Models for Additive and Sufficient Cause Interaction

DANIEL BERGLUND

Licentiate Thesis
KTH Royal Institute of Technology
School of Engineering Sciences
Department of Mathematics
Applied and Computational Mathematics
Stockholm, Sweden 2019

Department of Mathematics
KTH Royal Institute of Technology
100 44 Stockholm, Sweden

TRITA-SCI-FOU 2019;43
ISBN 978-91-7873-308-8

Akademisk avhandling som med tillstånd av Kungliga Tekniska Högskolan framlägges till offentlig granskning för avläggande av teknologie licentiatexamen, torsdagen den 10 oktober 2019 klockan 10.00 i sal F11, Lindstedtsvägen 22, Kungliga Tekniska Högskolan, Stockholm.

Print: Universitetsservice US AB, Stockholm, 2019

Abstract

The aim of this thesis is to develop and explore models in, and related to, the sufficient cause framework, and additive interaction. Additive interaction is closely connected with public health interventions and can be used to make inferences about the sufficient causes in order to find the mechanisms behind an outcome, for instance a disease.

In paper A we extend the additive interaction, and interventions, to include continuous exposures. We show that there does not exist a model that does not lead to inconsistent conclusions about the interaction.

The sufficient cause framework can also be expressed using Boolean functions, which is expanded upon in paper B. In this paper we define a new model based on the multifactor potential outcome model (MFPO) and independence of causal influence models (ICI).

In paper C we discuss the modeling and estimation of additive interaction in relation to if the exposures are harmful or protective conditioned on some other exposure. If there is uncertainty about the effects direction there can be errors in the testing of the interaction effect.

Keywords: Causal Inference; Sufficient Cause; Potential Outcomes; Counterfactual; Additive Interaction; Interaction; MFPO; ICI; Logistic Regression; Linear Odds; Public Health; Interventions; Probabilistic Potential Outcome

Sammanfattning

Målet med denna avhandling är att utveckla, och utforska modeller i det så kallade sufficent cause ramverket, och additiv interaktion. Additiv interaktion är nära kopplat till interventioner inom epidemiology och sociologi, men kan också användas för statistiska tester för sufficient causes för att förstå mekanimser bakom ett utfall, tex en sjukdom.

I artikel A så expanderar vi modellen för additiv interaktion och interventioner till att också inkludera kontinuerliga variabler. Vi visar att det inte finns någon modell som inte leder till motsägelser i slutsatsen om interaktionen.

Sufficient cause ramverket kan också utryckas via Boolska funktioner, vilket byggs vidare på i artikel B. I den artikeln definerar vi en modell baserad på mutltifactor potential outcome modellen (MFPO) och independence of causal influence modellen (ICI).

I artikel C diskuterar vi modelleringen och estimering av additiv interaktion i relation till om variablerna har skadlig eller skyddande effekt betingat på någon annan variabel. Om det finns osäkerhet kring en effekts riktning så kan det leda till fel i testerna för den additiva interaktionen.

Acknowledgments

I would like to thank my supervisors Professor Timo Koski and Assistant Professor Helga Westerlind for their support and sharing their thoughts on various topics.

I would also like to thank all of my colleagues at KTH and KI for the support, as well as all the interesting discussions.

Lastly I wish to thank my parents and my sister for all the continuing love and support over the years.

Table of Contents

Part I: Introduction

Part II: Scientific Papers

Part: I

Introduction

1 Conterfactual Causality

Cause and effect are fundamental parts of most areas of scientific research. However, from centuries of debate there is no consensus of the meaning of causality [7, 38].

The methods and models in this thesis are based in the *counterfactual* framework also called *potential outcome* framework, which has become a common framework for causality in epidemiology and social science. The terms 'potential outcome' and 'counterfactual outcome' can be used mostly interchangeably[1].

The idea for causality as defined by counterfactuals goes at least as far back as 1748 with Hume writing: [24]

> We may define a cause to be an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second. Or in other words where, if the first object had not been, the second never had existed.

In other words, one event caused another event if the occurance of the former was necessary for the occurance of the latter [18, 20]. This naturally leads to that the difference

[1] 'Potential outcome' and 'counterfactual outcome' have somewhat different linguistic meanings and some authors prefer one term over the other. Counterfactual outcome tends to be used in the philosophical litterature, while potential outcome is more common in the statistical literature. The term potential outcome is more accurate for the models in this thesis, since it can be argued that the outcome that was in fact observed is not counter to the fact, i.e., an observed, or future, outcome is not counterfactual. For more details see p. 461 in [52].

between the outcome if some events did or did not occur is the basic condition for if the events are a cause of the outcome. For example, suppose we have an outcome D and an event X, then X is causal if $D_X - D_{\bar{X}} \neq 0$.

The framework first got formalized in mathematical terms in 1923 with Neyman writing about potential outcomes and experiments in agriculture [49]. The model was later expanded upon by Rubin [43–46]. It also seems that the same concepts and ideas have appeared several times independently in various field such as economics [35, 42], and computer science [32, 33].

1.1 The Fundamental Problem of Causal Inference

However, there is a problem, we can always only observe the outcome that has happened. Observing the counterfactual outcome that would had happened if some other action had been taken is not possible, i.e., we can only observe one of D_X or $D_{\bar{X}}$, not both. In essence, the counterfactual view of causality is based on the difference between these two outcomes, but we can only know one of the outcomes [20]. This is known as *the fundamental problem of causal inference*. There are two main approaches to solving the problem, the *scientific solution*, and the *statistical solution* [20].

The *scientific solution* uses assumptions about homogeneity or invariance in order to observe both outcomes [20]. One method is to assume that the experiment is invariant to time, for instance we expect a light switch to have the same behavior independent of its previous state and time. The same light switch can then be used to make causal claims about the statement if it turns on a light or not. Another possible method is to use a homogeneity assumption in order to be able to treat different objects as the 'same' object. With this assumption we can take an action on a subset of the objects and compare those objects against the objects where the action was not taken. Continuing with the light switch example we would then have a number of switches connected to different lights and flick some of the switches.

The *statistical solution* instead uses population averages combined with assumptions to estimate the average causal effect [20]. This solution is used for the models in this thesis and will be discussed in more detail in Section 4.

2 Interaction

For some outcomes it can be the case that a combination of events are required for the outcome to occur. This is referred to as interaction.

Example 2.1: Example from [3]. Phenylketonuria (PKU) is a metabolic disorder where a genetic mutation when combined with a particular amino acid in the diet causes mental retardation. Since both the mutation and the amino acid are required PKU can be prevented by testing the infants and restricting the diet for the ones with the mutation. The mutation and the amino acid are interacting to cause the mental retardation. ■

The meaning of the word interaction is not always clear [3] and interaction can mean different things in different contexts. For instance, it could be a physical reaction between two molecules, a mechanism for a disease such as in the example above, or a coefficient in a mathematical model. In many contexts the term interaction is hard to define, however, statistical interaction is well defined. It is when the effect from one exposure depends on the level of another exposure [40, 52]. Statistical interaction is scale dependent and different scales can lead to different conclusions. An example is the interaction between smoking and asbestos on the risk of lung cancer where the risk from asbestos is higher for nonsmokers than for smokers on the ratio scale, however, the opposite is true on the additive scale [40].

The two most common scales are multiplicative (ratio scale) and additive. The multiplicative scale is the inclusion of the product term between variables in a model, but lacks the connection with a causality framework [54]. For multiplicative interaction, or interaction in general without an exact definition, there are a vast number of different methods [11, 15, 39, 59, 60]. This area of research tends to focus on computational speed and finding potential interaction that can be studied later, rather than some more precise form of interaction based in a causality framework. However, these types of methods and models are outside the scope of this thesis.

In contrast, additive interaction is connected to the sufficient cause model for causality, which we will define in Section 5 [40, 41, 52, 55]. It also plays an important role in public health interventions [40, 41]. As described earlier, the fundamental part of the counterfactual model for causality is the difference between the outcome with and without some action, or event. The effect from an intervention is also the difference between outcomes with and without the intervention/action. We illustrate with an example of how additive interaction can be used to draw conclusions about interventions.

Example 2.2: Example from [54]. Consider the data in Table 1 from a study on the effect of smoking and asbestos exposure on the risk of lung cancer.

	No asbestos	Asbestos
Non-smoker	0.0011	0.0067
Smoker	0.0095	0.0450

Table 1: Risk of lung cancer based on smoking status and asbestos exposure.

Suppose we want to do an intervention on smokers with or without asbestos exposure, on which group does the intervention have the greatest effect? The differences in risk between smokers and non-smokers in the two groups is $0.0095 - 0.0011 = 0.0084$ for workers that have not be exposed to asbestos, but for workers with asbestos exposure the difference is $0.0450 - 0.0067 = 0.0383$. Thus, the difference is largest for workers that have been exposed to asbestos, so the intervention has the most effect in this group. ■

This is one reason for why the additive scale can be argued to be the preferred scale for interventions and public health, as it leads to the most effective interventions when choosing which group to intervene upon [4, 6, 40, 41, 47]. Additive interaction, and how to estimate it, will be explained in more detail in Section 6.

3 Notation

Before we can formally define the concepts discussed in the introduction we need to introduce the notation. We will use the same notation as in [55].

An *event* is a binary (Boolean) variable, indicated by uppercase (X). Boldface uppercase is *set of events* ($\mathbf{C}$). The *complement* or *negation* of some event X is $\overline{X} \equiv 1 - X$. Lowercase is used for *specific values* of events, $X = x$, or sets of events, $\{\mathbf{C} = \mathbf{c}\} \equiv \{\forall i, (\mathbf{C})_i = (\mathbf{c})_i\}$. The *cardinality* of a set is denoted by $|\mathbf{C}|$. $\mathbf{C}_1 \dot{\cup} \mathbf{C}_2$ is the disjoint union between $\mathbf{C}_1$ and $\mathbf{C}_2$. Fraktur/gothic typeset ($\mathfrak{B}$) is used for *collections of sets of events*.

A *literal event* associated with X, is either X or $\overline{X}$. We form for a given set of events $\mathbf{C}$, $\mathbb{L}(\mathbf{C})$ the associated set of literal events

$$\mathbb{L}(\mathbf{C}) \equiv \mathbf{C} \cup \{\overline{X} \mid X \in \mathbf{C}\}. \tag{3.1}$$

For a literal $L \in \mathbb{L}(\mathbf{C})$, $(L)_{\mathbf{c}}$ denotes the value set by an assignment $\mathbf{C} = \mathbf{c}$.

The *conjunction* of a set of literal events $\mathbf{B} = \{F_1, \ldots, F_m\} \subseteq \mathbb{L}(\mathbf{C})$ is defined as:

$$\bigwedge(\mathbf{B}) \equiv \prod_{i=1}^{m} F_i = \min\{F_1, \ldots, F_m\}; \tag{3.2}$$

In terms of Boolean functions $\bigwedge(\mathbf{B})$ is *and* [31], i.e. $\bigwedge(\mathbf{B}) = 1$ if and only if for all i, $F_i = 1$. The *disjunction* of a set of binary events is defined as:

$$\bigvee(\{Z_1, \ldots, Z_p\}) \equiv \max\{Z_1, \ldots, Z_p\}; \tag{3.3}$$

This is equivalent to the Boolean function *or* [31], so $\bigvee(\{Z_1, \ldots, Z_p\}) = 1$ if and only if for some j, $Z_j = 1$.

For a collection of sets of literals $\mathfrak{B} = \{\mathbf{C}_1, \ldots, \mathbf{C}_q\}$ define

$$\bigvee\bigwedge(\mathfrak{B}) \equiv \bigvee_i (\bigwedge(\mathbf{C}_i)). \tag{3.4}$$

which is the Boolean function *tribes* [31], i.e. $\bigvee\bigwedge(\mathfrak{B}) = 1$ if and only if for some j $\bigwedge(\mathbf{C}_j) = 1$.

$\mathbb{P}(\mathbb{L}(\mathbf{C}))$ is the set of subsets of $\mathbb{L}(\mathbf{C})$ that do not contain both X and $\overline{X}$ for any $X \in \mathbf{C}$. Formally,

$$\mathbb{P}(\mathbb{L}(\mathbf{C})) \equiv \{\mathbf{B} \mid \mathbf{B} \subset \mathbb{L}(\mathbf{C}) \; \forall \; X \in \mathbf{C}, \{X, \overline{X}\} \not\subseteq \mathbf{B}\}. \tag{3.5}$$

4 Potential Outcome Models

Suppose that we have a potential outcome model [20, 49, 55] with d binary events, $\mathbf{C} = \{X_1, \ldots, X_d\}$, and an binary outcome D, e.g., disease. We have a large but finite population sample space, U, and u denotes a particular individual in the sample space.

Definition 4.1: *Potential outcome*
For an individual $u \in U$ let $D_{\mathbf{x}}(u)$ be the potential outcome if the binary events are set as $\mathbf{X} = \mathbf{x}$. ■

Definition 4.2: *Observed outcome*
For an individual $u \in U$ let $D(u)$ be the observed outcome. ■

We also define

$$p_{\mathbf{x}} \equiv P(D = 1 \mid \mathbf{X} = \mathbf{x}). \tag{4.1}$$

We will use $\mathfrak{D}(\mathbf{C}; \mathfrak{u})$ to denote the set of all possible outcomes for an individual u in the population U, i.e. all the individuals outcomes for all possible assignments to $\mathbf{C}$. For $d = 2$ there are four potential outcomes per individual corresponding to a row in Table 2. $\mathfrak{D}(\mathbf{C}; U)$ is the set of all possible outcomes for the population. The number of possible types of individuals is 2^{2^s}. All the possible types of individuals for $d = 2$ are shown in Table 2.

4.1 Statistical Solution to the Fundamental Problem of Causal Inference

As explained in the introduction, the fundamental problem of causal inference is that it is not possible to observe all the potential outcomes for an individual. The statistical solution uses the difference between averages instead of the actual difference, so we can estimate the average effect. This solution to the fundamental problem of causal inference requires two assumptions, consistency and conditional exchangeability.

Assumption 4.1: *Consistency*
For all individuals $u \in U$ with $\mathbf{C} = \mathbf{c}$ it holds that

$$D_{\mathbf{c}}(u) = D(u). \tag{4.2}$$

■

The consistency assumption means that an individual's observed outcome is the same as the potential outcome for individual's actual exposures. It also implies that, if we take some action that changes the individuals exposures the observed outcome also changes.

This assumption may look trivial, however, it can easily be broken in several situations. For example, it does not hold if different variants of the exposure have different effects on the outcome, such as that income from work might not have the same effect as income

Individual response type	D_{11}	D_{01}	D_{10}	D_{00}
1	1	1	1	1
2^-	1	1	1	0
3^+	1	1	0	1
4	1	1	0	0
5^+	1	0	1	1
6	1	0	1	0
7^+	1	0	0	1
8^+	1	0	0	0
9^-	0	1	1	1
10^-	0	1	1	0
11	0	1	0	1
12^-	0	1	0	0
13	0	0	1	1
14^-	0	0	1	0
15^+	0	0	0	1
16	0	0	0	0

Table 2: All possible response types for two events. The response types marked with $+$ corresponds to superadditive interaction and the ones marked with $-$ subadditive interaction.

from a lottery winning [37]. Interventions can also have issues with the assumption, since the intervention might not be the same as the original effect. For example, the difference between income from work or as a social program [37]. The assumption also fails if the treatment has an effect on other individuals than the one treated, such as herd immunity in vaccine trials [20].

Assumption 4.2: *Conditional exchangeability*
Given a set of events, $\mathbf{C}$, and a set of confounding events, $\mathbf{W}$, the potential outcome $D_{\mathbf{c}}$ is conditionally independent of $\mathbf{C}$ given $\mathbf{W}$

$$D_{\mathbf{c}} \perp\!\!\!\perp \mathbf{C} \mid \mathbf{W}. \tag{4.3}$$

■

The conditional exchangeability assumption implies that given the covariates W then the potential outcome $D_{\mathbf{c}}$ is independent of the value of $\mathbf{C}$. We need this assumption to be able to compare the potential outcomes across groups with different covariates. The assumption is also known as ignorable treatment assignment, no unmeasured confounding or exogeneity[52].

Based on the two assumptions it is possible to prove the following theorem [52].

Theorem 4.1: *Suppose for a set of events* $\mathbf{C}$ *we have two different sets of possible instances of these events,* $\mathbf{C} = \mathbf{c_1}$ *and* $\mathbf{C} = \mathbf{c_2}$*, and their potential outcomes,* $D_{\mathbf{c_1}}(u)$*,* $D_{\mathbf{c_2}}(u)$*, and also a set of covariates* $\mathbf{W} = \mathbf{w}$*. Assuming consistency and conditional exchangeability assumptions hold, then*

$$E[D_{\mathbf{c_1}} - D_{\mathbf{c_2}} \mid \mathbf{W} = \mathbf{w}] = E[D \mid \mathbf{C} = \mathbf{c_1}, \mathbf{W} = \mathbf{w}] - E[D \mid \mathbf{C} = \mathbf{c_2}, \mathbf{W} = \mathbf{w}]. \quad (4.4)$$

Proof:

$$\begin{aligned} E[D_{\mathbf{c_1}} - D_{\mathbf{c_2}} \mid \mathbf{W} = \mathbf{w}] &= E[D_{\mathbf{c_1}} \mid \mathbf{W} = \mathbf{w}] - E[D_{\mathbf{c_2}} \mid \mathbf{W} = \mathbf{w}] = \\ \text{(conditional exchangeability)} &= E[D_{\mathbf{c_1}} \mid \mathbf{C} = \mathbf{c_1}, \mathbf{W} = \mathbf{w}] - E[D_{\mathbf{c_2}} \mid \mathbf{C} = \mathbf{c_2}, \mathbf{W} = \mathbf{w}] = \\ \text{(consistency)} &= E[D \mid \mathbf{C} = \mathbf{c_1}, \mathbf{W} = \mathbf{w}] - E[D \mid \mathbf{C} = \mathbf{c_2}, \mathbf{W} = \mathbf{w}] \end{aligned} \quad (4.6)$$

■

In other words, based on the assumptions we can estimate the average causal effect with the difference of the average observed outcomes between two sub-populations with different assignments to the events.

4.2 Other Assumptions

Depending on the context, further assumptions can be required in addition, or instead of, the assumptions already made in the previous section [7, 12, 20].

For the causal effects to be identifiable we require the positivity assumption, which implies that all sub-populations are possible and not too rare [34]:

Assumption 4.3: *Positivity*

$$\forall \mathbf{X} = \mathbf{x} : \; p_{\mathbf{x}} \neq 0, p_{\mathbf{x}} \neq 1 \quad (4.7)$$

■

In some cases and for some models it can also be required to make the rare disease assumption:

Assumption 4.4: *Rare disease*
The outcome in all strata being studied is rare, i.e. all $p_{\mathbf{x}}$ are small ■

How small the probabilities should be to be small enough is hard to say, and depends on how large error one tolerates. Some authors have considered probabilties below 10% as small [40, 54]. A square root transformation of the estimate can in some cases be used to decrease the error that this assumption introduces [50].

Note that, the set $\mathbf{x}$ is a set of specific values for the set of events, $\mathbf{X}$. The implication is that it is not enough that the outcome for all the events by themselves is rare, but all combinations of the events are also required to be rare. This means that it is harder to

know than it might seem if the assumption holds. For instance, we might know that some disease is rare in the general population, but we will likely not know if the disease is rare in some particular combination of the events in the study. Especially, since part of the goals for a study often is to estimate the risk of disease in that subpopulation or the risk relative to some other subpopulation.

5 Sufficient Cause Models

Based on the potential outcome model introduced in the previous section we will now define the sufficient cause model and summarize parts of the results in [55], which defines a general theory for the framework with arbitrary d.

The sufficient cause model is closely connected to additive interaction since the presence of interaction on the additive scale implies interaction in the sufficient cause model [52, 53, 56, 57]. Additive interaction can be used to draw conclusions if some sufficient cause is present, i.e., if some individuals require the presence of a combination of the exposures to get the outcome. However, the interpretation of the presence of additive interaction is not necessarily that there is some actual mechanism, e.g. chemical reaction, present. The interaction can occur from a large number of different sources such as competing causes, confounding, and transformations [40, 52].

Definition 5.1: *Sufficient cause*
A set $\mathbf{B} \in \mathbb{L}(\mathbf{C})$ for D forms a *sufficient cause for D relative to $\mathbf{C}$ in sub-population* U^* if for all $\mathbf{c} \in \{0,1\}^{|\mathbf{C}|}$ such that $(\bigwedge(\mathbf{B}))_{\mathbf{c}} = 1$, we have that $D_{\mathbf{c}}(u) = 1$ for all $u \in U^* \subseteq U$. ■

In other words, $\mathbf{B}$ is a sufficient cause if there is a subset of the population for which they get the outcome if all the events in $\mathbf{B}$ are true. Based on this definition and the potential outcome model, any intervention setting the events $\mathbf{C} = \mathbf{c}$ with $(\bigwedge(\mathbf{B}))_{\mathbf{c}} = 1$ will cause $D = 1$ for all $u \in U^*$.

Definition 5.2: *Minimal sufficient cause*
A set $\mathbf{B} \subset \mathbb{L}(\mathbf{C})$ is a *mimimal sufficient cause for D relative to $\mathbf{C}$ in sub-population* U^* if $\mathbf{B}$ is a sufficient cause for D in U^*, but no proper subset $\mathbf{B}^* \subset \mathbf{B}$ is also a sufficient cause for D in U^*. ■

Example 5.1: Continuation of Example 2.1. The combination of the mutation and the amino acid is a minimal sufficient cause, since neither the mutation or the amino acid alone are sufficient causes, but they are *component causes*. ■

The sufficient cause model has a strong connection with Boolean functions [31] and related fields, such as digital circuit theory [8]. The sufficient causes are equivalent to implicants, while minimal sufficient causes are equivalent to prime implicants [55].

We can form sets of sufficient causes, and if such a set explains all outcomes in a subpopulation we say that the set is a determinative set of sufficient causes for that subpopulation.

Definition 5.3: *Determinative set of sufficient causes*
A set of sufficient causes for D, $\mathfrak{B} = \{\mathbf{B}_1, \ldots \mathbf{B}_n\} \in \mathbb{P}(\mathbb{L}(\mathbf{C}))$, is said to be *determinative for D (relative to* $\mathbf{C}$*) in subpopulation* U^* if for all $u \in U^*$ and for all $\mathbf{c}$, $D_{\mathbf{c}}(u) = 1$ if and only if $(\bigvee \bigwedge(\mathbf{B}))_{\mathbf{c}} = 1$. ■

A determinative set of sufficient causes is also referred to as a *sufficient cause model* [55].

In most settings it is unlikely that a single determinative set of sufficient causes will be able to explain all outcomes for the whole population. Therefore, different subsets of the population can be required to have different determinative sets of sufficient causes. The sufficient cause representation is the set of those subpopulations and their corresponding determinative sets.

Definition 5.4: *Sufficient cause representation*
A *sufficient cause representation* $(\mathbf{A}, \mathfrak{B})$ for $\mathfrak{D}(\mathbf{C}; U)$ is an ordered set $\mathbf{A} = \langle A_1, \ldots, A_p \rangle$ of binary random variables with $(A_i)_{\mathbf{c}} = A_i$ for all $i, \mathbf{c}$, and a set $\mathfrak{B} = \langle \mathbf{B}_1, \ldots, \mathbf{B}_p \rangle$, with $\mathbf{B}_i \in \mathbb{P}(\mathbb{L}(\mathbf{C}))$, such that for all $u, \mathbf{c}$, $D_{\mathbf{c}(u)} = 1 \Leftrightarrow$ for some j, $A_j(u) = 1$ and $(\bigwedge(\mathbf{B}_j))_{\mathbf{c}} = 1$. ■

Note that an individual can be associated with more than one of the random variables A_i. A_i and the sets $\mathbf{B}_i$ are paired via the orderings of $\mathbf{A}$ and $\mathfrak{B}$. So A_i sets up different pre-existing sub-populations with particular sets of potential outcomes for D. There can be multiple possible sufficient cause representations that describes a population, however, some certain conjunctions can be present in every representation.

Definition 5.5: *Irreducible*
$\mathbf{B} \in \mathbb{P}(\mathbb{L}(\mathbf{C}))$ is *irreducible* for $\mathfrak{D}(\mathbf{C}; U)$ if in every representation $(\mathbf{A}, \mathfrak{B})$ for $\mathfrak{D}(\mathbf{C}; U)$, there exists $\mathbf{B}_i \in \mathfrak{B}$, with $\mathbf{B} \subseteq \mathbf{B}_i$. ■

Irreducible is sometimes referred to as 'sufficient cause interactions' between the components of $\mathfrak{B}$ [57]. However, note that if $\mathbf{B}$ is irreducible then in general it is not true that $\mathbf{B}$ is a sufficient cause, only that there is a sufficient cause that contains $\mathbf{B}$. If a sufficient cause is both irreducible and minimal sufficient it is the same as essential prime implicant [8, 55]. Note that if $|\mathbf{B}| = |\mathbf{C}|$ then $\mathbf{B}$ is a minimal sufficient cause if and only if $\mathbf{B}$ is irreducible.

If some certain conditions for a set of events are met then there has to be irreducibility for that set in the population.

Theorem 5.1: *Let* $\mathbf{C} = \mathbf{C}_1 \dot{\cup} \mathbf{C}_2$, $\mathbf{B} \in \mathbb{P}(\mathbb{L}(\mathbf{C}))$, $|\mathbf{B}| = |\mathbf{C}_1|$. *Then* $\mathbf{B}$ *is irreducible for* $\mathfrak{D}(\mathbf{C}; U)$ *if and only if there exists* $u^* \in U$ *and values* $\mathbf{c}_2$ *for* $\mathbf{C}_2$ *such that:*

(i) $D_{\mathbf{B}=\mathbf{1},\mathbf{C}_2=\mathbf{c}_2^*}(u^*) = 1$

(ii) for all $L \in \mathbf{B}$, $D_{\mathbf{B}\setminus\{L\}} = \mathbf{1}, L = 0, \mathbf{C}_2 = \mathbf{c}_2^*(u^*) = 0$

The conditions (i) *and* (ii) *are equivalent to*

$$D_{\mathbf{B}=\mathbf{1},\mathbf{C}_2=\mathbf{c}_2^*}(u^*) - \sum_{L\in\mathbf{B}} D_{\mathbf{B}\setminus\{L\}} = \mathbf{1}, L = 0, \mathbf{C}_2 = \mathbf{c}_2^*(u^*) > 0. \tag{5.1}$$

Proof: See Theorem 3.2 in [55]. ■

Thus, if the condition is met, there exists at least one individual, $u^* \in U$, that have the outcome $D = 1$ if *every* literal in $\mathbf{B}$ is set to 1, but $D = 0$ if *one* of the literals in $\mathbf{B}$ is set to 0 and the other literals are set to 1. However, the theorem uses the potential outcomes, which as explained earlier can not be directly observed. We can solve this by using the assumptions made in Section 4 and Theorem 4.1. This leads to the next theorem, which can be used in practice.

Theorem 5.2: *Let* $\mathbf{C} = \mathbf{C}_1 \dot\cup \mathbf{C}_2$, $\mathbf{B} \in \mathbb{P}(\mathbb{L}(\mathbf{C}))$, $|\mathbf{B}| = |\mathbf{C}_1|$. *If* $\mathbf{W}$ *is sufficient to adjust for confounding of* $\mathbf{C}$ *on* D*, and for some* $\mathbf{c}_2$*, and* $\mathbf{w}$*,*

$$\begin{aligned} &E[D|\mathbf{B} = 1, \mathbf{C}_2 = \mathbf{c}_2, \mathbf{W} = \mathbf{w}] \\ &- \sum_{L\in\mathbf{B}} E[D \mid \mathbf{B} \setminus \{L\} = 1, L = 0, \mathbf{C}_2 = \mathbf{c}_2, \mathbf{W} = \mathbf{w}] > 0, \end{aligned} \tag{5.3}$$

then $\mathbf{B}$ *is irreducible for* $\mathfrak{D}(\mathbf{C}; U)$.

Proof: See Theorem 4.3 in [55]. ■

6 Additive Interaction

In this section we will summarize some results related to additive interaction and its' estimation. For simplicity and since some results have not yet been generalized to d number of exposures we will be using $d = 2$.

Definition 6.1: For two binary exposures, the interaction contrast (IC) is defined as

$$\text{IC} = p_{11} - p_{10} - p_{01} + p_{00}. \tag{6.1}$$

■

Example 6.1: Continuation of Example 2.2. With the probabilities as in Table 1 IC is

$$\text{IC} = 0.0450 - 0.0067 - 0.0095 + 0.0011 = 0.0299, \tag{6.2}$$

which means that the intervention's effect is 0.0299 higher from the intervention upon smokers with absestos exposure compared to smokers without the exposure. ■

Instead of using the probabilities, it is common to use relative risks, or approximations with odds ratios, with some reference group, $\mathbf{ref}$, which are defined as

$$\text{RR}_{\mathbf{x}} = \frac{p_{\mathbf{x}}}{p_{\mathbf{ref}}}, \tag{6.3}$$

and

$$\text{OR}_{\mathbf{x}} = \frac{p_{\mathbf{x}}}{1-p_{\mathbf{x}}} / \frac{p_{\mathbf{ref}}}{1-p_{\mathbf{ref}}}. \tag{6.4}$$

The relative risks are often approximated with odds ratios using the rare disease assumption, Assumption 4.4, because the relative risk can not be estimated from case control data while odds ratios can. However, with some study designs the rare disease assumption is not required for the odds ratios to approximate the relative risks [28].

Additive interaction is measured using differences between ratios, which causes problems if some ratios are harmful (above 1) while others are protective (below 1) [29]. The reference group then needs to be chosen so that all ratios are harmful.

