RESEARCH ARTICLE

On the collapsibility of measures of effect in the counterfactual causal framework

Anders Huitfeldt[1]* , Mats J. Stensrud[2,4] and Etsuji Suzuki[3,4]

Abstract

The relationship between collapsibility and confounding has been subject to an extensive and ongoing discussion in the methodological literature. We discuss two subtly different definitions of collapsibility, and show that by considering causal effect measures based on counterfactual variables (rather than measures of association based on observed variables) it is possible to separate out the component of non-collapsibility which is due to the mathematical properties of the effect measure, from the components that are due to structural bias such as confounding. We provide new weights such that the causal risk ratio is collapsible over arbitrary baseline covariates. In the absence of confounding, these weights may be used for standardization of the risk ratio.

Introduction

A measure of association (such as the risk difference or the risk ratio) is said to be collapsible if the marginal measure of association is equal to a weighted average of the stratum-specific measures of association [1]. The relationship between collapsibility and confounding has been subject to an extensive and ongoing discussion in the literature [2]. In this paper, we argue that the concept of collapsibility can be made clearer if we frame the discussion in terms of causal effect measures based on counterfactual variables.

In all the examples, we are interested in the effect of a binary exposure A (e.g. a drug), on a binary outcome Y (e.g. a side effect). We use superscript to denote counterfactual variables [3]. For example, $Y^{a=1}$ is an indicator for whether an individual would have got the outcome if, possibly contrary to fact, she had been exposed to the drug. We will make a distinction between measures of association, which compare the distribution of the outcome in the exposed with the distribution of outcome in the unexposed; and causal measures of effect, which compare the counterfactual distribution of the outcome under exposure (that is, the distribution of the outcome if everyone is exposed) with the counterfactual distribution of

the outcome under the absence of exposure (that is, the distribution of the outcome if everyone is unexposed). For example, the associational risk difference is $\Pr(Y = 1|A = 1) - \Pr(Y = 1|A = 0)$ whereas the causal risk difference (RD) is $\Pr(Y^{a=1} = 1) - \Pr(Y^{a=0} = 1)$. These effect measures may be defined within levels of covariates V.

This paper is organized as follows. In second section, we discuss two definitions of collapsibility, and show how established results about collapsibility depend on which definition is considered. In third section, we provide an application of these definitions to three measures of causal effects (risk difference, risk ratio, and odds ratio) and discuss the weights that make the risk difference and the risk ratio collapsible over an arbitrary set of baseline covariates. These weights can be used to standardize the effect measure to a population with any distribution of V, and we explicitly introduce such weights for the risk ratio. In fourth section, we show that while the weights that are required for collapsibility of the risk ratio involve counterfactual (and therefore unobservable) components, these weights are identified from observed data in the absence of confounding. In fifth section, we discuss the implications of these results.

Definitions of collapsibility

We will adopt Pearl's definition of collapsibility for measures of association [4]:

*Correspondence: anders@huitfeldt.net
[1] Norwegian Institute of Public Health, Oslo, Norway
Full list of author information is available at the end of the article

Huitfeldt *et al. Emerg Themes Epidemiol*   (2019) 16:1

Page 2 of 5

Definition 1 (*Collapsibility of a Measure of Association*) Let $g[f(A, Y)]$ be any functional that measures the association between A and Y in the joint distribution $f(A, Y)$. We say that g is collapsible on a variable V with weights w_v if $\frac{\sum_v \{g[f(A,Y|V=v)] \times w_v\}}{\sum_v w_v} = g[f(A, Y)]$

Newman [5] showed conditions under which the associational risk difference, risk ratio, and odds ratio are collapsible (or averageable) according to this definition. He also provided corresponding weights. Briefly, we note that: the associational risk difference is collapsible with weights $\Pr(V = v|A = 1)$ if V is not associated with the outcome in the unexposed, or if V is not associated with the exposure; the associational risk ratio is collapsible with weights $\Pr(V = v|A = 1) \times \Pr(Y = 1|A = 0, V = v)$ under similar conditions; and the associational odds ratio is collapsible with weights $\Pr(V = v|A = 1) \times \Pr(Y = 0|A = 1, V = v)$ $\times \frac{\Pr(Y=1|A=0,V=v)}{1-\Pr(Y=1|A=0,V=v)}$ under certain very limiting conditions, for example if $\Pr(Y = 1|A = 0, V = v)$ is equal for all values of v. A full discussion of the graphical and probabilistic conditions that lead to collapsibility under this definition is provided by Greenland and Pearl [6].

