchers were to report solely absolute and baseline risks, they would still need outcome measures to summarize trial data such as to form evidence for causal inferences (see section 2).

better when provided with effect sizes rather than with absolute and baseline risks. Assessing such considerations is beyond the scope of this paper. Instead, I will simply take the practice of reporting effect sizes seriously. To do so, I will henceforth assume that we as deciding agents do not know the absolute risks and, unless noted otherwise, we do not know the baseline risks either. These assumptions allow us to identify the conditions under which different outcome measures provide us with decision-relevant information.

## 4. The Decision-theoretic Dominance of Absolute Measures

Under what conditions do absolute or relative outcome measures provide sufficient probabilistic information to choose rationally between treatments? In choices involving outcome measures from a single trial, only absolute measures always do so (section 4.1). This is established by Sprenger & Stegenga (2017). In choices involving outcome measures from distinct trials, depending on our knowledge of baseline risks, both absolute and relative measures provide the decision-relevant information, neither do or only absolute ones always do (section 4.2). Overall, absolute measures dominate relative ones.

### 4.1. Choices involving Outcome Measures from a Single Trial

Sprenger & Stegenga (2017) use the decision model (3) to argue that absolute measures but not relative ones always provide sufficient probabilistic information to decide between treatments given costs and utilities. These authors fail to distinguish a choice involving outcome measures from a single trial from one involving outcome measures from distinct trials. As I will show in section 4.2, this failure poses a problem for applying their argument to the latter case. However, the authors' argument still applies to the former case. To see this, I briefly recap their argument (Sprenger and Stegenga 2017, 845–48).

From (3) and assuming without loss of generality $u > u'$ one can derive

$$(6)\ \mathrm{EU(A)} > \mathrm{EU(A')} \text{ iff } \mathrm{P(E|A)} - \mathrm{P(E|A')} > \frac{a - a'}{u - u'}$$

The red term in (6) is equivalent to $RD_A$. As a result, given costs, utilities and $RD_A$ one always knows whether $\mathrm{EU(A)} > \mathrm{EU(A')}$. The same does not hold for relative measures. To see this, we can derive

$$(7)\ \mathrm{P(E|A)} - \mathrm{P(E|A')} = \mathrm{P(E|A')}\left(\frac{\mathrm{P(E|A)}}{\mathrm{P(E|A')}} - 1\right) = \mathrm{P(E|A')}\,(\mathrm{RR_A} - 1)$$

From (6) and (7), we get

$$(8)\ \mathrm{EU(A)} > \mathrm{EU(A')} \text{ iff } \mathrm{P(E|A')}\,(\mathrm{RR_A} - 1) > \frac{\mathrm{a}-\mathrm{a'}}{\mathrm{u}-\mathrm{u'}}$$

As Sprenger and Stegenga note, costs and utilities do not determine $\mathrm{P(E|A')}$. Nor does a given $\mathrm{RR_A}$. As a result, assuming $\mathrm{a} \neq \mathrm{a'}$ as the authors do, one cannot always decide whether $\mathrm{EU(A)} > \mathrm{EU(A')}$ given costs, utilities and $\mathrm{RR_A}$. $\mathrm{P(E|A')}$ could be such that $\mathrm{EU(A)} > \mathrm{EU(A')}$ or such that $\mathrm{EU(A)} < \mathrm{EU(A')}$. Suppose $\mathrm{RR_A} = 1.25$ and $\frac{\mathrm{a}-\mathrm{a'}}{\mathrm{u}-\mathrm{u'}} = 0.1$. Then, $\mathrm{EU(A)} > \mathrm{EU(A')}$ if $\mathrm{P(E|A')} = 0.5$, but $\mathrm{EU(A)} < \mathrm{EU(A')}$ if $\mathrm{P(E|A')} = 0.3$. Both baseline risks are compatible with but unknown given these costs, utilities and $\mathrm{RR_A}$. Hence, in contrast to absolute measures, given costs, utilities and $\mathrm{RR_A}$ one cannot always decide whether $\mathrm{EU(A)} > \mathrm{EU(A')}$.[8]