Example 6.2: Suppose we have a one exposure, X, and that its relative risk is $\text{RR}_X = 2$, with reference group $X = 0$. If we change the reference to $X = 1$, i.e. if look at the exposure of $\bar{X}$, then $\text{RR}_{\bar{X}} = \frac{1}{2}$. The difference between these two ratios is not zero even though the effect is the same. ■

Using the risk ratios four measures of interaction can be dervied from IC [40]. The relative excess risk due to interaction (RERI) is

$$\text{RERI} = \text{RR}_{11} - \text{RR}_{10} - \text{RR}_{01} + 1. \tag{6.5}$$

The intepretation of RERI is the same as IC but on an ratio scale instead of the scale of the probabilities.

The other three measures are proportional measures, two attributable proportional measures, AP and AP*, and the synergi index, SI. They are defined as

$$\text{AP} = \frac{\text{RERI}}{\text{RR}_{11}}, \tag{6.6}$$

$$\text{AP}^* = \frac{\text{RERI}}{\text{RR}_{11} - 1}, \tag{6.7}$$

and

$$\text{SI} = \frac{\text{RR}_{11} - 1}{(\text{RR}_{10} - 1) + (\text{RR}_{01} - 1)}. \tag{6.8}$$

The attributable proportions both measures the proportion of interaction in the group with both exposures, but they differ in which proportion is used. AP is the proportion of the *disease* in the doubly exposed group that is due to the interaction, while AP* is the *effect* that is due to interaction in the doubly exposed group [51]. AP^* more closely follows the intuitive interpretation one would expect as demonstrated in the following example.

Example 6.3: Let $\text{RR}_{11} = 2$, $\text{RR}_{10} = 1$, and $\text{RR}_{01} = 1$. Then $\text{AP} = \frac{1}{2} = 50\%$, while $\text{AP}^* = \frac{1}{1} = 100\%$. One would expect 100% from AP since $\text{RR}_{10} = 1$, $\text{RR}_{01} = 1$, but this is not case since AP is the proportion of the disease in the doubly exposed group, and not the effect.

In other words, AP is the ratio of the number of individuals with the outcome that are doubly exposed compared to the number of all other individuals with the outcome. Even with $\text{RR}_{10} = 1$, and $\text{RR}_{01} = 1$ there are individuals in the population that are exposed to only one of the exposures which have the outcome since $p_{10} = p_{01} = p_{00}$ and due to the positivity assumption $p_{00} \neq 0$. ■

The synergy index is similar to the attributable proportions, and it measures how much more the effect in the doubly exposed group exceeds 1, i.e. the ratio when there is no effect, compared to how much more the exposures separately exceeds 1 together [54].

6.1 Connection with Sufficient Cause Models

We can use the measures of additive interaction define above with Theorem 5.2 to test for irreducibility as shown in the following example.

Example 6.4: Suppose we have a population, two events, $\mathbf{C} = \{\mathbf{C}_1, \mathbf{C}_2\}$, and no confounders. From Equation 5.3 with $\mathbf{B} = \{\mathbf{C}_1, \mathbf{C}_2\}$, $\mathbf{B}$ is irreducible, and also minimal sufficient since $|\mathbf{B}| = |\mathbf{C}|$, if

$$\begin{aligned} &E[D \mid \mathbf{C}_1 = 1, \mathbf{C}_2 = 1] - E[D \mid \mathbf{C}_1 = 1, \mathbf{C}_2 = 0] \\ &- E[D \mid \mathbf{C}_1 = 0, \mathbf{C}_2 = 1] > 0 \end{aligned} \tag{6.10}$$

Meaning that if the condition is met, some individuals in the population require both events in $\mathbf{C}$ to get the outcome. Using the probabilities as estimates for the potential outcome averages the condition is equivalent to

$$p_{11} - p_{10} - p_{01} > 0, \tag{6.11}$$

which can be written as

$$\text{IC} > p_{00}. \tag{6.12}$$

We can also express the condition using the interaction measures from the relative risks, such as RERI,

$$\text{RERI} > 1. \tag{6.13}$$

■

But the condition for presence of additive interaction is $\text{IC} \neq 0$, so what does additive interaction in general imply for the sufficient cause model? The implication is somewhat unexpected, it implies synergism between A and B, where any combination of A, B, and the outcome can be in the form of its negation [53]. Hence, some types of additive interaction are connected with synergism between causes for $\bar{D}$, for instance response pattern 2 in Table 2.

Example 6.5: Suppose we have two exposures, A and B. An individual with subadditive interaction and belonging to response type 2 in Table 2 is shown in Table 3 with the potential outcomes for both D and $\bar{D}$.

For D the individual have the outcome if any of the exposures are present. However, the sufficient causes for D are the two *separate* causes, A and B. I.e. there is no sufficient cause synergism between A and B. However, if we examine $\bar{D}$ instead, the individual only gets the outcome $\bar{D}$ if both exposures are not present. Then the sufficient cause for $\bar{D}$ is the combination of $\bar{\text{A}}$ and $\bar{\text{B}}$. Thus, there is synergism between $\bar{\text{A}}$ and $\bar{\text{B}}$ for $\bar{D}$.

The two exposures are in this case competing to cause the outcome, which causes the subadditive interaction since an individual can only get the outcome once. For two exposures there are three more competing types, 3, 5, and 9 [17, 53]. ■

Outcome \ Exposures	11	01	10	00
D	1	1	1	0
$\bar{D}$	0	0	0	1

Table 3: Potential outcomes for an individual of type 2.

6.2 Estimation and Modeling of Additive Interaction

There are several statistical methods that can be used to estimate the ratios, each with their own advantages and disadvantages. For estimating additive interaction, the most common is logistic regression [52, 61]

$$\ln\left(\frac{p_{xyz}}{1-p_{xyz}}\right) = \alpha + \beta_1 x + \beta_2 y + \beta_3 xy + \gamma_z z. \tag{6.14}$$

An alternative model is the linear odds model [48, 61],

$$\frac{p_{xyz}}{1-p_{xyz}} = a^* + b_1^* x + b_2^* y + b_3^* xy + g_z^* z. \tag{6.15}$$

Both models as expressed here are not fully saturated due to the absence of coefficients corresponding to interaction between the exposures and the covariates. The logistic model is a multiplicative model, while the linear odds model is additive; which means that the coefficient for the interaction term in the logistic model corresponds to multiplicative interaction, not additive [16, 48]. Nonetheless, the logistic model can be used for estimating additive interaction in the true underlying model in some situations, but there can be issues if covariates are included [16, 48]. If the terms for the interaction between the exposures and covariates are not included then it is implied that there is additive interaction between

the exposures and the covariates [18, 48]. This means that the interaction estimated from the logistic model can be incorrect [48].

Suppose that true underlying model is linear, i.e.

$$p_{xyz} = a + b_1 x + b_2 y + b_3 xy + g_z z, \tag{6.16}$$

then it follows that the true relative risk is

$$\frac{p_{xyz}}{p_{00z}} = \frac{a + b_1 x + b_2 y + b_3 xy + g_z z}{a + g_z z}, \tag{6.17}$$

which depends on the covariate z. This is referred to as the *misspecification problem* [48]. The relative risks and the odds ratios from the linear model depend on the covariate, but the odds ratio from the logistic model does not, unless the model is fully saturated.

RERI based on the true linear model is also dependent on the covariates,

$$\text{RERI}_{linear} = \frac{b_3}{a + g_z z} \tag{6.18}$$

although, AP^* and SI are not,

$$\text{AP}^*_{linear} = \frac{b_3}{b_1 + b_2 + b_3}, \tag{6.19}$$

$$\text{SI}_{linear} = \frac{b_1 + b_2 + b_3}{b_1 + b_2}. \tag{6.20}$$

This is referred to as the *uniqueness problem* [48], the value of RERI depends on the covariates while the additive interaction measures from the logistic model does not.

Both problems can be solved by using the linear odds model together with AP^* or SI [48]. However, it is not a perfect solution since the linear odds model can lead to negative odds and fail to converge [48, 58, 61], also the maximum likelihood estimators can have problems when used with continuous covariates [58].

These issues and combined with the fact that the distributions for the interaction measures are complicated means that it is not trivial to estimate the measures. The common approach to estimate the variance and confidence intervals for odds ratios is to use a log transformation to avoid issues with estimating the variance [1, 63]. However, the log transformation can not be used for the interaction measures because they are based on the difference between ratios. Instead approximative methods have to be used. The issue is related to the problem of estimation of the variance of the binomial proportion, i.e., the estimation of the variance for the estimate of p in a binomial distribution [2, 9, 10, 30].

The first approach to estimate the confidence interval of RERI used the delta method [21], which has also been used in several later papers [4, 22, 23, 25]. However, the methods using the delta method do not have the expected coverage of the confidence interval; likely because of the problem the variance for the binomial proportion [5, 64]. Other methods such as bootstrap [5] and MOVER [64] do not have this issue.

7 Summary of Papers

7.1 Paper A, On the Existence of Suitable Models for Additive Interaction with Continuous Exposures

Suppose that the exposures are continuous instead of binary, is the estimation of additive interaction still useful? The sufficient cause model is not defined for continuous exposures, however additive interaction can also be used for interventions as shown in Example 2.2 and 6.1.

Interventions on continuous exposures are more complex than in the binary case, since there are more possibilities for what the intervention does rather than just setting the binary exposure as 1 or 0.

Definition 7.1: *Intervention on continuous exposures*
The intervention on the continuous exposures G and E are functions, $I_G(g)$ and $I_E(e)$, that transforms the exposures for all individuals in the population. Let g_0, e_0 be the individual's exposure levels. The effect of the intervention is then defined as

$$p_{I_G(g_0),I_E(e_0)} - p_{g_0,e_0}. \tag{7.1}$$

■

Definition 7.2: *Marginal effect*
The marginal effect from an intervention, $I_G(g_0)$, on a *single* continuous exposure G is defined as

$$p_{I_G(g_0)} - p_{g_0}. \tag{7.2}$$

An intervention, $I_G(g)$, has no marginal effect if

$$\forall g_0 \in \mathbb{R}\ p_{I_G(g_0)} - p_{g_0} = 0. \tag{7.3}$$

■

We can define IC in the same manner as for the binary case, though, it is no longer unique, since instead of choosing between intervening on two different groups, based on a binary exposure, the scale is now continuous and there are multiple ways to define the groups. Hence, there are several possible types of additive interaction.

We say that there is *single additive interaction* present for a value of an exposure, $G = g$, if additive interaction is present between g and some values, $E = e_a$, $E = e_b$, for the other exposure E. If there, for all values of the exposure is no single additive interaction then there is no *exposure additive interaction* for that particular exposure.

We set the functions for interventions, $I_G(g)$ and $I_E(e)$, as strictly decreasing functions and show that the there is no model can fulfill both of the following criteria:

i. There exist some values for the model parameters for which both exposures have marginal effects and no additive interaction on the exposure level for both exposures.

ii. There exist some values for the model parameters for which there are no marginal effects and there is additive interaction between the exposures.

The main result is a theorem that shows that there is no model for the probabilities that can fullfill the criteria given above.

Theorem 7.1: *With the interventions set as a strictly decreasing function there is no model that can fulfill both of the criteria for the model.*

Our result also implies that the methods for additive interaction with continuous interaction in [26, 27, 54] are flawed.

The candidate's contribution: The candidate suggested topic, showed the proofs and wrote the manuscript.

7.2 Paper B, On Probabilistic Multifactor Potential Outcome Models

The sufficient cause framework is connected to Boolean functions as mentioned in Section 5. In this paper we formally define the sufficient cause model based on Boolean functions, and show how Fourier expansion and Blake Canonical form can be used for new insights about the sufficient causes. We also present a new model based on the multifactor potential outcome model (MFPO) [14] and independence of causal influence models (ICI) [13, 19, 62].

The MFPO model introduces unknown complementary causes, ξ_i, to the sufficient causes, $\mathbf{B}_i$. The sufficient cause is therefore not complete unless ξ_i is also present in addition to the events in $\mathbf{B}_i$. We show that the MFPO model is equivalen to the sufficient cause representation.

Theorem 7.2: *Let* $\mathbf{A} = \langle \xi_1, \ldots, \xi_k \rangle$ *and* $\mathfrak{B}$ *as defined above. Then* $(\mathbf{A}, \mathfrak{B})$ *is a sufficient cause model representation of some* $D(\mathbf{C}; U)$.

The model so far has been deterministic. Based on the probabilistic potential outcomes in [36], and the ICI model we express a probabilistic version of the model. In the ICI model the effect from the events $\mathbf{X}$ on D are mediated through a layer of binary variables ω. The model is a form a of a Bayeseian network, and a graphical representation is shown in Figure 1. Since it is a Bayeseian network it can be written as

$$P_\alpha(D = \delta | \underline{x}) \equiv \sum_{\omega | \alpha(\omega) = \delta} \prod_{j=1}^{d} p(\omega_j \mid x_j), \tag{7.4}$$

where α is the interaction function. In this case the Boolean function representing the sufficient cause model.

The probabilities for the potential outcomes in the noisy MFPO model can then be derived using the ICI results which leads to the main result of the paper.

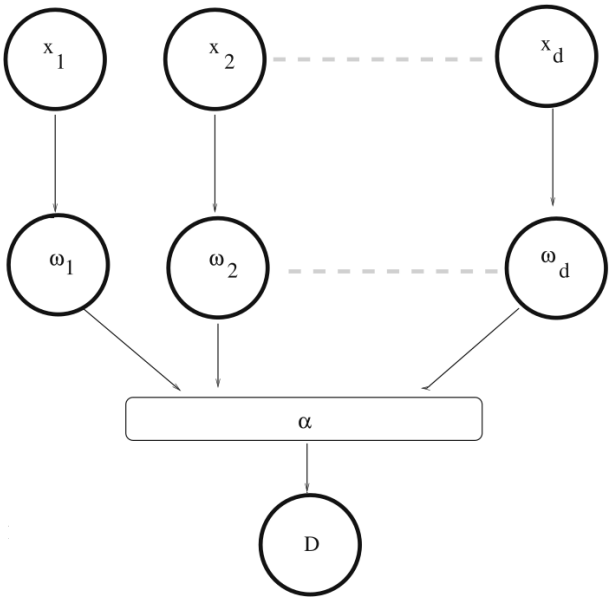


Figure 1: Graphical representation of the ICI model.

Proposition 7.3: *It holds for the noisy MFPO$_k$, $k > 2$ that*

$$P_{MFPO_k}(D = 1|\underline{x}) = g_1 \prod_{j=2}^{k}(1 - g_j) + \sum_{j=2}^{k-1} g_j \prod_{i=j+1}^{k} (1 - g_i) + g_k \tag{7.5}$$

and for $k = 2$

$$P_{MFPO_2}(D = 1|\underline{x}) = (1 - g_2)g_1 + g_2, \tag{7.6}$$

and $p_1 = g_1$, where for $i = 1, 2, \ldots, k$

$$g_i = P_{\beta_i}(D = 1|\underline{x}) \cdot \theta_i \tag{7.7}$$

An interesting consequence of the Fourier expansion of the Boolean function is that the probabilities for the potential outcomes follow a linear model. It can therefore be argued that for binary exposures the linear odds model, Equation 6.15, is closer to the true model than logistic regression, Equation 6.14. Though, as discussed in Section 6, the linear odds model is not without significant drawbacks.

The candidate's contribution: Co-author suggested the topic and wrote most of the manuscript. The candidate's work was mostly on the connections between the models and related to the linear model, which roughly corresponding to sections 3.3-3.4 and 10.

7.3 Paper C, Measures of Additive Interaction and Effect Direction

The measures of additive interaction defined in Section 6 are defined using risk ratios, and as described in that section the reference group needs to be chosen so that all the ratios are above one. However, if a ratio is above or below one in the data is random, which means that the choice of reference also becomes random. This can impact the estimation of the variance of the interaction.

We define the direction of effects as the following:

Definition 7.3: *Direction of effect*
A set of exposures, x, is a harmful exposure and has risk direction of effect if $p_x > p_{\bar{x}}$. If $p_x < p_{\bar{x}}$ then the effect is protective. ■

Definition 7.4: *Conditional direction of effect*
A set of exposures, x, is a risk or protective exposure condtional on a set of exposures c, $c \not\subset e$, if $p_{x,c} > p_{\bar{x},c}$ respectively $p_{x,c} < p_{\bar{x},c}$ ■

If all ratios are risk then, with reference group $r = \{r_1, r_2\}$ it is equivalent to that all three of the following effects are harmful: $\{\bar{r_1}, \bar{r_2}\}$, $\{\bar{r_1} \mid r_2\}$, and $\{\bar{r_2} \mid r_1\}$.

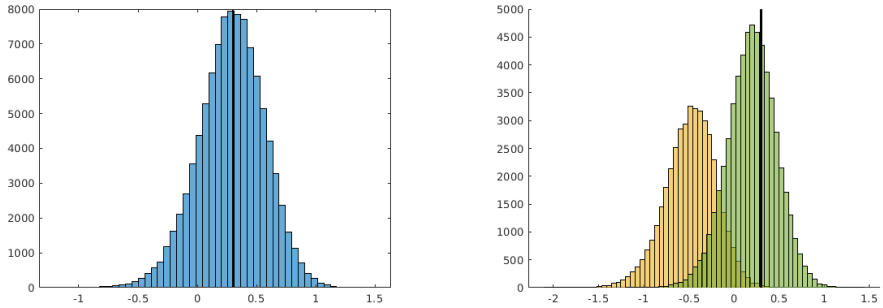
We illustrate the problem of the variance of the interaction measures with an example.

Example 7.1: 100 000 cohorts were simulated with 10 000 individuals in each. The true probabilities for the outcome were set as $p_{11} = 0.07$, $p_{10} = 0.06$, $p_{01} = 0.031$, $p_{00} = 0.03$ so there is high uncertainty in the cohorts if the second exposure is harmful or protective. For each cohort RERI was calculated using either the true reference group, i.e. the second exposure is harmful, or the reference group as estimated in the cohort. The histograms for the estimated RERI are shown in Figure 2.

In Figure 2b there are two different distributions; The left distribution is when the second exposure is estimated as protective, and the right distribution when the exposure is estimated as harmful. The high uncertainty about the second exposure's conditional effect direction means we do not know which distribution is the true one. Note that the right distribution is not centered on the true value of RERI even though its reference group is the true one because the underlying distributions for the ratios are truncated, and the truncation skews the distribution to the left in this case. ■

However, the effect of the uncertainty of reference group is smaller than it might seem from the example. The left peak in Figure 2b is estimated with viewing the second exposure as protective so the interpretation of RERI has to account for that. In other words the interpretation for the distribution to the right with superadditive interaction is the same as for subadditive interaction in the left distribution. Though, this can lead to errors in the confidence interval and hypothesis testing depending on hypothesis, and which model is used.

The candidate's contribution: The candidate suggested topic, showed the proofs and wrote the manuscript.



(a) Classically assumed distribution for RERI (b) True distribution for RERI including the randomness of the reference group

Figure 2: Histograms from simulating cohorts and estimating RERI with and without accounting for the reference group. Vertical black line is RERI calculated from the true probabilities.

References

[1] A. Agresti. *Categorical Data Analysis (2nd Edition)*. John Wiley & Sons, 2002.

[2] A. Agresti and B. A. Coull. Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, 52(2):119–126, 1998.

[3] A. Ahlbom and L. Alfredsson. Interaction: a word with two meanings creates confusion. *European Journal of Epidemiology*, 20(7):563–564, 2005.

[4] T. Andersson, L. Alfredsson, H. Källberg, S. Zdravkovic, and A. Ahlbom. Calculating measures of biological interaction. *European Journal of Epidemiology*, 20(7):575–579, 2005.

[5] S. F. Assmann, D. W. Hosmer, S. Lemeshow, and K. A. Mundt. Confidence intervals for measures of interaction. *Epidemiology*, pages 286–290, 1996.

[6] W. J. Blot and N. E. Day. Synergism and interaction: are they equivalent? *American Journal of Epidemiology*, 110(1):99–100, 1979.

[7] H. E. Brady. Models of causal inference: Going beyond the Neyman-Rubin-Holland theory. In *Annual Meetings of the Political Methodology Group*, 2002.

[8] F. M. Brown. *Boolean reasoning: the logic of Boolean equations*. Springer Science & Business Media, 2012.

[9] L. D. Brown, T. T. Cai, and A. DasGupta. Interval estimation for a binomial proportion. *Statistical Science*, pages 101–117, 2001.

[10] L. D. Brown, T. T. Cai, A. Dasgupta, et al. Confidence intervals for a binomial proportion and asymptotic expansions. *The Annals of Statistics*, 30(1):160–201, 2002.

[11] H. J. Cordell. Detecting gene–gene interactions that underlie human diseases. *Nature Reviews Genetics*, 10(6):392, 2009.

[12] A. P. Dawid. Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95(450):407–424, 2000.

[13] F. J. Dıez and M. J. Druzdzel. Canonical probabilistic models for knowledge engineering. *UNED, Madrid, Spain, Technical Report CISIAD-06-01*, 2006.

[14] W. D. Flanders. On the relationship of sufficient component cause models with potential outcome (counterfactual) models. *European Journal of Epidemiology*, 21(12):847–853, 2006.

[15] B. Goudey, D. Rawlinson, Q. Wang, F. Shi, H. Ferra, R. M. Campbell, L. Stern, M. T. Inouye, C. S. Ong, and A. Kowalczyk. GWIS-model-free, fast and exhaustive search for epistatic interactions in case-control gwas. *BMC genomics*, 14(3):S10, 2013.

[16] S. Greenland. Additive risk versus additive relative risk models. *Epidemiology*, pages 32–36, 1993.

[17] S. Greenland and C. Poole. Invariants and noninvariants in the concept of interdependent effects. *Scandinavian Journal of Work, Environment & Health*, pages 125–129, 1988.

[18] S. Greenland, J. M. Robins, and J. Pearl. Confounding and collapsibility in causal inference. *Statistical Science*, pages 29–46, 1999.

[19] D. Heckerman and J. S. Breese. Causal independence for probability assessment and inference using bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 26(6):826–831, 1996.

[20] P. W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

[21] D. W. Hosmer and S. Lemeshow. Confidence interval estimation of interaction. *Epidemiology*, 3(5):452–456, 1992.

[22] O. Hössjer, L. Alfredsson, A. K. Hedström, M. Lekman, I. Kockum, and T. Olsson. Quantifying and estimating additive measures of interaction from case-control data. *Modern Stochastics: Theory and Applications*, 4(2):109–125, 2017.

[23] O. Hössjer, I. Kockum, L. Alfredsson, A. K. Hedström, T. Olsson, and M. Lekman. A general framework for and new normalization of attributable proportion. *Epidemiologic Methods*, 6(1), 2017.

[24] D. Hume. *An enquiry concerning human understanding*. 1748. Reprinted in 1878.

[25] H. Källberg, A. Ahlbom, and L. Alfredsson. Calculating measures of biological interaction using R. *European Journal of Epidemiology*, 21(8):571–573, 2006.

[26] M. Katsoulis and C. Bamia. Additive interaction between continuous risk factors using logistic regression. *Epidemiology*, 25(3):462–464, 2014.

[27] M. J. Knol, I. van der Tweel, D. E. Grobbee, M. E. Numans, and M. I. Geerlings. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. *International Journal of Epidemiology*, 36(5):1111–1118, 2007.

[28] M. J. Knol, J. P. Vandenbroucke, P. Scott, and M. Egger. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *American Journal of Epidemiology*, 168(9):1073–1081, 2008.

[29] M. J. Knol, T. J. VanderWeele, R. H. Groenwold, O. H. Klungel, M. M. Rovers, and D. E. Grobbee. Estimating measures of interaction on an additive scale for preventive exposures. *European Journal of Epidemiology*, 26(6):433–438, 2011.

[30] R. G. Newcombe. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine*, 17(8):873–890, 1998.

[31] R. O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[32] J. Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

[33] J. Pearl. Direct and indirect effects. In *Proceedings of the seventeenth conference on uncertainty in artificial intelligence*, pages 411–420. Morgan Kaufmann Publishers Inc., 2001.

[34] M. L. Petersen, K. E. Porter, S. Gruber, Y. Wang, and M. J. van der Laan. Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research*, 21(1):31–54, 2012.

[35] R. E. Quandt. A new approach to estimating switching regressions. *Journal of the American Statistical Association*, 67(338):306–310, 1972.

[36] R. R. Ramsahai. Probabilistic causality and detecting collections of interdependence patterns. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):705–723, 2013.

[37] D. H. Rehkopf, M. M. Glymour, and T. L. Osypuk. The consistency assumption for causal inference in social epidemiology: When a rose is not a rose. *Current Epidemiology Reports*, 3(1):63–71, 2016.

[38] J. Reiss. Causation in the social sciences: Evidence, inference, and purpose. *Philosophy of the Social Sciences*, 39(1):20–40, 2009.

[39] M. D. Ritchie, L. W. Hahn, N. Roodi, L. R. Bailey, W. D. Dupont, F. F. Parl, and J. H. Moore. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *The American Journal of Human Genetics*, 69(1):138–147, 2001.

[40] K. J. Rothman, S. Greenland, and T. L. Lash. *Modern epidemiology*. Lippincott Williams & Wilkins, 2008.

[41] K. J. Rothman, S. Greenland, and A. M. Walker. Concepts of interaction. *American Journal of Epidemiology*, 112(4):467–470, 1980.

[42] A. D. Roy. Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2):135–146, 1951.

[43] D. B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688, 1974.

[44] D. B. Rubin. Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, pages 34–58, 1978.

[45] D. B. Rubin. Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 74(366a):318–328, 1979.

[46] D. B. Rubin. Formal mode of statistical inference for causal effects. *Journal of Statistical Planning and Inference*, 25(3):279–292, 1990.

[47] R. Saracci. Interaction and synergism. *American Journal of Epidemiology*, 112(4):465–466, 1980.

[48] A. Skrondal. Interaction as departure from additivity in case-control studies: a cautionary note. *American Journal of Epidemiology*, 158(3):251–258, 2003.

[49] J. Splawa-Neyman, D. M. Dabrowska, T. P. Speed, et al. On the application of probability theory to agricultural experiments. *Statistical Science*, 5(4):465–472, 1990. Reprint.

[50] T. VanderWeele. On a square-root transformation of the odds ratio for a common outcome. *Epidemiology*, 28(6):e58, 2017.

[51] T. J. VanderWeele. Reconsidering the denominator of the attributable proportion for interaction. *European Journal of Epidemiology*, 28(10):779–784, 2013.

[52] T. J. VanderWeele. *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press, 2016.

[53] T. J. VanderWeele and M. J. Knol. Remarks on antagonism. *American Journal of Epidemiology*, 173(10):1140–1147, 2011.

[54] T. J. VanderWeele and M. J. Knol. A tutorial on interaction. *Epidemiologic Methods*, 3(1):33–72, 2014.

[55] T. J. VanderWeele and T. S. Richardson. General theory for interactions in sufficient cause models with dichotomous exposures. *Annals of Statistics*, 40(4):2128, 2012.

[56] T. J. VanderWeele and J. M. Robins. The identification of synergism in the sufficient-component-cause framework. *Epidemiology*, 18(3):329–339, 2007.

[57] T. J. VanderWeele and J. M. Robins. Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika*, 95(1):49–61, 2008.

[58] S. Wachholder. Binomial regression in GLIM: Estimating risk ratios and risk differences. *American Journal of Epidemiology*, 123(1):174–184, 1986.

[59] X. Wan, C. Yang, Q. Yang, H. Xue, X. Fan, N. L. Tang, and W. Yu. BOOST: A fast approach to detecting gene-gene interactions in genome-wide case-control studies. *The American Journal of Human Genetics*, 87(3):325–340, 2010.

[60] S. J. Winham and J. M. Biernacka. Gene–environment interactions in genome-wide association studies: current approaches and new directions. *Journal of Child Psychology and Psychiatry*, 54(10):1120–1134, 2013.

[61] L. N. Yelland, A. B. Salter, and P. Ryan. Relative risk estimation in randomized controlled trials: a comparison of methods for independent observations. *The International Journal of Biostatistics*, 7(1):1–31, 2011.

[62] N. L. Zhang and D. Poole. Exploiting causal independence in bayesian network inference. *Journal of Artificial Intelligence Research*, 5:301–328, 1996.

[63] G. Zou and A. Donner. Construction of confidence limits about effect measures: a general approach. *Statistics in Medicine*, 27(10):1693–1702, 2008.

[64] G. Y. Zou. On the estimation of additive interaction by use of the four-by-two table and beyond. *American Journal of Epidemiology*, 168(2):212–224, 2008.

Part II: Scientific Papers

Paper A

On the Existence of Suitable Models for Additive Interaction with Continuous Exposures

On the Existence of Suitable Models for Additive Interaction with Continuous Exposures

by

Daniel Berglund, Claudia Carlucci, Helga Westerlind,
and Timo Koski

Abstract

Additive interaction can be of importance for public health interventions and it is commonly defined using binary exposures. There has been expansions of the models to also include continuous exposures, which could lead to better and more precise estimations of the effect of interventions. In this paper we define the intervention for a continuous exposure as a monotonic function. Based on this function for the interventions we prove that there is no model for estimating additive interactions with continuous exposures for which it holds that; (i) both exposures have marginal effects and no additive interaction on the exposure level for both exposures, (ii) neither exposure has marginal effect and there is additive interaction between the exposures. We also show that a logistic regression model for continuous exposures will always produce additive interaction if both exposures have marginal effects.

Keywords: Additive Interaction; Multiplicative Interaction; Logistic Regression; Linear Odds; Continuous Exposures; Public Health; Interventions

1 Introduction

For some diseases and traits, a combination of factors is required for the diseases to occur, and factors can also modify each others effect, and significantly increase or decrease in strength. In epidemiological and social science research much work has been focused on finding and estimating such effects [20, 28, 32]. For example smoking combined with the genotype HLA-DRB1 SE substantially increases the risk for Rheumatoid Arthritis [11]. However, the exact mechanisms and potential mechanistic interactions are in general hard to define and impossible to estimate from data. This is referred to as biological, or sufficient

cause, interaction among others, and does not necessarily imply that some mechanisms behind the factors are directly interacting, such as a chemical reaction [20, 28].