From these results, it follows that general statements about the collapsibility properties of effect measures (e.g. "the risk difference is collapsible") must either be qualified by a specification of the conditions that are being assumed, or alternatively taken to refer to some other definition of collapsibility. We therefore suggest a suitable alternative definition: a measure of causal effect is collapsible if the marginal effect measure is equal to a weighted average of the stratum-specific causal effect measures. This is a formalization of the definition used in Fine Point 4.3 in Hernan and Robins textbook *Causal Inference*, and can mathematically be stated as follows:

Definition 2 (*Collapsibility of a Measure of Causal Effect*) Let $h[f(Y^{a=0}, Y^{a=1})]$ be any function of the parameters $Y^{a=0}$ and $Y^{a=1}$ in the joint distribution $f(Y^{a=0}, Y^{a=1})$. We say that h is collapsible on a variable V with weights w_v if $\frac{\sum_v \{\{h[f(Y^{a=0}, Y^{a=1}|V=v)] \times w_v\}}{\sum_v w_v} = h[f(Y^{a=0}, Y^{a=1})]$.

As we will show in the next section, under Definition 2 it is possible to provide results that guarantee collapsibility of certain effect measures *for any data set*. Collapsibility can then be understood as a mathematical property of the effect measure, rather than a consequence of certain graphical or probabilistic structures in the particular data set. Consequently, results from

Greenland and Pearl do not apply under Definition 2, and measures of effect may be collapsible over V even if V is a confounder. Definitions 1 and 2 are not generally equivalent: a set of weights that satisfies Definition 1 may not satisfy Definition 2, and conversely a set of weights that satisfies Definition 2 may not satisfy Definition 1. The definitions are however equivalent if there is neither confounding unconditionally, nor confounding conditional on V (i.e. if $Y^a \perp\!\!\!\perp A$ and $Y^a \perp\!\!\!\perp A|V$ for all values of a, respectively).

Finally, we consider a third related concept, discussed by Miettinen [7], who stated (correctly, but without proof) that the "standardized risk ratio" (SRR), which is constructed by standardizing the risk in the exposed and the risk in the unexposed separately with weights $\Pr(V = v)$ and reporting the ratio of these measures (Formula 4 in Miettinen), is equal to a weighted average of the stratum-specific risk ratios under the weights $\Pr(V = v) \times \Pr(Y = 1|A = 0, V = v)$. (Formula 6 in Miettinen). Since Miettinen's SRR is equal to the causal risk ratio if there is no unmeasured confounding, Miettinen's weights satisfy Definition 2 in the special case of no confounding conditional on V.

Collapsibility of measures of causal effect
Risk difference
The causal risk difference is collapsible over covariates V with respect to weights w_v if $\frac{\sum_v \{[\Pr(Y^{a=1}=1|V=v) - \Pr(Y^{a=0}=1|V=v)] \times w_v\}}{\sum_v w_v} = \Pr(Y^{a=1} = 1) - \Pr(Y^{a=0} = 1)$. We next proceed to show that the causal risk difference is collapsible over arbitrary covariates V if we use the weights $w_v = \Pr(V = v)$.

First note that the sum of the weights is 1, allowing the denominator to be ignored. Next,

$$\sum_v [\Pr(Y^{a=1} = 1|V = v) - \Pr(Y^{a=0} = 1|V = v)]$$
$$\times \Pr(V = v)$$
$$= \sum_v \Pr(Y^{a=1} = 1|V = v) \times \Pr(V = v)$$
$$- \sum_v \Pr(Y^{a=0} = 1|V = v) \times \Pr(V = v)$$
$$= \Pr(Y^{a=1} = 1) - \Pr(Y^{a=0} = 1) \tag{1}$$

Also note that in the absence of effect modification (i.e. if the risk difference is the same in every stratum) the stratum-specific risk differences will also be equal to the marginal risk difference, and the risk difference is collapsible with any weights. It can be shown that this is true for any measure of effect for which there exist weights that guarantee collapsibility over arbitrary covariates.

Risk ratio

The risk ratio is asymmetric with respect to coding of the outcome, so it is necessary to consider each risk ratio model separately. These are defined as follows:

$$RR(-) = \frac{\Pr(Y^{a=1} = 1)}{\Pr(Y^{a=0} = 1)}$$

$$RR(+) = \frac{\Pr(Y^{a=1} = 0)}{\Pr(Y^{a=0} = 0)}$$

The two risk ratio models require different sets of weights for collapsibility. We next show that the causal risk ratio $RR(-)$ is collapsible over arbitrary covariates V if we use the weights $w_v = \Pr(V = v | Y^{a=0} = 1)$, i.e. weights determined by the distribution of the baseline covariates among those individuals who would have been cases if they, possibly contrary to fact, were not treated with drug A:

