Journal of the Royal Statistical Society Series A: Statistics in Society, 2024, **187**, 636–655
https://doi.org/10.1093/jrsssa/qnae074

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society

Original Article

'Does God toss logistic coins?' and other questions that motivate regression by composition

Rhian M. Daniel[1], Daniel M. Farewell[1] and Anders Huitfeldt[2]

[1]Division of Population Medicine, School of Medicine, Cardiff University, Cardiff, UK
[2]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway

Address for correspondence: Rhian M. Daniel, Division of Population Medicine, School of Medicine, Cardiff University, Cardiff, CF14 4YS, UK. Email: danielr8@cardiff.ac.uk

Abstract

Regression by composition is a new and flexible toolkit for building and understanding statistical models. Focusing here on regression models for a binary outcome conditional on a binary treatment and other covariates, we motivate the need for regression by composition. We do this first by exhibiting—using L'Abbé plots—the families of relationships between untreated and treated conditional outcome risks that emerge from generalized linear models for many different link functions. These are compared with the relationships (between untreated and treated risks) that arise from mechanistic sufficient component cause models, which are first principles causal models for binary outcomes. By considering mechanistic models that allow for non-monotone causal effects and by allowing sufficient causes to be associated, we expand upon similar discussions in the recent literature. We discuss conditions under which commonly used statistical models for binary data, such as logistic regression, arise from mechanistic models where the sufficient causes are associated in a particular way, as well as other situations in which the statistical models arising do not correspond to a generalized linear model but can be naturally expressed as a regression by composition model.

Keywords: binary regression, causal models, generalized linear models, mechanistic models, regression by composition

1 Introduction

The spirited debate among statisticians and other users of statistics over how to analyse binary outcome data—and whether to report results as estimated odds ratios, risk ratios, risk differences, or other measures—has been going on for decades (Altman et al., 1998; Choi, 2016; Cook, 2002; Doi et al., 2022; Huitfeldt, 2017, 2023; Huitfeldt et al., 2021, 2018; Permutt, 2020; Rohrer & Arslan, 2021; Rudolph et al., 2023; Senn, 2011; Sheps, 1958; Sonis, 2018; Van Der Laan et al., 2007; White et al., 2021; Xiao et al., 2022; Zipkin et al., 2014). Views on this controversy focus on many different aspects, including interpretability, transportability, parsimony, and the potential for robust estimation of marginal estimands. Contributions come from various angles, including philosophical, statistical, and empirical viewpoints.

Recent contributions by Huitfeldt et al. (2021) and Huitfeldt (2023) have approached the question of transportability from first principles, using sufficient component cause ('causal pie') models (Rothman, 1976) for the heterogeneity of exposure effects on binary outcomes. They advocate for

Received: June 12, 2024. **Revised:** July 1, 2024. **Accepted:** July 4, 2024

an effect measure known as the *switch relative risk* (Van Der Laan et al., 2007) in a wide (but by no means exhaustive) range of settings.

In Section 3, we take Huitfeldt's argument further by allowing the sufficient causes to be associated and by relaxing the assumption that treatment effects are monotonic (i.e. that all individuals affected by the treatment are affected in the same direction). The former relaxation leads us to uncover a form of association between sufficient causes that would make logistic regression the correct model and the odds ratio a transportable measure of effect: the conditions under which *God tosses logistic coins*. We describe similar conditions for other link functions.

We also exhibit novel families of effects that do not align with a generalized linear model (GLM) for any link function. In Section 4, we briefly describe a general modelling framework, regression by composition, that admits such effects.

We conclude in Section 5 with a discussion that briefly touches on how these ideas extend beyond binary regression, and begin in Section 2 with a view of GLMs as relevant to the rest of the paper.

2 Risk transformations implied by generalized linear models

Consider a binary outcome Y, and a parametric model for the conditional distribution of Y given binary exposure or treatment X and baseline covariates $\mathbf{C}$. A consequence of such a model for the conditional dependence on X can be viewed as the implied transformation that maps the conditional outcome risk (given $\mathbf{C}$) in the untreated to the conditional risk in the treated. Let

$$P_x := \Pr(Y = 1 \mid \mathbf{C}, X = x),$$

where the choice of a capital letter for P_x is made to emphasize that P_x is a random variable through its dependence on $\mathbf{C}$, and the bold-face letter $\mathbf{C}$ reminds us that the covariates will typically be multidimensional.

The aforementioned transformation maps P_0 to P_1. A choice of parametric model confers a particular *shape* on this transformation, with the *magnitude* of the transformation typically estimated from data. For example, a GLM (McCullagh & Nelder, 1989) with link function g is of the form

$$g(P_x) = \alpha(\mathbf{C}) + \beta(\mathbf{C})\, x \tag{1}$$

where, often, $\alpha(\mathbf{C})$ is a linear combination of the variables in $\mathbf{C}$, with unknown coefficients, and $\beta(\mathbf{C})$ is either a function of only a few variables in $\mathbf{C}$ (the few by which *effect measure modification* (Greenland, 2014) are being considered on this particular scale), or is assumed to be a constant, β.

Model (1) coincides with the following transformation of conditional risks, p:

$$f(p; v) = g^{-1}\{g(p) + v\}$$

with $P_1 = f(P_0; \beta(\mathbf{C}))$ under this model.

For example, for logistic regression (Cox, 1958), g is logit ($g(p) = \log\{p/(1-p)\}$) and we have:

$$f(p; v) = \frac{pe^v}{1 - p + pe^v}. \tag{2}$$

The first panel in Figure 1 shows this transformation for several different values of a constant conditional log odds ratio $v = \beta(\mathbf{C}) = \beta$ in what is known as a L'Abbé plot (L'Abbé et al., 1987). For our purposes, a L'Abbé plot shows all pairs (P_0, P_1) of untreated and treated conditional risks across the range of possible values of $\mathbf{C}$.

Model (1) when $g = \log$ leads instead to a proportional risks model, with the following transformation:

$$f(p; v) = pe^v.$$

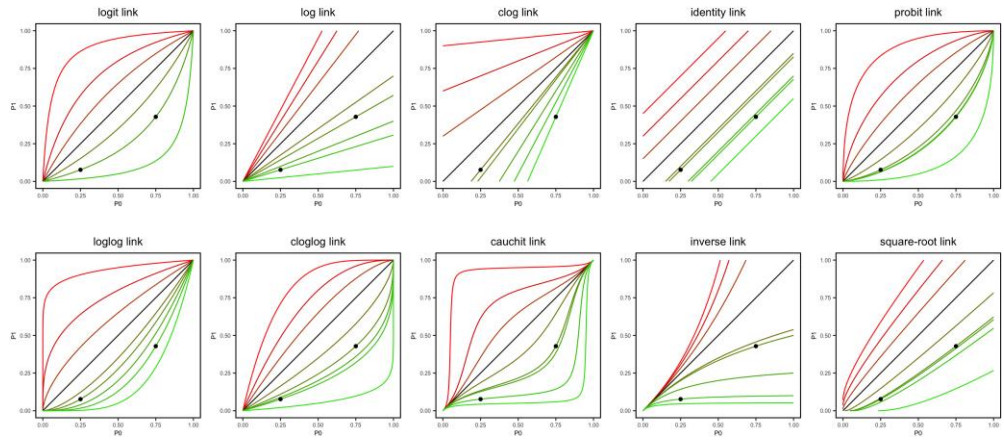

Figure 1. L'Abbé plots for GLMs with 10 different link functions. For each link function, the L'Abbé plot is given for many different values of $\beta(\mathbf{C}) = \beta$, not dependent on $\mathbf{C}$. The increasingly bright red lines correspond to risk-raising choices of β of increasing magnitude, and the increasingly bright green lines correspond to risk-lowering choices of β of increasing magnitude. For each panel, seven values of β were initially chosen for illustration, such that, for each panel, an untreated conditional risk of 0.5 would be transformed to treated risks of each of 0.95, 0.80, 0.65, 0.50 (the identity transformation, shown in black), 0.35, 0.20, and 0.05, respectively. The two points denote the same two pairs of untreated and treated risks in each panel, namely (0.25, 0.077) and (0.75, 0.429). The risks 0.077 and 0.429 are the transformed values of 0.25 and 0.75, respectively, under logistic regression (panel 1) for the sixth choice of β (namely that which transforms 0.5 to 0.2). By construction, therefore, the points lie on the same curve for the logit link, but on different curves for all the other link functions, and two additional green curves, one through each of the points, have been added to each panel to illustrate this. We have used the conventional definition of each link function, so that $g = \Phi$ for the probit link, where Φ is the cumulative distribution function of a standard normal variate, and $g(p)$ is respectively, $\log(-\log(p))$, $\log(-\log(1-p))$, $\tan(\pi(p-0.5))$, p^{-1}, and $\sqrt{p}$ for the loglog, cloglog, Cauchit, inverse and square-root link.

If g is the complementary log link ($g(p) = \log(1-p)$) this leads to a proportional survival probability model:

$$f(p; v) = 1 - (1-p)e^{v}.$$

And for an identity link—i.e. the linear probability model (Battey et al., 2019)—the transformation is simply:

$$f(p; v) = p + v.$$

The corresponding L'Abbé plots for a range of $v = \beta$ (not dependent on $\mathbf{C}$) are shown in panels two, three, and four of Figure 1. In these panels, the different values of β giving rise to the different lines, are respectively interpreted as log risk ratios, log survival probability ratios, and risk differences. The remaining six panels show similar plots for six further GLM link functions, namely probit, log–log, complementary log–log, Cauchit, inverse, and square-root.