Statistical interaction on the other hand has a clear definition, it is when the effect from one exposure depends on the level of another exposure [20, 28]. Statistical interactions are scale dependent and different scales can lead to different conclusions. An example is the interaction between smoking and asbestos on the risk of lung cancer where the risk from asbestos is higher for nonsmokers than for smokers on the ratio scale, however the opposite is true on the additive scale [20].

For public health it can be argued that the additive scale is the preferred scale as it leads to the most effective interventions when choosing which group to intervene upon [1, 4, 20, 21, 23, 33]. The additive scale is also connected to the sufficient-cause model, where the presence of interaction on the additive scale implies interaction in the sufficient-cause model, and thereby that both factors are involved in the causal pathway [28, 30, 32–34].

Most research on additive interaction have focused on binary exposures with few papers in the literature related to additive interaction using continuous exposures [3, 12, 13, 29, 31]. Instead of dichotomizing the continuous exposures, as in [3, 29], in the ideal case the ideal model would use the continuous exposures directly. Such an approach could improve the models and lead to new insights, as dichotomizing can cause a loss of information [22].

We will show in this paper that there is no model for additive interaction with continuous exposures that met the criteria for a suitable model. Our result also implies that the models with continuous exposures for additive interaction in [12, 13, 31] are flawed.

In Section 2 we summarize some of the background for interaction when the exposures are binary. Then in Section 3 we will define interventions and interaction for continuous exposures. In that section we also show the main result; that there is no model that can meet the criteria for a suitable model.

2 Interaction with Binary Exposures

We are going to start with a summary of the background for additive and multiplicative interaction with binary exposures. Let D denote a binary outcome, e.g., disease, and let X and Y be two binary exposures. We assume that there is adjustment for confounding, and that after the adjustment there is no confounding, this is know as the conditional exchangeability assumption [28].

Using the potential outcomes model [8, 25], and with exposure levels set to $X = x$ and $Y = y$ let

$$p_{x,y} = P(D = 1|X = x, Y = y). \tag{2.1}$$

We make two additional assumptions; First, for the causal effects to be identifiable we require the positivity assumption to be true, i.e., that no $p_{x,y}$ is either zero or one [17, 28]. Second, the effect of the exposures needs to be consistent for all individuals [18, 28];

Formally, if we would intervene upon an exposure for an individual the effect on D would be the same as if the individual would have had that exposure level in the first place [18].

To make it easier to interpret and to be able to estimate the additive interaction in case-control studies relative risks (RR) and odds ratios (OR) are used instead of the probabilities. The relative risks can be approximated with odds ratios either if case-control data is used and the rare disease assumption is met or if certain study designs are used [14]. Note that the rare disease assumption applies to the risk in all stratas studied, and not only the prevalence in the general population as a rare disease in the general population might not be rare in some specific strata.

The relative risks and odds ratios for the exposures $X = x$ and $Y = y$ with the reference group *ref* are

$$\text{RR}_{x,y}^{ref} = \frac{p_{x,y}}{p_{ref}} \tag{2.2}$$

and

$$\text{OR}_{x,y}^{ref} = \frac{\frac{p_{x,y}}{1-p_{x,y}}}{\frac{p_{ref}}{1-p_{ref}}}. \tag{2.3}$$

We now define what interventions on binary exposures are based on previous literature [20].

Definition 2.1: *Intervention on binary exposure*
We define the intervention on the binary exposures X and Y as changing the population's exposures with some binary values x_d and y_d respectively from the reference level x_0, y_0. Then

$$p_{x_0+x_d,y_0+y_d} - p_{x_0,y_0} \tag{2.4}$$

is the effect of the intervention. ■

As mentioned in the Introduction, interactions are scale dependent, and the two most common scales are multiplicative (ratio scale) and additive. As the name suggests, in the multiplicative model for interaction there is interaction if the effect for the doubly exposed group does not follow multiplicative scaling, while in the additive model there is interaction if the effect does not follow additive scaling. The main criticism against multiplicative interaction is that it is not connected to any actual interaction on the biological level [20, 31]. However one could make similar arguments against additive interaction, as even if additive interaction implies sufficient cause interaction, sufficient cause interaction does not imply any chemical or biological interaction [20]. Additive interaction can occur without interaction, for instance from competition between exposures, e.g., an individual can not die from cancer if they died in a car crash [6]. It is recommended that both additive and multiplicative measures are estimated in studies [15, 31].

2.1 Multiplicative Interaction

The multiplicative interaction measure between two exposures is defined as the following [20, 31].

Definition 2.2: *Multiplicative interaction*
The multiplicative interaction measure for two exposures is

$$\mathrm{M} = \frac{p_{11}p_{00}}{p_{10}p_{01}}. \tag{2.5}$$

■

If the measure is one then there is no multiplicative interaction. Expressed using relative risks, multiplicative interaction is

$$\mathrm{M} = \frac{\mathrm{RR}_{11}}{\mathrm{RR}_{10}\mathrm{RR}_{01}}. \tag{2.6}$$

The reference group for the relative risks are cancelled out so the choice of reference does not matter, and the value of the measure also is the same as when probabilities are used.

2.2 Additive Interaction

Additive interaction can be derived based on interventions [20]. If we would intervene on Y by setting $Y = 1$, for either the individuals with $X = 0$ or with $X = 1$, in which group would the intervention have the most effect? The effect for $X = 0$ is $p_{01} - p_{00}$ and the effect for $X = 1$ is $p_{11} - p_{10}$; thus the difference between these two effects is $(p_{01} - p_{00}) - (p_{11} - p_{10})$. Rearranging leads to the interaction contrast (IC),

$$\mathrm{IC} = p_{11} - p_{10} - p_{01} + p_{00}. \tag{2.7}$$

IC measures the amount of additive interaction: if there is no difference between the effects (i.e., no additive interaction) then the contrast is zero. If there is interaction on the additive scale then IC represents the effect that is lost (or gained) by intervening on the wrong (or correct group). Negative IC is referred to as subadditive interaction while positive IC is referred to as superadditive interaction [30]. An interesting note is that it does not matter whether the intervention is considered on X or Y, or if the intervention is to set the exposure to zero or one, the value of IC is the same, only its sign changes.

Additive interaction can also be derived from the sufficient cause model and different inequalities using IC can also imply various types of sufficient cause interaction, such as synergism between X and $\bar{Y}$ [20, 28, 32]. $IC \neq 0$ implies the presence of interaction, but is not a necessary condition, as there can be interaction in the true underlying sufficient cause model even if $\mathrm{IC} = 0$ [20].

Using relative risks for additive interaction the following measures of additive interaction can be derived from the IC [20].

$$\mathrm{RERI} = \mathrm{RR}_{11}^{00} - \mathrm{RR}_{10}^{00} - \mathrm{RR}_{01}^{00} + 1 \tag{2.8}$$

$$\text{AP} = \frac{\text{RERI}}{\text{RR}_{11}^{00}} \tag{2.9}$$

$$\text{AP}^* = \frac{\text{RERI}}{\text{RR}_{11}^{00} - 1} \tag{2.10}$$

The interpretation of the relative excess risk due to interaction (RERI) is similar to that of IC. RERI measures how much more, or less, risk there is in the doubly exposed group compared to if there was no additive interaction [20]. The attributable proportion(AP) is the proportion of the *disease* in the group with both exposures that is due to the interaction, while AP* is the proportion of the *effect* that is due to interaction [27]. In other words the two AP measures tells us what proportion of the outcome is caused by the interaction. Similar measures can also be calculated for the exposures themselves, e.g., how much of the disease is caused by exposure to x alone [10, 27].

The measures are defined using p_{00} as the reference group and the exposures as harmful. In the case any exposure is protective the reference group needs to be switched so that all the relative risks are risk ratios, otherwise the interaction measure can be incorrect and misleading [16].

2.3 Modeling Relative Risks and Odds Ratios

There are several possible ways to model the probabilities, relative risks or odds ratios. Following [5] with some changes and some abuse of notation we define two possible forms for the models. With $r(x)$ as a risk, rate, odds or similar, the additive form and multiplicative form of the model are

$$r(x) = \alpha + x\beta \tag{2.11}$$

and

$$r(x) = \alpha(1 + x\beta). \tag{2.12}$$

The difference between the models is the effect of the parameters; if $r(x)$ changes in an additive way (first equation), or multiplicative way (second equation). Note that the additive form of the model in Equation 2.11 is the one implied by "no causal interaction" in Rothman's model for causality [5, 20]. Under certain conditions, depending on the covariates, models of the multiplicative form can be used to make inferences even if the true model's form is additive [5].

Logistic Regression

Odds ratios are often modeled and adjusted for confounding by using logistic regression with the model shown below [2, 9, 20, 31].

$$\ln\left(\frac{p_{x,y}}{1 - p_{x,y}}\right) = \alpha + \beta_1 x + \beta_2 y + \beta_3 xy \tag{2.13}$$

The logistic regression model is a multiplicative model since the logarithmic transformation means that the coefficients have a multiplicative effect on the odds. With the logistic model the corresponding measures, with the unexposed group as reference, for additive interaction are:

$$\text{RERI}_{\text{LR}} = e^{\beta_1+\beta_2+\beta_3} - e^{\beta_1} - e^{\beta_2} + 1 \tag{2.14}$$

$$\text{AP}_{\text{LR}} = \frac{e^{\beta_1+\beta_2+\beta_3} - e^{\beta_1} - e^{\beta_2} + 1}{e^{\beta_1+\beta_2+\beta_3}} \tag{2.15}$$

$$\text{AP}^*_{\text{LR}} = \frac{e^{\beta_1+\beta_2+\beta_3} - e^{\beta_1} - e^{\beta_2} + 1}{e^{\beta_1+\beta_2+\beta_3} - 1} \tag{2.16}$$

Testing for the presence of additive interaction using logistic regression can not be directly done by testing the interaction coefficient, β_3, since $\beta_3 = 0$ does not have to imply that interaction on the additive scale is not present [9, 24]. Instead other methods such as the delta method [9], bootstrap methods [2] or MOVER [39] can be used. It is worth noting that methods using the delta method have shown worse accuracy for the confidence interval in simulations than other methods [2, 39]. Also, the logistic model can have problems when including covariates without also including the interaction terms between the exposures and the covariates. If this is the case, it implies there is additive interaction between the exposures and the covariates [7, 24].

Linear Odds

The linear odds model, Equation 2.17, is an alternative to logistic regression, which is additive instead of multiplicative. The linear odds model uses the odds instead of the log odds like the logistic model,

$$\frac{p_{x,y}}{1-p_{x,y}} = a + b_1 x + b_2 y + b_3 xy. \tag{2.17}$$

The linear odds model is useful for assessing additive interaction since when used together with AP^* the additive interaction measures does not have the problems of misspecification and uniqueness [24]. However, the downside is that this model can lead to negative odds and fail to converge [24, 37, 38]. Furthermore, the maximum likelihood estimators can have problems when used with continuous covariates [37].

With the linear odds model, and the unexposed group as reference, the additive interaction measures are:

$$\text{RERI}_{\text{LO}} = \frac{b_3}{a} \tag{2.18}$$

$$\text{AP}_{\text{LO}} = \frac{b_3}{a + b_1 + b_2 + b_3} \tag{2.19}$$

$$\text{AP}^*_{\text{LO}} = \frac{b_3}{b_1 + b_2 + b_3} \tag{2.20}$$

The presence of additive interaction in the measures only depend on b_3. Testing $b_3 \neq 0$ is simple since the coefficients are normally distributed and the test can be done with a Wald test or using bootstrap methods [19, 24, 36].

3 Interaction with Continuous Exposures

For this section we will be using the continuous exposures G and E instead of the binary exposures X and Y as in the previous section. As we will see multiplicative interaction is largely the same as in the binary case. However, we will show later in the section 3.2 that for additive interaction there is no suitable model for the probabilities when using continuous exposures.

3.1 Multiplicative Interaction

The extension of multiplicative interaction to also cover continuous variables is straightforward. We are going to use logistic regression, since the multiplicative interaction is independent of the covariates. Adapting Equation 2.6 with the logistic regression model with the continuous exposures leads to

$$\begin{aligned}\mathrm{M}_{g,e} =& \frac{e^{\beta_1(g_1-g_0)+\beta_2(e_1-e_0)+\beta_3(g_1e_1-g_0e_0)}}{e^{b_1(g_1-g_0)+b_3e_0(g_1-g_0)}e^{b_2(e_1-e_0)+b_3g_0(e_1-e_0)}} \\ =& \frac{e^{\beta_3(g_1e_1-g_0e_0)}}{e^{\beta_3(e_0(g_1-g_0)+g_0(e_1-e_0))}} \\ =& e^{\beta_3(g_1-g_0)(e_1-e_0)}.\end{aligned} \tag{3.1}$$

$\mathrm{M}_{g,e}$ only depends on β_3 assuming that $g_1 - g_0 \neq 0$ and $e_1 - e_0 \neq 0$ $M_{g,e}$. Therefore, the presence of interaction and its direction can be tested with a Wald test or bootstrap methods. Assuming that (g, e) is not random then because $\hat{\beta} \sim N(\mu_\beta, \Sigma_\beta)$ the confidence interval with confidence level α for $ln(M_{g,e})$ is:

$$\mathrm{CI}_{\ln(\mathrm{M}_{g,e})} = (g_1 - g_0)(e_1 - e_0)(\hat{\beta}_3 \pm \lambda_{\alpha/2}\hat{s}_{\beta_3}), \tag{3.2}$$

where the parameters, $\hat{\beta}_3$, $\hat{s}_{\beta_3}$ are estimated from the data, and $\lambda_{\alpha/2}$ is the normal quantile for the confidence level α. Since the exponential function is strictly increasing it follows that

$$\mathrm{CI}_{\mathrm{M}_{g,e}} = e^{(g_1-g_0)(e_1-e_0)(\hat{\beta}_3 \pm \lambda_{\alpha/2}\hat{s}_{\beta_3})}. \tag{3.3}$$

The confidence interval is the smallest when the point g_1, e_1 is close to the reference point and increases exponentially as it moves away from the reference point.

3.2 Additive Interaction

Additive interaction for continuous exposures can be derived in a similar manner as for the binary case. However, the connection with the sufficient cause model no longer holds, since the sufficient cause model is undefined for continuous exposures [32].

We start by defining the interventions on the exposures, which now have to be more general instead of just setting the exposure to 1 or 0.

Definition 3.1: *Intervention on continuous exposures*
The intervention on the continuous exposures G and E are functions, $I_G(g)$ and $I_E(e)$, that transforms the exposures for all individuals in the population. Let g_0, e_0 be the individual's exposure levels. The effect of the intervention is then defined as

$$p_{I_G(g_0),I_E(e_0)} - p_{g_0,e_0}. \tag{3.4}$$

■

If G and E are binary then this would be equivalent to the binary definition since the only two possible functions are to either remove the exposure, or expose everyone. In the continuous case, a possible function for $I_G(g)$ and $I_E(e)$ could for instance be that the intervention on $X = x_0$ for some constant x_d is $I_X(x_0) = x_0 - x_d$ as used in [12, 13, 31]. In this paper we will unless otherwise noted, set the interventions as some strictly decreasing function.

Before defining additive interaction for continuous exposures we will also define the marginal effect of an intervention.

Definition 3.2: *Marginal effect*
The marginal effect from an intervention, $I_G(g_0)$, on a continuous exposure G is defined as

$$p_{I_G(g_0)} - p_{g_0}. \tag{3.5}$$

An intervention, $I_G(g)$, has no marginal effect if

$$\forall g_0 \in \mathbb{R}\ p_{I_G(g_0)} - p_{g_0} = 0. \tag{3.6}$$

■

Just as in the binary case we can derive the IC for the continuous case as the effect difference between intervening on G on either the groups defined by $E = e_A$ or $E = e_B$, or intervening on E on the groups defined by $G = g_A$ or $G = g_B$. Formally

$$IC_{I_G(g_0),e_A,g_0,e_B} = (p_{I_G(g_0),e_A} - p_{g_0,e_A}) - (p_{I_G(g_0),e_B} - p_{g_0,e_B}) \tag{3.7}$$

and

$$IC_{g_A,I_E(e_0),g_B,e_0} = (p_{g_A,I_E(e_0)} - p_{g_A,e_0}) - (p_{g_B,I_E(e_0)} - p_{g_B,e_0}). \tag{3.8}$$

However, in contrast to the binary case the additive interaction measure is no longer unique, since instead of choosing between intervening on two different groups, based on

a binary exposure, the scale is now continuous and there are multiple ways to define the groups. Hence, there are several possible types of additive interaction.

We will start with the simplest case; the case when there is no additive interaction for some *single* value of the exposure intervened upon. In other words, for that value of the exposure it does not matter for the intervention what value to the other exposure takes.

Definition 3.3: *Single additive interaction*
There is single additive interaction for $G = g_0$ with the intervention $I_G(g)$ if

$$\exists e_A, e_B \in \mathbb{R}\ IC_{I_G(g_0),e_A,g_0,e_B} \neq 0, \tag{3.9}$$

and for $E = e_0$ with the intervention $I_E(e)$,

$$\exists g_A, g_B \in \mathbb{R}\ IC_{g_A,I_E(e_0),g_B,e_0} \neq 0. \tag{3.10}$$

With no single additive interaction being the opposite,

$$\forall e_A, e_B \in \mathbb{R}\ IC_{I_G(g_0),e_A,g_0,e_B} = 0, \tag{3.11}$$

and

$$\forall g_A, g_B \in \mathbb{R}\ IC_{g_A,I_E(e_0),g_B,e_0} = 0. \tag{3.12}$$

■

If there, for all values of the exposure is no single additive interaction we say that there is no exposure additive interaction for that particular exposure.

Definition 3.4: *Exposure additive interaction*
There is exposure additive interaction for G with the intervention $I_G(g)$ if

$$\exists g_0, e_A, e_B \in \mathbb{R}\ IC_{I_G(g_0),e_A,g_0,e_B} \neq 0, \tag{3.13}$$

and for E with the intervention $I_E(e)$,

$$\exists g_A, g_B, e_0 \in \mathbb{R}\ IC_{g_A,I_E(e_0),g_B,e_0} \neq 0. \tag{3.14}$$

So there is no exposure additive interaction for G if

$$\forall g_0, e_A, e_B \in \mathbb{R}\ IC_{I_G(g_0),e_A,g_0,e_B} = 0, \tag{3.15}$$

and for E if

$$\forall g_A, g_B, e_0 \in \mathbb{R}\ IC_{g_A,I_E(e_0),g_B,e_0} = 0. \tag{3.16}$$

■

For the binary model both definitions correspond to absence of additive interaction. Exposure additive interaction is what we would consider closest to being equivalent to the intuitive idea of additive interaction with binary exposures, namely that the effect of the intervention on an exposure does not depend on the level of another exposure no matter what the levels of the first and second exposures are.

Existence of a Suitable Model for the Probabilities

So far with the continuous exposures we have not defined any model for the probabilities, $p_{g,e}$. In this section we are going to show the main result of this paper, Theorem 3.4, which shows that there is no model for the probabilities that can fulfill the criteria we consider required for the model.

We propose the following two criteria that the model should be able to fulfill:

i. There exist some values for the model parameters for which both exposures have marginal effects and no additive interaction on the exposure level for both exposures.

ii. There exist some values for the model parameters for which there are no marginal effects and there is additive interaction between the exposures.

which, formally based on the definitions of marginal effect and exposure additive interaction, and with interventions $I_G(g_0)$, $I_E(e_0)$, is given in the following definition.

Definition 3.5: *Suitable model*
A suitable model meet both of the following criteria:

i. There exist some values for the model parameters, K_1, for which all of the following statements hold:

$$\exists g_0 \in \mathbb{R}\ p_{I_G(g_0)} - p_{g_0} \neq 0 \tag{3.17}$$

$$\exists e_0 \in \mathbb{R}\ p_{I_E(e_0)} - p_{e_0} \neq 0 \tag{3.18}$$

$$\forall g_0, e_A, e_B \in \mathbb{R}\ IC_{I_G(g_0),e_A,g_0,e_B} = 0 \tag{3.19}$$

$$\forall g_A, g_B, e_0 \in \mathbb{R}\ IC_{g_A,I_E(e_0),g_B,e_0} = 0 \tag{3.20}$$

ii. There exist some values for the model parameters, K_2, for which all of the following statements hold:

$$\forall g_0 \in \mathbb{R}\ p_{I_G(g_0)} - p_{g_0} = 0 \tag{3.21}$$

$$\forall e_0 \in \mathbb{R}\ p_{I_E(e_0)} - p_{e_0} = 0 \tag{3.22}$$

$$\exists g_0, e_A, e_B \in \mathbb{R}\ IC_{I_G(g_0),e_A,g_0,e_B} \neq 0 \tag{3.23}$$

$$\exists g_A, g_B, e_0 \in \mathbb{R}\ IC_{g_A,I_E(e_0),g_B,e_0} \neq 0 \tag{3.24}$$

■

Note that the model need to fulfill both criteria to be considered as suitable, but not with the same values for the model parameters. E.g., if for some model k_i is the parameter for the interaction, then criterion (i) means that $k_i = 0$, while criterion (ii) implies that $k_i \neq 0$.

We will now show that with the intervention functions as strictly decreasing there is no model for the probabilities such that both criteria can be fulfilled. To prove this result in Theorem 3.4 we first require some additional lemmas. To start with we are going to derive alternative conditions for the conditions in Definition 3.3 and 3.4.

Lemma 3.1: *There is no single additive interaction for $I_G(g_o)$ if*

$$\forall e \in \mathbb{R}: \ \frac{\partial p(I_G(g_0), e)}{\partial e} = \frac{\partial p(g_0, e)}{\partial e}, \tag{3.25}$$

and correspondingly for $I_E(e_o)$,

$$\forall g \in \mathbb{R}: \ \frac{\partial p(g, I_E(e_0))}{\partial g} = \frac{\partial p(g, e_0)}{\partial g}. \tag{3.26}$$

Proof: Proving the case with intervention on G. The condition $IC_{I_G(g_0),e_A,g_0,e_B} = 0$ can be written as

$$p_{I_G(g_0),e_A} - p_{g_0,e_A} = p_{I_G(g_0),e_B} - p_{g_0,e_B}, \tag{3.27}$$

which means that since the condition in Definition 3.3 is *for all* e_A, e_B for some constant α we can instead write the condition as

$$\forall e \in \mathbb{R} \ p_{I_G(g_0),e} - p_{g_0,e} = \alpha. \tag{3.28}$$

Then since $p_{I_G(g_0),e} - p_{g_0,e}$ is a constant function of e its derivative has to be

$$\forall e \in \mathbb{R} \ \frac{\partial}{\partial e}(p_{I_G(g_0),e} - p_{g_0,e}) = 0, \tag{3.29}$$

which leads to

$$\forall e \in \mathbb{R}: \ \frac{\partial p(I_G(g_0), e)}{\partial e} = \frac{\partial p(g_0, e)}{\partial e} \tag{3.30}$$

as a condition equivalent to Equation 3.11. ■

Lemma 3.2: *There is no exposure additive interaction for G if*

$$\forall e, g_0 \in \mathbb{R}: \ \frac{\partial p(I_G(g_0), e)}{\partial e} = \frac{\partial p(g_0, e)}{\partial e}, \tag{3.31}$$

and correspondingly for E,

$$\forall g, e_0 \in \mathbb{R}: \ \frac{\partial p(g, I_E(e_0))}{\partial g} = \frac{\partial p(g, e_0)}{\partial g}. \tag{3.32}$$

Proof: The result follows directly from using the the definition of exposure additive interaction with the proof in Lemma 3.1. ■

Based on the above lemmas we can derive the form which the model for $p_{g,e}$ must have to be able to show no exposure additive interaction.

Theorem 3.3: *Let k_e and m be constants, h some function of G, and $f(g, e)$ is the interaction function and k_i is the parameter for the interaction. For the model to be able to show no exposure additive interaction for an exposure $G = g$ relative to some other exposure $E = e$ it is required that the model can be written on the form*

$$p(g, e) = m + k_e e + h(g) + k_i f(g, e). \tag{3.33}$$

If and only if $k_i = 0$ then there is no additive interaction on the exposure level.

Proof: For Equation 3.31 in the proof of Lemma 3.2 to hold it must be true that for a constant, k_e,

$$\frac{\partial p(g, e)}{\partial e} = k_e. \tag{3.34}$$

Then with constants k_e, m and function $h(g)$, the model for $p(g, e)$ can be written on the form of

$$p(g, e) = m + k_e e + h(g). \tag{3.35}$$

More generally and since we want to model the interaction there could be some parameter for an interaction function that is zero. I.e., with k_i as the parameter for the function f which represents the interaction, $p(g, e)$ can be written as

$$p(g, e) = m + k_e e + h(g) + k_i f(g, e). \tag{3.36}$$

From Lemma 3.2 we know that for there to be no additive interaction on the exposure level, then k_i must be 0. ■

From these results it also follows that for two exposures with both being able to show no exposure additive interaction the model's form has to be

$$p(g, e) = m + k_g g + k_e e + k_i f(g, e). \tag{3.37}$$

Using the form of the model above that is required for the model to be able to meet criterion (i) we will now show that the model can not also fulfill criterion (ii).

Theorem 3.4: *Let the intervention functions be strictly decreasing, then there is no model which can fulfill both of the criteria in Definition 3.5.*

Proof: We prove the theorem by showing that the combination of critera (i) and (ii) leads to a contradiction. Assume that the model meet both criteria (i) and (ii). From the first criterion we know from Theorem 3.3 that the model has to have the form of Equation 3.37. From Definition 3.1 the marginal effects for g and e are,

$$p_{I_G(g_0),e} - p_{g_0,e} = k_g(I_G(g_0) - g_0) + k_i(f(I_G(g_0), e) - f(g_0, e)), \tag{3.38}$$

and

$$p_{g,I_E(e_0)} - p_{g,e_0} = k_e(I_E(e_0) - e_0) + k_i(f(g, I_E(e_0)) - f(g, e_0)). \tag{3.39}$$

No marginal effects for an exposure means that the effect from the intervention on that exposure is zero,

$$\forall g_0, e: \; p_{I_G(g_0),e} - p_{g_0,e} = 0, \tag{3.40}$$

and

$$\forall g, e_0: \; p_{g,I_E(e_0)} - p_{g,e_0} = 0. \tag{3.41}$$

We will be focusing on the exposure G in the next part, but the result also holds for E. Criterion (ii) requires that there is interaction, formally, $k_i \neq 0$ and $\exists g, e : f(g,e) \neq 0$. Then Equation 3.38 have to depend on the interaction function, unless it holds that

$$\forall g_0, e: \; f(I_G(g_0), e) - f(g_0, e) = 0. \tag{3.42}$$

If Equation 3.42 is true then it implies that the value of g does not matter for $f(g,e)$ for some given e. The function $f(g,e)$ can then not represent the interaction, since the definition of statistical interaction requires that the effect from one exposure depends on the level of another exposure. This in turn implies that $p(g,e)$ has no interaction. However, this is a contradiction since criterion (ii) states that there is interaction. ■

Model Example: Logistic Regression

In [12] the authors investigated additive interaction with continuous exposures by modeling RERI with logistic regression. The intervention on an exposure G with some constant g_d was defined as $I_G(g_0) = g_0 + g_d$. The logistic model was

$$\frac{p_{g,e}}{1 - p_{g,e}} = e^{\alpha + \beta_1 g + \beta_2 e + \beta_3 ge}, \tag{3.43}$$

and their RERI, based on the intervention, was derived as

$$\begin{aligned} \text{RERI}^{\text{LR}}_{g,e} = {} & e^{\beta_1 g_d + \beta_2 e_d + \beta_3((g_0+g_d)(e_0+e_d) - g_0 e_0)} \\ & - e^{g_d(\beta_1 + \beta_3 e_0)} - e^{e_d(\beta_2 + \beta_3 g_0)} + 1. \end{aligned} \tag{3.44}$$

Let us now examine this model for continuous additive interaction using Theorem 3.3. Using the rare disease assumption then for the requirements of Theorem 3.3 to be met Equation 3.43 has to be able to be written in the form of Equation 3.37. This is not possible since exponential functions are multiplicative. It then follows that the model can not show no exposure additive interaction if there are marginal effects, i.e., $\beta_1 \neq 0$ and $\beta_2 \neq 0$.