Moving beyond Sprenger & Stegenga (2017), it is worth noting that relative measures provide sufficient information for choosing if both treatments come at equal costs. From (8) and assuming $\mathrm{a} = \mathrm{a'}$ we can derive

$$(9)\ \mathrm{EU(A)} > \mathrm{EU(A')} \text{ iff } \mathrm{P(E|A')}(\mathrm{RR_A} - 1) > 0$$

If $\mathrm{RR_A} < 1$, then (9) demands to choose A′. This is because a well-defined $\mathrm{RR_A} < 1$ implies that $\mathrm{P(E|A')} > 0$, and thus that $\mathrm{P(E|A')}(\mathrm{RR_A} - 1) < 0$. Analogously, if $\mathrm{RR_A} > 1$, (9) demands to choose A. In other words, by knowing $\mathrm{RR_A}$, utilities and equality of costs we can always settle which treatment to take. However, we are rarely if ever in a situation in which treatments come at equal costs. Thus, I will henceforth not mention this special case.

[8] Note that from (8) it follows that reporting a relative measure *and* the baseline risk to the deciding agent is always sufficient for a choice between A and A′.

To summarize, Sprenger & Stegenga (2017) establish that absolute but not relative measures always provide sufficient information to choose between treatments from a single trial. In the next section, I show that this argument fails to apply to choices between treatments tested in distinct trials. I then argue that absolute measures still dominate relative ones for informing such choices.

### 4.2. Choices involving Outcome Measures from Distinct Trials

We cannot apply Sprenger & Stegenga's (2017) argument to the decision model for a choice between treatments from distinct trials (5) as we have to the one for a choice involving a single trial (3). To see this, note that from (5) and assuming without loss of generality $u > u'$ one can derive

$$(10)\ \mathrm{EU(A)} > \mathrm{EU(B)} \text{ iff } \mathrm{P(E|A)} - \mathrm{P(E|B)} > \frac{a-b}{u-u'}$$

In (10), we cannot interpret the red difference in probabilities as $RD_A$ or as $RD_B$, as we have done in (6). The same holds for the case of RR. This can be seen by noting that

$$(11)\ \mathrm{P(E|A)} - \mathrm{P(E|B)} = \mathrm{P(E|B)} \left( \frac{\mathrm{P(E|A)}}{\mathrm{P(E|B)}} - 1 \right)$$

The red term in (11) is neither equal to $RR_A$ nor $RR_B$, contrary to the previous case (7). Hence, we cannot rely on Sprenger & Stegenga's (2017) argument here.

As can be seen in (10), for choices between treatments from distinct trials, we need information about the difference in absolute risks, i.e. the difference between $\mathrm{P(E|A)}$ and $\mathrm{P(E|B)}$. Correspondingly, reporting these absolute risks rather than outcome measures is a straightforward option to inform such choices. This importance of absolute risks is obscured in Sprenger & Stegenga (2017) since they fail to distinguish a choice involving several trials from one involving a single trial.

Still, as discussed in section 3, I will here acknowledge the practice of reporting outcome measures. Hence, I assume that we as deciding agents choose between A and B without knowing the absolute risks, but rather only the reported absolute and relative outcome measures. This assumption allows us to see when absolute or relative outcome measures

can still inform choices, in the sense of providing information about the decision-relevant difference in absolute risks. To answer this question, let us distinguish three epistemic situations we could be in when deciding between treatments from distinct trials, differing in how much we as deciding agents know about the baseline risks:

**Ignorance:** We know nothing about $P(E|A')$ and $P(E|B')$.

**Full knowledge:** We know $P(E|A')$ and $P(E|B')$.

**Partial knowledge:** We know that $P(E|A') = P(E|B')$, though we know neither $P(E|A')$ nor $P(E|B')$.

Let us examine each case in turn.

### 4.2.1. Ignorance

In the case of ignorance about baseline risks, knowing absolute or relative outcome measures is not sufficient to have any information about the difference in $P(E|A)$ and $P(E|B)$. This is because both a given absolute or a given relative measure is compatible with a range of values for $P(E|A)$ and $P(E|B)$, absent any information about $P(E|A')$ and $P(E|B')$. This result is important. It shows that unless we have some information about baseline risks outcome measures from distinct trials do not provide the information needed for choosing between the treatments tested in these trials.