Remarks

Many features of binary regression can be understood from Figure 1. First, some models (logistic regression is an example, as are GLMs with probit, log–log, complementary log–log and Cauchit links) are *closed* in the sense that any probability P_0 (in [0, 1]) is mapped to another probability P_1 (in [0, 1]) no matter the value of β. This is not the case for the other five models shown. For example, for the proportional risks model (log link), some untreated risks are transformed to risks greater than 1 for $\beta > 0$ and for the proportional survival probabilities model (clog link), some untreated risks are transformed to negative risks for $\beta > 0$. Second, some transformations are affine (straight lines), whereas most (such as logistic regression) are not. Affinity of the transformation leads to collapsibility

of the associated effect measure (Daniel et al., 2021; Farewell et al., under revision). Third, some transformation families have no fixed points (e.g. identity and square-root links), others have one fixed point (for the log and inverse links, an untreated risk of 0 is always mapped to a treated risk of 0; likewise an untreated risk of 1 is unaltered by a GLM with the clog link), and others (such as GLMs with logit, probit, loglog, cloglog, and Cauchit links) have two fixed points (0 and 1).

Another aspect illustrated by the L'Abbé plots is that (as long as P_0 varies with $\mathbf{C}$ and $P_1 \not\equiv P_0$), at most one model can be correct with a constant β. This is equivalent to the well-known observation that if 'no effect modification' holds on, say, the proportional odds scale, then it will not hold on any other scale (Miettinen, 1974). This is seen from the L'Abbé plots by observing that if two different points lie on the same line for one of the link functions (e.g. those shown in the first panel) they will be on different lines for each of the other link functions.

Finally, the plots illustrate the importance of model choice for extrapolation, and hence for considerations of generalizability and the transportability of effects. For example, at low risks, panels 1 and 2 contain similar transformations; at high risks, panels 1 and 3 contain similar transformations; and at medium risks, panels 1 and 4 contain similar transformations. Thus, if a study contains only healthy participants at relatively low risk of the outcome, a proportional odds and proportional risks model, each assuming a constant odds or risk ratio, respectively, would lead to similar fits to the observed data, but would imply very different predictions for the transformed risks of less healthy patients, outside the study, at higher risk of the outcome. It is common for the published results of clinical trials conducted on highly selected patient populations to be used by clinicians to inform treatment choice in practice for patients with a wider variety of untreated risks. Whether the results are published as an estimated odds ratio, risk ratio, survival probability ratio, risk difference, or any other choice of measure, is therefore of great importance, and the data—especially when the trial participants are relatively homogeneous—are not always useful in informing this choice.

3 Fundamental causal models and the risk transformations they imply

In this section, we discuss situations in which biological or other subject-matter knowledge about how outcomes arise may guide the choice of model, starting with two examples from Huitfeldt et al. (2021).

3.1 Example 1: penicillin and the prevention of rheumatic fever

Consider an RCT comparing treatment (X) with penicillin vs. placebo for the prevention of acute rheumatic fever (Y) in UK children with suspected group A streptococcus pharyngitis (strep throat) (Huitfeldt et al., 2021). Let D be 1 if all three of the following occur: (1) the child takes the assigned medication as directed, (2) the child's particular infection strain is susceptible to the action of penicillin, and (3) the child can metabolize penicillin normally at the assigned dose; let D be 0 otherwise. Let W be 1 if the child's immune system would on its own clear the infection irrespective of antibiotic treatment, and thus a prolonged infection leading to rheumatic fever would be avoided, and let W be 0 otherwise. Both D and W are typically unobserved, but at least in principle it could be argued that

$$\neg Y = W \vee (X \wedge D),$$

where $\neg Y = 1 - Y$, and where $\vee$ and $\wedge$ denote maximum and minimum, corresponding to OR and AND for binary variables, respectively. That is, rheumatic fever would be avoided ($Y = 0$) either if the child's own immune system would clear the infection ($W = 1$), or if the child is assigned to the penicillin arm, takes it as directed, metabolizes it normally and it is effective against the child's infection strain ($X = D = 1$), or both ($W = X = D = 1$). This implies that

$$Y = \neg W \wedge (\neg X \vee \neg D),$$

which leads (by writing down the probability of each side being equal to 1 given **C** and given $X = 1$ and 0, in turn) to

$$P_1 = \Pr(W = 0 \mid X = 1, \mathbf{C})\Pr(D = 0 \mid W = 0, X = 1, \mathbf{C})$$

and

$$P_0 = \Pr(W = 0 \mid X = 0, \mathbf{C}).$$

This is an example of a sufficient component cause model as introduced by Rothman (1976); see also Chapter 5 of Hernán and Robins (2020).

An aside: a brief introduction to Rothman's sufficient component cause models

A sufficient component cause model assumes the existence of a set of sufficient binary causes that bring about a binary outcome, where each sufficient cause comprises a set of necessary binary components. For example, suppose the binary outcome Y can arise in five ways: either (1) U_1 happens (that is, the binary U_1 takes value 1), or (2) U_2, U_3 and U_4 all happen, or (3) X, A_1 and A_2 all happen, or (4) X, A_2, A_3 and A_4 all happen, or (5) X does not happen, and B_1 and B_2 both happen. If any two or more of (1)–(5) happen, then Y still happens. This is an example of a sufficient component cause model for Y, with:

$$Y = U_1 \vee (U_2 \wedge U_3 \wedge U_4) \vee (X \wedge A_1 \wedge A_2) \vee (X \wedge A_2 \wedge A_3 \wedge A_4) \vee (\neg X \wedge B_1 \wedge B_2). \quad (3)$$

This model for Y is depicted in the top row of Figure 2, with shaded examples for different hypothetical individuals depicted in subsequent rows. The model states that the outcome Y happens if and only if at least one of the sufficient causes (depicted as circles, or 'pies') happens, and a sufficient cause happens only if all of its components (or 'slices') happen. When studying the effect of X on Y, there is particular interest in pies that contain X (or $\neg X$), since the slices that accompany X (or $\neg X$) in such pies explain why X causes (or prevents) Y for some individuals but not others. Moreover, if the same sufficient component cause model holds for two different populations, but the probability of the U-, A-, and B-slices differ between the two populations, then this explains the heterogeneity of the treatment effect between the two populations.

In some settings, it may be more natural to think of sufficient component cause models for avoiding the outcome, i.e. for $\neg Y$ as opposed to Y (although one can always be constructed from the other); this is why we include 'For Y' in Figure 2. Although sufficient component cause models can be extremely complex (many pies, each containing many slices), they can be simplified by defining summary binary variables as follows. Let U be 1 if at least one of the U-pies has all its slices filled (and $U = 0$ otherwise), let A be 1 if at least one of the X-pies has all of its non-X slices filled ($A = 0$ otherwise) and let B be 1 if at least one of the $\neg X$-pies has all of its non-$\neg X$ slices filled. In the above example, U, A, and B would be defined as follows:

$$\begin{aligned} U &= U_1 \vee (U_2 \wedge U_3 \wedge U_4) \\ A &= A_2 \wedge (A_1 \vee (A_3 \wedge A_4)) \\ B &= B_1 \wedge B_2. \end{aligned}$$

Thus if (3) holds, so does:

$$Y = U \vee (X \wedge A) \vee (\neg X \wedge B),$$

which is depicted in Figure 3a. Similarly, any sufficient component cause model for $\neg Y$ can be summarized as:

$$\neg Y = W \vee (X \wedge D) \vee (\neg X \wedge E)$$

and is depicted in Figure 3b.

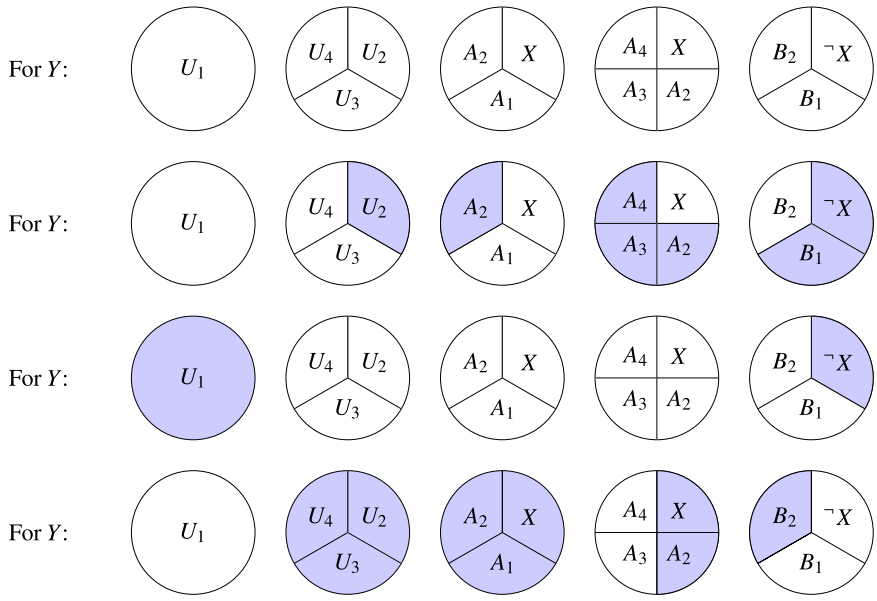

Figure 2. An example of a sufficient component cause model for Y (top row). The subsequent rows show the same example specifically for different individuals where the shading indicates that the relevant binary component is 1, and unshaded components are 0. The untreated individual in the second row does not experience the outcome Y (no pie has all slices shaded for them), whereas the untreated and treated individuals in rows 3 and 4, respectively, both experience the outcome (there is one fully-shaded pie in row 3 and two fully-shaded pies in row 4).