The model can also not fulfill criterion (ii) since with $\beta_1 = 0$, $\beta_2 = 0$ the odds ratio for the marginal effect depend on β_3, for instance the marginal effect for G is:

$$\text{OR}^{g_0,e_0}_{g_0+g_d,e_0} = e^{\beta_3 g_d e_0} \tag{3.45}$$

Model Example: Linear Odds

A model that does fulfill the requirements for criterion (i) is the linear odds model,

$$\frac{p_{g,e}}{1-p_{g,e}} = a + b_1 g + b_2 e + b_3 ge. \tag{3.46}$$

In contrast to the logistic model the linear odds model can be written on the form of Equation 3.37, meaning that it can show no exposure additive interaction with no interaction corresponding to $b_3 = 0$. The linear odds model leads to $\text{RERI}^{\text{LO}}_{g,e}$ and $\text{AP}^{*\text{LO}}_{g,e}$ as:

$$\text{RERI}^{\text{LO}}_{g,e} = \frac{b_3(I_G(g_0) - g_0)(I_E(e_0) - e_0)}{a + b_1 g_0 + b_2 e_0 + b_3 g_0 e_0} \tag{3.47}$$

$$\text{AP}^{*\text{LO}}_{g,e} = \frac{b_3(I_G(g_0) - g_0)(I_E(e_0) - e_0)}{b_1(I_G(g_0) - g_0) + b_2(I_E(e_0) - e_0) + b_3(I_G(g_0)I_E(e_0) - g_0 e_0)} \tag{3.48}$$

The presence or direction of interaction in this model can be tested by testing if b_3 deviates from 0. However, as shown in Theorem 3.4, models of the form in Equation 3.37 will have issues with criterion (ii), that is no marginal effects and the presence of additive interaction at the same time. In the linear odds model we can see this by setting the coefficients corresponding to the marginal effects to zero, i.e., $b_1 = 0$, $b_2 = 0$, for $\text{AP}^{*\text{LO}}_{g,e}$ which becomes

$$\text{AP}^{*\text{LO}}_{g,e} = \frac{(I_G(g_0) - g_0)(I_E(e_0) - e_0)}{(I_G(g_0)I_E(e_0) - g_0 e_0)}. \tag{3.49}$$

For there to be only interaction effects and no marginal effects then it should be true that $\text{AP}^{*\text{LO}}_{g,e} = 1$ for all values of g and e, which is clearly not true unless $I_G(g_0) = I_E(e_0) = 0$ or $g_0 = e_0 = 0$. It also holds if the intervention functions are $I_G(g) = I_E(e) = 0$ for all g and e, which would mean that the interventions removes the exposures, and the linear odds model with these interventions does fulfill both criteria.

4 Discussion

The purpose of this paper is to explore models for additive interaction with continuous exposures. We have shown that with the intervention defined as a monotonic function there is no model that can fulfill both of the following criteria:

i. There exist some values for the model parameters for which both exposures have marginal effects and no additive interaction on the exposure level for both exposures.

ii. There exist some values for the model parameters for which there are no marginal effects and there is additive interaction between the exposures.

Our finding means that the previously suggested methods in the literature ([12, 13, 31]) for additive interaction and continuous exposures do not work in practice, since those methods are based on logistic regression and are unable to fulfill any of the two criteria. If there are marginal effects, those model's estimates will always conclude there is additive interaction, given that there is sufficient statistical power.

Note that the results are dependent on the definition of the intervention on the continuous exposures. The definition of the intervention is more complicated than in the binary case. Partially because the groups that can be intervened upon can be more complex, and partially because the interventions themselves can be modeled in more detail, and what an intervention is can also have different implications. In this paper we defined the intervention as a monotonic function on the exposure, other definitions could lead to different conclusions. In the linear odds example in Section 3.2.3 with the intervention defined as setting the exposure to zero both criteria do hold.

However, our results do not change the fact that the risk differences are an important part of causal inference and that the IC is the measure of interest when comparing interventions on different groups. With the correct context and model for the estimation of the odds or probabilities, the IC is still be useful.

Mechanistic interpretations of interaction with continuous exposures is different though, and would require a different causality model than the sufficient cause model, since the sufficient cause model is binary and there is no obvious way to extend the model that is not simply a transformation from continuous to binary. A continuous deterministic causality model will always have some corresponding binary sufficient cause model, since a deterministic continuous causality model can always be transformed into a binary sufficient cause model given the cutoffs for the transformation from continuous to binary is set properly, such as done in [3, 35]. The most trivial model transformation is to use the same function as for the continuous model itself, i.e., the new binary variable is set so that it is one for all values of the continuous exposures where the outcome is one in the continuous causality model and zero for all other areas. However, such a model is not trivial to estimate in practice since the true cutoffs are unknown [3, 35] Another motivation is that a deterministic model could not fulfill some of the properties that would be useful to take advantage of the extra information provided by the continuous exposures, e.g., a model where the probability increases proportionally with the exposure. However, this continuous non-deterministic causality model is not the same kind of stochastic model used in [3, 32, 35] since the exposures themselves do not have a random effect on the outcome in those models.

An investigator in a study about additive interaction with continuous exposures would need to be careful about the choice of model for the probabilities or odds, since both the aforementioned criteria can not be true for the same model. For instance if the interest of the study is masking effects then criterion (ii) is important since it tells us that one effect is masking the other.

Even though the linear odds model does fulfill criterion (i), it can be unsuitable in practice because convergence problems as explained in Section 3.2, or because a linear

increase of the risk might be unrealistic. However, in some cases a linear risk model can be realistic. One example is the linear no-threshold model for cancer induced by radiation which models the cancer risk as proportional to the amount of radiation [26].

With more general interventions, and also possibly comparisons between different possible intervention functions, since no model can give a simple answer as in the binary case using decision theory could be useful for finding the best intervention and which group to target, although to a large extent such methods would likely depend on the study's context.

References

[1] T. Andersson, L. Alfredsson, H. Källberg, S. Zdravkovic, and A. Ahlbom. Calculating measures of biological interaction. *European Journal of Epidemiology*, 20(7):575–579, 2005.

[2] S. F. Assmann, D. W. Hosmer, S. Lemeshow, and K. A. Mundt. Confidence intervals for measures of interaction. *Epidemiology*, pages 286–290, 1996.

[3] C. Berzuini and A. P. Dawid. Stochastic mechanistic interaction. *Biometrika*, 103(1):89–102, 2016.

[4] W. J. Blot and N. E. Day. Synergism and interaction: are they equivalent? *American Journal of Epidemiology*, 110(1):99–100, 1979.

[5] S. Greenland. Additive risk versus additive relative risk models. *Epidemiology*, pages 32–36, 1993.

[6] S. Greenland. Basic problems in interaction assessment. *Environmental Health Perspectives*, 101(Suppl 4):59, 1993.

[7] S. Greenland, J. M. Robins, and J. Pearl. Confounding and collapsibility in causal inference. *Statistical Science*, pages 29–46, 1999.

[8] P. W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

[9] D. W. Hosmer and S. Lemeshow. Confidence interval estimation of interaction. *Epidemiology*, 3(5):452–456, 1992.

[10] O. Hössjer, I. Kockum, L. Alfredsson, A. K. Hedström, T. Olsson, and M. Lekman. A general framework for and new normalization of attributable proportion. *Epidemiologic Methods*, 6(1), 2017.

[11] H. Källberg, B. Ding, L. Padyukov, C. Bengtsson, J. Rönnelid, L. Klareskog, L. Alfredsson, E. S. Group, et al. Smoking is a major preventable risk factor for rheumatoid arthritis: estimations of risks after various exposures to cigarette smoke. *Annals of the Rheumatic Diseases*, 70(3):508–511, 2011.

[12] M. Katsoulis and C. Bamia. Additive interaction between continuous risk factors using logistic regression. *Epidemiology*, 25(3):462–464, 2014.

[13] M. J. Knol, I. van der Tweel, D. E. Grobbee, M. E. Numans, and M. I. Geerlings. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. *International Journal of Epidemiology*, 36(5):1111–1118, 2007.

[14] M. J. Knol, J. P. Vandenbroucke, P. Scott, and M. Egger. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *American Journal of Epidemiology*, 168(9):1073–1081, 2008.

[15] M. J. Knol and T. J. VanderWeele. Recommendations for presenting analyses of effect modification and interaction. *International Journal of Epidemiology*, 41(2):514–520, 2012.

[16] M. J. Knol, T. J. VanderWeele, R. H. Groenwold, O. H. Klungel, M. M. Rovers, and D. E. Grobbee. Estimating measures of interaction on an additive scale for preventive exposures. *European Journal of Epidemiology*, 26(6):433–438, 2011.

[17] M. L. Petersen, K. E. Porter, S. Gruber, Y. Wang, and M. J. van der Laan. Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research*, 21(1):31–54, 2012.

[18] D. H. Rehkopf, M. M. Glymour, and T. L. Osypuk. The consistency assumption for causal inference in social epidemiology: When a rose is not a rose. *Current Epidemiology Reports*, 3(1):63–71, 2016.

[19] D. B. Richardson and J. S. Kaufman. Estimation of the relative excess risk due to interaction and associated confidence bounds. *American Journal of Epidemiology*, 169(6):756–760, 2009.

[20] K. J. Rothman, S. Greenland, and T. L. Lash. *Modern epidemiology*. Lippincott Williams & Wilkins, 2008.

[21] K. J. Rothman, S. Greenland, and A. M. Walker. Concepts of interaction. *American Journal of Epidemiology*, 112(4):467–470, 1980.

[22] P. Royston, D. G. Altman, and W. Sauerbrei. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine*, 25(1):127–141, 2006.

[23] R. Saracci. Interaction and synergism. *American Journal of Epidemiology*, 112(4):465–466, 1980.

[24] A. Skrondal. Interaction as departure from additivity in case-control studies: a cautionary note. *American Journal of Epidemiology*, 158(3):251–258, 2003.

[25] J. Splawa-Neyman, D. M. Dabrowska, T. P. Speed, et al. On the application of probability theory to agricultural experiments. *Statistical Science*, 5(4):465–472, 1990. Reprint.

[26] M. Tubiana, L. E. Feinendegen, C. Yang, and J. M. Kaminski. The linear no-threshold relationship is inconsistent with radiation biologic and experimental data. *Radiology*, 251(1):13–22, 2009.

[27] T. J. VanderWeele. Reconsidering the denominator of the attributable proportion for interaction. *European Journal of Epidemiology*, 28(10):779–784, 2013.

[28] T. J. VanderWeele. *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press, 2016.

[29] T. J. VanderWeele, Y. Chen, and H. Ahsan. Inference for causal interactions for continuous exposures under dichotomization. *Biometrics*, 67(4):1414–1421, 2011.

[30] T. J. VanderWeele and M. J. Knol. Remarks on antagonism. *American Journal of Epidemiology*, 173(10):1140–1147, 2011.

[31] T. J. VanderWeele and M. J. Knol. A tutorial on interaction. *Epidemiologic Methods*, 3(1):33–72, 2014.

[32] T. J. VanderWeele and T. S. Richardson. General theory for interactions in sufficient cause models with dichotomous exposures. *Annals of Statistics*, 40(4):2128, 2012.

[33] T. J. VanderWeele and J. M. Robins. The identification of synergism in the sufficient-component-cause framework. *Epidemiology*, 18(3):329–339, 2007.

[34] T. J. VanderWeele and J. M. Robins. Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika*, 95(1):49–61, 2008.

[35] T. J. VanderWeele and J. M. Robins. Stochastic counterfactuals and stochastic sufficient causes. *Statistica Sinica*, 22(1):379, 2012.

[36] T. J. VanderWeele and S. Vansteelandt. A weighting approach to causal effects and additive interaction in case-control studies: marginal structural linear odds models. *American Journal of Epidemiology*, 174(10):1197–1203, 2011.

[37] S. Wachholder. Binomial regression in GLIM: Estimating risk ratios and risk differences. *American Journal of Epidemiology*, 123(1):174–184, 1986.

[38] L. N. Yelland, A. B. Salter, and P. Ryan. Relative risk estimation in randomized controlled trials: a comparison of methods for independent observations. *The International Journal of Biostatistics*, 7(1):1–31, 2011.

[39] G. Y. Zou. On the estimation of additive interaction by use of the four-by-two table and beyond. *American Journal of Epidemiology*, 168(2):212–224, 2008.

Paper B

On Probabilistic Multifactor Potential Outcome Models

On Probabilistic Multifactor Potential Outcome Models

by

Daniel Berglund, and Timo Koski

Abstract

The sufficient cause framework describes how sets of sufficient causes are responsible for causing some event or outcome. It is known that it is closely connected with Boolean functions. In this paper we define this relation formally, and show how it can be used together with Fourier expansion of the Boolean functions to lead to new insights. The main result is a probibalistic version of the multifactor potential outcome model based on independence of causal influence models and Bayesian networks.

Keywords: Causal Inference; Sufficient Cause; Potential Outcome; Boolean Function; MFPO; ICI; BCF; Probabilistic Potential Outcome; Qualitative Bayesian Network; Additive Interaction

1 Introduction

1.1 Potential Outcomes and Sufficient Cause Models

An important idea of genetic epidemiology is the **potential outcome** also known as the Rubin-Neyman causal model, see e.g., [16, 29, 41], which setups a model for causality based on the idea of counterfactuals. For each individual, individual response types indicates whether or not a certain disease would develop under the different possible combinations of a set of exposures (risk factors), or events, that are being considered. The individual response type is thus an entity offering potential outcomes, as we observe an individual exposed to one combination of the exposures.

The potential outcome variable is counterfactual, in the sense that only one exposure condition is actually possible at any one time, and the potential outcome variable specifies disease occurrence under other conditions not all of which actually occur [16, 41]. A pioneering study of potential outcomes is [39] that first appeared the year 1923 in Polish.

In **sufficient cause** (SC) models each exposure is regarded as a component of a collection of causes, see [36, ch. 1]. Each such collection forms a minimal set of conditions

of exposure that yield disease and is sufficient in this sense. This is in the deterministic realm: once all component causes/exposures of a sufficient cause are present, that particular sufficient cause is complete and disease occurs.

For example, P. Kearns et.al. report in [18] a three component sufficient cause for multiple sclerosis. The sufficient cause in loc.cit. contains a fourth component, genetic susceptibility, which signifies an increased likelihood of developing a particular disease based on a person's genetic makeup. Moreover, Kearns et.al. require even certain order in time for the three exposures for the sufficient cause to be completed.

Sufficient cause synergism between a number of exposures corresponds, e.g., to the situation when they act together in causing disease, reflected in a sufficient cause that includes all of them as components. The SC model is a set of interacting causal components and is therefore distinct from the classical statistical view of interaction as the inclusion of a product term in a model [35]. In the medical and public health education the SC model is known as (Rothman's) causal pie [34, ch. 1] due to the natural graphical illustrations by pie chart diagrams.

There is a connection to potential outcomes; the basic units of SC are the mechanisms that determine the potential outcomes of individuals. Many different sets of mechanisms will lead to the same pattern of potential outcomes for an individual; as will become evident later.

The potential outcome models and SC models are not to be seen as innately deterministic. A random element is quite natural, see [11, 45]. The computational techniques and details in this paper differ from those of [45], we rely more on the calculus of Boolean functions, but in a sense their conclusion holds; the empirical conditions that suffice to conclude a sufficient cause interaction in the deterministic sufficient cause framework will also lead to a sufficient cause interaction in the stochastic sufficient cause models presented here.

1.2 Organization of the Paper

The basic notation used in this paper is shown in Section 1.3. We then introduce the potential outcomes model in Section 2. Based on the potential outcomes we continue in that section with a formal definition of sufficient cause, sufficient model, and sufficient cause representation, summarized from [42].

The sufficient cause model's connection with Boolean functions is formally shown in Section 3. In that section we also summarize the MFPO model from [11], and show that it is equivalen to the sufficient cause representation. Results related to the connection between the sufficient cause model and the Blake canonical form are also shown.

In Section 4 we have summary of the probabilistic potential outcomes and the probabilistic sufficient causes theory from [31]. We then express the probabilistic model in the form of the ICI model and Bayesian networks in Section 5.

Section 6 applies the results from Section 5 on different response profiles and find their P-sufficient forms. The model is then expanded to more than two events in Section 7. In

Section 8 we compute the probability of a potential outcome for MFPO based on the ICI model.

The practical part of the model is at the population level, for which some results are shown in Section 9. The sufficent cause model's connection with the linear risk model is proved in Section 10.

1.3 Basic notation

Based on the notation in [42] we will now introduce the basic notation for this paper.

An **event/exposure** is a binary (Boolean) variable, indicated by uppercase (X). Boldface uppercase is **set of events** ($\mathbf{C}$). The **complement** of some event X is $\overline{X} \equiv 1 - X$, for simplicity we also use $\neg$. Lowercase used for **specific values/instantiations** of events, $X = x$, or sets of events, $\{\mathbf{C} = \mathbf{c}\} \equiv \{\forall i, (\mathbf{C})_i = (\mathbf{c})_i\}$. The **cardinality** of a set is denoted by $|\mathbf{C}|$. Fraktur/gothic typeset ($\mathfrak{B}$) is used for **collections of sets of events**. D denotes the binary outcome or response of interest. The outcome $D = 1$ is the onset of an effect (a disease).

A **literal event** associated with X, is either X or $\overline{X}$. For a given set of events $\mathbf{C}$, $\mathbb{L}(\mathbf{C})$ is the associated set of literal events

$$\mathbb{L}(\mathbf{C}) \equiv \mathbf{C} \cup \{\overline{X} | X \in \mathbf{C}\}. \tag{1.1}$$

For a literal $L \in \mathbb{L}(\mathbf{C})$ $(L)_{\mathbf{c}}$ denotes the value set by an assignment $\mathbf{C} = \mathbf{c}$. The **conjunction** of a set of literal events $\mathbf{B} = \{F_1, \ldots, F_m\} \subseteq \mathbb{L}(\mathbf{C})$ is defined as:

$$\bigwedge(\mathbf{B}) \equiv \prod_{i=1}^{m} F_i = \min\{F_1, \ldots, F_m\}; \tag{1.2}$$

Hence $\bigwedge(\mathbf{B}) = 1$ if and only if for all i, $F_i = 1$. The Boolean form $\bigwedge(\mathbf{B})$ represents the Boolean function AND_m [26]. The **disjunction** of a set of binary events is defined as:

$$\bigvee(\{Z_1, \ldots, Z_p\}) \equiv \max\{Z_1, \ldots, Z_p\}; \tag{1.3}$$

so $\bigvee(\{Z_1, \ldots, Z_p\}) = 1$ if and only if for some j, $Z_j = 1$. This Boolean form represents the Boolean function OR_p [26].

The set of literals corresponding to assignment $\mathbf{C} = \mathbf{c}$ is defined as:

$$\mathbf{B}^{[\mathbf{c}]} \equiv \{L | L \in \mathbb{L}(\mathbf{C}), (L)_{\mathbf{c}} = 1\}. \tag{1.4}$$

For a collection of sets of literals $\mathfrak{B} = \{\mathbf{C}_1, \ldots, \mathbf{C}_q\}$ define the Boolean form:

$$\bigvee\bigwedge(\mathfrak{B}) \equiv \bigvee_{i=1}^{q} (\bigwedge(\mathbf{C}_i)), \tag{1.5}$$

i.e. $\bigvee\bigwedge(\mathfrak{B}) = 1$ if and only if for some j it holds that $\bigwedge(\mathbf{C}_j) = 1$. This is a disjunctive normal form (DNF), a canonical normal form of a Boolean form consisting of a disjunction

of conjunctives; it represents a Boolean function, c.f. the remarks above, an OR of ANDs. It is in addition known as a sum of products (SOP). A case of Boolean function, which is OR of ANDs, will in the sequel be called **tribes**, c.f., [26] and Section 7.2.

$\mathbb{P}(\mathbb{L}(\mathbf{C}))$ is the set of subsets of $\mathbb{L}(\mathbf{C})$ that do not contain both X and $\overline{X}$ for any $X \in \mathbf{C}$. Formally,

$$\mathbb{P}(\mathbb{L}(\mathbf{C})) \equiv \{\mathbf{B} | \mathbf{B} \subset \mathbb{L}(\mathbf{C}) \ \forall \ X \in \mathbf{C}, \{X, \overline{X}\} \not\subseteq \mathbf{B}\}. \tag{1.6}$$

2 Potential Outcomes and the Sufficient Cause Framework

In this section we will introduce the potential outcome model [16, 37, 39] and and summarize certain parts of the framework and results for the general sufficient cause framework in [42].

2.1 Potential Outcomes

Consider a finite but large population U of individuals designated by u. Hereby u can be seen as the conjunction of latent variables or intrinsic factors [8, p.372] comprising all that characterizes the individual for some purposes. The population U of individuals renders the deterministic theory an element of randomness as the statistical sample space. D is a binary outcome, e.g., disease.

We have d binary variables $\mathbf{C} = \{X_1, \ldots, X_d\}$. By convention, $x_i = 1$ means exposure to X_i and $x_i = 0$ means no exposure. These represent exposures, which can appear as hypothetical interventions or as natural events. The *specific values/instantiations* of $\mathbf{C}$ are given as $\mathbf{c} = \{x_1, \ldots, x_d\}$.

The binary potential outcome variable D ($\in \{0, 1\}$) for an individual u in the population U, when the exposures are $\mathbf{C} = \mathbf{c}$, is denoted by $D_{\mathbf{C}=\mathbf{c}}(u)$. Then

$$D_{\mathbf{C}=\mathbf{c}}(u) = D_{\mathbf{c}}(u) \tag{2.1}$$

denotes the **potential outcome** of D for the individual u. We assume in addition that $D_{\mathbf{c}}(u)$ is independent of the assignments for another individual u^*. We are thus replacing a single outcome variable D with the multitude of variables $D_{\mathbf{C}=\mathbf{c}}$.

> Expanding on [7, p.20] we illustrate the multitude of variables with a polyptych. This is a painting, typically an altarpiece from Renaissance, consisting of more than three sections or panels joined by hinges or folds. The hinged panels can be varied in arrangement to show different views in the polyptych. In each panel one instantiation $\mathbf{c}$ is inscribed. With exposures $\mathbf{C}$ instantiated at $\mathbf{c}$, just one of the views, the one associated to $\mathbf{c}$, is shown. From this view we get the only measurable real entity, the actual $D_{\mathbf{c}}$. We are not permitted to open another view in the polyptych.

In this setting there are 2^d different configurations of $\mathbf{c}$ and the corresponding potential outcomes for each individual. We let $\mathfrak{D}(\mathbf{C}; u)$ denote this set of potential outcomes, and will later be called an individual response profile. $\mathfrak{D}(\mathbf{C}; U)$ is the set of potential outcomes for the whole population.

Next, $\sigma(X_1) = x_1, \ldots, \sigma(X_n) = x_n$ designates an intervention/external decision that sets the exposures to $x_1, \ldots, x_n$. Then, for individual u, $D_{\sigma(\mathbf{C})=\mathbf{c}} = D_{\sigma(X_1)=x_1,\ldots,\sigma(X_n)=x_n}$ is the value of D, when $\mathbf{C}$ has been forced (by an external actor) to be $\mathbf{c}$. The same policy combination of exposures may bring about a different outcome imposed on different individuals (having their respective different response type). For a particular individual u the combination of exposures $\mathbf{c}$ fixes the outcome $D_{\mathbf{c}}(u)$ uniquely.

$D_{\mathbf{C}=\mathbf{c}}(u)$ is the value which was observed for individual u, if $\mathbf{C}$ is instantiated at $\mathbf{c}$ due to a natural course of events. We make the **consistency assumption**

$$D_{\sigma(\mathbf{C})=\mathbf{c}}(u) = D_{\mathbf{C}=\mathbf{c}}(u). \tag{2.2}$$

In words, the potential response $D(u)$ of individual u to the hypothetical interventionist exposure $\sigma(\mathbf{C}) = \mathbf{c}$ must coincide with $D_{\mathbf{C}}(u)$, whenever the actual exposure, $\mathbf{C}$, happened to be $\mathbf{c}$, [29].

The consistency looks perhaps trivial to assume, however, this assumption is violated, if different variants of the same exposure have different effects on the outcome. For instance, an increase of income from more or less permanent work might not have the same effect on the spending during next weekend as a windfall gain from a lottery winning. It also applies to some measures of disease severity, since different sets of symptoms can lead to the same disease severity according to the chosen measure. In addition, the effect of an intervention might not be the same as the original effect. Consider, e.g., the difference between earnings from a job or benefits from a social welfare program [32].

2.2 Sufficient Cause Models

Using the potential outcome model introduced above we will now introduce, and summarize some results for, the sufficient cause model in [42], which provides a rigorously stated general theory for the sufficient cause model casting the various intuitive and graphical designs in a united mathematical framework for arbitrary d.

Definition 2.1: A set $\mathbf{B} \in \mathbb{L}(\mathbf{C})$ for D forms a **sufficient cause** for D relative to $\mathbf{C}$ in subpopulation U^* if for all $\mathbf{c} \in \{0,1\}^{|\mathbf{C}|}$ such that $(\bigwedge(\mathbf{B}))_{\mathbf{c}} = 1$, we have that $D_{\mathbf{c}}(u) = 1$ for all $u \in U^* \subseteq U$. ■

In other words, $\mathbf{B}$ is a sufficient cause if there is a subset of the population for which $D = 1$, when all the factors in $\mathbf{B}$ are instantiated at $\mathbf{c}$ as true. Based on this definition and the potential outcome model, then any intervention $\sigma(\mathbf{C}) = \mathbf{c}$ when $(\bigwedge(\mathbf{B}))_{\mathbf{c}} = 1$ will cause $D = 1$ for all $u \in U^*$.

Definition 2.2: A set $\mathbf{B} \subset \mathbb{L}(\mathbf{C})$ is a **minimal sufficient cause** for D relative to $\mathbf{C}$ in sub-population U^* if $\mathbf{B}$ is a sufficient cause for D in U^*, but no proper subset $\mathbf{B}^* \subset \mathbf{B}$ is also a sufficient cause for D in U^*. ■

The phrase 'minimal sufficient cause' has been first used in [47], that seems to be an impulse for [22], an influence for the sufficient causes model in epidemiology.

Sufficient and minimal sufficient cause have equivalent concepts in Boolean functions, Boolean reasoning and in digitial circuit theory [5, 20]. Sufficient cause is the same as implicant, while minimal sufficient cause is the same as prime implicant. This will be discussed in Section 3.

Definition 2.3: A set $\mathfrak{B} = \{\mathbf{B}_1, \ldots, \mathbf{B}_n\} \subseteq \mathbb{P}(\mathbb{L}(\mathbf{C}))$ is said to be **determinative** for D (relative to $\mathbf{C}$) in *sub-population* U^*, if for all $u \in U^*$ and all $\mathbf{c}$, $D_{\mathbf{c}} = 1$, if and only if $(\bigvee \bigwedge(\mathfrak{B}))_{\mathbf{c}} = 1$.

A determinative set of sufficient causes for D will also be referred to as a **sufficient cause model**. ■

In most settings it is unlikely that a single set $\mathbf{B}$ will be able to explain all cases of individuals with $D = 1$. Therefore, different sets of sufficient cause models are needed. The sufficient cause representation for $\mathfrak{D}(\mathbf{C}; U)$ is the collection of these subsets of the population and their corresponding sufficient cause models.

Definition 2.4: A **sufficient cause representation** $(\mathbf{A}, \mathfrak{B})$ for $\mathfrak{D}(\mathbf{C}; U)$ is an ordered set $\mathbf{A} = \langle A_1, \ldots, A_p \rangle$ of binary random variables with $(A_i)_{\mathbf{c}} = A_i$ for all $i, \mathbf{c}$, and a set of ordered causes $\mathfrak{B} = \langle \mathbf{B}_1, \ldots, \mathbf{B}_p \rangle$, such that for all $u, \mathbf{c}$, $D_{\mathbf{c}(u)} = 1 \Leftrightarrow$ for some j, $A_j(u) = 1$ and $(\bigvee \bigwedge(\mathfrak{B}_j))_{\mathbf{c}} = 1$. ■

It has been shown [42, Thm. 2.10] that a sufficient cause representation exists for all $\mathfrak{D}(\mathbf{C}; U)$. However, there seem to be two slightly different notions of sufficient cause representation in [42, pp. 2133-2134], and we have chosen as definition 2.4 the one that seems more useful.

The sufficent causes in $\mathfrak{B}$ can be shown to be determinative. A single indidividual can be associated with more than one of the random variables A_i. The condition $(A_i)_{\mathbf{c}} = A_i$ means that no intervention $\sigma(\mathbf{C}) = \mathbf{c}$ on the sub-population can change A_i. The variables A_i and the sufficient cause models $\mathbf{B}_i$ are paired via the orderings of $\mathbf{A}$ and $\mathfrak{B}$. So A_i sets up different pre-existing sub-populations with particular sets of sufficient causes and potential outcomes for D. There can be multiple possible sufficient cause representations that describe a population, however, certain conjunctions can be present in every representation.

Definition 2.5: $\mathbf{B} \in \mathbb{P}(\mathbb{L}(\mathbf{C}))$ is **irreducible** for $\mathfrak{D}(\mathbf{C}; U)$ if in every representation $(\mathbf{A}, \mathfrak{B})$ for $\mathfrak{D}(\mathbf{C}; U)$, there exists $\mathbf{B}_i \in \mathfrak{B}$, with $\mathbf{B} \subseteq \mathbf{B}_i$. ■

Irreducibility in this sense is sometimes referred to as 'sufficient cause interactions' between the components of $\mathbf{B}$, e.g., in [44]. However, if $\mathbf{B}$ is irreducible then in general it is not true that $\mathbf{B}$ is a sufficient cause, only that there is a sufficient cause that contains $\mathbf{B}$. If $|\mathbf{B}| = |\mathbf{C}|$, then $\mathbf{B}$ is a minimal sufficient cause if and only if $\mathbf{B}$ is irreducible.

It is possible to test for irreducibility as the following theorem shows. $\mathbf{C}_1 \dot{\cup} \mathbf{C}_2$ is the disjoint union between $\mathbf{C}_1$ and $\mathbf{C}_2$. The set $\mathbf{1}$ contains only ones(1). The symbol $\setminus$ is set difference.