### 4.2.2. Full Knowledge

When knowing the baseline risks, absolute or relative measures can both be used to calculate the absolute risks, and thus their difference. If one knows $RD_A$ and $RD_B$ as well as $P(E|A')$ and $P(E|B')$ then one knows $P(E|A)$ and $P(E|B)$. If one knows $RR_A$ and $RR_B$ as well as $P(E|A')$ and $P(E|B')$ then one knows $P(E|A)$ and $P(E|B)$. This result shows that relative measures can sometimes provide equally valuable information for decisions as absolute ones.

### 4.2.3. Partial Knowledge

In the case of partial knowledge about baseline risks, absolute measures but not relative ones always provide sufficient probabilistic information to choose between treatments. To see this, note that given $\mathrm{P(E|A')} = \mathrm{P(E|B')}$ we can derive

$$(12)\ \mathrm{P(E|A)} - \mathrm{P(E|B)} = \big(\mathrm{P(E|A)} - \mathrm{P(E|A')}\big) - \big(\mathrm{P(E|B)} - \mathrm{P(E|B')}\big) = \mathrm{RD_A} - \mathrm{RD_B}$$

As can be seen in (12), under partial knowledge of baseline risks, absolute measures always provide the difference in absolute risks that is sufficient to decide between A and B. Moreover, the same does not hold for relative measures. Given $\mathrm{P(E|A')} = \mathrm{P(E|B')}$ we get

$$(13)\ \mathrm{P(E|A)} - \mathrm{P(E|B)} = \frac{\mathrm{P(E|A)}*\mathrm{P(E|A')}}{\mathrm{P(E|A')}} - \frac{\mathrm{P(E|B)}*\mathrm{P(E|B')}}{\mathrm{P(E|B')}} = \mathrm{P(E|A')}\,(\,\mathrm{RR_A} - \mathrm{RR_B})$$

From (13) and (10), we can derive

$$(14)\ \mathrm{EU(A)} > \mathrm{EU(B)} \text{ iff } \mathrm{P(E|A')}\,(\,\mathrm{RR_A} - \mathrm{RR_B}) > \frac{a-b}{u-u'}$$

Just as in the case of a choice involving a single trial, neither costs, utilities nor $\mathrm{RR_A}$ and $\mathrm{RR_B}$ determine $\mathrm{P(E|A')}$. As a result, assuming $a \neq b$, knowing costs, utilities as well as $\mathrm{RR_A}$ and $\mathrm{RR_B}$ does not always provide us with sufficient information to decide between A and B. For example, suppose $\mathrm{RR_A} = 1.25$, $\mathrm{RR_B} = 1.02$ and $\frac{a-b}{u-u'} = 0.15$. Then, $\mathrm{EU(A)} > \mathrm{EU(B)}$ if $\mathrm{P(E|A')} = \mathrm{P(E|B')} = 0.7$ , but $\mathrm{EU(A)} < \mathrm{EU(B)}$ if $\mathrm{P(E|A')} = \mathrm{P(E|B')} = 0.5$. Both baseline risks are compatible with but unknown given costs, utilities, $\mathrm{RR_A}$ and $\mathrm{RR_B}$. We cannot decide whether A or B maximizes expected utility here. Relative measures are not always sufficient to decide between treatments under partial knowledge of baseline risks.

It is worth noting that relative measures are apt to inform choices given equal costs of treatments and partial knowledge of baseline risks. Using (14) and assuming $a = b$ we can derive

$$(15)\ \mathrm{EU(A)} > \mathrm{EU(B)} \text{ iff } \mathrm{P(E|A')}\,(\,\mathrm{RR_A} - \mathrm{RR_B}) > 0$$

We know that $P(E|A') = P(E|B') > 0$ if $RR_A$ and $RR_B$ are well-defined. Thus, if we know $RR_A$ and $RR_B$ we know whether $P(E|A')\,(RR_A - RR_B) > 0$ and therefore whether $EU(A) > EU(B)$. Again, I will henceforth bracket this unusual case. Overall, in the case of partial knowledge, absolute measures but not relative ones always provide the decision-relevant probabilistic information.