Returning now to Example 1, this is an example of a sufficient component cause model for $\neg Y$ with two sufficient causes: W and $X \wedge D$. In this example, there are no sufficient causes that contain $\neg X$ as a component, which is why E is not defined. We return to this in Section 3.3.

Figure 4d shows the random variables corresponding to Figure 3b (with the bold red arrows signifying a determinism from (X, D, W) to Y) together with the baseline covariates $\mathbf{C}$ under a particular set of additional assumptions, namely that (X, D, W) are independent, and that (X, D) are independent of $\mathbf{C}$ so that these covariates affect Y only via their effect on W. Under the assumptions in Figure 4d, $\Pr(W = 0 \mid X = 1, \mathbf{C}) = \Pr(W = 0 \mid X = 0, \mathbf{C}) = \Pr(W = 0 \mid \mathbf{C})$ and $\Pr(D = 0 \mid W = 0, X = 1, \mathbf{C}) = \Pr(D = 0)$. Thus the risk ratio P_1/P_0 is equal to a constant, $\Pr(D = 0)$, suggesting that a GLM with a log link would be an appropriate choice, at least from the point of view of the dependence on X. We discuss relaxations of these assumptions in Section 3.3 and return to the issue of model choice for the $\mathbf{C}$-dependence in Section 4.

3.2 Example 2: penicillin and anaphylaxis

Again taken from Huitfeldt et al. (2021), consider the same example but with a different outcome Y, namely the occurrence of anaphylactic shock on the same day as treatment initiation. Let A be 1 if the assigned treatment is taken and the child has an underlying severe allergy to penicillin (with $A = 0$ otherwise), and let U be 1 if both (1) the child has an underlying severe allergy to some other allergen, e.g. peanuts or bee sting, and that (2) the child is exposed to such an allergen on the same day as treatment initiation (with $U = 0$ otherwise). Then it could be argued, again with U and A potentially unobserved, that:

$$Y = U \vee (X \wedge A)$$

that is, that anaphylactic shock occurs on the day of treatment initiation either if the child is assigned to the penicillin arm, takes it, and has a severe reaction to it, or the child has a severe allergic reaction to a different allergen on the same day, or both. This is an example of a sufficient component cause model for Y as shown in Figure 3a, and leads to

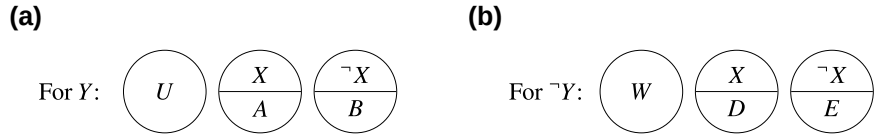

Figure 3. Summary versions of Rothman's sufficient component causes for Y (in panel a) and for $\neg Y$ (in b).

$$1 - P_1 = \Pr(U = 0 \mid X = 1, \mathbf{C})\Pr(A = 0 \mid U = 0, X = 1, \mathbf{C})$$

and

$$1 - P_0 = \Pr(U = 0 \mid X = 0, \mathbf{C}).$$

Under the assumptions shown in Figure 4a, $\Pr(U = 0 \mid X = 1, \mathbf{C}) = \Pr(U = 0 \mid X = 0, \mathbf{C}) = \Pr(U = 0 \mid \mathbf{C})$, and $\Pr(A = 0 \mid U = 0, X = 1, \mathbf{C}) = \Pr(A = 0)$. Thus the ratio of survival probabilities $(1 - P_1)/(1 - P_0)$ is equal to a constant, $\Pr(A = 0)$, suggesting that a GLM with a complementary log link would be an appropriate choice, at least for the X-dependence.

As with Example 1, the plausibility of these assumptions are questionable: there are likely unmeasured (**F**) as well as measured (**C**) reasons why an individual prone to a penicillin allergy would also be prone to other allergies. We consider relaxations to these assumptions below.

3.3 Remarks on Examples 1 and 2

Monotonicity and the absence of B *and* E

In the two examples discussed thus far, there were no sufficient causes with $\neg X$ as a component. Thus, there was no E in Example 1, and no B in Example 2. Penicillin was thought to be sufficient to prevent rheumatic fever in some children and sufficient to cause anaphylactic shock in some children, but the absence of penicillin was not thought to be sufficient to prevent rheumatic fever nor to cause anaphylactic shock in any children. In both examples, the action of pencillin was taken to be *monotonic*. Monotonicity is a counterfactual concept and states that if a treatment is protective for at least one individual, then for all other individuals it must either have no effect or also be protective; it cannot be harmful for any individual. Similarly, if the treatment is harmful for at least one individual, then it cannot be protective for anyone (Angrist & Pischke, 2008).

Although probably a reasonable assumption in these examples, we consider in Section 3.6 settings where monotonicity is not assumed.

The choice between Figure 3a and b and the switch relative risk

If a sufficient component cause model for Y (Figure 3a) exists, then so does a sufficient component cause model for $\neg Y$ (Figure 3b), e.g. by choosing $W = \neg U \wedge \neg A \wedge \neg B$, $D = \neg U \wedge \neg A$, and $E = \neg U \wedge \neg B$. Due to their nested nature, W and D cannot in general be independent when defined in this way, nor can W and E. However, sufficient component cause models are not unique, and so this does not rule out there being an alternative such model for $\neg Y$ for which the alternatively defined W and D are independent. Huitfeldt et al. (2021) argue that for a reasonably large class of biological examples (that by no means covers all settings), a sufficient component cause model that satisfies the independence assumptions (depicted in Figure 4a and d) is more plausible for models for Y (Figure 3a) when X raises the risk of Y and for models for $\neg Y$ (Figure 3b) when X lowers the risk of Y. This leads them to favour an effect measure known as the *switch relative risk* (Van Der Laan et al., 2007), that chooses the survival probability ratio for harmful treatments and the risk ratio for protective treatments. We return to this effect measure, which also has the two desirable properties of being both closed and collapsible, in Section 3.6.

Relaxing the independence assumptions

In certain settings, some of the independence assumptions in Figure 4a and d can be justified, e.g. in a randomized trial, the absence of any arrow into X is justified. Others, however, such as the

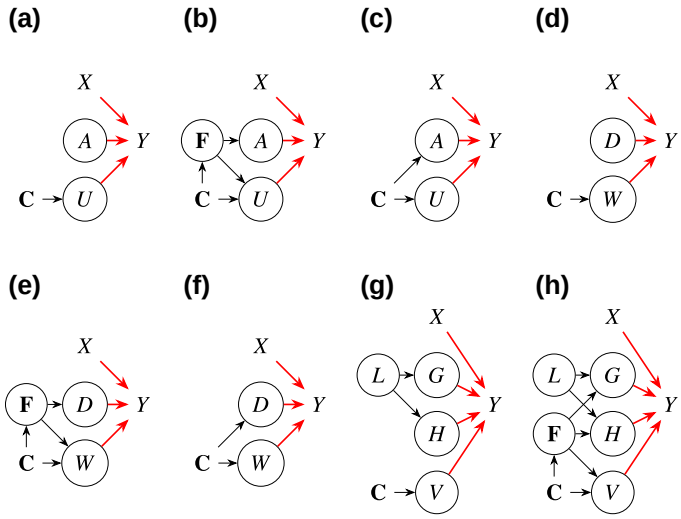

Figure 4. Causal directed acyclic graphs (DAGs) depicting how the component causes from the mechanistic models discussed in Section 3 may depend on covariates **C**. The DAGs in (a), (b), and (c) refer to Example 2 (anaphylaxis); the DAGs in (d), (e), and (f) refer to Example 1 (rheumatic fever); and the DAGs in (g) and (h) refer to Example 3 (COVID-19). Bold (red) arrows are used to depict deterministic relationships and circled variables are typically unobserved.

independence of A and U in Figure 4a or of D and W in Figure 4d, may be difficult to justify, even when the most amenable choice between the two options (i.e. whether to consider sufficient component causes of Y or $\neg Y$) has been made. Relaxing these independence assumptions gives rise to non-affine L'Abbé plots, as we see in the next sections.

Non-affine risk transformations

Suppose there are common causes **F** of D and W, as shown in Figure 4e for Example 1. Note that we allow **F** to depend on **C** but we suppose that D is conditionally independent of **C** given **F**. We also suppose that **F** is unmeasured, so that its inclusion in **C** is not possible. For simplicity of notation, we will also take **F** to be discrete, but extensions to continuous **F** (replacing sums by integrals and mass functions by density functions) are natural. It is straightforward to show that under the assumptions in Figure 4e,

$$P_1 = \sum_{\mathbf{f}} \Pr(D = 0 \mid \mathbf{F} = \mathbf{f}) \Pr(W = 0 \mid \mathbf{F} = \mathbf{f}, \mathbf{C}) \Pr(\mathbf{F} = \mathbf{f} \mid \mathbf{C})$$
$$P_0 = \sum_{\mathbf{f}} \Pr(W = 0 \mid \mathbf{F} = \mathbf{f}, \mathbf{C}) \Pr(\mathbf{F} = \mathbf{f} \mid \mathbf{C}).$$

When $\Pr(D = 0 \mid \mathbf{F} = \mathbf{f})$ is the same for all $\mathbf{f}$, this is the setting considered in Example 1, leading to the constant risk ratio $\Pr(D = 0)$. But whenever $\Pr(D = 0 \mid \mathbf{F} = \mathbf{f})$ depends on $\mathbf{f}$, the ratio P_1/P_0 depends, in general, on **C**, leading to a non-affine L'Abbé plot. Note that $P_0 = 0$ only if $\Pr(W = 0 \mid \mathbf{F} = \mathbf{f}, \mathbf{C}) = 0$ for all $\mathbf{f}$, which implies that $P_1 = 0$ whenever $P_0 = 0$, thus $p = 0$ is a fixed point of this transformation, whether or not D depends on **F**. If $P_0 = 1$ then $\Pr(W = 0 \mid \mathbf{F} = \mathbf{f}, \mathbf{C}) = 1$ for all $\mathbf{f}$, which implies that $P_1 = \Pr(D = 0)$ whenever $P_0 = 1$, again, whether or not D depends on **F**.