Theorem 2.1: *Let* $\mathbf{C} = \mathbf{C}_1 \dot{\cup} \mathbf{C}_2$, $\mathbf{B} \in \mathbb{P}(\mathbb{L}(\mathbf{C}))$, $|\mathbf{B}| = |\mathbf{C}_1|$. *Then* $\mathbf{B}$ *is irreducible for* $\mathfrak{D}(\mathbf{C}; U)$ *if and only if there exists* $u^* \in U$ *and values* $\mathbf{c}_2$ *for* $\mathbf{C}_2$ *such that:*

(i) $D_{\mathbf{B}=\mathbf{1}, \mathbf{C}_2 = \mathbf{c}_2^*}(u*) = 1$

(ii) for all $L \in \mathbf{B}$, $D_{\mathbf{B}\setminus\{L\}=\mathbf{1}, L=0, \mathbf{C}_2=\mathbf{c}_2^*}(u^*) = 0$

The conditions (i) *and* (ii) *are equivalent to:*

$$D_{\mathbf{B}=\mathbf{1}, \mathbf{C}_2=\mathbf{c}_2^*}(u^*) - \sum_{L \in \mathbf{B}} D_{\mathbf{B}\setminus\{L\}} = \mathbf{1}, L = 0, \mathbf{C}_2 = \mathbf{c}_2^*(u^*) > 0. \tag{2.3}$$

Proof: See Theorem 3.2 in [42]. ■

This theorem implies the existence of an indidividual $u^* \in U$ that has outcome $D = 1$ if every literal in $\mathbf{B}$ is set to 1, but $D = 0$ if one of the literals in $\mathbf{B}$ is set to 0 and the other ones are set to 1.

3 Potential Cause Models expressed with Boolean Functions

We will now introduce the Boolean functions and express the sufficient cause models using them, and show how the sufficient cause representation is related to the Blake canonical form. We also show that the sufficient cause representation is also equivalent to the MFPO model in [11].

3.1 Boolean functions and Response Types

We are going to start with some notations for the boolean functions. Let $\underline{x} = (x_j)_{j=1}^d \in \{0,1\}^d$ and α be a Boolean function of d variables, i.e., $\alpha(\underline{x}) \in \{0,1\}$. We have denoted by $\mathfrak{D}(\mathbf{C}; u)$ the set of all potential outcomes for individual u. Hence $|\mathfrak{D}(\mathbf{C}; u)| = 2^d$ and there is a unique Boolean function α for each $\mathfrak{D}(\mathbf{C}; u)$, such that the set $\mathfrak{D}(\mathbf{C}; u)$ is the image of $\{0,1\}^d$ under action of α or

$$\forall u \in U \quad \exists! \alpha \quad \text{s.t. } \{\alpha(\underline{x})\}_{\underline{x} \in \{0,1\}^d} = \mathfrak{D}(\mathbf{C}; u). \tag{3.1}$$

Next we let $\underline{\alpha}$ denote a set, or table, given by

$$\underline{\alpha} = \{\alpha(\underline{x})\}_{\underline{x}\in\{0,1\}^d}. \tag{3.2}$$

If an individual $u \in U$ is such that

$$\alpha = \mathfrak{D}(\mathbf{C}; u), \tag{3.3}$$

then we say that $\alpha = \mathfrak{D}(\mathbf{C}; u)$ is the **individual response type (of u) induced by (the Boolean function)** α. The following example exhausts the important special case with $d = 2$.

Example 3.1: We have $\mathbf{C} = \{X_1, X_2\}$. The set of potential outcomes $\mathfrak{D}(\mathbf{C}; u)$ has four elements. There are $2^{2^2} = 16$ different types. We relate explicitly these $\mathfrak{D}(\mathbf{C}; u)$ to the sixteen different Boolean functions $\alpha_i(x_1, x_2)$ on $\{0, 1\}^2$.

Then $\underline{\alpha}$ defined as in (3.2) is an individual response type induced by one of the sixteen the Boolean functions. The 16 different response types are the rows in Table 1. Their indexing i is as given in [36, chap.5], c.f., [41, ch. 10.2-10.3], originally in [24]. Note that we traverse (x_1, x_2) in the same order for each i. Here and elsewhere the order is from $2^d - 1$ to 0, where $\underline{x}$ is seen as the binary representation of $i \in \{0, 1, \ldots, 2^d - 1\}$. ■

With $1 =$ true, and $0 =$ false the rows of Table 1 are the truth tables of the sixteen Boolean functions now in the role of individual response types. We can then express the rows of the table as: $\neg x_2 \mapsto \underline{\alpha}_{11}$, $\neg x_1 \mapsto \underline{\alpha}_{13}(x_1, x_2)$, $x_1 \vee x_2 \mapsto \underline{\alpha}_2(x_1, x_2)$, $x_1 \wedge x_2 \mapsto \underline{\alpha}_8(x_1, x_2)$. The function α_1, which is $= 1$ for all (x_1, x_2), is known as **tautology** and denoted by $\top$. Additional descriptive names of logical connectives and their respective definitions in terms of the basis ($\neg$, $\vee$, $\wedge$) will appear in the sequel.

Example 3.2: The symbol $\uparrow$ is defined as

$$\alpha \uparrow \beta \equiv \neg(\alpha \wedge \beta) \tag{3.4}$$

and is called Sheffer stroke and is known as NAND in the context of digital gates [20, p. 38] and is found as $\underline{\alpha}_9$ in the Table 1. Thus $\alpha_9 = x_1 \uparrow x_2$ induces

$$\underline{\alpha}_9 = \{\alpha(11) \quad \alpha(10) \quad \alpha(01) \quad \alpha(00)\} = \{0 \quad 1 \quad 1 \quad 1\} \tag{3.5}$$

■

3.2 To Sufficient Cause Models from Potential Cause Models

We consider for $\underline{e} = (e_j)_{j=1}^d \in \{0, 1\}^d$ the Boolean literals X_i and $\bar{X}_i (= 1 - X_i = \neg X_i)$ (the complement) so that

$$\tilde{X}(e_i) = \begin{cases} X_i & \text{if } e_i = 1 \\ \bar{X}_i & \text{if } e_i = 0. \end{cases} \tag{3.6}$$

Response Type	$\alpha(1,1)$	$\alpha(1,0)$	$\alpha(0,1)$	$\alpha(0,0)$
$\underline{\alpha}_1$	1	1	1	1
$\underline{\alpha}_2$	1	1	1	0
$\underline{\alpha}_3$	1	1	0	1
$\underline{\alpha}_4$	1	1	0	0
$\underline{\alpha}_5$	1	0	1	1
$\underline{\alpha}_6$	1	0	1	0
$\underline{\alpha}_7$	1	0	0	1
$\underline{\alpha}_8$	1	0	0	0
$\underline{\alpha}_9$	0	1	1	1
$\underline{\alpha}_{10}$	0	1	1	0
$\underline{\alpha}_{11}$	0	1	0	1
$\underline{\alpha}_{12}$	0	1	0	0
$\underline{\alpha}_{13}$	0	0	1	1
$\underline{\alpha}_{14}$	0	0	1	0
$\underline{\alpha}_{15}$	0	0	0	1
$\underline{\alpha}_{16}$	0	0	0	0

Table 1: The sixteen Individual Response Types $\underline{\alpha}_i \in \{0,1\}^{2^2}$ for two binary exposures.

Let next α be a Boolean function on $\{0,1\}^d$. Then we define the Boolean form

$$Q_\alpha(X) \equiv \bigvee_{\underline{e}:\alpha(\underline{e})=1} \bigwedge_{i=1}^{d} \tilde{X}(e_i). \tag{3.7}$$

Then we can use the representation map $<>$, [14, pp. 52-55] [33], on the Boolean form $Q_\alpha(X)$ to obtain uniquely the Boolean function $< Q_\alpha(X) >$,

$$\alpha(\underline{x}) =< Q_\alpha(X) > . \tag{3.8}$$

In terms of [5, ch. 3 and Appendix 2] we note the following. Consider any of the terms $p_e(\underline{x}) = \bigwedge_{i=1}^{d} \tilde{X}(e_i)$. If $p_e(\underline{x}) = 1$, then $\alpha(\underline{x}) = \bigvee_{\underline{e}:\alpha(\underline{e})=1} p_e(\underline{x}) = 1$. We say that p_e is a product term that implies α, i.e., $p_e = 1 \Rightarrow \alpha = 1$ and p_e is called an implicant of α. If no proper subterm of any p_e is an implicant, then p_e is a prime implicant of α. In this case (3.7) can be reduced to a BCF(α) **Blake canonical form**, a special DNF representation of a Boolean α by a disjunction of prime implicants. Hence, sufficient causes are implicants and minimal sufficient causes are prime implicants of D. BCF(α) and its relationship with the sufficent cause representation is explored later in this section.

Example 3.3: Take $d = 2$ and $\alpha_9 = \alpha_9(x_1, x_2) = x_1 \uparrow x_2$. Assume that there is a subpopulation U^* such that α_9 induces the individual response type of every $u \in U^*$, i.e.,

$$D(u)_{x_1,x_2} = \alpha_9(x_1, x_2) \quad \text{for all } (x_1, x_2) \in \{0,1\}^2.$$

From (3.4) we get $\{\underline{e} \in \{0,1\}^4 : \alpha_9(\underline{e}) = 1\} = \{(1,0),(0,1),(0,0)\}$. Then we have in (3.7), parentheses augmented for clarity (" $\bigwedge$ binds stronger than $\bigvee$", [33]),

$$Q_{\alpha_9}(X) = (X_1 \bigwedge \bar{X}_2) \bigvee (\bar{X}_1 \bigwedge X_2) \bigvee (\bar{X}_1 \bigwedge \bar{X}_2).$$

If $\mathbf{B}_1 = \{X_1, \bar{X}_2\}$, then with $\mathbf{c} = (1,0)$

$$(\bigwedge(\mathbf{B}_1))_{\mathbf{c}} = (X_1 \bigwedge \bar{X}_2)_{\mathbf{c}} = 1$$

and $D_{\mathbf{c}} = \alpha_9(1,0) = 1$. Hence, by definition 2.1, $\mathbf{B}_1$ is a sufficient cause for D (relative to some sub-population). In the same way we can show that $\mathbf{B}_2 = \{\bar{X}_1, X_2\}$ and $\mathbf{B}_3 = \{\bar{X}_1, \bar{X}_2\}$ are sufficient causes for D.

Then we take $\mathfrak{B} = \{\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3\}$ We see that $\mathfrak{B} \subseteq \mathbb{P}(\mathbb{L}(\{\mathbf{X_1}, \mathbf{X_2}\}))$ as required in Definition 2.3 and check that $(\bigvee \bigwedge(\mathfrak{B}))_{\mathbf{c}} = 1$ if and only if $D_{\mathbf{c}} = 1$. Hence $\mathfrak{B} = \{\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3\}$ is a sufficient cause model for $D \equiv \alpha_9$. $D \equiv \alpha_9$ means that $D_{x_1,x_2} = 1$ if and only if $\alpha_9(x_1, x_2) = 1$. ■

Theorem 3.1: *Let* $\underline{e} = (e_j)_{j=1}^d \in \{0,1\}^{2^d}$ *and define Boolean literals*

$$\tilde{X}_i(e_i) = \begin{cases} X_i & \text{if } e_i = 1 \\ \bar{X}_i & \text{if } e_i = 0. \end{cases} \tag{3.9}$$

Let α *be a Boolean function of* $\{0,1\}^d$. *Take all* $\underline{e}$ *such that* $\alpha(\underline{e}) = 1$ *and index these by* $\underline{e}^{(j)}$, $j = 1, \ldots, n_\alpha$. *Set for* $j = 1, \ldots n_\alpha$

$$\mathbf{B}_j = \{\tilde{X}_1(e_1^{(j)}), \tilde{X}_2(e_2^{(j)}), \ldots, \tilde{X}_d(e_d^{(j)})\}. \tag{3.10}$$

Then

$$\mathfrak{B} = \{\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_{n_\alpha}\} \tag{3.11}$$

is a sufficient cause model for D *in a subpopulation* U^* *such that* α *induces the individual response type of every* $u \in U^*$.

Proof: It follows in the same fashion as in Example 3.3 that each $\mathbf{B}_i$ is a sufficient cause. It holds clearly that $\mathfrak{B} \subseteq \mathbb{P}(\mathbb{L}(\{\mathbf{X_1}, \ldots, \mathbf{X_d}\}))$. We need to check the condition in definition 2.3.

Recall the assertion about subpopulations but supress u for simplicity. Hence we need therefore to verify that for all $\mathbf{c}$, $D_{\mathbf{c}} = 1$, if and only if $(\bigvee \bigwedge(\mathfrak{B}))_{\mathbf{c}} = 1$. We observe first that in the current setting that $D_{\mathbf{c}} = 1$ if and only if $\alpha(\mathbf{c}) = 1$.

$\Rightarrow$ Take now $\mathbf{c}$ such that $D_{\mathbf{c}} = 1$. Then insert the index of $\mathbf{c}$ by $\mathbf{c}^{(j)} = \left(c_i^{(j)}\right)_{i=1}^{d}$. Then $(\bigwedge(\mathbf{B}_j))_{\mathbf{c}^{(j)}} = 1$ by the construction in (3.9). Hence $(\bigvee\bigwedge(\mathfrak{B}))_{\mathbf{c}^{(j)}} = 1$ for the $\mathfrak{B}$ in (3.11).

$\Leftarrow$ Take any $\mathbf{c}$ so that $(\bigvee\bigwedge(\mathfrak{B}))_{\mathbf{c}} = 1$ for the $\mathfrak{B}$ in (3.11). Then there is at least one $\mathbf{B}_j$ in $\mathfrak{B}$ such that $(\bigwedge(\mathbf{B}_j))_{\mathbf{c}} = 1$ and this means by construction in (3.9) that $\alpha(\mathbf{c}) = 1$ ($\mathbf{c}$ is in the model of α). But then $D_{\mathbf{c}} = 1$. ■

In view of the preceding result we say for any individual u with the individual response profile $\underline{\alpha}$, that $Q_\alpha(X)$ in (3.7) is a **sufficient causes form** of $\underline{\alpha}$.

3.3 Blake Canonical Form and Sufficient Cause Representation

The BCF(α) for a Boolean function α is the function expressed as a *sum-of-products* (SOP) where each product is a prime implicant [5, 33]. In other words, the BCF(α) is the sum of all products of the variables in each minimal sufficient cause. It does not directly correspond to a sufficient cause representation, unless the representation consists only of minimal sufficient causes. In this section we define the minimal sufficient cause representation and show that it is functionally equivalent to the sufficient cause representation in relation to irreducibility.

Definition 3.1: *Minimal sufficient cause representation*
A sufficient cause representation $(\mathbf{A}_{min}, \mathfrak{B}_{min})$ for $\mathfrak{D}(\mathbf{C}; U)$ is a **minimal sufficient cause representation** if all $\mathbf{B}_i \in \mathfrak{B}_{min}$ are minimal sufficient causes. ■

Lemma 3.2: *Every sufficient cause representation* $(\mathbf{A}, \mathfrak{B})$ *for* $\mathfrak{D}(\mathbf{C}; U)$ *can be transformed into a minimal sufficient cause representation.*

Proof: For every $\mathbf{B}_i \in \mathfrak{B}$, let us form the corresponding minimal sufficient cause, $\mathbf{B}_i^* \in \mathbf{B}_i$, and set $A_i^* = A_i$, so that $\mathbf{B}_i^*$ is the minimal sufficient cause for the subpopulation defined by $A_i^* = 1$. Form the ordered sets $\mathbf{A}_{min} = \langle A_1^*, \ldots, A_p^* \rangle$ and $\mathfrak{B}_{min} = \langle \mathbf{B}_1^*, \ldots, \mathbf{B}_p^* \rangle$ and remove any duplicate minimal sufficient causes, and their corresponding A_i^*. $(\mathbf{A}_{min}, \mathfrak{B}_{min})$ is now a minimal sufficient cause representation for $\mathfrak{D}(\mathbf{C}; U)$. ■

Different sufficient cause representations can have the same minimal sufficient cause representation, since the sets in $\mathfrak{B}$ can have different events removed to form the sets in $\mathfrak{B}_{min}$. There is also not a unique minimal sufficient cause representation for $\mathfrak{D}(\mathbf{C}; U)$ as illustrated in the example below.

Example 3.4: For the population shown in Table 2 both representations $(\mathbf{A}^1_{min}, \mathfrak{B}^1_{min})$ and $(\mathbf{A}^2_{min}, \mathfrak{B}^2_{min})$ are minimal sufficient cause representations with:

$$\mathfrak{B}^1_{min} = \langle \{X_1, X_2, \bar{X}_3\}, \{X_1, X_2\}, \{X_2, \bar{X}_3\}, \{\bar{X}_2, X_3\} \rangle \tag{3.12}$$

$$\mathfrak{B}^2_{min} = \langle \{X_1, \bar{X}_2, X_3\}, \{X_1, X_3\}, \{X_2, \bar{X}_3\}, \{\bar{X}_2, X_3\} \rangle \tag{3.13}$$

$$A^1_{min} = A^2_{min} = \langle \mathbb{1}(u = 1), \mathbb{1}(u = 2), 1, 1 \rangle \tag{3.14}$$

All the sufficient causes in $\mathfrak{B}^1_{min}$ and $\mathfrak{B}^2_{min}$ are minimal sufficient causes in the sub-populations defined by A^1_{min} and A^2_{min} respectively. ■

Individual	D_{000}	D_{001}	D_{010}	D_{011}	D_{100}	D_{101}	D_{110}	D_{111}
1	0	1	1	0	0	1	1	0
2	0	1	1	0	0	1	1	1

Table 2: The potential outcomes for a population with two individuals and three events.

Theorem 3.3: $\mathbf{B} \in \mathbb{P}(\mathbb{L}(\mathbf{C}))$ *is irreducible for* $\mathfrak{D}(\mathbf{C}; U)$ *if and only if in every minimal sufficient cause representation* $(\mathbf{A}_{min}, \mathfrak{B}_{min})$ *for* $\mathfrak{D}(\mathbf{C}; U)$*, there exists* $\mathbf{B}^*_i \in \mathfrak{B}_{min}$*, with* $\mathbf{B} \subseteq \mathbf{B}^*_i$.

Proof: For all minimal sufficient causes, $\mathbf{B}^*_i \in \mathfrak{B}_{min}$, in the minimal sufficient cause representation $(\mathbf{A}_{min}, \mathfrak{B}_{min})$ with some non-minimal sufficient cause representation $(\mathbf{A}, \mathfrak{B})$ for $\mathfrak{D}(\mathbf{C}; U)$ and for some $\mathbf{B}_j \in \mathfrak{B}$ it must be true that $\mathbf{B}^*_i \subseteq \mathbf{B}_j$ with the same argument as in the proof of Lemma 3.2.

We know from Lemma 3.2 that all sufficient cause representations can be transformed into minimial sufficient cause representations. Based on the above then if for $(\mathbf{A}_{min}, \mathfrak{B}_{min})$ there exists $\mathbf{B}^*_i \in \mathfrak{B}_{min}$ with $\mathbf{B} \subseteq \mathbf{B}^*_i$, there must also exist $\mathbf{B}^*_j \in \mathfrak{B}$ with $\mathbf{B} \subseteq \mathbf{B}_j$ for all those non-minimal representations, $(\mathbf{A}, \mathfrak{B})$, that can be transformed into $(\mathbf{A}_{min}, \mathfrak{B}_{min})$.

Hence the above means that if there exists $\mathbf{B}^*_i \in \mathfrak{B}_{min}$ with $\mathbf{B} \subset \mathbf{B}^*_i$ for all minimal sufficient cause representations $(\mathbf{A}_{min}, \mathfrak{B}_{min})$ then there must also exist $\mathbf{B}^*_j \in \mathfrak{B}$ with $\mathbf{B} \subseteq \mathbf{B}_j$ for every sufficient cause representation $(\mathbf{A}, \mathfrak{B})$. But if for every representation $(\mathbf{A}, \mathfrak{B})$ there exists $\mathbf{B}_j \in \mathfrak{B}$ with $\mathbf{B} \subseteq \mathbf{B}_j$ then $\mathbf{B}$ is irreducible. ■

Based on the above theorem it is enough to check the minimal sufficient cause representations and not all representations for irreducibility. As the following theorem shows this has implications for finding irreducibility with BCF(α).

Theorem 3.4: *Let* $\mathbf{C} = \mathbf{C}_1 \dot{\cup} \mathbf{C}_2$, $\mathbf{B} \in \mathbb{P}(\mathbb{L}(\mathbf{C}))$, $|\mathbf{B}| = |\mathbf{C}_1|$. *For every minimal sufficient cause representations* $(\mathbf{A}, \mathfrak{B})$ *for* $\mathfrak{D}(\mathbf{C}; U)$ *form the corresponding Boolean function* α. *Then* $\mathbf{B}$ *is irreducible for* $\mathfrak{D}(\mathbf{C}; U)$ *if and only if there is a term consisting of the events in* $\mathbf{B}^*_i$*, with* $\mathbf{B} \subseteq \mathbf{B}^*_i$*, in every BCF(*$\alpha$*).*

Proof: $\Rightarrow$ Based on Theorem 3.3 $\mathbf{B}$ is irreducible if and only if in every minimal sufficient cause representation $(\mathbf{A}_{min}, \mathfrak{B}_{min})$ for $\mathfrak{D}(\mathbf{C}; U)$, there exists $\mathbf{B}_i^* \in \mathfrak{B}_{min}$, with $\mathbf{B} \subseteq \mathbf{B}_i^*$. Then each minimal sufficient cause representation, $(\mathbf{A}_{min}, \mathfrak{B}_{min})$, has a corresponding BCF in which the events $\mathbf{B}_i^*$ is a product term. Then every BCF has a product term that contains the events in $\mathbf{B}$ since $\mathbf{B} \subseteq \mathbf{B}_i^*$.

$\Leftarrow$ In every BCF there is a product term $\mathbf{B}_i^*$ with $\mathbf{B} \subseteq \mathbf{B}_i^*$. Each corresponding minimal sufficient cause representation, $(\mathbf{A}_{min}, \mathfrak{B}_{min})$, for the BCF then must have a minimal sufficient cause $\mathbf{B}_i^* \in \mathfrak{B}_{min}$, with $\mathbf{B} \subseteq \mathbf{B}_i^*$. Then it follows from Theorem 3.3 that $\mathbf{B}$ is irreducible.

3.4 Multifactor Potential Outcome and Sufficient Cause Representation

The work of [11] introduces the notion of complementary component causes, which in mathematical terms are binary random variables denoted by ξ_i in a conjunction with a Boolean function.

An individual is by definition at risk for sufficient cause i **if and only if** ξ_i **is present**, which, together with the appropriate combination of exposures in a certain set $\mathbf{B}$, **completes the sufficient cause**. A natural biologic idea about ξ_i would seem to be genetic susceptibility. If no individual in the population has or can have a particular ξ_i, then that sufficient cause is absent. The combination of sufficient causes for which an individual is at risk determines the potential outcome to each of the possible combinations of exposure factors.

Let us take k Boolean functions β_i on $\{0,1\}^d$, and let ξ_i be k independent $\text{Be}(\theta_i)$-distributed random variables. Let $\tau_i(\xi_i) = 1$, if $\xi_i = 1$, and $\tau_i(\xi_i) = 0$, if $\xi_i = 0$ and set

$$(\beta_i \wedge \tau_i)(\omega, \xi_i) \equiv \beta_i(\omega) \wedge \tau_i(\xi_i), \quad i = 1, \ldots, k. \tag{3.15}$$

These are well defined Boolean functions on $(\omega, \xi_i) \in \{0,1\}^{d+1}$. Then we define the multifactor potential outcome function by

$$\text{MFPO}_k \left((\beta_l)_{l=1}^k, (\tau_l)_{l=1}^k \right) \equiv \text{OR}_k \left(\beta_1 \wedge \tau_1, \ldots, \beta_k \wedge \tau_k \right), \tag{3.16}$$

where OR_k is the disjunction $\vee$ in k Boolean functions.

We show first that that the MFPO is a sufficient cause model representation for some $\mathfrak{D}(\mathbf{C}; U)$ in the sense of definition 2.4. For this we let $\mathbf{C} = \{X_1, \ldots, X_d\}$. Then we use the construction in Theorem 3.1. In view of this theorem we can construct for each β_i inside MFPO a sufficient cause model

$$\mathfrak{B}_{\beta_i} = \{\mathbf{B}_1^{(i)}, \mathbf{B}_2^{(i)}, \ldots, \mathbf{B}_{n_{\beta_i}}^{(i)}\}. \tag{3.17}$$

Next we consider the ordered set of sufficient cause models

$$\mathfrak{B} = \langle \mathfrak{B}_{\beta_1}, \ldots, \mathfrak{B}_{\beta_k} \rangle. \tag{3.18}$$

An individual $u \in U$ and $\mathbf{A} = \langle \xi_1, \ldots, \xi_k \rangle$ are linked as follows; for each u values of ξ_i, $i = 1, \ldots, k$ are independent samples of $\text{Be}(\theta_i)$, respectively, c.f., [11]. We denote an individual's sample value by $\xi_i(u)$. By this machinery an individual can assigned to more than one $\mathfrak{B}_{\beta_i}$, or $u \mapsto (\xi_1(u), \xi_1(u), \ldots, \xi_k(u))$.

Theorem 3.5: *Let* $\mathbf{A} = \langle \xi_1, \ldots, \xi_k \rangle$ *and* $\mathfrak{B}$ *as defined above. Then* $(\mathbf{A}, \mathfrak{B})$ *is a sufficient cause model representation of some* $D(\mathbf{C}; U)$.

Proof: By design

$$D(u) = 1 \Leftrightarrow \text{ MFPO}_k \left((\beta_l)_{l=1}^k, (\tau_l)_{l=1}^k \right)(u) = 1.$$

In view of definition 2.4 we must prove that for all u and $\mathbf{c} \in \{0,1\}^d$, $D_{\mathbf{c}}(u) = 1 \Leftrightarrow$ for some j, $\xi_j(u) = 1$ and $(\bigvee \bigwedge \mathfrak{B}_{\beta_j})_{\mathbf{c}} = 1$.

- $\Rightarrow$ $D_{\mathbf{c}}(u) = 1$ means that there is at least one j such that $(\beta_j \wedge \tau_j(u))_{\mathbf{c}} = 1$. It must be that $\xi_j(u) = 1$ and $\beta_j(\mathbf{c}) = 1$. But as $\mathfrak{B}_{\beta_j}$ is by the construction in Theorem 3.1 a sufficient cause of model of $D = \beta_j$, then $(\bigvee \bigwedge \mathfrak{B}_{\beta_j})_{\mathbf{c}} = 1$.
- $\Leftarrow$ Let for some j, and u $\xi_j(u) = 1$ and $(\bigvee \bigwedge (\mathfrak{B}_{\beta_j})_{\mathbf{c}} = 1$. By construction in Theorem 3.1 it follows that $\beta_j(\mathbf{c}) = 1$. Hence

$$D(u) = 1 \Leftrightarrow (\text{ MFPO}_k \left((\beta_l)_{l=1}^k, (\tau_l)_{l=1}^k \right)_{\mathbf{c}} (u) = 1.$$

3.5 Multifactor Potential Outcome and Sufficient Cause Representations with Two Events

The MFPO model with $d = 2$ in [11] deals in fact with a sufficient cause model representation in the sense of Theorem 3.5 based on $k = 9$ Boolean forms of $\mathbf{C} = \{X_1, X_2\}$. We have $\mathbf{A} = \langle \xi_1, \ldots, \xi_9 \rangle$.

The underlying Boolean functions are here denoted by β_i, $i = 1, \ldots, 9$, and correspond to the Boolean functions of Table 1 ($\underline{\alpha}_i \mapsto \alpha_i$) according to:

$$\begin{aligned} \beta_1 \equiv \top = \alpha_1, \quad \beta_2 \equiv \alpha_4 \qquad & \beta_3 \equiv \alpha_6 \\ \beta_4 \equiv \alpha_{13}, \beta_5 \equiv \alpha_{11}, \qquad & \beta_6 \equiv \alpha_8 \\ \beta_7 \equiv \alpha_{12}, \beta_8 \equiv \alpha_{14}, \qquad & \beta_9 \equiv \alpha_{15} \end{aligned} \tag{3.19}$$

These nine functions play an important role in clarifying the biologic meaning of sufficient cause [11], [12], [41], and [43, ch.10] and show explicitly the equivalence between the potential outcome model and a sufficient cause model representation.