Our goal is to show that

$$
\begin{aligned}
RR(-) &= \frac{\Pr(Y^{a=1} = 1)}{\Pr(Y^{a=0} = 1)} \\
&= \frac{\sum_v \left[\frac{\Pr(Y^{a=1} = 1 | V = v)}{\Pr(Y^{a=0} = 1 | V = v)} \times \Pr(V = v | (Y^{a=0} = 1)) \right]}{\sum_v \Pr(V = v | Y^{a=0} = 1)}
\end{aligned}
$$

Again, we note that the sum of the weights is 1, and that the denominator can therefore be ignored.

$$
\begin{aligned}
&\sum_v \frac{\Pr(Y^{a=1} = 1 | V = v) \times \Pr(V = v | Y^{a=0} = 1)}{\Pr(Y^{a=0} = 1 | V = v)} \\
&= \sum_v \frac{\Pr(Y^{a=1} = 1 | V = v) \times \Pr(Y^{a=0} = 1 | V = v) \times \Pr(V = v)}{\Pr(Y^{a=0} = 1 | V = v) \times \Pr(Y^{a=0} = 1)} \quad \text{(Bayes Theorem)} \\
&= \sum_v \frac{\Pr(Y^{a=1} = 1 | V = v) \times \Pr(V = v)}{\Pr(Y^{a=0} = 1)} \\
&= \frac{\sum_v \Pr(Y^{a=1} = 1 | V = v) \times \Pr(V = v)}{\Pr(Y^{a=0} = 1)} \\
&= \frac{\Pr(Y^{a=1} = 1)}{\Pr(Y^{a=0} = 1)}
\end{aligned}
\tag{2}
$$

This proof is not invariant to the coding of the exposure or outcome variables, and the correct weights will therefore depend on the exact specification of the risk ratio parameter. Analogous proofs can be provided to show that the weights for $RR(+)$ are given by $\Pr(V = v | Y^{a=0} = 0)$, the weights for $\frac{1}{RR(-)}$ are given by $\Pr(V = v | Y^{a=1} = 1)$, and that the weights for $\frac{1}{RR(+)}$ are given by $\Pr(V = v | Y^{a=1} = 0)$,

Note that the marginal causal risk ratio is generally not equal to a weighted average of the conditional causal risk ratios, if the weights are determined by the marginal distribution of the covariates V. Exceptions occur in special situations, such as when the risk ratio is equal in every stratum (i.e. when there is no effect modification on the risk ratio scale).

Odds ratio

For all the previously discussed parameters, we have shown that for any baseline covariates V, there exist weights such that the marginal effect measure is equal to a weighted average of the stratum-specific effects. We will now show that this does not hold for the odds ratio by considering the following simple counterexample:

Consider a population, with 25% men and 75% women, where a randomized trial is conducted on the effect of drug A. The hypothetical results are shown in Table 1. The randomization probability is equal in men and women and we have an infinite sample size, there is therefore no confounding.

This table shows that for the variable sex, the stratum-specific causal odds ratios are equal between men and women, but the overall causal odds ratio is different from the stratum-specific odds ratios. Moreover, since any weighted average of the stratum-specific odds ratios is 3, there does not exist any set of weights that makes the odds ratio collapsible over sex. This counterexample shows that no generally applicable weights such as those

Table 1 Conditional and marginal odds ratios

	Counterfactual risk of outcome (placebo)	Counterfactual risk of outcome (treatment)	Causal odds ratio
Men (25%)	0.5	0.75	3
Women (75%)	0.25	0.5	3
Overall	0.3125	0.5625	2.83

Huitfeldt *et al. Emerg Themes Epidemiol*   (2019) 16:1

Page 4 of 5

for the risk difference and the risk ratio can be provided for the odds ratio.

Identification of the weights

If the investigator intends to report an average of the stratum-specific effects as an estimate of the marginal effect, it is necessary to know not only that the effect is collapsible in principle, but also to construct appropriate weights, identify them from the data and apply them in the analysis. The weights for the causal risk ratio $RR(-)$, $\Pr(V = v|Y^{a=0} = 1)$, have a counterfactual variable in the conditioning event, and may not be identified from the data. However, we proceed to show that the weights are identified in the absence of unmeasured confounding, i.e if $Y^{a=0} \perp\!\!\!\perp A | V$

Proof

$$
\begin{aligned}
&\Pr(V = v|Y^{a=0} = 1) \\
&= \frac{\Pr(Y^{a=0} = 1|V = v) \times \Pr(V = v)}{\Pr(Y^{a=0} = 1)} \quad \text{Bayes Theorem} \\
&= \frac{\Pr(Y^{a=0} = 1|A = 0, V = v) \times \Pr(V = v)}{\Pr(Y^{a=0} = 1)} \quad \text{Partial exchangeability} \\
&= \frac{\Pr(Y = 1|A = 0, V = v) \times \Pr(V = v)}{\Pr(Y^{a=0} = 1)} \quad \text{Consistency} \qquad \Box
\end{aligned}
\tag{3}
$$