Figure 5 shows examples where $\mathbf{F} = F$ is a single binary variable, independent of **C**, for different combinations of the marginal probabilities $\Pr(F = 1)$ and $\Pr(D = 1)$. In each panel, the transformation from P_0 to P_1 is shown for different strengths (and directions) of association between D and W achieved by altering the dependencies of D and W on F. The straight line in the middle of each cluster of curves represents the situation where D and W are independent. The convex curves below this straight line represent situations with an increasingly strong negative association between D and W (induced by their mutual dependence on F), and the concave curves above the straight

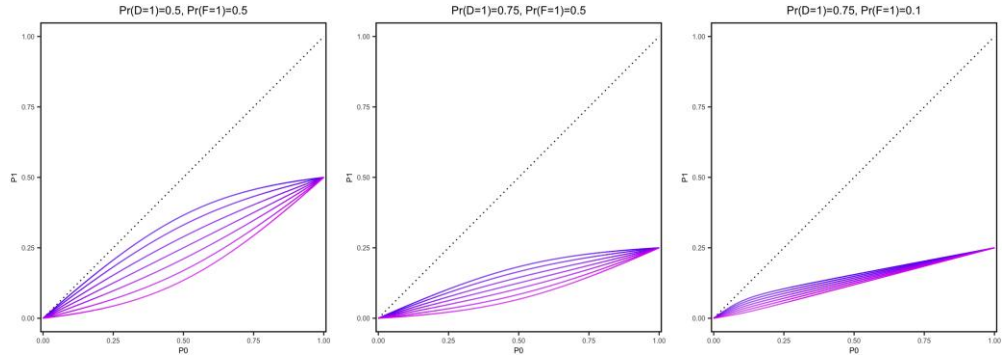

Figure 5. L'Abbé plots for Example 1 under the DAG shown in Figure 4e, varying the strength (and direction) of the association between D and W by introducing a common binary cause F. The panels illustrate examples with different values of $\Pr(F = 1)$ and $\Pr(D = 1)$, as indicated.

line represent situations of increasingly strong positive association. It may be the case, for example, that frailer children ($F = 1$) are less likely to be able to clear the infection without treatment and also that the treatment is less likely to work for them, in which case D and W would be positively associated via F. However, if frailer children were more likely to take the treatment as directed (due to closer management), then a negative association would be expected. The third panel in Figure 5 represents a somewhat plausible setting with 10% frailty and overall a proportion of 0.75 of children who would take the treatment as directed, would metabolize it normally and whose strain of infection would be susceptible to penicillin.

If F were measured, then the analysis could be stratified on F to obtain f-specific risk ratios $\Pr(D = 0 \mid F = f)$ for $f = 0, 1$, which would not depend on $\mathbf{C}$. The curves in Figure 5 arise from taking weighted averages (over the distribution of F) of the f-specific straight lines.

This is made clearer in Figure 6 where a number of marginal transformations (corresponding to different associations between F and W and between F and D) are shown on separate axes, together with the two stratum-specific risk ratio transformations and the two stratum-specific distributions of P_0, so that the nature and consequences of the averaging can be understood.

Similar non-affine L'Abbé plots can also be exhibited from Example 2 with associations between U and A as indicated in Figure 4b. Analogous to Figure 5, these transformations have a fixed point at $(P_0, P_1) = (1, 1)$ and $P_0 = 0$ is mapped to $P_1 = \Pr(A = 1)$ irrespective of the association between U and A. Such plots are included in the online supplementary material.

3.4 Logistic regression

A natural question when inspecting the non-affine L'Abbé plots discussed above (and the corresponding plots above the identity line included in the online supplementary material) is whether any association between W and D due to $\mathbf{F}$ (or between U and A due to $\mathbf{F}$ for the mechanistic models of the other type) could give rise to the risk transformations implied by logistic regression (first panel, Figure 1). The immediate answer is 'no', and the fact that all such plots pass through either $(P_0, P_1) = (1, \Pr(D = 0))$ or $(P_0, P_1) = (0, \Pr(A = 1))$, whereas transformations arising from logistic regression pass through both (0, 0) and (1, 1), is one way of demonstrating this. In fact, as we show in the Appendix, the family of non-affine plots discussed above does not arise from a GLM for any link function. This claim needs to made carefully, since we are considering one-parameter GLM models, and asking if they give rise to transformation families (such as those illustrated in Figure 5) that are clearly characterized by at least two parameters. A fairer question, therefore, is 'is there a GLM link function that gives rise to a subset of the types of transformations illustrated in Figure 5, where this subset exhibits some of the interesting features seen in Figure 5?', where what constitutes an 'interesting feature' is of course subjective. In the Appendix, we carefully describe such a claim, based on the existence of both everywhere-convex and everywhere-concave risk transformations on the same side of the identity transformation, and show that this feature cannot arise from a GLM for any link function.

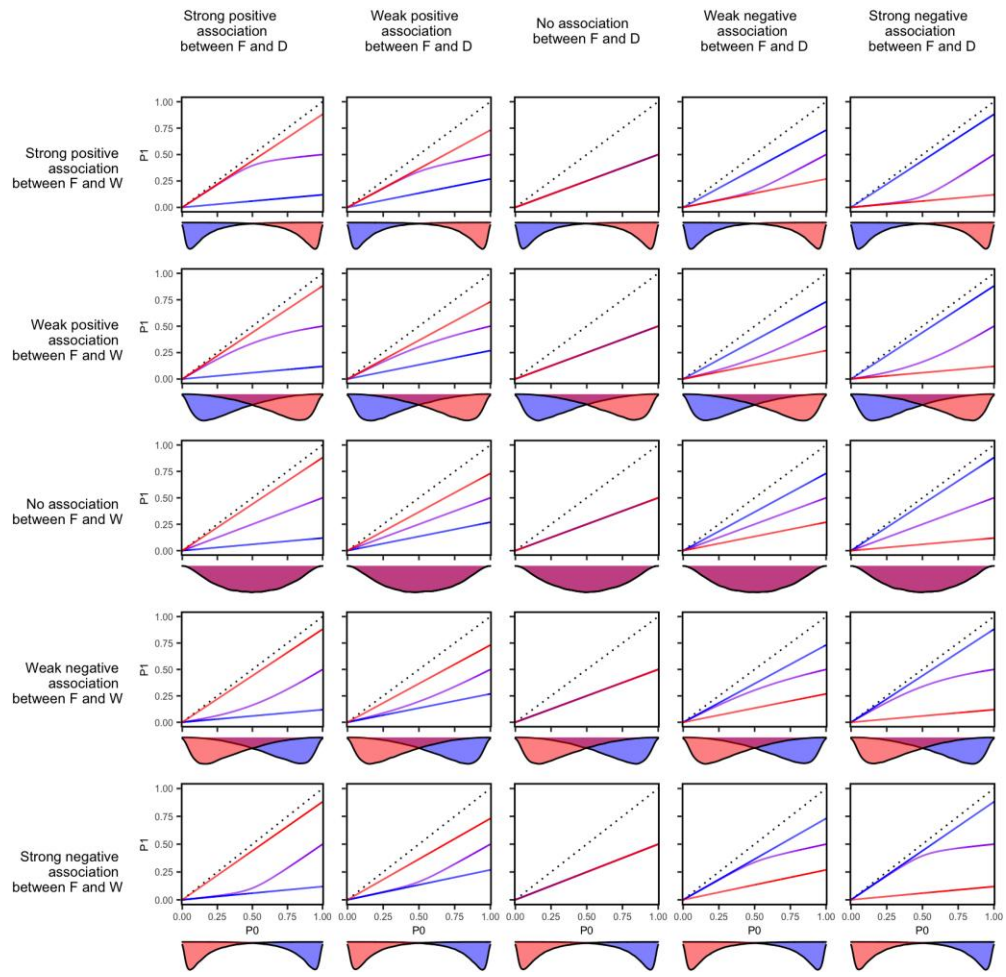

Figure 6. An illustration of how the curves in the left-most panel of Figure 5 can arise. The blue and red lines represent risk ratio transformations for $F = 1$ and $F = 0$, respectively. Density plots under each P_0-axis show the distribution of P_0 for $F = 1$ (blue) and $F = 0$ (red). The purple line shows the marginal risk transformation from appropriately averaging the two stratum-specific transformations. Moving down the rows and right through the columns, respectively, shows decreasing associations between F and W and between F and D from strong positive to strong negative. The nine panels on the central 'cross' depict situations where W and D are independent. The top left and bottom right quadrants show W and D positively associated, while the top right and bottom left quadrants show them negatively associated.