In [12] these are called sufficient component types and denoted by SC_i. When we use (3.19) and Table 1 we get the Boolean forms

$$\begin{aligned}
\mathrm{SC}_1 &\leftrightarrow G_{\beta_1}(X) = X_1 \bigvee \bar{X}_1 = X_2 \bigvee \bar{X}_2 \\
\mathrm{SC}_2 &\leftrightarrow G_{\beta_2}(X) = X_1 \\
\mathrm{SC}_3 &\leftrightarrow G_{\beta_3}(X) = X_2 \\
\mathrm{SC}_4 &\leftrightarrow G_{\beta_4}(X) = \bar{X}_1 \\
\mathrm{SC}_5 &\leftrightarrow G_{\beta_5}(X) = \bar{X}_2 \\
\mathrm{SC}_6 &\leftrightarrow G_{\beta_6}(X) = X_1 \bigwedge X_2 \\
\mathrm{SC}_7 &\leftrightarrow G_{\beta_7}(X) = X_1 \bigwedge \bar{X}_2 \\
\mathrm{SC}_8 &\leftrightarrow G_{\beta_8}(X) = \bar{X}_1 \bigwedge X_2 \\
\mathrm{SC}_9 &\leftrightarrow G_{\beta_9}(X) = \bar{X}_1 \bigwedge \bar{X}_2.
\end{aligned} \tag{3.20}$$

Then one can represent all the sufficient cause forms for $\mathbf{C} = \{X_1, X_2\}$ defined according to (3.7) by means of the nine forms G as follows:

$$\begin{aligned}
Q_{\alpha_1}(X) &= G_{\beta_1}(X) \\
Q_{\alpha_2}(X) &= G_{\beta_6}(X) \bigvee G_{\beta_7}(X) \bigvee G_{\beta_8}(X), \\
Q_{\alpha_3}(X) &= G_{\beta_6}(X) \bigvee G_{\beta_7}(X) \bigvee G_{\beta_9}(X), \\
Q_{\alpha_4}(X) &= G_{\beta_6}(X) \bigvee G_{\beta_7}(X) (\equiv G_{\beta_2}), \\
Q_{\alpha_5}(X) &= G_{\beta_6}(X) \bigvee G_{\beta_8}(X) \bigvee G_{\beta_9}(X), \\
Q_{\alpha_6}(X) &= G_{\beta_6}(X) \bigvee G_{\beta_8}(X) (\equiv G_{\alpha_6}), \\
Q_{\alpha_7}(X) &= G_{\beta_6}(X) \bigvee G_{\beta_9}(X), \\
Q_{\alpha_8}(X) &= G_{\beta_6}(X), \\
Q_{\alpha_9}(X) &= G_{\beta_7}(X) \bigvee G_{\beta_8}(X) \bigvee G_{\beta_9}(X), \\
Q_{\alpha_{10}}(X) &= G_{\beta_7}(X) \bigvee G_{\beta_8}(X), \\
Q_{\alpha_{11}}(X) &= G_{\beta_7}(X) \bigvee G_{\beta_9}(X) (\equiv G_{\beta_5}), \\
Q_{\alpha_{12}}(X) &= G_{\beta_7}(X), \\
Q_{\alpha_{13}}(X) &= G_{\beta_7}(X) \bigvee G_{\beta_9}(X) (\equiv G_{\beta_4}), \\
Q_{\alpha_{14}}(X) &= G_{\beta_8}(X), \\
Q_{\alpha_{15}}(X) &= G_{\beta_9}(X).
\end{aligned} \tag{3.21}$$

Here α_{16} is excluded as it has no natural sufficient cause form. As is obvious from the above, there are several ways of writing a form $Q_\alpha(X)$ as a disjunction of $G_\beta(X)$. This corresponds to the biologic statement that several sufficient cause models may lead to the same individual response type. For example

$$Q_{\alpha_2}(X) = G_{\beta_6}(X) \bigvee G_{\beta_7}(X) \bigvee G_{\beta_8}(X) = G_{\beta_2}(X) \bigvee G_{\beta_3}(X). \tag{3.22}$$

4 Probabilistic Potential Outcomes and Probabilistic Response Profile

The study in [13] considers exposure to smoking, $X_1 = 1$, and exposure to asbestos (for more than some span of time), $X_2 = 1$. Smoking will not cause cancer in everyone. There seem to be individuals, who by virtue of genetic makeup or other things, like exposure to asbestos, are susceptible to the effects of smoking and others are not, to paraphrase [35]. It could also be that it is not deterministic which indidividual that gets the disease. In this section we will continue by a summary the definitions and concepts in the probabilistic sufficient causes theory due to Ramsahai [31].

4.1 Probabilistic Potential Outcomes

In the probabilistic potential outcome framework each set of exposures corresponds to a probability distribution (a Bernoulli distribution) of the potential outcome $D(u)$ for each individual u. In this sense the probability of the potential response $D_X(u) = 1$ is denoted by

$$P(D_X(u) = 1 \mid X, u, \sigma(X)), \tag{4.1}$$

where $\sigma(X) = (\sigma(X_1), \ldots, \sigma(X_d))$ denotes a intervention/treatment variable as discussed in the preceding. If $\sigma(\underline{x}) = \emptyset$, the probability $P(D(u) = 1 \mid X, u)$ refers to an outcome under exposures that happened to be X.

We make the **conditional exchangeability assumption**, i.e.,

$$\begin{aligned} &P(D_{X(u)=\underline{x}}(u) = 1 \mid X = \underline{x}, U = u) = \\ &P(D_{X(u)=\underline{x}}(u) = 1 \mid X = \underline{x}, U = u, \sigma(X) = \underline{x}). \end{aligned} \tag{4.2}$$

or that D is conditionally independent of $\sigma(X)$, when X and u are given. This is also known by other names like ignorable treatment assignment, no unmeasured confounding or exogeneity, see [41]. Due to the consistency assumption (2.2) we get in addition

$$P(D(u) = 1 \mid X = \underline{x}, u) = P(D_{X(u)=\underline{x}}(u) = 1 \mid X = \underline{x}, u, \sigma(X) = \underline{x}) \tag{4.3}$$

and if in fact the individual response type of the individual u is induced by α we write

$$P_\alpha(D = 1 \mid \underline{x}) = P(D(u) = 1 \mid X = \underline{x}, u), \tag{4.4}$$

where the dependence on α will be constructed explicitly below.

It can be shown, see [48], that the consistency and conditional exchangeability assumptions do not restrict the observed data distribution and that there is a construction of counterfactuals as a function of the observed data distribution.

4.2 Probabilistic Response Profile

The potential outcomes D for a given individual response profile are given by some Boolean function α. Hence $P_\alpha(D = 1 \mid \underline{x})$ must be equal to the number $P(\alpha = 1 \mid \underline{x})$, somehow determined. Let us first take a straighforward but unpractical approach. We choose in some manner 2^d numbers $\mathbf{p}_{\underline{x}} \in [0, 1]$ and assign each of these as the value for respective $P(D = 1 \mid \underline{x})$:

$$\mathbf{p}_{\underline{x}} = P(D = 1 \mid \underline{x}). \tag{4.5}$$

This is a probability of one potential outcome of an individual when exposed to $\underline{x}$. We have dropped the variable u for ease of notation. We have

$$\mathbf{p} = \left(\mathbf{p}_{\underline{x}}\right)_{\underline{x}\in\{0,1\}^d}, \tag{4.6}$$

i.e., $\mathbf{p} \in [0, 1]^{2^d}$, and represents an **invidual (probabilistic) response profile**. Next, a condition for describing $\mathbf{p}$ in terms of a Boolean function α and its response type given in (3.2) is presented.

Definition 4.1: An individual response profile **p exhibits a causal response profile of the type** $\underline{\alpha}$ if

$$\mathbf{p}_{\underline{x}} > \mathbf{p}_{\underline{x}'} \tag{4.7}$$

for all $\underline{x}$ and $\underline{x}^{'} \in \{0, 1\}^d$ such that $\alpha(\underline{x}) = 1$ and $\alpha(\underline{x}^{'}) = 0$. ■

A single $\mathbf{p}$ can correspond to more than one response types. We let

$$\mathcal{Q}\,(\mathbf{p}) = \{\text{causal response profiles } \underline{\alpha} \text{ exhibited by } \mathbf{p}\,\} \tag{4.8}$$

The interpretation of probabilistic causation in the definition 4.1 below rests on the lectures by Patrick Suppes [40] and Nancy Cartwright [6]. The idea is to allow imperfect or probabilistic regularities between exposure $\underline{x}$ and effect D. Instead of requiring that events of exposure are invariably followed by a certain event of effect, an exposure of some kind raises the probability of the potential outcome $D_{\underline{x}}$. The next technical statement to follow is due to [31, p.712].

Definition 4.2: Let $\underline{\alpha}$ be a causal response profile exhibited by $\mathbf{p}$. Let $Q_\alpha(X)$ be a sufficient causes form of $\underline{\alpha}$. $Q_\alpha(X)$ is said to be a **P-sufficient causes form** of $\mathbf{p} \in [0, 1]^{2^d}$. ■

5 Qualitative Bayesian Networks

The application of the conditions in the previous section requires that one specifies the 2^d numbers $\mathbf{p}_{\underline{x}} \in [0,1]$, while at the same time trying to model potential outcomes in some given situation of biologic interaction. Suppose, however, that we were to know that $\alpha \in \mathcal{Q}(\mathbf{p})$ for a given $\mathbf{p}$. Then we shall construct numbers $\mathbf{p}^{\alpha}_{\underline{x}} \in [0,1]^{2^d}$ by means of what is known as qualitative Bayesian networks [9, 21],so that $\mathbf{p}^{\alpha}$ has the particular causal response profile $\underline{\alpha}$ exhibited by $\mathbf{p}$. This construction is sparse in the sense that we need only d numbers to determine $\mathbf{p}^{\alpha}$.

Thereby we show that there exists for a number of interesting classes of Boolean functions in d variables at least one $\mathbf{p}$ satisfying the condition of probabilistic causation in definition 4.1. In addition we find interesting insights in sufficient causes models in terms of definition 4.2.

5.1 Independence of Causal Influence Modeling and the Interaction Function

In Independence of Causal Influence (ICI) modeling the exposures $\mathbf{C} = \{X_1, X_2, \ldots X_d\}$ influence a given outcome/effect D through respective **mediating** (and hidden) binary variables $\omega_1, \omega_2, \ldots \omega_d$ and act independently of each other. That is, ω_i is considered to be a mediator of the corresponding exposure variable, X_i, to the common effect D. The causal independence model involves the functions $p(\omega_j \mid x_j)$, the probability mass functions of ω_j given the exposure x_j. The mediators are connected to each otherto yield the final effect D through a joint deterministic Boolean function α. The probability of the potential effect D given the exposures $\mathbf{C}$ is defined as

$$P_{\alpha}(D = \delta|\underline{x}) \equiv \sum_{\omega|\alpha(\omega)=\delta} \prod_{j=1}^{d} p(\omega_j \mid x_j). \tag{5.1}$$

Here $\delta = 1$ or 0. We take $\delta = 1$ as the base and then the summation is over all states Ω of the hidden variables that make α true (=1). It is also an SOP, like the BCF.

The ICI model (or family of models) in (5.1) has been introduced in the artificial intelligence literature [15] and [49], see also [9]. There α is called the **interaction function**. One can easily convince oneself that $P_{\alpha}(D = \delta|\underline{x})$ corresponds to a factorization of probability along the directed acyclic graph in Figure 1, and therefore we can talk about a Bayesian network. The epithet qualitative indicates that we can state qualitative properties of $P_{\alpha}(D = \delta|\underline{x})$ without specifying any numerical values for $p(\omega_j \mid x_j)$.

We find first a calculus of the individual response profiles yielded by $P_{\alpha}(D = \delta|\underline{x})$ connected to the proposition calculus of the interaction functions. Clearly the ultimate purpose here is to connect this to the notion of a causal response profile of the type $\underline{\alpha}$.

One may/should think of situations, where the model does not capture all possible exposures/causes. To take into accout this, one might expand the formula (5.1) by an

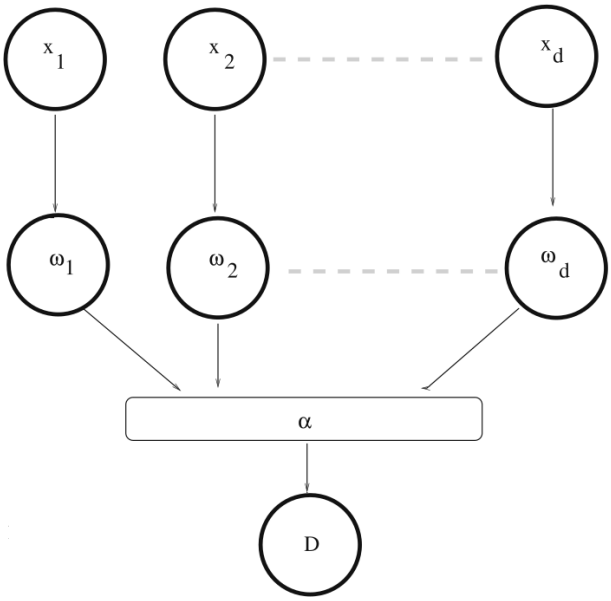


Figure 1: Graphical representation of the ICI model.

additional exposure, which summarizes the impacts of the unidentified causes influencing D, see Section 5.6.

5.2 Preliminaries

Let each ω be a binary d-string. Thereby we denote the binary hypercube as

$$\Omega = \{\omega | \omega = (\omega_j)_{j=1}^{d}, \omega_j \in \{0,1\}\}, \tag{5.2}$$

which will play the role of all outcomes for a probability space in the sequel. Let us take for $0 \leq q_j \leq 1$, where $x_j \in \{0,1\}$,

$$p(\omega_j \mid x_j) = (1 - q_j^{x_j})^{\omega_j} (q_j^{x_j})^{1-\omega_j}. \tag{5.3}$$

Then let us set

$$\mu(\omega; \underline{x}) \equiv \prod_{j=1}^{d} p(\omega_j \mid x_j) \tag{5.4}$$

and the probability mass function $\mu_{\underline{x}}$ is

$$\mu_{\underline{x}} = (\mu(\omega; \underline{x}))_{\omega \in \Omega}\,, \tag{5.5}$$

a multivariate Bernoulli distribution on Ω. Here $\underline{x} = (x_j)_{j=1}^d$ ($\in \{0,1\}^{2^d}$) enters as a parameter of $\mu_{\underline{x}}$.

Let $(\Omega, \mathcal{F}, \mu_{\underline{x}})$ be a probability space for any $\underline{x}$, where $\mathcal{F}$ is a sigma field of events in Ω. We define for any subset $A \in \mathcal{F}$ the probability

$$P_{\mu_{\underline{x}}}(A) \equiv \sum_{\omega \in A} \mu(\omega; \underline{x}). \tag{5.6}$$

Let

$$\alpha : \omega = (\omega_j)_{j=1}^d \mapsto \{0,1\}. \tag{5.7}$$

Consider now an individual u in a finite population. The probabilistic response profile $\underline{\alpha}$ of u turns out to be directly associated to as the response type $\underline{\alpha}$, when this α is acting on the exposures $\underline{x}$.

We define the binary random variable D in $(\Omega, \mathcal{F}, \mu_{\underline{x}})$ for α as

$$D(u, \omega) \equiv \alpha(\omega). \tag{5.8}$$

As is customary in probability theory we shall often drop ω from the explicit notation for D. Since the outcome space Ω is discrete and finite, we have no problems of measurability in defining D at this stage. Then it follows by (5.4) and (5.6) that

$$\begin{aligned} P_\alpha(D = 1 | \underline{x}) &= P_{\mu_{\underline{x}}}(\{\omega | \alpha(\omega) = 1\}) = \sum_{\omega : \alpha(\omega) = 1} \mu(\omega; \underline{x}) \\ &= \sum_{\omega | \alpha(\omega) = 1} \prod_{j=1}^d p(\omega_j \mid x_j). \end{aligned} \tag{5.9}$$

Then we get for (4.5)

$$\mathbf{p}_{\underline{x}}^{\alpha} = P_\alpha(D = 1 \mid \underline{x}) = \sum_{\omega | \alpha(\omega) = 1} \prod_{j=1}^d p(\omega_j \mid x_j), \tag{5.10}$$

and the probabilistic response profile,

$$\mathbf{p} = \left(\mathbf{p}_{\underline{x}} \right)_{\underline{x} \in \{0,1\}^d}, \tag{5.11}$$

for the individual u with the response type $\underline{\alpha}$ induced by α.

We note that

$$\sum_{\omega | \alpha(\omega) = 1} \prod_{j=1}^d p(\omega_j \mid x_j) = \sum_{\omega_1, \omega_2, \ldots, \omega_d : \alpha(\omega_1, \omega_2, \ldots, \omega_d) = 1} \prod_{j=1}^d p(\omega_j \mid x_j). \tag{5.12}$$

5.3 The Sum-Product Law

Several computations in the sequel evoking (5.12) and the corresponding later Fourier expansions involve some version of the so-called Sum-Product law or the generalized distributive law [2]. Let $f(\underline{x})$ be a real valued function of n variables x_i, where each x_i assumes values in a finite discrete set $\mathcal{A}_i$. Let $\{S_i\}_{i=1}^s$ be a partitioning of the set $\{1, 2, \ldots, n\}$ and let $\underline{x}^{(i)} = (x_{i_l})_{i_l \in \mathcal{S}_i}$ so that

$$\underline{x} = (x_1, \ldots, x_n) = (\underline{x}^{(1)}, \ldots, \underline{x}^{(s)}). \tag{5.13}$$

We have that $\underline{x}^{(i)} \in \mathcal{A}_{S_i} \equiv \times_{j \in \mathcal{S}_i} \mathcal{A}_{i_j}$. Suppose that $f(x_1, \ldots, x_n)$ factorizes as

$$f(\underline{x}) = \prod_{i=1}^{s} g_{S_i}(\underline{x}^{(i)}), \tag{5.14}$$

where $g_{S_i}(\underline{x}^{(i)})$ is a real valued function of the variables x_j with $j \in S_i$.

Then the **Sum-Product law** is

$$\begin{aligned} &\sum_{x_1 \in \mathcal{A}_1, \ldots, x_n \in \mathcal{A}_n} f(\underline{x}) \\ =& \sum_{x_1 \in \mathcal{A}_1, \ldots, x_n \in \mathcal{A}_n} \prod_{i=1}^{s} g_{S_i}(\underline{x}^{(i)}) \\ =& \prod_{i=1}^{s} \sum_{\underline{x}^{(i)} \in \mathcal{A}_{S_i}} g_{S_i}(\underline{x}^{(i)}). \end{aligned} \tag{5.15}$$

As a quick application of this law to the probabilities introduced in (5.10) we note ($\top(\omega) = 1$ for all ω)

$$\begin{aligned} &\sum_{\omega | \top(\omega)=1} \prod_{j=1}^{d} p(\omega_j \mid x_j) = \sum_{\omega_1, \omega_2, \ldots, \omega_d} \prod_{j=1}^{d} p(\omega_j \mid x_j) \\ =& \prod_{i=1}^{d} \sum_{\omega_j \in \{0,1\}} p(\omega_j \mid x_j) = 1, \end{aligned} \tag{5.16}$$

as should be. In addition, for $1 \leq k < d$ we have the probability marginalization,

$$\begin{aligned} &\sum_{\omega_1,\omega_2,\ldots,\omega_k} \prod_{j=1}^{d} p(\omega_j \mid x_j) \\ &= \prod_{j=k+1}^{d} p(\omega_j \mid x_j) \prod_{i=1}^{k} \sum_{\omega_j \in \{0,1\}} p(\omega_j \mid x_j) \\ &= \prod_{j=k+1}^{d} p(\omega_j \mid x_j). \end{aligned} \tag{5.17}$$

5.4 A Probability Calculus for ICI

As $P_{\mu_{\underline{x}}}$ is a probability measure on a finite set of events, one can find by standard probability calculus several useful rules tailored for computing with compounded individual response types. We set $\mathcal{L} = \{$ Boolean functions on $\Omega\}$. $\perp \in \mathcal{L}$ is called the contradiction, when $\perp(\omega) = 0$ for all ω. The following definition is found e.g. in [30, p. 176].

Definition 5.1: For every $\alpha \in \mathcal{L}$ and every $\beta \in \mathcal{L}$ we say that α **implies** β and write $\alpha \models \beta$, if the model of α is a subset of the model of β, i.e.,

$$\{\omega|\alpha(\omega) = 1\} \subseteq \{\omega|\beta(\omega) = 1\}. \tag{5.18}$$

$\alpha \equiv \beta$ means that

$$\{\omega|\alpha(\omega) = 1\} = \{\omega|\beta(\omega) = 1\}. \tag{5.19}$$

$\alpha \wedge \beta \equiv \perp$ means that

$$\{\omega|\alpha(\omega) = 1\} \cap \{\omega|\beta(\omega) = 1\} = \emptyset. \tag{5.20}$$

■

Inherent in all that follows is the following assumption:

Assumption 5.1: The probability functions $p(\omega_j \mid x_j)$ in (5.3) are for all j the same for all $\alpha \in \mathcal{L}$. ■

This can be justified by thinking of any $p(\omega_j \mid x_j)$ in (5.10) as a property of a mediator with biologic characteristics that are not adapted to individual response profiles.

Lemma 5.1: *For any $\alpha \in \mathcal{L}$ and any $\beta \in \mathcal{L}$, if $\alpha \models \beta$, then*

$$P_\alpha(D = 1|\underline{x}) \leq P_\beta(D = 1|\underline{x}). \tag{5.21}$$

for every $\underline{x}$.

Proof: In view of (5.6) and (5.9)

$$P_\alpha(D=1|\underline{x}) = P_{\mu_{\underline{x}}}(\{\omega|\alpha(\omega)=1\}) \tag{5.22}$$

and the assumption $\alpha \models \beta$ means that (5.18) holds, which gives by (5.6) and (5.9)

$$\leq P_{\mu_{\underline{x}}}(\{\omega|\beta(\omega)=1\}) = P_\beta(D=1|\underline{x}). \tag{5.23}$$

■

Proposition 5.2: *For all $\alpha \in \mathcal{L}$,*

$$0 \leq P_\alpha(D=1|\underline{x}) \leq P_\top(D=1|\underline{x}) = 1. \tag{5.24}$$

Proof: It is obvious by (5.12) that $0 \leq P_\alpha(D=1|\underline{x})$. The fact that $P_\top(D=1|\underline{x})=1$ is established in (5.16). We have by definition of $\top$ that $\alpha \models \top$ for any $\alpha \in \mathcal{L}$. Hence by lemma 5.1 $P_\alpha(D=1|\underline{x}) \leq P_\top(D=1|\underline{x})$, which establishes the remaining assertion. ■

The preceding result and the statements in proposition 5.3 below are parallels to the rules of probability on propositional languages, c.f., [17, 30]. However, the formulae below are in fact more of a practical interface between Boolean functions and probability calculus on events. Hence the case **g)** and later proposition 5.6 are not available in [17, 30].

Proposition 5.3: *For all α and $\beta \in \mathcal{L}$,*

a) *If $\alpha \equiv \beta$, then*

$$P_\alpha(D=1|\underline{x}) = P_\beta(D=1|\underline{x}). \tag{5.25}$$

b)

$$P_{\neg\alpha}(D=1|\underline{x}) = 1 - P_\alpha(D=1|\underline{x}). \tag{5.26}$$

c)

$$P_{\alpha\vee\beta}(D=1|\underline{x}) = P_\alpha(D=1|\underline{x}) + P_\beta(D=1|\underline{x}) - P_{\alpha\wedge\beta}(D=1|\underline{x}). \tag{5.27}$$

d) *If $\alpha \wedge \beta \equiv \bot$, then*

$$P_{\alpha\wedge\beta}(D=1|\underline{x}) = 0. \tag{5.28}$$

e)

$$P_\alpha(D=1|\underline{x}) = P_{\alpha\wedge\beta}(D=1|\underline{x}) + P_{\alpha\wedge\neg\beta}(D=1|\underline{x}). \tag{5.29}$$

f) *$\alpha \to \beta$ is material implication. Then*

$$P_{\alpha\to\beta}(D=1|\underline{x}) = 1 - P_\alpha(D=1|\underline{x}) + P_{\alpha\wedge\beta}(D=1|\underline{x}). \tag{5.30}$$

g)

$$P_{\alpha\to\beta}(D=1|\underline{x}) = 1 - P_\alpha(D=1|\underline{x})\left(1 - P_{\mu_{\underline{x}}}(\beta=1 \mid \alpha=1)\right). \tag{5.31}$$

Proof:

a) The equality (5.25) follows by (5.6), since $\alpha \equiv \beta$ means

$$\{\omega|\alpha(\omega)=1\} = \{\omega|\beta(\omega)=1\}.$$

b)

$$\begin{aligned} &P_{\neg\alpha}(D=1|\underline{x}) = P_{\mu_{\underline{x}}}(\{\omega|\neg\alpha(\omega)=1\}) = P_{\mu_{\underline{x}}}(\{\omega|\alpha(\omega)=0\}) \\ &= 1 - P_{\mu_{\underline{x}}}(\{\omega|\alpha(\omega)=1\}) = 1 - P_\alpha(D=1|\underline{x}), \end{aligned}$$

c) For any α and β in $\mathcal{L}$

$$\begin{aligned} &P_{\alpha\vee\beta}(D=1|\underline{x}) = P_{\mu_{\underline{x}}}(\{\omega|(\alpha\vee\beta)(\omega)=1\}) \\ &= P_{\mu_{\underline{x}}}(\{\omega|\alpha(\omega)=1\} \cup \{\omega|\beta(\omega)=1\}), \end{aligned}$$

where we invoked the well known correspondence between $\vee$ and the set operation $\cup$. This gives by the rule for probability $P_{\mu_{\underline{x}}}$ on a union of events

$$\begin{aligned} &= P_{\mu_{\underline{x}}}(\{\omega|\alpha(\omega)=1\}) + P_{\mu_{\underline{x}}}(\{\omega|\beta(\omega)=1\}) \\ &- P_{\mu_{\underline{x}}}((\{\omega|\alpha(\omega)=1\} \cap \{\omega|\beta(\omega)=1\})), \end{aligned}$$

and by the well known correspondence between $\wedge$ and the set operation $\cap$ we get

$$\begin{aligned} &= P_{\mu_{\underline{x}}}(\{\omega|\alpha(\omega)=1\}) + P_{\mu_{\underline{x}}}(\{\omega|\beta(\omega)=1\}) \\ &- P_{\mu_{\underline{x}}}((\{\omega|(\alpha\wedge\beta)(\omega)=1\}). \end{aligned}$$

d) If $\alpha \wedge \beta \equiv \bot$, we have by (5.25) and (5.20)

$$P_{\alpha\wedge\beta}(D=1|\underline{x}) = P_\bot(D=1|\underline{x}) = P_{\mu_{\underline{x}}}(\emptyset) = 0.$$

Alternatively, since $\bot \equiv \neg\top$, (5.20) and (5.26) yield $P_\bot(D=1|\underline{x}) = 0$.

e) For (5.29), we note that $\alpha \equiv \alpha \wedge \top \equiv \alpha \wedge (\beta \vee \neg\beta)$. Furthermore, $\alpha \wedge (\beta \vee \neg\beta) \equiv (\alpha \wedge \beta) \vee (\alpha \wedge \neg\beta)$. Thus by (5.25)

$$P_\alpha(D=1|\underline{x}) = P_{(\alpha\wedge\beta)\vee(\alpha\wedge\neg\beta)}(D=1|\underline{x})$$

and by (5.27)

$$= P_{\alpha\wedge\beta}(D=1|\underline{x}) + P_{\alpha\wedge\neg\beta}(D=1|\underline{x}),$$

by **d)**, since $(\alpha \wedge \beta) \wedge (\alpha \wedge \neg\beta) \equiv \bot$. Therefore (5.29) holds.

f) Finally we want to find

$$P_{\alpha\to\beta}(D=1|\underline{x}).$$

Due to due to the fact that $\alpha \to \beta \equiv \neg\alpha \vee \beta$ and (5.25)

$$P_{\alpha\to\beta}(D=1|\underline{x}) = P_{\neg\alpha\vee\beta}(D=1|\underline{x}).$$

We have $\neg\alpha \vee \beta \equiv \neg(\alpha \wedge \neg\beta)$ by De Morgan. Thus (5.26) gives

$$P_{\neg\alpha\vee\beta}(D=1|\underline{x}) = P_{\neg(\alpha\wedge\neg\beta)}(D=1|\underline{x}) = 1 - P_{\alpha\wedge\neg\beta}(D=1|\underline{x})$$

and (5.29) yields

$$= 1 - P_{\alpha}(D=1|\underline{x}) + P_{\alpha\wedge\beta}(D=1|\underline{x}).$$

g) Now,

$$\begin{aligned} &P_{\alpha\wedge\beta}(D=1|\underline{x}) = P_{\mu_{\underline{x}}}(\{\omega|(\alpha\wedge\beta)(\omega)=1\}) \\ &= P_{\mu_{\underline{x}}}(\{\omega|\alpha(\omega)=1\}\cap\{\omega|\beta(\omega)=1\}). \end{aligned}$$

Using the definition of conditional probability on the above we can write

$$\begin{aligned} &P_{\alpha\wedge\beta}(D=1|\underline{x}) \\ &= P_{\mu_{\underline{x}}}(\{\omega|\beta(\omega)=1\} \mid \{\omega|\alpha(\omega)=1\}) \cdot P_{\alpha}(D=1|\underline{x}), \end{aligned} \tag{5.32}$$

which yields (5.31) when inserted in (5.30). ■

Corollary 5.4: *If α and β are independent random variables on $(\Omega, \mathcal{F}, \mu_{\underline{x}})$, then*

$$P_{\alpha\wedge\beta}(D=1|\underline{x}) = P_{\beta}(D=1|\underline{x})P_{\alpha}(D=1|\underline{x}) \tag{5.33}$$

Proof: When α and β are independent random variables on $(\Omega, \mathcal{F}, \mu_{\underline{x}})$, then

$$P_{\mu_{\underline{x}}}(\{\omega|\beta(\omega)=1\} \mid \{\omega|\alpha(\omega)=1\}) = P_{\mu_{\underline{x}}}(\{\omega|\beta(\omega)=1\}).$$

The claim follows by (5.32). ■

In the sequel we will be need the notion of a symmetric Boolean function [26, p. 28]. Let π be a permutation of $\{1, 2, \ldots, d\}$. On $\underline{x} \in \{0,1\}^d$ a permutation acts as $\pi(\underline{x})_i = x_{\pi(i)}$. On functions we have $\alpha^{\pi}(\underline{x}) = \alpha(\pi(\underline{x}))$. A Boolean function α is **symmetric**, if $\alpha^{\pi}(\underline{x}) = \alpha(\underline{x})$ for all permutations π of $\{1, 2, \ldots, d\}$. We state the following obvious fact for ease of reference.

Proposition 5.5: *If α is symmetric, then*

$$P_{\alpha^\pi}(D=1|\underline{x}) = P_\alpha(D=1|\underline{x}). \tag{5.34}$$

Proof: By definition of a symmetric Boolean function on $\{0,1\}^d$ it holds for every permutation π of that $\{0,1\}^d$

$$\begin{aligned} P_{\alpha^\pi}(D=1|\underline{x}) &= \sum_{\omega\in\{0,1\}^d:\alpha^\pi(\omega)=1} \mu(\omega;\underline{x}) \\ &= \sum_{\omega\in\{0,1\}^d:\alpha(\omega)=1} \mu(\omega;\underline{x}) = P_\alpha(D=1|\underline{x}). \end{aligned}$$

■

5.5 Interpetations

There is, however, an issue with assigning a meaning in terms of exposure and potential outcome to the expressions just derived.