$\Pr(Y^{a=0} = 1)$ is constant over v and can therefore be factored out of the weights. In the absence of unmeasured confounders, the weights $\Pr(V = v|Y^{a=0} = 1)$ are therefore equivalent to Miettinen's weights $\Pr(V = v) \times \Pr(Y = 1|A = 0, V = v)$ as discussed earlier. These weights can be used both to control for confounding due to V, as suggested in a simple example in Table 2, or to standardize the results to a new target population (by taking the risk ratio from the study population, and information on $\Pr(V = v)$ and $P(Y = 1|A = 0, V = v)$ from the target population).

An alternative identification of the weights can be used if standardizing experimental results to a population in which everyone is unexposed. In such situations, $Y^{a=0} = Y$ in all individuals by consistency, and the weights in the target population are identified as $\Pr(V = v|Y = 1)$.

Discussion

We have reviewed well-established results on the collapsibility of measures of association, and shown corresponding results for measures of causal effect. With these causal effect measures, one can disentangle the components of non-collapsibility that are due to the mathematical properties of the effect measure from the components that are due to structural bias and the probabilistic

Table 2 Standardization of the risk ratio

	Men	Women
Exposed cases (n)	100	200
Exposed non-cases (n)	400	300
Unexposed cases (n)	400	300
Unexposed non-cases (n)	100	200
Risk Ratio	0.25	0.6667
$\Pr(V = v)$	0.5	0.5
$P(Y = 1|A = 0, V = v)$	0.8	0.6
Weight	0.4	0.3

Table shows a simple toy example of 2000 subjects, of whom 1000 are exposed to a drug. Here, the covariate V denotes gender, and we assume no unmeasured confounders. The marginal causal risk ratio can be calculated as $\frac{0.25 \times 0.4 + 0.6667 \times 0.3}{0.4 + 0.3} = 0.428$. Note that the weights are easy to obtain: $\Pr(V = v)$ is just the probability of belonging to the gender, and $P(Y = 1|A = 0, V = v)$ is the probability of the outcome among the unexposed in that gender

structure of the dataset. We have provided new, simple weights for the causal risk ratio, which guarantee collapsibility over arbitrary baseline covariates, and we showed that such weights do not exist for the causal odds ratio.

Our weights for the causal risk ratio $RR(-)$ are equivalent to the weights previously discussed by Miettinen when there is no unmeasured confounding; in other words, in all situations where standardizing over V provides a valid estimate of the causal effect. Our formulation allows a simpler presentation of the weights and of the proofs. Furthermore, our formulation highlights pitfalls of using weighted averages: When conditioning on V, the correct weights cannot be estimated from the data if unmeasured confounding is present. In such scenarios, using erroneous weights can possibly amplify the bias that is caused by unmeasured confounding within the strata.

Finally, we note that in many cases we do not need to use results on collapsibility al all, because we can standardize the distributions of $Y^{a=1}$ and $Y^{a=0}$ separately. One way to do this is by reporting the overall marginal risk ratio $RR(-)$ as

$$
\frac{\sum_v \Pr(Y^{a=1} = 1|V = v) \times \Pr(V = v)}{\sum_v \Pr(Y^{a=0} = 1|V = v) \times \Pr(V = v)}
$$

Huitfeldt *et al. Emerg Themes Epidemiol*   (2019) 16:1

Page 5 of 5

Since this simple procedure does not depend on collapsibility, analogous procedures are valid for any effect measure, including the odds ratio.

Author details
[1] Norwegian Institute of Public Health, Oslo, Norway. [2] Department of Biostatistics, University of Oslo, Oslo, Norway. [3] Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, Japan. [4] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, USA.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Whittemore AS. Collapsibility of multidimensional contingency tables. J R Stat Soc Ser B. 1978;40(3).
2. Greenland S, Pearl J, Robins JM. Confounding and collapsibility in causal inference. Stat Sci. 1999;14(1):29–46.
3. Hernán MA, Robins JM. Causal inference. Boca Raton: CRC Press; 2018 (forthcoming).
4. Pearl J. Causality: models, reasoning and inference. 2nd ed. Cambridge: Cambridge University Press; 2009.
5. Newman SC. Biostatistical methods in epidemiology. Hoboken: Wiley; 2001.
6. Greenland S, Pearl J. Adjustments and their consequences-collapsibility analysis using graphical models. Int Stat Rev. 2011;79(3):401–26.
7. Miettinen OS. Standardization of risk ratios. Am J Epidemiol. 1972;96(6):383–8.