The non-affine transformations considered thus far arose from an unmeasured common cause $\mathbf{F}$ of W and D (Figure 4e). If instead we remove $\mathbf{F}$ and include an arrow from $\mathbf{C}$ to D directly (so that the association between W and D is due to the covariates conditioned upon in the analysis), as in Figure 4f, then it *is* possible to recover logistic regression by carefully choosing the relationship between $\Pr(D = 0 \mid \mathbf{C})$ and $\Pr(W = 0 \mid \mathbf{C})$. This can similarly be achieved by the sufficient component cause model for Y by including an arrow directly from $\mathbf{C}$ to A (Figure 4c) and carefully choosing the relationship between $\Pr(A = 0 \mid \mathbf{C})$ and $\Pr(U = 0 \mid \mathbf{C})$ (see online supplementary material).

Under the directed acyclic graph (DAG) in Figure 4f,

$$P_1 = \Pr(D = 0 \mid \mathbf{C})P_0 = \Pr(D = 0 \mid \mathbf{C})\Pr(W = 0 \mid \mathbf{C})$$

and by comparing this with (2), we see that if

$$\Pr(D = 0 \mid \mathbf{C}) = \frac{e^{\beta}}{1 - \Pr(W = 0 \mid \mathbf{C})(1 - e^{\beta})} \tag{4}$$

then the transformation from P_0 to P_1 corresponds to logistic regression with a constant conditional (on $\mathbf{C}$) log odds ratio of β.

This can be understood intuitively by imagining a single discrete C. If C were binary, say, with $\Pr(Y = 1 \mid X = 0, C = 0) = 0.25$ and $\Pr(Y = 1 \mid X = 0, C = 1) = 0.75$ then $\Pr(W = 0 \mid C = 0) = 0.25$ and $\Pr(W = 0 \mid C = 1) = 0.75$ so that were the two stratum-specific risk ratios to be

$$\Pr(D = 0 \mid C = 0) = \frac{1}{0.75e^2 + 0.25}$$

and

$$\Pr(D = 0 \mid C = 1) = \frac{1}{0.25e^2 + 0.75}$$

then the transformation from P_0 to P_1 would coincide with logistic regression with a conditional log odds ratio of -2 (Figure 7, left).

Figure 7 illustrates the corresponding stratum-specific risk ratios for a discrete C with more levels, intuitively leading to logistic regression for a continuous C in the limit.

The relationship (4) for nine different values of β is shown in the first panel of Figure 8.

If we believe that the variation in binary outcomes given binary exposures can be understood in terms of Rothman's sufficient component cause model, then the above shows that a logistic regression model can be correct (and the odds ratio transportable), but only if the sufficient causes are associated, and if this association takes a very particular form, notable in at least three ways:

(a) Any of the measured covariates $\mathbf{C}$ (those on which the analysis will be conditioned) associated with one sufficient cause must also be associated with the other. This is seen directly from (4): D can only be independent of a variable, C_1 say, in $\mathbf{C}$ if W is also independent of C_1.
(b) There can be no other sources of association between the sufficient causes. Intuitively, this is because associations due to $\mathbf{F}$ (not in $\mathbf{C}$), as seen above, remove one of logistic regression's fixed points.
(c) The strength of the association between the sufficient causes is a function of the value of the conditional log odds ratio, β.

Even in a setting where (a) and (b) could be deemed plausible, it seems unlikely that the quantitative relationship (4) for logistic regression could be deemed a priori plausible without knowledge of β.

This does not mean that a more natural (i.e. qualitative rather than quantitative) first principles justification of logistic regression does not exist. It is possible that a more complex mechanistic model—with more binary variables used to determine Y together with X in different combinations of $\wedge$ and $\vee$ operations—could lead to logistic regression under assumptions of a more qualitative nature, e.g. that certain binary variables in that more complex generating causal model are independent, or even that logistic regression could be justified using a mechanistic model other than Rothman's sufficient component cause model.

3.5 Other link functions

Using similar reasoning, we can uncover the nature of the association between W and D due to $\mathbf{C}$ that would lead to other forms for the dependence of P_1 on P_0. For example, for a GLM model with link function g and constant treatment coefficient ν (not dependent on $\mathbf{C}$), the relationship between $\Pr(W = 0 \mid \mathbf{C})$ and $\Pr(D = 0 \mid \mathbf{C})$ leading to that model (given that treatment is randomized) is:

$$\Pr(D = 0 \mid \mathbf{C}) = \frac{g^{-1}\left[g\{\Pr(W = 0 \mid \mathbf{C})\} + \nu\right]}{\Pr(W = 0 \mid \mathbf{C})}. \tag{5}$$

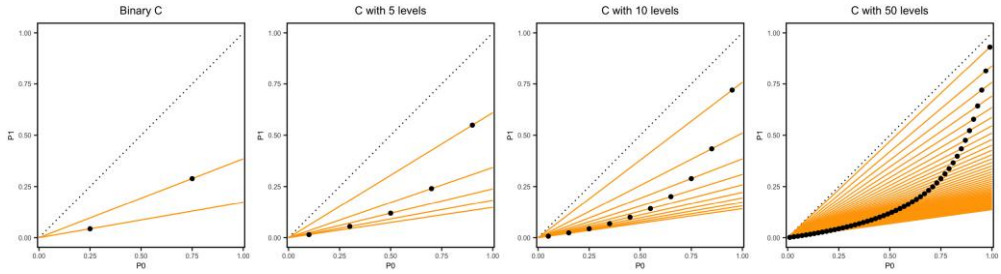

Figure 7. L'Abbé plots illustrating how logistic regression for a risk-lowering treatment can arise from the sufficient component cause model in Figure 3b, the DAG in Figure 4f, and the particular relationship shown in (4).

As we know, if $g = \log$, $\Pr(W = 0 \mid \mathbf{C})$ disappears from the right-hand side, and $\Pr(D = 0 \mid \mathbf{C})$ is a constant, e^{ν}. Other choices of g lead to non-trivial relationships as was discussed in Section 3.4 for $g = \text{logit}$.

Figure 8 shows the relationship between $\Pr(W = 0 \mid \mathbf{C})$, and $\Pr(D = 0 \mid \mathbf{C})$ for 9 choices of a risk-lowering ν and for 10 different link functions g. The nine horizontal lines for the log link indicate that $\Pr(D = 0 \mid \mathbf{C}) = \Pr(D = 0) = e^{\nu}$ is a constant (and here illustrated for nine different values of ν), and hence unrelated to $\Pr(W = 0 \mid \mathbf{C})$, which is not true for any of the other link functions. Similar figures for risk-raising ν and for the sufficient component cause model of the other type (Figure 3a) are given in the online supplementary material.

Of note are the qualitative differences between the relationships shown in Figure 8 for different link functions, even though all refer to risk-lowering treatment effects. The implied function mapping $\Pr(W = 0 \mid \mathbf{C})$ to $\Pr(D = 0 \mid \mathbf{C})$ can be a constant, strictly increasing, strictly decreasing, or none of these, with a variety of different curvature patterns and fixed points.

At least in some situations, a data analyst may know that some of these are (im)plausible based on subject-matter knowledge of how the binary variables D and W are likely to be associated via their common dependence on $\mathbf{C}$, including when the dependence can plausibly be deemed to be null. It is difficult, however, to imagine that a data analyst could commit to any one particular link function, other than $g = \log$, via such knowledge, for the reasons discussed in Section 3.4 in the particular case of logistic regression.

3.6 Non-monotone risk transformations, including Example 3: vaccination and hospitalization from COVID-19

Thus far, we have discussed settings in which the underlying mechanistic causal model is monotonic, in the sense described in Section 3.3. As a consequence, we have discussed only L'Abbé plots that either lie entirely above, entirely below, or entirely on the identity line ($P_1 = P_0$); no L'Abbé plot considered thus far has *crossed* the identity line. We now relax this, starting with an example, which we label Example 3.

Consider the period in 2021 when vaccinations against COVID-19 were first being rolled out. Let X be the binary treatment of receiving the vaccine, let Y be the binary outcome of hospitalization due to a severe COVID-19 infection, and let G, H, and V be binary latent variables defined as follows. G denotes whether or not a participant would have a normal immune response to the vaccine if vaccinated, H whether or not they would substantially increase the extent to which they mix with others if vaccinated, and V whether or not they would be hospitalized due to COVID-19 infection if unvaccinated.

It could then be argued that

$$Y = (V \wedge \neg X) \vee [\neg G \wedge (H \vee V) \wedge X] \tag{6}$$

i.e. that hospitalization occurs in the vaccinated if (1) they have no immune response to the vaccine and (2) they either would have been hospitalized if unvaccinated or their social mixing was

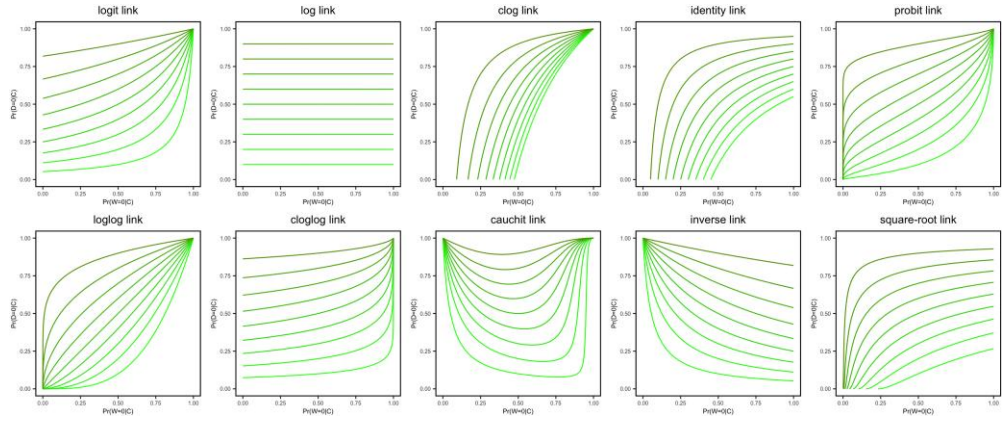

Figure 8. The relationship between $\Pr(W = 0 \mid \mathbf{C})$ and $\Pr(D = 0 \mid \mathbf{C})$ for 9 choices of a risk-lowering v and for 10 different link functions g, as described by (5). The link functions are defined as described in the caption of Figure 1 and the nine values of v were chosen such that a baseline risk of 0.5 would be transformed, respectively, to 0.45, 0.4, 0.35, 0.3, 0.25, 0.2, 0.15, 0.1, and 0.05, by each different v for each link function.

substantially increased (i.e. sufficiently so to cause serious infection) by the knowledge that they were vaccinated, (or both).