The inequality (5.21) means that an individual with the response profile induced by β has a greater risk than an individual with the response profile induced by α, as soon as $\alpha \models \beta$. In the next proposition $\top$ is the tautology, $\top(\omega) = 1$ for every $\omega \in \Omega$, which correspond to the doomed respones profile.

The statement in (5.25) makes sense, as it says that if an individual u has the response profile $\underline{\alpha}$ and another individual $u^{'}$ has the response profile $\underline{\beta}$, and $\alpha \equiv \beta$, then u and $u^{'}$ must have the same probability (risk) of the outcome $D = 1$.

The rule (5.26) is useful for many purposes, e.g., in checks of computations. In (5.26), D in the left hand side is given by $D(\omega) = \neg\alpha(\omega)$ and D in the right hand side is given by $D(\omega) = \alpha(\omega)$, i.e., we are facing two different random variables rendering the notational apparatus potentially misleading. Hence (5.26) says that if an individual u has the response profile $\underline{\alpha}$ and another individual $u^{'}$ has the response profile $\neg\underline{\alpha}$, then risk for $u^{'}$ of the outcome $D = D(u^{'}) = 1$ is one minus the risk of the outcome $D = D(u) = 1$. Same comment on D is valid for the rest of the formulas stated in the proposition above.

5.6 Expansion Formula for Probability of Potential Outcome

The following result corresponds to an expansion formula for Boolean functions. We need some notations. Let $\alpha = \alpha(\omega, \omega_{d+1})$, where $\omega \in \{0,1\}^d$ and $\omega_{d+1} \in \{0,1\}$. We extend (5.9) by

$$P_\alpha(D=1 \mid (\underline{x}, x_{d+1})) = \sum_{\omega|\alpha(\omega)=1} \prod_{j=1}^{d+1} p(\omega_j \mid x_j),$$

where $p(\omega_{d+1} \mid x_{d+1}) = (1-q_{d+1}^{x_{d+1}})^{\omega_{d+1}}(q_{d+1}^{x_{d+1}})^{1-\omega_{d+1}}$.

We need also (c.f. [26, p.30]) the subfunction defined as

$$\alpha^{(d+1)\mapsto o}(\omega) \equiv \alpha(\omega, o), \quad o \in \{0,1\}. \tag{5.35}$$

Proposition 5.6: *For a Boolean function $\alpha = \alpha(\omega, \omega_{d+1})$ it holds that*

$$\begin{aligned} P_\alpha(D=1|(\underline{x}, x_{d+1})) &= (1-q_{d+1}^{x_{d+1}}) P_{\mu_{\underline{x}}}\left(\{\omega|\alpha^{(d+1)\mapsto 1}(\omega)=1\}\right) \\ &+ q_{d+1}^{x_{d+1}} P_{\mu_{\underline{x}}}\left(\{\omega|\alpha^{(d+1)\mapsto 0}(\omega)=1\}\right). \end{aligned} \tag{5.36}$$

Proof: Let β be a Boolean function on $\{0,1\}^{d+1}$ defined as the projection $\beta(\omega, \omega_{d+1}) = \omega_{d+1}$. Then **e)**, or (5.29), in proposition (5.3) entails

$$P_\alpha(D=1|(\underline{x}, x_{d+1})) = P_{\alpha\wedge\beta}(D=1|(\underline{x}, x_{d+1})) + P_{\alpha\wedge\neg\beta}(D=1|\ (\underline{x}, x_{d+1})), \tag{5.37}$$

where

$$P_{\alpha\wedge\beta}(D=1|(\underline{x}, x_{d+1})) = P_{\mu_{(\underline{x}, x_{d+1})}}(\{(\omega, \omega_{d+1})|(\alpha\wedge\beta)((\omega, \omega_{d+1})=1\})$$

and

$$P_{\alpha\wedge\neg\beta}(D=1|\ (\underline{x}, x_{d+1})) = P_{\mu_{(\underline{x}, x_{d+1})}}(\{(\omega, \omega_{d+1})|(\alpha\wedge\neg\beta)((\omega, \omega_{d+1})=1\}).$$

Here

$$(\alpha\wedge\beta)(\omega, \omega_{d+1}) = 1$$

is equivalent to

$$(\alpha\wedge\beta)(\omega, \omega_{d+1}) = \alpha(\omega, \omega_{d+1}) \wedge \beta(\omega, \omega_{d+1}) = \alpha(\omega, \omega_{d+1}) \wedge \omega_{d+1} = 1.$$

Next, $\alpha(\omega, \omega_{d+1}) \wedge \omega_{d+1} = 1$ holds if and only if $\omega_{d+1} = 1$ and $\alpha(\omega, \omega_{d+1}) = 1$. Hence, we are in fact dealing with

$$\alpha(\omega, 1) \wedge \omega_{d+1} = 1$$

or with the notation in (5.35)

$$\alpha^{(d+1)\mapsto 1}(\omega) \wedge \omega_{d+1} = 1.$$

Thus

$$\begin{aligned} &P_{\alpha\wedge\beta}(D=1|(\underline{x}, x_{d+1})) = P_{\alpha^{(d+1)\mapsto 1}\wedge\omega_{d+1}}(D=1|(\underline{x}, x_{d+1})) \\ &= \sum_{(\omega,\omega_{d+1})=1|\alpha^{(d+1)\mapsto 1}\wedge 1=1} \prod_{j=1}^{d+1} p(\omega_j \mid x_j) = (1-q_{d+1}^{x_{d+1}}) P_{\mu_{\underline{x}}}\left(\{\omega|\alpha^{(d+1)\mapsto 1}(\omega)=1\}\right). \end{aligned}$$

The asserted expression for the second term in the right hand side of (5.37) is obtained analogously, after the observation that

$$(\alpha \wedge \neg\beta)(\omega, \omega_{d+1}) = 1$$

if and only if

$$\alpha^{(d+1)\mapsto 0} \wedge \neg\omega_{d+1} = 1.$$

■

The result in proposition 5.6 can be regarded as the general formula for ICI and impacts of unidentified causes, when we use

$$\begin{aligned} P_\alpha(D = 1|(\underline{x}, 1)) &= (1 - q_{d+1}) P_{\mu_{\underline{x}}}\left(\{\omega|\alpha^{(d+1)\mapsto 1}(\omega) = 1\}\right) \\ &+ q_{d+1} P_{\mu_{\underline{x}}}\left(\{\omega|\alpha^{(d+1)\mapsto 0}(\omega) = 1\}\right). \end{aligned} \tag{5.38}$$

Examples will be given in the sequel. Note that

$$P_\alpha(D = 1|(\underline{x}, 0)) = P_{\mu_{\underline{x}}}\left(\{\omega|\alpha^{(d+1)\mapsto 0}(\omega) = 1\}\right).$$

in which case no background variable was added. In the course of the preceding proof we provided most of the (short) proof a so-called **expansion formula for Boolean functions**, [20, Thm 2.1.3 p. 50], i.e.,

$$\alpha(\omega_1, \ldots, \omega_{d+1}) = (\alpha^{(d+1)\mapsto 1}(\omega) \wedge \omega_{d+1}) \vee (\alpha^{(d+1)\mapsto 0}(\omega) \wedge \neg\omega_{d+1}). \tag{5.39}$$

Alternatively, (5.36) could have been etablished by directly starting from the expansion in (5.39).

5.7 Probabilistic Individual Response Profiles with ICI and Two Events

We compute for Boolean functions α_i the repsonse patterns in in Table 1

$$P_{\alpha_i}(D = 1 \mid x_1, x_2) \equiv \sum_{\omega_1,\omega_2:\alpha_i(\omega_1,\omega_2)=1} \prod_{j=1}^{2} \mu(\omega_j \mid x_j).$$

This yileds 16 possible probabilistic response profiles for an individual u, u is droppoed from the notation.

By the straighforward summations of products and by (5.26) we get the following expressions.

α_1**:** $= \tau(\omega_1, \omega_2) = 1$ tautology or doomed

$$P_{\alpha_1}(D = 1 \mid x_1, x_2) = 1$$

α_2**:** $= (\omega_1 \vee \omega_2)$ Noisy -Or

$$P_{\alpha_2}(D = 1 \mid x_1, x_2) = 1 - q_1^{x_1} q_2^{x_2}$$

α_3**:** $= (\omega_2 \to \omega_1)$

$$P_{\alpha_3}(D = 1 \mid x_1, x_2) = 1 - q_1^{x_1}(1 - q_2^{x_2})$$

α_4**:** $= \omega_1$

$$P_{\alpha_4}(D = 1 \mid x_1, x_2) = 1 - q_1^{x_1}$$

α_5**:** $= (\omega_1 \to \omega_2)$

$$P_{\alpha_5}(D = 1 \mid x_1, x_2) = 1 - q_2^{x_2}(1 - q_1^{x_1})$$

α_6**:** $= \omega_2$

$$P_{\alpha_6}(D = 1 \mid x_1, x_2) = 1 - q_2^{x_2}$$

α_7**:** $= (\omega_1 \leftrightarrow \omega_2)$

$$P_{\alpha_7}(D = 1 \mid x_1, x_2) = 1 - q_2^{x_2}(1 - q_1^{x_1}) - q_1^{x_1}(1 - q_2^{x_2})$$

α_8**:** $= (\omega_1 \wedge \omega_2)$ Noisy-And,

$$P_{\alpha_8}(D = 1 \mid x_1, x_2) = 1 - q_1^{x_1} - q_2^{x_2} + q_1^{x_1} q_2^{x_2}$$

α_9**:** $= (\omega_1 \uparrow \omega_2)$ the noisy Sheffer stroke or noisy NAND

$$P_{\alpha_9}(D = 1 \mid x_1, x_2) = q_1^{x_1}(1 - q_2^{x_2}) + q_2^{x_2}$$

$$= 1 - P_{\alpha_8}(D = 1 \mid x_1, x_2)$$

α_{10}**:** $= (\omega_1 \oplus \omega_2)$ Noisy Exclusive - OR

$$P_{\alpha_{10}}(D = 1 \mid x_1, x_2) = q_1^{x_1}(1 - q_2^{x_2}) + q_2^{x_2}(1 - q_1^{x_1})$$

$$= 1 - P_{\alpha_7}(D = 1 \mid x_1, x_2)$$

α_{11}**:** $= \neg\omega_2$

$$P_{\alpha_{11}}(D = 1 \mid x_1, x_2) = q_2^{x_2}$$

$$= 1 - P_{\alpha_6}(D = 1 \mid x_1, x_2)$$

α_{12}**:** $= \neg(\omega_1 \to \omega_2)$ Noisy-And-Not

$$P_{\alpha_{12}}(D = 1 \mid x_1, x_2) = q_2^{x_2}(1 - q_1^{x_1})$$
$$= 1 - P_{\alpha_5}(D = 1 \mid x_1, x_2)$$

α_{13}**:** $\neg\omega_1$

$$P_{\alpha_{13}}(D = 1 \mid x_1, x_2) = q_1^{x_1}$$
$$= 1 - P_{\alpha_4}(D = 1 \mid x_1, x_2)$$

α_{14}**:** $= (\neg\omega_1 \wedge \omega_2) \equiv \neg(\omega_2 \to \omega_1)$

$$P_{\alpha_{14}}(D = 1 \mid x_1, x_2) = q_1^{x_1}\,(1 - q_2^{x_2})$$

α_{15}**:** $= \neg(\omega_1 \vee \omega_2)$

$$P_{\alpha_{15}}(D = 1 \mid x_1, x_2) = q_1^{x_1} q_2^{x_2}$$
$$= 1 - P_{\alpha_7}(D = 1 \mid x_1, x_2)$$

α_{16}**:** $= \bot(\omega_1, \omega_2)$ contradiction or immunity.

$$P_{\alpha_{16}}(D = 1 \mid x_1, x_2) = 0$$
$$= 1 - P_{\alpha_1}(D = 1 \mid x_1, x_2)$$

The individual response type $\underline{\alpha}$ is with ICI a degenerate case of the probabilistic response profile $\mathbf{p}$. Let us consider the case of two exposures indexed as 1 and 2 and $\mathbf{p}^{\alpha_i}_{x_1,x_2} = P_{\alpha_i}(D = 1 \mid x_1, x_2)$. Then $\mathbf{p}^{\alpha_i}_{x_1,x_2} \to \alpha_i(x_1, x_2)$ as $q_1 \to 0$ and $q_2 \to 0$. This is demonstrated by the Example below.

Example 5.1: Let us consider

$$\mathbf{p}^{\alpha_9}_{x_1,x_2} = q_1^{x_1}(1 - q_2^{x_2}) + q_2^{x_2}.$$

The table of values of this generalized Boolean function (Noisy Sheffer stroke a.k.a. Noisy NAND) is

x_1	x_2	$\mathbf{p}^{\alpha_9}_{x_1,x_2}$
1	1	$q_1(1 - q_2) + q_2$
1	0	1
0	1	1
0	0	1

If we now set $q_1 = q_2 = 0$, we recover the indivual response profile $\underline{\alpha}_9$ of the Sheffer stroke $x_1 \uparrow x_2$, c.f., Example 3.4. ■

The question is whether these probabilistic profiles exhibit causal response profile of the corresponding α_i. We check only a portion of the cases.

6 Probabilistic Sufficient Causes and ICI

In this section we apply the general setting of (5.3), (5.4), and (5.9), with different individual response profiles $\underline{\alpha}$ and d and find the P-sufficient forms.

6.1 Examples: ICI with Two Exposures, P-sufficient causes

Example 6.1 (Noisy Material Implication): Take $\alpha_5 = \omega_i \to \omega_j$ or equivalentely with projections $\alpha(\omega) = \omega_i$, $\beta(\omega) = \omega_j$, $i \neq j$ and $\alpha \to \beta$. We compute by means of (5.30). First we get by our construction in (5.9)

$$P_\alpha(D = 1|\underline{x}) = P_{\mu_{\underline{x}}}(\omega_i = 1)$$

$$= \sum_{\omega_1,\ldots,\omega_{i-1},\omega_{i+1},\ldots\omega_d} P_{\mu_{\underline{x}}}(\omega_1, \ldots, \omega_{i-1}, \omega_i = 1, \omega_{i+1}, \ldots \omega_d)$$

$$= p(\omega_i = 1|x_i) \sum_{\omega_1,\ldots,\omega_{i-1},\omega_{i+1},\ldots\omega_d} \prod_{l\neq i} p(\omega_l|x_l)) = p(\omega_i = 1|x_i),$$

by the Sum-Product Law. In the same way

$$P_\beta(D = 1|\underline{x}) = p(\omega_j = 1|x_j).$$

Thus we get in (5.30)

$$P_{\omega_i\to\omega_j}(D = 1|\underline{x}) = 1 - P_{\omega_i}(D = 1|x_i) + P_{\omega_i\wedge\omega_j}(D = 1|\underline{x})$$

and by the above and by definition of conjunction and independence of ω_i and ω_j in (5.4)

$$= 1 - p(\omega_i = 1 \mid x_i) + p(\omega_i = 1 \mid x_i)p(\omega_j = 1 \mid x_j)$$

$$= 1 - (1 - q_i^{x_i}) + (1 - q_i^{x_i})(1 - q_j^{x_j})$$
$$= q_i^{x_i} + 1 - q_j^{x_j} - q_i^{x_i} + q_j^{x_j} q_j^{x_j},$$

i.e.,

$$P_{\omega_i\to\omega_j}(D = 1|\underline{x}) = 1 - q_j^{x_j}(1 - q_i^{x_i}). \tag{6.1}$$

which is the expression sought for. By Table 1

$$\underline{\alpha}_5 = (1 \quad 0 \quad 1 \quad 1)$$

and by the above

$$\mathbf{p}^{\alpha_5} = (1 - q_2^{x_2}(1 - q_1^{x_1}))_{\underline{x}\in\{0,1\}^2}\,.$$

We have

$$\mathbf{p}^{\alpha_5}_{(0,1)} = \mathbf{p}^{\alpha_5}_{(0,0)} = 1 > \mathbf{p}^{\alpha_5}_{(1,0)} = q_1, \quad \mathbf{p}^{\alpha_5}_{(0,0)} = 1 > \mathbf{p}^{\alpha_5}_{(1,0)} = q_1,$$

and $\mathbf{p}^{\alpha_5}_{(1,1)} = 1 - q_2(1 - q_1)$.

It remains to show that $\mathbf{p}^{\alpha_5}_{(1,1)} > \mathbf{p}^{\alpha_5}_{(1,0)}$. Let us note that $1 - q_2 > (1 - q_2)q_1$ so that

$$1 - q_2 > (1 - q_2)q_1 \Leftrightarrow 1 - q_2 + q_2q_1 > q_1 \Leftrightarrow 1 - q_2(1 - q_1) > q_1. \tag{6.2}$$

Hence $\mathbf{p}^{\alpha_5}_{(1,1)} > \mathbf{p}^{\alpha_5}_{(1,0)}$ and thus we have checked that Definition 4.1, i.e., (4.7), is satisfied and therefore $Q_{\alpha_5}(X)$ in Section 3.5 is the P -sufficient cause for the individual response type $\underline{\alpha}_5$. ■

Example 6.2 (Two-variable Noisy Equivalence): With $\alpha_7 = \omega_j \leftrightarrow \omega_i$ or equivalently $\alpha = (\omega) = \omega_i$, $\beta(\omega) = \omega_j$, $i \neq j$ and $\alpha \rightarrow \beta$.

$$P_{\omega_j \leftrightarrow \omega_i = 1}(D = 1|\underline{x}) = \sum_{\omega:(\omega_i \leftrightarrow \omega_j)=1} \mu(\omega); \underline{x})$$

by marginalization with the Sum-Product law

$$= \mu((i = 1, j = 1); \underline{x}) + \mu((i = 0, j = 0); \underline{x}) = (1 - q_i^{x_i})(1 - q_j^{x_j}) + q_i^{x_i} q_j^{x_j}.$$

In addition, $\underline{\alpha}_7 = (1 \quad 0 \quad 0 \quad 1)$ from Table 1 and

$$\mathbf{p}^{\alpha_7} = \left((1 - q_i^{x_i})(1 - q_j^{x_j}) + q_i^{x_i} q_j^{x_j}\right)_{(x^i, x^j) \in \{0,1\}^4}.$$

We have

$$\mathbf{p}^{\alpha_7}_{(1,0)} = q_i, \mathbf{p}^{\alpha_7}_{(0,1)} = q_j, \mathbf{p}^{\alpha_7}_{(0,0)} = 1,$$

and

$$\mathbf{p}^{\alpha_7}_{(1,1)} = (1 - q_i)(1 - q_j) + q_i q_j.$$

It remains to check that $\mathbf{p}^{\alpha_7}_{(1,1)} > \mathbf{p}^{\alpha_7}_{(1,0)}$ and $\mathbf{p}^{\alpha_7}_{(1,1)} > \mathbf{p}^{\alpha_7}_{(0,1)}$.

$$\begin{aligned}
&\mathbf{p}^{\alpha_7}_{(1,1)} \geq q_i \Leftrightarrow (1 - q_i)(1 - q_j) + q_i q_j \geq q_i \\
&\Leftrightarrow (1 - q_i)(1 - q_j) - (1 - q_j)q_i \geq 0 \\
&\Leftrightarrow (1 - q_j)(1 - 2q_i) \geq 0 \Leftrightarrow q_i \leq 1/2.
\end{aligned}$$

In the same way one sees also that $q_j \leq 1/2$ is needed for $\mathbf{p}^{\alpha_7}_{(1,1)} \geq q_j$. Hence, if $q_i \leq 1/2$ and $q_j \leq 1/2$ the P -sufficient cause for $D = 1$ is $Q_{\alpha_7}(X)$ in Section 3.5. ■

Example 6.3 (Noisy-And-Not): With the indexing in Table 1 we set $\alpha = \alpha_{12} = \neg(\omega_i \to \omega_j) = \neg\alpha_5$ and therefore (5.26) and (6.1) entail

$$P_{\alpha_{12}}(D = 1|\underline{x}) = 1 - P_{\alpha_5}(D = 1|x_i, x_j) = q_j^{x_j}(1 - q_i^{x_i}).$$

Directly from the definitions we find by De Morgan $\neg(\neg\omega_i \vee \omega_j) \equiv \omega_i \wedge \neg\omega_j$, and (5.25) and

$$P_{\alpha_{12}}(D = 1|\underline{x}) = \sum_{\omega:\omega_i \wedge \neg\omega_j = 1} \mu((\omega); \underline{x}) = \mu((\omega_i = 1, \omega_j = 0); \underline{x}) = (1 - q_i^{x_i})q_j^{x_j}.$$

Or,

$$\mathbf{p}_{\underline{x}}^{\alpha_{12}} = P_{\alpha_{12}}(D = 1|x_1, x_2) = (1 - q_i^{x_1})q_j^{x_2}. \tag{6.3}$$

Therefore by Table 1

$$\underline{\alpha}_{12} = (0 \quad 1 \quad 0 \quad 0)$$

and by the above

$$\mathbf{p}^{\alpha_{12}} = \left((1 - q_i^{x_1})q_j^{x_2}\right)_{(x_1,x_2)\in\{0,1\}^4}.$$

This entails

$$\mathbf{p}_{(1,1)}^{\alpha_{12}} = (1 - q_i)q_j < (1 - q_i) = \mathbf{p}_{(1,0)}^{\alpha_{12}} \tag{6.4}$$

and

$$\mathbf{p}_{(0,1)}^{\alpha_{12}} = \mathbf{p}_{(0,0)}^{\alpha_{12}} = 0.$$

Thus we have found that Definition 4.1, i.e., (4.7), is satisfied and thereby $Q_{\alpha_{12}}(X) = X_1 \bigwedge \bar{X}_2$ is the P -sufficient cause form for $D_{\alpha_{12}}$, as found in 3.5.

Let us regard q_j as the probabilistic strength of x_j to prevent via the mediator ω_j the outcome $D = 1$ on its own and $1 - q_i$ is the probabilistic strength of x_i to cause the outcome $D = 1$ on its own for an individual with response profile $\underline{\alpha}_{12}$. Hence, we can say in view of (6.4) that $x_2 = 1$ blocks in a probabilistic sense the effect of $x_1 = 1$. This is associated with a terminology about causal antagonisms, c.f. [24]. ■

Example 6.4 (Noisy-Or & Noisy-Or with Unkown Variables): With the numbering in Table 1 we consider $\alpha(= \alpha_2) = \omega_i \vee \omega_j)$. We take, as before, $\alpha(\omega) = \omega_i$, $\beta(\omega) = \omega_j$, $i \neq j$, $i \neq j$ and consider ICI with $\alpha \vee \beta$. We compute by means of (5.27) to get

$$P_{\alpha\vee\beta}(D = 1|\underline{x}) = P_\alpha(D = 1|\underline{x}) + P_\beta(D = 1|\underline{x}) - P_{\alpha\wedge\beta}(D = 1|\underline{x}).$$

and by the marginalization arguments and the independence of ω_i and ω_j

$$= (1 - q_i^{x_i}) + (1 - q_j^{x_i} - (1 - q_i^{x_i}) \cdot (1 - q_j^{x_j}) = 1 - q_i^{x_i} \cdot q_j^{x_j}.$$

$$P_{\omega_i\to\omega_j}(D = 1|\underline{x}) = 1 - q_j^{x_j}q_i^{x_i}. \tag{6.5}$$

With re-indexing, we have here found

$$\mathbf{p}^{\alpha_2}_{x_1,x_2} = 1 - q_1^{x_1} q_2^{x_2}.$$

This is single plus joint causation, [12]. It holds that $\alpha_2(x_1, x_2) = 0$ if and only if $(x_1, x_2) = (0, 0)$. Then $\mathbf{p}^{\alpha_2}_{0,0} = 0$ and hence $\mathbf{p}^{\alpha_2}_{x_1,x_2} > \mathbf{p}^{\alpha_2}_{0,0}$ and in view of (3.21) $Q_{\alpha_2}(X) = G_{\beta_6}(X) \bigvee G_{\beta_7}(X) \bigvee G_{\beta_8}(X)$ is the P-sufficient causes formulation for α_2.

If we consider a leaky version of this, by (5.36) and $\alpha = \omega_i \vee \omega_j \vee \omega_{d+1}$, it holds that

$$P_\alpha(D = 1|(\underline{x}, x_{d+1} = 1)) = (1 - q_{d+1}) + q_{d+1}(1 - q_j^{x_j} q_i^{x_i}) = 1 - q_{d+1} q_j^{x_j} q_i^{x_i}. \quad (6.6)$$

This holds, since $\omega_i \vee \omega_j \vee 1 \equiv \top$.

The Noisy-Or with two exposures was first considered by Arthur C. Cailey in 1853, see [19]. Another relatively early study is [10]. Later Noisy-Or has played a large role in artificial intelligence, see [9] and [46]. In a causal power theory of cognitive psychology the equation (6.6) appears in [25, p. 463, eq. (6)], where it is interpreted as the probability of $D = 1$, when two observable generative causes and an unobservable cause are present and cause $D = 1$. ■

Example 6.5 (MFPO and Sufficient Causes Forms with $d = 2$): Let us recall the notations in in Section 3.4 and 3.5. If the condition (4.7) is true for

$$\mathbf{p}^{\beta_i} \times \text{Be}(\theta_i) \equiv P_{\mu_{\underline{x}},\xi}(\{(\omega, o)|\beta_i(\omega) \wedge \xi_i(o) = 1\}) = P_{\beta_i}(D = 1|\underline{x}) \cdot \theta_i. \quad (6.7)$$

then we have by Theorem 3.1 a P-sufficient causes form of $\mathbf{p}^{\beta_i} \times \text{Be}(\theta_i)$. It follows immediately, with the completing literal

$$\tilde{\Xi}(\xi_i) = \begin{cases} \Xi_i & \text{if } \xi_i = 1 \\ \bar{\Xi}_i & \text{if } \xi_i = 0, \end{cases} \quad (6.8)$$

that

$$Q_{\beta_i,\xi_i}(X) = G_{\beta_i}(X) \bigwedge \Xi_i$$

is a P -sufficient causes form of $\mathbf{p}^{\beta_i} \times \text{Be}(\theta_i)$. By this example it is demonstrated that the sufficient cause representation of D. Flanders [11] is also exactly captured by the nine P -sufficient causes forms $G_{\beta_i}(X) \bigwedge \Xi_i$, when we use $P_{\mu_{\underline{x}},\xi}$ in ICI. Also, the P-sufficient causal forms $Q_{\beta_1,\xi_1}(X)$ correspond to the graphical presentation on [11, p. 848] or [36, p.81]. ■

7 Probabilities of Potential Outcome and SC with ICI for more than two Exposures

Decomposability of a Boolean function means that there are ways of computing it by sequential composing of Boolean functions with two arguments. This turns out to be useful in the current setting.

In technical terms there are disjunctive and nondisjunctive decompositions. A Boolean function $\alpha \in \mathcal{L}$ has a disjunctive decomposition, if it can be written as

$$\alpha(\omega) = F\left(\beta_1(\omega^{(1)}), \ldots, \beta_k(\omega^{(k)})\right),$$

$$\alpha(\omega) = F\left(\beta(\omega^{(1)}), \omega^{(2)}\right)$$

where $\omega^{(1)}$ has at least two elements (the domain of β and $\omega^{(2)}$ need not be disjoint).

We shall also evoke the nondisjunctive decompositions, where there are $d-1$ Boolean functions g_i with two arguments such that

$$\alpha\left(\omega_1, \omega_2, \ldots, \omega_d\right) = g_d\left(\omega_d, g_{d-1}\right) = g_d\left(\omega_d, g_{d-1}(\omega_2, g_{d-2})\right) = \ldots \quad \text{e.t.c..} \tag{7.1}$$

7.1 Noisy AND$_d$

We shall next study the probability of potential outcome of classes of ICI in (5.9) by means of the rules in proposition 5.3. Let us set

$$\text{AND}_d(\omega) \equiv \omega_1 \wedge \ldots \wedge \omega_d. \tag{7.2}$$

Then we obtain (the Noisy-And probability)

Lemma 7.1: *For AND$_d$ defined in (7.2)*

$$P_{AND_d}(D = 1|\underline{x}) = \prod_{j=1}^{d}(1 - q_j^{x_j}). \tag{7.3}$$

Proof: AND$_d$ $= \omega_1 \wedge \omega_2 \ldots \wedge \omega_d$ is decomposable in the sense of (7.1), AND$_d$ $= g_d(g_{d-1}, \omega_d) = g_{d-1} \wedge \omega_d$. By the truth table of $\wedge$

$$P_{\text{AND}_d}(D = 1|\underline{x}) = \sum_{g_{d-1} \wedge \omega_d = 1} \mu(\omega; \underline{x}) = P_{\mu_{\underline{x}}}(g_{d-1} = 1, \omega_d = 1)$$

and the independence of ωs under $P_{\mu_{\underline{x}}}$ entails

$$= P_{\mu_{\underline{x}}}(g_{d-1} = 1) P_{\mu_{\underline{x}}}(\omega_d = 1) = P_{\mu_{\underline{x}}}(g_{d-1} = 1) \cdot (1 - q_d^{x_d}).$$

The obvious Ansatz

$$P_{\mu_{\underline{x}}}(g_{d-1} = 1) = \prod_{j=1}^{d-1}(1 - q_j^{x_j})$$

entails immediately (7.3). ■

7.2 Noisy Tribes

For the Boolean functions $\text{Tribes}_{w,s}$ for non-negative integers w and s called **tribes of width** w **and size** s we split up $\omega \in \Omega = \{0,1\}^{sw}$ into s blocks $\omega^{(i)}$, called tribes, of equal size w so that

$$\omega^{(1)} = (\omega_1, \ldots, \omega_w)\,, \tag{7.4}$$

$$\omega^{(i)} = \left(\omega_{(i-1)w+1}, \ldots, \omega_{iw}\right), i = 2, \ldots, s, \tag{7.5}$$

and

$$\omega = \left(\omega^{(1)}, \omega^{(2)}, \ldots, \omega^{(s)}\right). \tag{7.6}$$

Then we take

$$\text{AND}_w(\omega^{(i)}) = \omega_{(i-1)w+1} \wedge \ldots \wedge \omega_{iw}. \tag{7.7}$$

OR_s is the disjunction $\vee$ in s variables

$$\text{OR}_s\left(\text{AND}_w(\omega^{(1)}), \ldots, \text{AND}_w(\omega^{(s)})\right) = \text{AND}_w(\omega^{(1)}) \vee \ldots \vee \text{AND}_w(\omega^{(s)}) \tag{7.8}$$

so that

$$\text{Tribes}_{w,s}(\omega_1, \omega_2, \ldots, \omega_d) = \text{OR}_s\left(\text{AND}_w(\omega^{(1)}), \ldots, \text{AND}_w(\omega^{(s)})\right). \tag{7.9}$$

Therefore, Tribes is $= 1$ if and only if at least one AND_w is $= 1$.