These results can also contribute to debates on using placebos versus so-called active comparators, i.e. already used treatments, as control group treatments in trials. Reviews suggest that placebo-controlled studies or studies with no treatment for the control group are more common than ones using active comparators (Hochman and McCormick 2010; Cipriani et al. 2020). Yet, the use of placebos is often criticized for ethical reasons; if an effective treatment exists giving a placebo to a trial participant implies withholding this treatment from her (Emanuel and Milller 2001; European Medicines Agency 2001). Indeed, research guidelines only allow using placebos under specific conditions (World Medical Association 2013). Moreover, more active comparator trials have been demanded on grounds of them establishing the comparative effectiveness of treatments that matters for decision-making (Cipriani et al. 2020; Naci et al. 2020).

The above results add a decision-theoretic nuance to this debate. Unless baseline risks are known to decision-makers, to compare treatments between trials using absolute measures we need to establish equality of baseline risks. One way to ensure good grounds for such equality would be to use the same control group treatment across trials, for example, the same active comparator or the same placebo. On the one hand, a plea for more active comparator trials to inform decision-making then ought to recognize the importance of using the same comparator across trials to inform choices between treatments tested in these trials. On the other hand, using the same placebo across studies also suggests equality of baseline risks. Note that this decision-theoretic nuance only applies to the debate on control group treatments if absolute risks are not reported in addition to or instead of outcome measures. If they are, then no matter how the trials are designed, we can compare the reported absolute risks for deciding between treatments. This insensitivity to trial

design could pose an advantage for reporting absolute risks over reporting effect sizes for informing decisions between treatments tested in different trials.

## 5. Conclusion

Let us return to our example of the Heart Protection Study: how should we measure the effect size of simvastatin to inform a rational choice on taking this drug? I have argued that absolute measures such as the risk difference of 1.2 % are at least as good as, if not better than, relative ones such as the relative risk reduction of 18 % for informing such a decision across choice scenarios. When choosing between a treatment, for instance, simvastatin, and the corresponding control group treatment, absolute measures but not relative ones always provide sufficient information for choosing rationally. When choosing between treatments tested in distinct trials, we need information about the difference in absolute risks. Absent some knowledge about the baseline risks in the considered trials, neither an absolute nor a relative outcome measure provides such information. If we know these baseline risks, then absolute risks can be calculated from both measures. If we only have partial knowledge of the baseline risks, then only absolute measures always provide sufficient information about the decision-relevant difference in absolute risks. Finally, instead of relying on outcome measures, we can also use absolute and baseline risks to decide between treatments. Overall, absolute measures dominate relative ones for informing rational decision-making. These results support the following options for reporting outcome measures:

**Option 1:** Report absolute measures and either absolute risks or baseline risks or ensure equality of baseline risks.
**Option 2:** Report relative measures and baseline risks.
**Option 3:** Instead of reporting outcome measures, report absolute and baseline risks.

Conclusions drawn from idealized decision-theoretic models cannot by themselves defend any reporting principle for outcome measures aimed at informing actual decision-makers. Important further considerations that should be discussed to justify reporting principles based on the three described options include ethical aspects (see Schroeder forthcoming) and insights on how to best communicate risks to people (see Spiegelhalter 2017).

Nevertheless, the decision-theoretic dominance of absolute measures challenges both the current practice of only reporting relative measures and suggestions to always report both (Hoefer and Krauss 2021). Instead, a safer conjecture is to always report absolute measures or absolute risks and baseline risks.

**References**

Bradley, Richard. 2017. *Decision Theory with a Human Face*. Cambridge: Cambridge University Press.

Broadbent, Alex. 2013. *Philosophy of Epidemiology*. London: Palgrave Macmillan.

Cipriani, Andrea, John P. A. Ioannidis, Peter M. Rothwell, Paul Glasziou, Tianjing Li, Adrian F. Hernandez, Anneka Tomlinson, John Simes, and Huseyin Naci. 2020. "Generating Comparative Evidence on New Drugs and Devices after Approval". *Lancet* 395 (10228): 998–1010. https://doi.org/10.1016/S0140-6736(19)33177-0.