Suppose that the DAG shown in Figure 4g holds, then (6) implies that:

$$P_0 = \Pr(V = 1 \mid \mathbf{C})$$

and

$$\begin{aligned} P_1 &= \Pr(G = 0)\{1 - \Pr(H = 0 \mid G = 0)\Pr(V = 0 \mid \mathbf{C})\} \\ &= \Pr(G = 0)[1 - \Pr(H = 0 \mid G = 0)\{1 - P_0\}]. \end{aligned}$$

P_1 can thus be obtained from P_0 first by multiplying the complement of P_0 by $\Pr(H = 0 \mid G = 0)$ and finding its complement (a GLM transformation with a complementary log link) and then multiplying the result by $\Pr(G = 0)$ (a GLM transformation with a log link). That is, P_1 is obtained from P_0 by composing two GLM transformations: those corresponding to a clog and a log link.

By composing *any* GLM-clog transformation with *any* GLM-log transformation, the resulting L'Abbé plots are *all* the straight lines with positive slope:

$$\begin{aligned} P_1 &= e^{\beta_2}\left\{1 - e^{\beta_1}(1 - P_0)\right\} \\ &= e^{\beta_2}\left(1 - e^{\beta_1}\right) + e^{\beta_1 + \beta_2} P_0. \end{aligned}$$

Figure 9a shows a selection of such plots.

In fact, since the survival ratio in the first transformation is $\Pr(H = 0 \mid G = 0)$ and the risk ratio in the second transformation is $\Pr(G = 0)$, both probabilities, the resulting dual transformation (the composition) arising from (6) together with Figure 4g can be shown to give rise to straight line L'Abbé plots where both intercept and slope are in the interval [0, 1], and where the slope is always less than or equal to the complement of the intercept; that is, these transformations are the subset of all the clog-log compositions that are *closed* in the sense described in Section 2. A selection is shown in Figure 9b.

As in Section 3.3, there may often be no good reason to believe that G and H are independent of V, and thus that a more plausible DAG is the one shown in Figure 4h, which, together with (6), leads to:

$$P_1 = \Pr(G = 0) - \sum_{\mathbf{f}} \Pr(G = 0, H = 0 \mid \mathbf{F} = \mathbf{f})\Pr(V = 0 \mid \mathbf{C}, \mathbf{F} = \mathbf{f})\Pr(\mathbf{F} = \mathbf{f} \mid \mathbf{C})$$

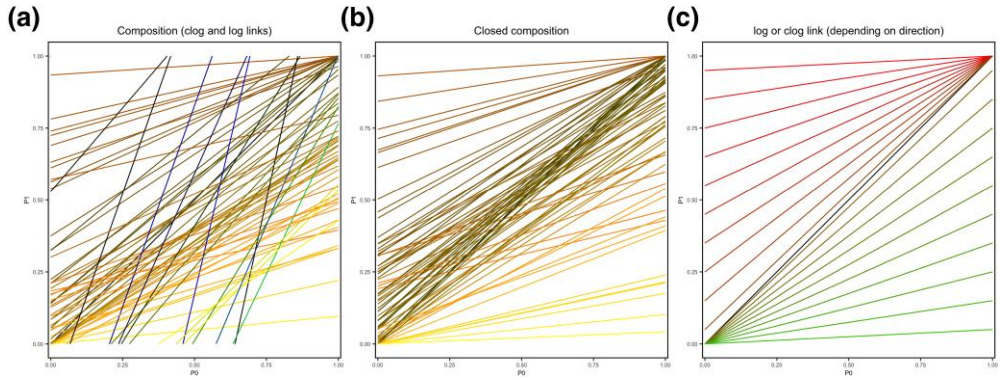

Figure 9. L'Abbé plots corresponding to, in panel (a), the general compositions of clog- and log-link GLMs, in (b) the closed subset of composed clog- and log-link GLM transformations implied by mechanistic model (6) for our Example 3 (COVID-19), and in (c) the subset of these corresponding to a switch relative risk transformation.

and

$$P_0 = \Pr(V = 1 \mid \mathbf{C}).$$

As in Section 3.3, this leads to non-affine extensions of the transformations in Figure 9b.

Remarks

All the lines in Figure 9b meet the identity line. When $\Pr(G = 0) = 1$ and $\Pr(H = 0 \mid G = 0) = 1$ then $P_1 \equiv P_0$ and the transformation is the identity transformation. Otherwise, it can be shown that the line meets the identity line at

$$P_0 = \frac{\Pr(G = 0)\{1 - \Pr(H = 0 \mid G = 0)\}}{1 - \Pr(G = 0)\Pr(H = 0 \mid G = 0)}$$

If $\Pr(H = 0 \mid G = 0) = 1$, then the point at which the line meets the identity is $P_0 = P_1 = 0$ and if $\Pr(G = 0) = 1$, then the point at which the line meets the identity is $P_0 = 1$. For all other values of these two probabilities, P_0 is strictly between 0 and 1 at the point at which the line meets the identity, i.e. the line *crosses* the identity line. This feature is not shared by a GLM for any link function, as we show in the Appendix.

The aforementioned special cases, where one or other of $\Pr(H = 0 \mid G = 0)$ and $\Pr(G = 0)$ is 1, correspond to the subset of transformations possible under the switch relative risk parameter: GLM transformations corresponding to both a log- and clog-link with a negative β are composed, but at most one of them is non-null. This subset of transformations is shown in Figure 9c.

It is important to note that non-monotonicity, as described in Section 3.3, does not imply that the L'Abbé plot must cross the identity line; however the converse is true, namely that if the L'Abbé plot crosses the identity line then monotonicity cannot hold. This can be understood by noting that monotonicity is defined on an individual level, whereas the L'Abbé plots show averages across individuals with the same values of **C**.

Although we moved in this section to an example with causal non-monotonicity, note that we did not choose simply to express our fundamental causal model as Figure 3a with both A and B or as Figure 3b with both D and E. Put another way, our model (6) is in fact of *both* types, with $U = W = \emptyset$, $A = \neg D = \neg G \wedge (H \vee V)$, and $B = \neg E = V$. The reason for this is that the first principles thinking led us directly to (6), but it was also useful since an assumption that (G, H) be independent of V then led us to an interesting new model of all the closed increasing affine risk transformations. Had we re-expressed the model (as we could have done) using either (U, A, B) or (W, D, E), and then assumed that (A, B) be independent of U or that (D, E) be independent of W (which are different from assuming that (G, H) be independent of V), these

independence constraints would have led us back to the GLM with a clog or a log link, respectively.

Although we motivated this section with an example from COVID-19 vaccination, similar examples arise in many different settings. Whenever an intervention is designed to protect, but is not efficacious in everyone, and knowledge of its receipt (without knowledge of its efficacy) may lead to the relaxation of other protective measures, then—under independence—relationships such as those shown in Figure 9b are plausible. For example, it may apply to the effect of safety measures such as protective clothing on the risk of serious injury in sport, or to the effect of energy-saving initiatives on the risk of unplanned outages in electricity systems.

4 Regression by composition: allowing for non-GLM risk transformations in a statistical model

In earlier sections of this manuscript, we have uncovered families of risk transformations (i.e. how the risk of a binary outcome conditional on covariates is transformed by a binary treatment) that arise from plausible fundamental causal models but that do not correspond to a GLM for any link function.

We have already seen how the non-monotone set of transformations (inspired by Example 3) arises from composing two different GLM transformations, namely those corresponding to a clog and log link.

In recent work (Farewell et al., under revision), we have proposed a novel and general framework for regression modelling, called regression by composition, that admits models such as those that compose both a clog-link and a log-link GLM transformation for the dependence on a particular variable in the model.

Let $\mathcal{P}$ be a set of distributions of Y. A *regression by composition* consists of

(a) a deterministic reference distribution $p_0 \in \mathcal{P}$, and
(b) a tuple $\eta_1, \eta_2, \eta_3, \ldots \in \mathbb{V}_1 \times \mathbb{V}_2 \times \mathbb{V}_3 \cdots$ of covariate-dependent linear predictors, in which the vector spaces $\mathbb{V}_1, \mathbb{V}_2, \mathbb{V}_3, \ldots$ have prescribed group actions that transform the set $\mathcal{P}$.

Regression by composition models the conditional distribution of Y, given covariates, as

$$P = ((p_0 \cdot \eta_1) \cdot \eta_2) \cdot \eta_3 \cdots,$$

writing group action on the right, so that $(p_0 \cdot \eta_1) \in \mathcal{P}$ is the result of the action of the linear predictor η_1 on the distribution p_0, and so on. In this general formulation, the word 'covariate' is used to include both $\mathbf{C}$ and X from our earlier discussions. The unknown parameters in each of the linear predictors η_j are estimated by maximum likelihood.

As well as including most existing regression models, new flavours of regression models can be formed by composing different transformation types (more formally, different group actions) for different covariates, or more than one group action for the same covariate. For example, with a binary Y as in the examples discussed thus far, p_0 could be the Bernoulli distribution with probability 0.5, $\eta_1 \in \mathbb{R}^+$ a $\mathbf{C}$-dependent linear predictor that transforms this distribution by scaling the odds, $\eta_2 \in \mathbb{R}^+$ an X-dependent linear predictor that scales the survival probability, and $\eta_3 \in \mathbb{R}^+$ an X-dependent linear predictor that scales the risk. The model could then be thought of as layered as follows:

$$\begin{aligned}\text{Bern}(0.5) &\to \text{Bern}\left(\frac{\eta_1}{1+\eta_1}\right) \to \text{Bern}\left(1 - \left\{1 - \frac{\eta_1}{1+\eta_1}\right\}\eta_2\right) \\ &\to \text{Bern}\left(\eta_3\left[1 - \left\{1 - \frac{\eta_1}{1+\eta_1}\right\}\eta_2\right]\right).\end{aligned}$$

We refer any reader interested in the theoretical underpinnings and wider scope of regression by composition to Farewell et al. (under revision), and instead here conclude with a simple illustrative simulated example to demonstrate the potential of the approach.

We simulated data in accordance with $\neg Y = W \vee (X \wedge D)$ and Figure 4e. A single continuous covariate C was simulated from a standard normal distribution and both X and F were simulated independently from Bernoulli distributions with mean 0.5. W was simulated, dependent on C and F, from a Bernoulli distribution with mean expit$(1 - C - 2F)$ and D was simulated, dependent on F, from a Bernoulli distribution with mean 0.88 or 0.12, depending on whether F was 0 or 1, respectively. Y was generated deterministically from W, D, and X, using $\neg Y = W \vee (X \wedge D)$.

Three models were compared for $Y \mid C, X$. First a standard logistic regression model (without a product term in CX), then a regression by composition model formed from a proportional odds (i.e. logistic regression) component for C composed with a proportional risks component for X, and finally a regression by composition formed from a proportional odds component for C composed with both a proportional risks component and a proportional odds component for X. Although the final model does not correspond exactly to the non-affine transformation implied by the dependence between W and D in the data-generating process, the hope is that the composition employed for X will approximate this reasonably well.

Although such an approach may successfully approximate the whole transformation when the data include individuals across the spectrum of untreated risks, it is unlikely to do so beyond the range of the data when only individuals with very low outcome risks, say, are observed. To investigate this, we repeated the three analyses across five datasets. In the first dataset, only individuals with a true untreated conditional risk of the outcome less than 0.2 are included, to mimic an RCT, say, with eligibility criteria that only admit the healthiest participants into the trial. In the second dataset, this threshold is increased to 0.4, and so on, and the fifth dataset includes all participants.

The results in Figure 10 show that the composition of the proportional risks and proportional odds transformations does indeed approximate the non-affine relationship between P_0 and P_1 quite well when data across the range of untreated risks are included, whereas the same is not true for either of the other models that do not admit transformations of approximately the correct form. Also of note is the stark difference in the width of confidence intervals across the different models when the datasets are limited to low-risk participants. Although none of the models successfully estimates the shape of the transformation when only low-risk participants are included, the model with the dual-transformation for treatment is the only one that appropriately reflects its uncertainty about this form. By relying on untestable assumptions, the other two models make precise (but incorrect) predictions about the nature of the risk transformation for risks very dissimilar to the ones on which the estimation was based.

5 Discussion

When introducing logistic regression in 1958, the late Professor Sir David Cox wrote:

> 'A linear relation is unsuitable, except over a narrow range, because of the restriction that [the outcome probability] must lie in [0, 1] and, in the absence of special considerations for a particular problem, the best form seems to be the logistic law.'

In this paper, we have attempted to describe situations in which the 'special considerations for a particular problem' would steer us away from logistic regression and towards more flexible binary regression models that better align with the plausible underlying causal mechanisms that generated the data.

To this end, we have followed Huitfeldt et al. (2021) in finding Rothman's sufficient component cause framework (Rothman, 1976), particularly when allowing the sufficient causes to be associated, to be a helpful guide towards a suitable choice of model. In so doing, we have partially answered the following call by Hernán & Robins in their recent book on causal inference:

> 'Though the sufficient component cause framework is useful from a pedagogic standpoint, its relevance to actual data analysis is yet to be determined.'

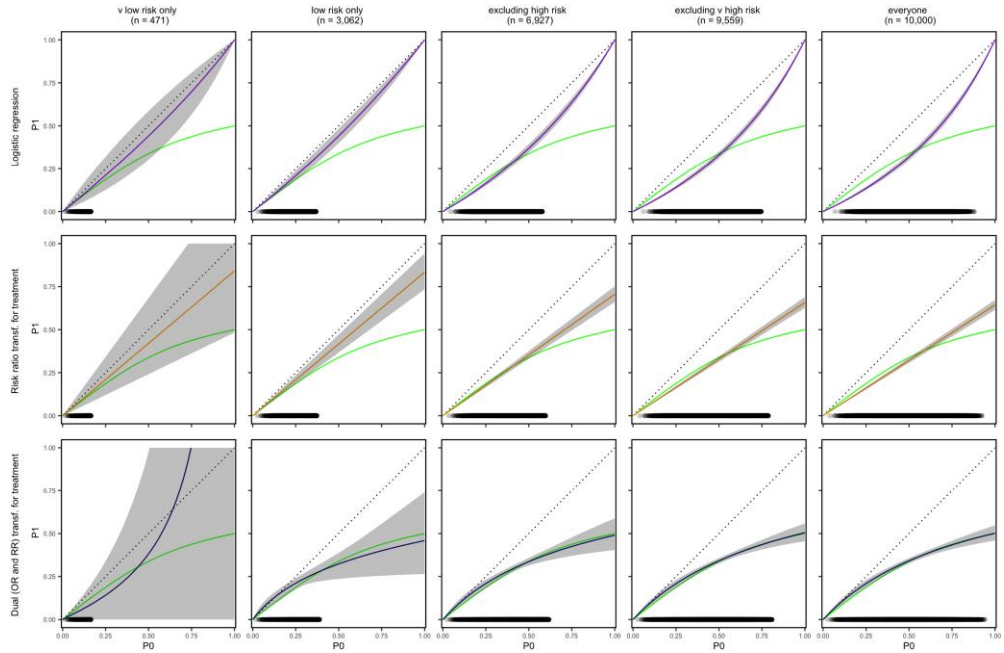

Figure 10. The results of the analysis of the simulated data. In each plot, the green curve shows the true relationship between P_0 and P_1 from the data-generating process. The purple curves in the top row are the risk transformations for the treatment as estimated from logistic regression, the orange lines in the middle row are the estimated risk transformations for the treatment from a regression by composition model that composes a logistic regression model for the dependence on **C** with a log-link GLM dependence (risk ratio transformation) on X, and the navy curves in the bottom row are from a similar model that includes both an odds ratio and a risk ratio transformation for X. 95% confidence regions are shaded in grey. The columns correspond to the increasing subsets of the data included in the analyses—from only low-risk individuals in the left-most column to all individuals in the right-most column. The points on the horizontal axis show the estimated untreated conditional risks from each model for the individuals included in the analysis.

Although space has not permitted further exposition here, the idea of composing transformations of binary-data distributions can be extended to, for example, time-to-event data, with similarly attractive results. Just as proportional risks and proportional survival probability transformations can be composed to form flexible non-monotone affine models for binary outcomes, the parametric version of the Aalen additive hazards model (Aalen, 1989) can be composed with its complement (a scaling of the survivor function) to form fairly flexible parametric models that exhibit features such as crossing survivor curves with low-dimensional parameters.

Conflicts of interest: We have no conflicts of interest nor funding sources to declare.

Data availability

The code for generating and analysing the simulated data presented in Section 4 is available by following the link given in the online supplementary material.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series A*.

Appendix. Two claims about generalized linear models

In Section 3 of this paper, we exhibited families of risk transformations arising from first principles causal models. In particular, in Section 3.3, we discussed non-affine risk-lowering transformations arising from causal models of the type $\neg Y = W \vee (X \wedge D)$ where W and D are associated due to their dependence on an unmeasured binary F. And in Section 3.6, we discussed non-monotone

affine transformations arising from the causal model $Y = (V \wedge \neg X) \vee [\neg G \wedge (H \vee V) \wedge X]$ when (G, H) are independent of V.

Both families are inherently characterized by more than one parameter for the treatment effect and it would thus be unreasonable to ask (and trivial to answer with 'no') if either family corresponds to a GLM (with a single treatment parameter) for some link function. As noted in Section 3.4, therefore, a more reasonable question to ask is whether a GLM link function exists such that the family of risk transformations arising from it constitutes an 'interesting' subset of either of our first principles families.

We contend that an important feature of our family of non-affine transformations (see Figure 5) is that it includes both convex and concave risk transformations on the same side (both risk-lowering) of the identity. The main feature of our non-monotone family (see Figure 9b) is that each transformation (other than those passing through (0, 0) or (1, 1)) crosses the identity line. In Claims 1 and 2, we show that these are features not exhibited by a GLM with a single treatment parameter, for any link function.