Elliott, Marissa H., Joshua J. Skydel, Sanket S. Dhruva, Joseph S. Ross, and Joshua D. Wallach. 2021. "Characteristics and Reporting of Number Needed to Treat, Number Needed to Harm, and Absolute Risk Reduction in Controlled Clinical Trials, 2001-2019". *JAMA Internal Medicine* 181 (2): 282–84. https://doi.org/10.1001/jamainternmed.2020.4799.

Emanuel, Ezekiel J., and Franklin G. Milller. 2001. "The Ethics of Placebo-Controlled Trials – a Middle Ground". *New England Journal of Medicine* 345 (12): 915–19. https://doi.org/10.1056/NEJM200109203451211.

European Medicines Agency. 2001. "ICH Topic E10: Choice of Control Group in Clinical Trials. Note for Guidance on Choice of Control Group in Clinical Trials". CPMP/ICH/364/96. European Medicines Agency. https://www.ema.europa.eu/en/ich-e10-choice-control-group-clinical-trials.

Fuller, Jonathan, and Luis J. Flores. 2015. "The Risk GP Model: The Standard Model of Prediction in Medicine". *Studies in History and Philosophy of Biological and Biomedical Sciences* 54: 49–61. https://doi.org/10.1016/j.shpsc.2015.06.006.

Heart Protection Study Collaborative Group. 2002. "MRC/BHF Heart Protection Study of Cholesterol Lowering with Simvastatin in 20,536 High-Risk Individuals: A

Randomised Placebo-Controlled Trial". *Lancet* 360 (9326): 7–22. https://doi.org/10.1016/S0140-6736(02)09327-3.

Hochman, Michael, and Danny McCormick. 2010. "Characteristics of Published Comparative Effectiveness Studies of Medications". *JAMA* 303 (10): 951–58. https://doi.org/10.1001/jama.2010.240.

Hoefer, Carl, and Alexander Krauss. 2021. "Measures of Effectiveness in Medical Research: Reporting Both Absolute and Relative Measures". *Studies in History and Philosophy of Science Part A* 88: 280–83. https://doi.org/10.1016/j.shpsa.2021.06.012.

Naci, Huseyin, Maximilian Salcher-Konrad, Aaron S. Kesselheim, Beate Wieseler, Lise Rochaix, Rita F. Redberg, Georgia Salanti, et al. 2020. "Generating Comparative Evidence on New Drugs and Devices before Approval". *Lancet* 395 (10228): 986–97. https://doi.org/10.1016/S0140-6736(19)33178-2.

Schroeder, Andrew. forthcoming. "An Ethical Framework for Presenting Scientific Results to Policy-Makers". *Kennedy Institute of Ethics Journal* 32 (1).

Spiegelhalter, David. 2017. "Risk and Uncertainty Communication". *Annual Review of Statistics and Its Application* 4 (1): 31–60. https://doi.org/10.1146/annurev-statistics-010814-020148.

Sprenger, Jan, and Jacob Stegenga. 2017. "Three Arguments for Absolute Outcome Measures". *Philosophy of Science* 84 (5): 840–52. https://doi.org/10.1086/693930.

Stegenga, Jacob. 2015. "Measuring Effectiveness". *Studies in History and Philosophy of Biological and Biomedical Sciences* 54: 62–71. https://doi.org/10.1016/j.shpsc.2015.06.003.

———. 2018. *Medical Nihilism*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198747048.001.0001.

Stegenga, Jacob, and Aaron Kenna. 2017. "Absolute Measures of Effectiveness". In *Measurement in Medicine: Philosophical Essays on Assessment and Evaluation*, ed. Leah McClimans, 35–51. London: Rowman & Littlefield.

Stewart, Antony. 2016. *Basic Statistics and Epidemiology: A Practical Guide*. 4th edition. Boca Raton: Taylor & Francis Ltd.

World Medical Association. 2013. "World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects". *JAMA* 310 (20): 2191–94. https://doi.org/10.1001/jama.2013.281053.

Worrall, John. 2010. "Do We Need Some Large, Simple Randomized Trials in Medicine?" In *EPSA Philosophical Issues in the Sciences*, ed. Mauricio Suarez, Mauro Dorato, and Miklos Redei, 289–301. Dordrecht: Springer.