Claim 1 Suppose a GLM for a binary outcome (with conditional outcome risk p, and treatment coefficient v) gives rise to the transformation $f(p; v) = g^{-1}\{g(p) + v\}$ for some monotonic link function g, sufficiently smooth for the necessary derivatives of f to exist. Suppose also that there exists some $v^* < 0$ such that $f''(p; v) \leq 0 \ \forall \ p \ \forall \ v \in [v^*, 0)$, where differentiation is with respect to p. Then, for any $v \in (-\infty, v^*)$, $f''(p; v) \leq 0 \ \forall \ p$. That is, if $f(p; v)$ is everywhere-concave in p for all negative v between some v^* and 0, then it is everywhere-concave in p for all negative v.

Proof. For any $v \in (-\infty, v^*)$, let m be the smallest positive integer such that

$$\frac{|v - v^*|}{m} < |v^*|$$

and let

$$u = \frac{v - v^*}{m}$$

so that by construction, $v = v^* + mu$ and $u \in (v^*, 0)$.

For a given v^* and u, write

$$h_k(p) := f(p; v^* + ku)$$

for $k \in \mathbb{N} \cup \{0\}$. Thus

$$h_{k+1}(p) \equiv f(h_k(p); u)$$

and it can easily be shown that

$$h''_{k+1}(p) = f''(h_k(p); u)\{h'_k(p)\}^2 + f'(h_k(p); u)h''_k(p).$$

By supposition, $f''(h_k(p); u) \leq 0 \ \forall \ p$, and by monotonicity of g, $f'(h_k(p); u) \geq 0 \ \forall \ p$. Thus, if $h''_k(p) \leq 0 \ \forall \ p$ then $h''_{k+1}(p) \leq 0 \ \forall \ p$ also. Since $h_0(p) \equiv f(p; v^*)$, we have by supposition that $h''_0(p) \leq 0 \ \forall \ p$, and thus $h''_k(p) \leq 0 \ \forall \ p$ for all $k \in \mathbb{N} \cup \{0\}$ by induction.

In particular, $h_m(p) \equiv f(p; v)$, which proves the claim. □

Claim 2 Suppose a GLM for a binary outcome (with conditional outcome risk p, and treatment coefficient v) gives rise to the transformation $f(p; v) = g^{-1}\{g(p) + v\}$

for some monotonic link function g. If there exists p^* such that $f(p^*; v) > p^*$, then $f(p; v) > p$ for all p, and if there exists a p^* such that $f(p^*; v) < p^*$ for some p^*, then $f(p; v) < p$ for all p.

Proof. Write $p_1 = f(p_0; v)$ for some value p_0 of p. If g is increasing, then

$$\begin{aligned} p_1 > p_0 &\Leftrightarrow g(p_1) > g(p_0) \\ &\Leftrightarrow g(p_0) + v > g(p_0) \\ &\Leftrightarrow v > 0 \end{aligned}$$

and similarly $p_1 < p_0 \Leftrightarrow g(p_1) < g(p_0) \Leftrightarrow v < 0$. If g is decreasing, then $p_1 > p_0 \Leftrightarrow g(p_1) < g(p_0) \Leftrightarrow v < 0$ and $p_1 < p_0 \Leftrightarrow g(p_1) > g(p_0) \Leftrightarrow v > 0$. In all cases, the sign of v dictates the sign of the difference between p_0 and p_1 for any such pair, which proves the claim. □

References

Aalen O. O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine*, *8*(8), 907–925. https://doi.org/10.1002/sim.v8:8

Altman D. G., Deeks J. J., & Sackett D. L. (1998). Odds ratios should be avoided when events are common. *BMJ*, *317*(7168), 1318. https://doi.org/10.1136/bmj.317.7168.1318

Angrist J. D., & Pischke J.-S. (2008). *Mostly harmless econometrics*. Princeton University Press.

Battey H. S., Cox D. R., & Jackson M. V. (2019). On the linear in probability model for binary data. *Royal Society Open Science*, *6*(5), 190067. https://doi.org/10.1098/rsos.190067

Choi S. W. (2016). Odds, risks, and other numbers needed to complicate things. *Anaesthesia*, *71*(10), 1234–1236. https://doi.org/10.1111/anae.2016.71.issue-10

Cook T. D. (2002). Advanced statistics: Up with odds ratios! a case for odds ratios when outcomes are common. *Academic Emergency Medicine*, *9*(12), 1430–1434. https://doi.org/10.1197/acem.2002.9.issue-12

Cox D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *20*(2), 215–232. https://doi.org/10.1111/j.2517-6161.1958.tb00292.x

Daniel R., Zhang J., & Farewell D. (2021). Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biometrical Journal*, *63*(3), 528–557. https://doi.org/10.1002/bimj.v63.3

Doi S. A., Furuya-Kanamori L., Xu C., Lin L., Chivese T., & Thalib L. (2022). Controversy and debate: Questionable utility of the relative risk in clinical research: Paper 1: A call for change to practice. *Journal of Clinical Epidemiology*, *142*, 271–279. https://doi.org/10.1016/j.jclinepi.2020.08.019

Farewell D. M., Daniel R. M., Stensrud M. J., & Huitfeldt A. (under revision). Regression by composition.

Greenland S. (2014). *Effect modification and interaction*. John Wiley & Sons, Ltd.

Hernán M., & Robins J. (2020). *Causal inference: What if*. Chapman & Hall/CRC.

Huitfeldt A. (2017). Odds ratios are not conditional risk ratios. *Journal of Clinical Epidemiology*, *84*, 191. https://doi.org/10.1016/j.jclinepi.2016.12.013

Huitfeldt A. (2023). Mindel C. Sheps: Counted, dead or alive. *Epidemiology*, *34*(3), 396–399. https://doi.org/10.1097/EDE.0000000000001591

Huitfeldt A., Fox M. P., Murray E. J., Hróbjartsson A., & Daniel R. M. (2021). 'Shall we count the living or the dead?', arXiv, arXiv:2106.06316, preprint: not peer reviewed. https://doi.org//10.48550/arXiv.2106.06316

Huitfeldt A., Goldstein A., & Swanson S. A. (2018). The choice of effect measure for binary outcomes: Introducing counterfactual outcome state transition parameters. *Epidemiologic Methods*, *7*(1), 20160014. https://doi.org/10.1515/em-2016-0014

L'Abbé K. A., Detsky A. S., & O'Rourke K. (1987). Meta-analysis in clinical research. *Annals of Internal Medicine*, *107*(2), 224–233. https://doi.org/10.7326/0003-4819-107-2-224

McCullagh P., & Nelder J. A. (1989). *Generalized linear models* (2nd ed.). Chapman & Hall/CRC.

Miettinen O. (1974). Confounding and effect-modification. *American Journal of Epidemiology*, *100*(5), 350–353. https://doi.org/10.1093/oxfordjournals.aje.a112044

Permutt T. (2020). Do covariates change the estimand? *Statistics in Biopharmaceutical Research*, *12*(1), 45–53. https://doi.org/10.1080/19466315.2019.1647874

Rohrer J. M., & Arslan R. C. (2021). Precise answers to vague questions: Issues with interactions. *Advances in Methods and Practices in Psychological Science*, *4*(2), 251524592110073. https://doi.org/10.1177/25152459211007368

Rothman K. J. (1976). Causes. *American Journal of Epidemiology*, *104*(6), 587–592. https://doi.org/10.1093/oxfordjournals.aje.a112335

Rudolph K. E., Williams N. T., Miles C. H., Antonelli J., & Diaz I. (2023). All models are wrong, but which are useful? comparing parametric and nonparametric estimation of causal effects in finite samples. *Journal of Causal Inference*, *11*(1). https://doi.org/10.1515/jci-2023-0022

Senn S. (2011). U is for unease: Reasons for mistrusting overlap measures for reporting clinical trials. *Statistics in Biopharmaceutical Research*, *3*(2), 302–309. https://doi.org/10.1198/sbr.2010.10024

Sheps M. C. (1958). Shall we count the living or the dead? *New England Journal of Medicine*, *259*(25), 1210–1214. https://doi.org/10.1056/NEJM195812182592505

Sonis J. (2018). Odds ratios vs risk ratios. *JAMA*, *320*(19), 2041. https://doi.org/10.1001/jama.2018.14417

Van Der Laan M. J., Hubbard A., & Jewell N. P. (2007). Estimation of treatment effects in randomized trials with non-compliance and a dichotomous outcome. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *69*(3), 463–482. https://doi.org/10.1111/j.1467-9868.2007.00598.x

White I. R., Morris T. P., & Williamson E. (2021). 'Covariate adjustment in randomised trials: canonical link functions protect against model mis-specification', arXiv, arXiv:2107.07278, preprint: not peer reviewed. https://doi.org//10.48550/arXiv.2107.07278

Xiao M., Chen Y., Cole S. R., MacLehose R. F., Richardson D. B., & Chu H. (2022). Controversy and debate: Questionable utility of the relative risk in clinical research: Paper 2: Is the odds ratio "portable" in meta-analysis? time to consider bivariate generalized linear mixed model. *Journal of Clinical Epidemiology*, *142*, 280–287. https://doi.org/10.1016/j.jclinepi.2021.08.004

Zipkin D. A., Umscheid C. A., Keating N. L., Allen E., Aung K., Beyth R., Kaatz S., Mann D. M., Sussman J. B., Korenstein D., Schardt C., Nagi A., Sloane R., & Feldstein D. A. (2014). Evidence-based risk communication. *Annals of Internal Medicine*, *161*(4), 270. https://doi.org/10.7326/M14-0295