By definition of symmetric Boolean functions every OR_s is symmetric as a function of the s Boolean functions AND_w. However, $\text{Tribes}_{w,s}$ is not symmetric in $\omega \in \Omega = \{0,1\}^{sw}$. Hence there is some physical constraint on the blocks $\omega^{(i)}$: you can permute the variables inside any of them, and you can permute the s blocks inside OR_s, but you can not move a variable from one block to another without changing $\text{Tribes}_{w,s}$. There is also a decomposable property or a recursion:

$$\text{Tribes}_{w,s}(\omega_1, \omega_2, \ldots, \omega_d) = \text{Tribes}_{w,s-1}(\omega_1, \omega_2, \ldots, \omega_{(s-1)w}) \vee \text{AND}_w(\omega^{(s)}). \tag{7.10}$$

Proposition 7.2: *It holds for the noisy* $\textit{Tribes}_{w,s}$, $s > 2$ *that*

$$P_{\textit{Tribes}_{w,s}}(D = 1|\underline{x}) = g_1 \prod_{j=2}^{s}(1 - g_j) + \sum_{j=2}^{s-1} g_j \prod_{i=j+1}^{s}(1 - g_i) + g_s \tag{7.11}$$

and for $s = 2$

$$P_{\textit{Tribes}_{w,2}}(D = 1|\underline{x}) = (1 - g_2)g_1 + g_2, \tag{7.12}$$

where for $i = 1, 2, \ldots, s$

$$g_i = \prod_{j=w(i-1)+1}^{wi}(1 - q_j^{x_j}). \tag{7.13}$$

Proof: A permutation of $\{1, 2, \ldots, s\}$ is denoted by π_s. Since OR_s is a symmetric function of its s binary arguments $\text{AND}_w (= \text{AND}_w(\omega))$ as in proposition 5.5 we get that

$$P_{\text{OR}_s^{\pi^s}}(D = 1|\underline{x}) = P_{\text{OR}_s}(D = 1|\underline{x}).$$

Hence $P_{\text{Tribes}_{w,s}}(D = 1|\underline{x})$ has the same value for all permutations of AND_w. Thus (7.10) is without loss of generality taken w.r.t. to an arbitrary ordering of AND_ws.

First we compute the probability

$$P_{\mu_{\underline{x}}}(\omega|\ \text{AND}_w(\omega^{(s)}) = 1) = \sum_{\omega_1,\omega_2,\ldots,\omega_d:\ \text{AND}_w(\omega^{(s)})=1} \mu(\omega; \underline{x}).$$

Let any $\underline{x}^{(j)}$ be a quantity written in the appropriate xs like $\omega^{(j)}$ in (7.4) and (7.5). Then $\mu(\omega; \underline{x})$ can by the construction in (5.4) be formally recast as

$$\mu(\omega; \underline{x}) = \prod_{j=1}^{s} p(\omega^{(j)} \mid \underline{x}^{(j)}). \tag{7.14}$$

Then we use the Sum-Product law to get the margial distribution

$$P_{\mu_{\underline{x}}}(\omega|\ \text{AND}_w(\omega^{(s)}) = 1) = \sum_{\omega_1,\omega_2,\ldots,\omega_{(s-1)w}} \prod_{j=1}^{s-1} p(\omega^{(j)} \mid \underline{x}^{(j)}) \sum_{\omega^{(s)}:\ \text{AND}_w(\omega^{(s)})=1} p(\omega^{(s)} \mid \underline{x}^{(s)}).$$

It follows by the construction in (5.4) and the Sum-Product law (5.17) that

$$\sum_{\omega_1,\omega_2,\ldots,\omega_{(s-1)w}} \prod_{j=1}^{s-1} p(\omega^{(j)} \mid \underline{x}^{(j)}) = 1. \tag{7.15}$$

Next, the Lemma 7.1, i.e., (7.3), gives that

$$P_{\mu_{\underline{x}}}(\omega|\ \text{AND}_w(\omega^{(s)}) = 1) = \sum_{\omega^{(s)}:\ \text{AND}_w(\omega^{(s)})=1} p(\omega^{(s)} \mid \underline{x}^{(s)}) = \prod_{j=w(s-1)+1}^{ws} (1 - q_j^{x_j}). \tag{7.16}$$

Of course, the argument above gives also for any $i = 1, 2, \ldots, s$

$$P_{\mu_{\underline{x}}}(\omega|\ \text{AND}_w(\omega^{(i)}) = 1) = \sum_{\omega^{(i)}:\ \text{AND}_w(\omega^{(i)})=1} \mu(\omega^{(s)}); \underline{x}^{(i)}) = \prod_{j=w(i-1)+1}^{wi} (1 - q_j^{x_j}). \tag{7.17}$$

The union rule **c)**, (5.27) in proposition 5.3 gives in view of (7.10), since different $\omega^{(j)}$ contain disjoint sets of variables and are independent blocks in view of (7.14), after some

rearrangement, that

$$\begin{aligned} P_{\mu_{\underline{x}}}(\text{ Tribes}_{w,s}=1) &= P_{\mu_{\underline{x}}}(\text{ AND}_w(\omega^{(s)})=1)(1-P_{\mu_{\underline{x}}}(\text{ Tribes}_{s-1}=1)) \\ &+ P_{\mu_{\underline{x}}}(\text{ Tribes}_{w,s-1}=1). \end{aligned} \tag{7.18}$$

Let us write for convenience of handling the expressions

$$p_s = P_{\mu_{\underline{x}}}(\text{ Tribes}_{w,s}=1), g_s = P_{\mu_{\underline{x}}}(\text{ AND}_w(\omega^{(s)})=1). \tag{7.19}$$

Then (7.18) is the non-homogeneous first order difference equation for the numbers $\{p_j\}_{j=1}^{s}$ with variable coefficients

$$p_s = (1-g_s)p_{s-1} + g_s, \tag{7.20}$$

where $p_0 = 0$ (= the probability that the empty Tribes$_0$ is true). By successive iterations, that is, by underlying successive applications of (7.10), we suggest the solution of (7.20) with $p_1 = g_1$ with the Ansatz

$$p_k = g_1 \prod_{j=2}^{k}(1-g_j) + \sum_{j=2}^{k-1} g_j \prod_{i=j+1}^{k}(1-g_i) + g_k.$$

This is readily verified against (7.20) by computing $(1-g_k)p_{k-1}$. Indeed, by the Ansatz above

$$(1-g_k)p_{k-1} = (1-g_k)[g_1 \prod_{j=2}^{k-1}(1-g_j) + \sum_{j=2}^{k-2} g_j \prod_{i=j+1}^{k-1}(1-g_i) + g_{k-1}]$$

$$= g_1 \prod_{j=2}^{k}(1-g_j) + \sum_{j=2}^{k-2} g_j \prod_{i=j+1}^{k}(1-g_i) + (1-g_k)g_{k-1} =$$

$$= g_1 \prod_{j=2}^{k}(1-g_j) + \sum_{j=2}^{k-1} g_j \prod_{i=j+1}^{k}(1-g_i)$$

$$= p_k - g_k.$$

Finally, when we recall (7.17) and (7.19) we get that

$$g_i = \prod_{j=w(i-1)+1}^{wi}(1-q_j^{x_j})$$

and the claims in the proposition follow. ■

Example 7.1:

$$\text{Tribes}_{2,2}(\omega_1,\omega_2,\omega_3,\omega_4) = \text{OR}_2\left(\text{AND}_2(\omega^{(1)}),\ \text{AND}_2(\omega^{(2)})\right). \tag{7.21}$$

The union rule **c)**, (5.27), in proposition 5.3 gives,

$$\begin{aligned} P_{\mu_{\underline{x}}}(\text{Tribes}_{2,2}=1) &= P_{\mu_{\underline{x}}}(\text{AND}_2(\omega^{(1)})=1) + P_{\mu_{\underline{x}}}(\text{AND}_2(\omega^{(2)})=1) \\ &- P_{\mu_{\underline{x}}}(\text{AND}_2(\omega^{(1)})=1)\cdot(P_{\mu_{\underline{x}}}(\text{AND}_2(\omega^{(2)})=1), \end{aligned} \tag{7.22}$$

since $\omega^{(1)}$ and $\omega^{(2)}$ are independent under $P_{\mu_{\underline{x}}}$. This agrees with (7.11), i.e., we have for any $\underline{x} \in \{0,1\}^4$

$$P_{\text{Tribes}_{2,2}}(D=1|\underline{x}) = g_1 + g_2 - g_1 g_2, \tag{7.23}$$

where from (7.13)

$$g_1 = \prod_{j=1}^{2}(1-q_j^{x_j}), g_2 = \prod_{j=3}^{4}(1-q_j^{x_j}).$$

Next, $\text{Tribes}_{2,2}(\underline{x}) = 1$ for

$$\underline{x}_1 = (1,1,1,1), \underline{x}_2 = (1,1,0,1), \underline{x}_3 = (1,1,1,0), \underline{x}_4 = (1,1,0,0)$$

and

$$\underline{x}_5 = (1,0,1,1), \underline{x}_6 = (0,1,1,1), \underline{x}_7 = (0,0,1,1).$$

Then by (7.23)

$$P_{\text{Tribes}_{2,2}}(D=1|\underline{x}_1) = \prod_{j=1}^{2}(1-q_j) + \prod_{j=3}^{4}(1-q_j) - \prod_{j=1}^{2}(1-q_j)\prod_{j=3}^{4}(1-q_j) \tag{7.24}$$

and for $i = 2,3,4$

$$P_{\text{Tribes}_{2,2}}(D=1|\underline{x}_i) = \prod_{j=1}^{2}(1-q_j) \tag{7.25}$$

and for $i = 5,6,7$

$$P_{\text{Tribes}_{2,2}}(D=1|\underline{x}_i) = \prod_{j=3}^{4}(1-q_j). \tag{7.26}$$

It holds also by (7.23) for any $\underline{x} \neq \underline{x}_i$ $i = 1, 2, 3, 4, 5, 6, 7$ that

$$P_{\text{Tribes}_{2,2}}(D = 1|\underline{x}) = 0. \tag{7.27}$$

Clearly (4.7) holds for all $\underline{x}$ and $\underline{x}^{'} \in \{0, 1\}^4$ such that $\text{Tribes}_{2,2}(\underline{x}) = 1$ and $\text{Tribes}_{2,2}(\underline{x}^{'}) = 0$. Thus

$$\begin{aligned} Q_{\text{Tribes}_{2,2}}(X) &= (X_1 \wedge X_2 \wedge X_3 \wedge X_4) \\ &\vee (X_1 \wedge X_2 \wedge \bar{X}_3 \wedge X_4) \\ &\vee (X_1 \wedge X_2 \wedge X_3 \wedge \bar{X}_4) \\ &\vee (X_1 \wedge X_2 \wedge \bar{X}_3 \wedge \bar{X}_4) \\ &\vee (\bar{X}_1 \wedge X_2 \wedge X_3 \wedge X_4) \\ &\vee (X_1 \wedge \bar{X}_2 \wedge X_3 \wedge X_4) \\ &\vee (\bar{X}_1 \wedge \bar{X}_2 \wedge X_3 \wedge X_4) \end{aligned} \tag{7.28}$$

is P-sufficient causes form for $\text{Tribes}_{2,2}$. ■

7.3 Noisy $\text{Tribes}_{w,1}$ and Noisy $\text{Tribes}_{1,s}$

Noisy-And

As a check, we have for $s = 1$ and $w > 1$ in (7.11) or in (7.20) that

$$P_{\text{Tribes}_{w,1}}(D = 1|\underline{x}) = g_1 = \prod_{j=1}^{w}(1 - q_j^{x_j}). \tag{7.29}$$

On the other hand by (7.9) it holds that $\text{Tribes}_{w,1}(\omega_1, \omega_2, \ldots, \omega_d) = \text{AND}_w(\omega^{(1)})$. Hence (7.29) shows that what ensues is Noisy AND_w for w binary exposures, since (7.11) agrees with (7.3) in proposition 7.1 for $d = w$. For obvious reasons $\text{Tribes}_{w,1}$ is known as Noisy-And. Here

$$\underline{\text{AND}_d} = (1, 0, \ldots, 0)$$

and

$$\mathbf{p} = \left(\prod_{j=1}^{d}(1 - q_j^{x_j}) \right)_{\underline{x} \in \{0,1\}^{2^d}}.$$

Let $\underline{1} = (1, 1, \ldots, 1)$ denote the d-string of ones. Thus $\text{AND}_d(\underline{1}) = 1$, and $\text{AND}_d(\underline{x}) = 0$, for $\underline{x} \neq \underline{1}$. Then it follows that

$$\mathbf{p}_{\underline{1}} = \prod_{j=1}^{d}(1 - q_j) > \mathbf{p}_{\underline{x}} = 0,$$

as soon as $\underline{x} \neq \underline{1}$, as any $x_j = 0$ gives a factor $= 0$ in the product $\prod_{j=1}^{d}(1 - q_j^{x_j})$. Then $X_1 \wedge X_2 \wedge \ldots \wedge X_d$ is the P-sufficient cause for noisy AND$_d$.

Noisy-Or

If $s > 1$ and $w = 1$ we have Tribes$_{1,s}$ = OR$_s$ and hence (7.11) should reduce to (6.5) in Example 6.4 when $s = 2$. When $w = 1$ we have in (7.17)

$$g_i = (1 - q_i^{x_i}), i = 1, 2, \ldots, s.$$

When this is substituted in (7.11) we get

$$p_s = g_1 \prod_{j=2}^{s}(1 - g_j) + \sum_{j=2}^{s-1} g_j \prod_{i=j+1}^{s}(1 - g_i) + g_s$$

$$= (1 - q_1^{x_1}) \prod_{j=2}^{s} q_i^{x_i} + \sum_{j=2}^{s-1}(1 - q_j^{x_j}) \prod_{i=j+1}^{s} q_i^{x_i} + g_s$$

$$= \prod_{j=2}^{s} q_i^{x_i} - \prod_{j=1}^{s} q_i^{x_i} + \left[\sum_{j=2}^{s-1}(1 - q_j^{x_j}) \prod_{i=j+1}^{s} q_i^{x_i} + g_s\right].$$

We expand term by term the sum inside the parenthesis in the right hand side to get

$$\sum_{j=2}^{s-1}(1 - q_j^{x_j}) \prod_{i=j+1}^{s} q_i^{x_i} + g_s = \underbrace{\prod_{i=3}^{s} q_i^{x_i} - \prod_{i=2}^{s} q_i^{x_i}}_{j=2} + \underbrace{\prod_{i=4}^{s} q_i^{x_i} - \prod_{i=3}^{s} q_i^{x_i}}_{j=3}$$

$$\vdots$$

$$+ \underbrace{\prod_{i=s-1}^{s} q_i^{x_i} - \prod_{i=s-2}^{s} q_i^{x_i}}_{j=s-2} + \underbrace{q_s^{x_s} - \prod_{i=s-1}^{s} q_i^{x_i}}_{j=s-1} + \underbrace{1 - q_s^{x_s}}_{=g_s}$$

and by a telescoping summation

$$p_s = 1 - \prod_{j=1}^{s} q_i^{x_i}.$$

Hence we have proved with $d = s$ the following proposition.

Proposition 7.3: *For* $Tribes_{1,d} = OR_d$ *we have*

$$P_{OR_d}(D = 1|\underline{x}) = 1 - \prod_{j=1}^{d} q_j^{x_j}. \tag{7.30}$$

This formula is derived, with a different argument, in [28, pp. 184−187], where Noisy -Or is called a canonical model of multicausal interaction. To reconnect to the notions of Section 4 we have in (4.6) and (5.10)

$$\underline{\mathrm{OR}_d} = (\ \mathrm{OR}_d(\underline{x}))_{\underline{x}\in\{0,1\}^{2^n}}\ , \tag{7.31}$$

$$\mathbf{p} = \left(1 - \prod_{j=1}^{d} q_j^{x_j}\right)_{\underline{x}\in\{0,1\}^{2^n}} \tag{7.32}$$

But $\mathrm{OR}_d(x_1, x_2, \ldots, x_d) = 0$ if and only if $x_1 = x_2 = \ldots = x_d = 0$ i.e. $\underline{x} = \underline{0}$. Hence w.r.t. (3.2) the response profile is

$$\underline{\mathrm{OR}_d} = (1 \quad 1 \ldots 1 \quad 0). \tag{7.33}$$

But therefore with regard to the probabilistic causation (4.7) we have

$$\mathbf{p}_{\underline{x}} > \mathbf{p}_{\underline{0}} = 1 - \prod_{j=1}^{d} q_j^0 = 1 - 1 = 0 \tag{7.34}$$

for any $\underline{x} \neq \underline{0}$. Hence the individual response profile in (7.31) exhibits a causal response profile $\underline{\mathrm{OR}_d}$ in (7.33). In addition

$$Q_{\,\mathrm{OR}_d}(X) = \vee_{\underline{e}\in\{0,1\}^{2^d}:\underline{\alpha}_{\underline{e}}=1} \wedge_{i=1}^{d} \tilde{X}_i(e_i) = \vee_{\underline{e}\in\{0,1\}^{2^d}:\underline{e}\neq\underline{0}} \wedge_{i=1}^{d} \tilde{X}_i(e_i) \tag{7.35}$$

is a sufficient causes form of OR_d.

As a special case and check of agreement with previous findings we take $d = 2$. The P -sufficient causes of $D = \mathrm{OR}_2$ are by the preceding $X_1 \wedge X_2$, $X_1 \wedge \bar{X}_2$ and $\bar{X}_1 \wedge X_2$. Note that $\mathrm{OR}_2 = \alpha_2$ in Table 1. But then $Q_{\alpha_2}(X)$ was already found as disjunction of these forms in Section 3.5. If

$$\underline{x}_k = (0 \ldots 0 \underbrace{1}_{\text{position k}} 0 \ldots 0),$$

then in (6.4)

$$P_{\mathrm{OR}_d}(D = 1|\underline{x}_k) = 1 - q_k. \tag{7.36}$$

By this we can interpet q_k as the probabilistic strength of x_k to prevent the outcome $D = 1$ on its own. Therefore $1 - q_k$ is the probabilistic strength of P -sufficient cause to yield $D = 1$ on its own.

Example 7.2 (Leaky Noisy-Or): Let us now apply proposition 5.6 when $\alpha = \text{OR}_{d+1}$. Hence by (5.35) and with the notation $o \in \{0, 1\}$

$$\alpha^{(d+1)\mapsto o}(\omega) = \text{OR}_d(\omega) \vee o \tag{7.37}$$

Here, by definition of $\text{OR}_d(\omega)$

$$\alpha^{(d+1)\mapsto 1}(\omega) = \top$$

and

$$\alpha^{(d+1)\mapsto 0}(\omega) = \text{OR}_d(\omega).$$

These facts give in proposition 5.6, i.e., (5.36) that

$$P_{\text{OR}_{d+1}}(D = 1|(\underline{x}, x_{d+1})) = (1 - q_{d+1}^{x_{d+1}}) + q_{d+1}^{x_{d+1}} P_{\mu_{\underline{x}}}(\{\omega|\ \text{OR}_d(\omega) = 1\}).$$

But here we recall proposition 7.3, (6.4) and get

$$P_{\text{OR}_{d+1}}(D = 1|(\underline{x}, x_{d+1})) = (1 - q_{d+1}^{x_{d+1}}) + q_{d+1}^{x_{d+1}} \left(1 - \prod_{j=1}^{d} q_j^{x_j}\right)$$

$$= 1 - q_{d+1}^{x_{d+1}} \prod_{j=1}^{d} q_j^{x_j}.$$

Hence, as the background factor is always on,

$$P_{\text{OR}_{d+1}}(D = 1|(\underline{x}, 1)) = 1 - q_{d+1} \prod_{j=1}^{d} q_j^{x_j} \tag{7.38}$$

is the ICI expression of what is called Leaky Noisy-Or [9]. This tells that

$$P_{\text{OR}_{d+1}}(D = 1|(\underline{0}, 1)) = 1 - q_{d+1}.$$

This expresses the probability of the potential outcome $D = 1$ due to the unknown background, when there is no exposure to the known causes. In the absence of the leaky factor we have $P_{\text{OR}_{d+1}}(D = 1|\underline{0}) = 0$. ■

Example 7.3 (Recoding): If we recode $\text{AND}_d = \omega_1 \wedge \omega_2 \ldots \wedge \omega_d$ as $\overline{\text{AND}_d} = \neg\omega_1 \wedge \neg\omega_2 \ldots \wedge \neg\omega_d$, then by the proof of (7.3) we get

$$P_{\overline{\text{AND}_d}}(D = 1|\underline{x}) = \prod_{j=1}^{d} q_j^{x_j}. \tag{7.39}$$

On the other hand, by De Morgan,

$$\neg\omega_1 \wedge \neg\omega_2 \ldots \neg \wedge \neg\omega_d \equiv \neg(\omega_1 \vee \omega_2 \ldots \vee \omega_d).$$

Therefore, by (5.25)

$$P_{\overline{\text{AND}_d}}(D = 1 \mid \underline{x}) = P_{\neg(\omega_1 \vee \omega_2 \ldots \vee \omega_d)}(D = 1 \mid \underline{x})$$

and by (5.26) the right hand side equals

$$= 1 - P_{\omega_1 \vee \omega_2 \ldots \vee \omega_d}(D = 1 \mid \underline{x})$$

and by the Noisy-Or formula (6.4)

$$= 1 - (1 - \prod_{j=1}^{d} q_j^{x_j}) = \prod_{j=1}^{d} q_j^{x_j}.$$

Hence the recoding of Noisy AND_d yields the complement of Noisy OR_d. ■

7.4 General Noisy-Or and Noisy And

We consider

$$\alpha(\omega_1, \omega_2, \ldots, \omega_d) = \text{OR}_k(\omega_1, \omega_2, \ldots, \omega_k) \wedge \text{AND}_{d-k}(\omega_{k+1}, \omega_{k+2}, \ldots, \omega_d)$$

and write this as $\alpha = \text{OR}_k \wedge \text{AND}_{d-k}$. Then

$$P_{\text{OR}_k \wedge \text{AND}_{d-k}}(D = 1|\underline{x}) = P_{\mu_{\underline{x}}}(\{\omega|\ \text{OR}_k(\omega) = 1\}) P_{\mu_{\underline{x}}}(\{\omega|\ \text{AND}_{d-k}(\omega) = 1\}),$$

since OR_k and AND_{d-k} depend here on disjoint subsets of ω and are therefore independent under $P_{\mu_{\underline{x}}}$. Then (6.4) and (7.3) yield

$$P_{\text{OR}_k \wedge \text{AND}_{d-k}}(D = 1|\underline{x}) = (1 - \prod_{j=1}^{k} q_j^{x_j}) \prod_{j=k+1}^{d} (1 - q_j^{x_j}). \tag{7.40}$$

Let us next introduce the auxiliary quantities corresponding to the respective (4.5)

$$\mathbf{p}^{\text{OR}_k}_{(x_1,x_2,\ldots,x_d)} = (1 - \prod_{j=1}^{k} q_j^{x_j}), \mathbf{p}^{\text{AND}_{d-k}}_{(x_1,x_2,\ldots,x_d)} = \prod_{j=k+1}^{d} (1 - q_j^{x_j}).$$

From the preceding examples we get that if $j \in \mathcal{S}$ $x_j = 1$, $j = 1, \ldots, k$, then

$$\mathbf{p}^{\text{OR}_k}_{(x_1,x_2,\ldots,x_k,1,1,\ldots,1)} \cdot \mathbf{p}^{\text{AND}_{d-k}}_{(x_1,x_2,\ldots,x_k,1,1,\ldots,1)} = (1 - \prod_{j \in \mathcal{S}} q_j) \cdot \prod_{j=k+1}^{d} (1 - q_j) > 0.$$

for any $(x_1, x_2, \ldots, x_k, \underbrace{1, 1, \ldots, 1}_{(d-k \quad \text{ones})})$, where $(x_1, x_2, \ldots, x_k) \neq (0, 0, \ldots, 0)$ satisfies (4.7).

Hence

$$Q_{\text{OR}_k \wedge \text{AND}_{d-k}}(X) = \vee_{\underline{e} \in \{0,1\}^{2^k} : \underline{e} \neq \underline{0}} \wedge_{i=1}^{k} \tilde{X}_i(e_i) \wedge (X_{k+1} \wedge X_2 \wedge \ldots \wedge X_d) \quad (7.41)$$

is a sufficient causes form of $\text{OR}_k \wedge \text{AND}_{d-k}$.

7.5 Majority in $d = 3$

Let $\omega = (\omega_1, \omega_2, \omega_3) \in \Omega = \{\omega = (\omega)_{j=1}^3 | \omega_j \in \{0, 1\}\}$. Then we introduce

$$\alpha(\omega) = \begin{cases} 1 & \text{if } \omega_1 + \omega_2 + \omega_3 \geq 2 \\ 0 & \text{otherwise.} \end{cases} \quad (7.42)$$

This is the majority function given in Table 3, denoted by Maj_3 in [26, pp. 26−27], and in the sequel here.

ω_1	ω_2	ω_2	$Maj_3(\omega)$
1	1	1	1
1	0	1	1
0	1	1	1
0	0	1	0
1	1	0	1
1	0	0	0
0	1	0	0
0	0	0	0

Table 3: Majority function for $d = 3$

Then for $\underline{x}_3 = (x_1, x_2, x_3)$ we get

$$P_{Maj_3}(D = 1 | \underline{x}_3) = \sum_{\omega | \sum_{i=1}^3 \omega_i = 3} p(\omega | \underline{x}_3) + \sum_{\omega | \sum_{i=1}^3 \omega_i = 2} p(\omega | \underline{x}_3)$$

$$= p((1,1,1)|\underline{x}_3) + p((1,0,1)|\underline{x}_3) + p((0,1,1)|\underline{x}_3) + p((1,1,0)|\underline{x}_3)$$

$$= p(1|x_1)p(1|x_2)p(1|x_3) + p(1|x_1)p(0|x_2)p(1|x_3) + p(0|x_1)p((1|x_2)p(1|x_3) + p(1|x_1)p(1|x_2)p(0|x_3)$$

$$= (1-q_1^{x_1})(1-q_2^{x_2})(1-q_3^{x_3}) + (1-q_1^{x_1})q_2^{x_2}(1-q_3^{x_3}) + q_1^{x_1}(1-q_2^{x_2})(1-q_3^{x_3}) + (1-q_1^{x_1})(1-q_2^{x_2})q_3^{x_3}.$$

The last expression can be simplified in a number of ways, e.g., as

$$P_{Maj_3}(D=1|\underline{x}_3) = (1-q_2^{x_2})(1-q_3^{x_3}) + (1-q_1^{x_1})q_2^{x_2}(1-q_3^{x_3}) + (1-q_1^{x_1})(1-q_2^{x_2})q_3^{x_3}. \tag{7.43}$$

Due to (7.43) it is obtained that

$$P_{Maj_3}(D=1|(0,0,0)) = 0, P_{Maj_3}(D=1|(0,0,1)) = 0,$$

$$P_{Maj_3}(D=1|(0,1,0)) = 0, P_{Maj_3}(D=1|(1,0,0)) = 0$$

and

$$P_{Maj_3}(D=1|(1,1,1)) = (1-q_2)(1-q_3) + (1-q_1)q_2(1-q_3) + (1-q_1)((1-q_2)q_3,$$

$$P_{Maj_3}(D=1|(1,1,0)) = (1-q_1)((1-q_2),$$

$$P_{Maj_3}(D=1|(1,0,1)) = (1-q_1)(1-q_3),$$

and

$$P_{Maj_3}(D=1|(0,1,1)) = (1-q_2)(1-q_3).$$

Here we note in the passing that $P_{Maj_3}(D=1|\underline{x}_3) \to Maj_3(\underline{x}_3)$ for all $\underline{x}_3$, as q_1, q_2 and q_3 decrease to zero. Since the response type vector is

$$\underline{Maj_3} = (1 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0),$$

we find immediately that (4.7) is satisfied and therefore

$$Q_{Maj_3}(X) = (X_1 \wedge X_2 \wedge X_3) \vee (X_1 \wedge \bar{X}_2 \wedge X_3) \vee (\bar{X}_1 \wedge X_2 \wedge X_3) \vee (X_1 \wedge X_2 \wedge \bar{X}_3))$$

is the P -sufficient cause for $D=1$ for the individual response type $\underline{Maj_3}$.

The random variable $U = \sum_{i=1}^{3} \omega_i$ under the multivariate Bernoulli distribution $p(\omega|\underline{x}_3)$ on $\Omega = \{\omega = (\omega_{j=1}^3\}$ with the probability mass function

$$P(U=k) = \sum_{\omega|\sum_{i=1}^{3}\omega_i = k} p(\omega|\underline{x}_3), k = 0,1,2,3$$

is known as a *Poisson-Binomial* random variable with the vector of success parameters $((1-q_1^{x_1}),(1-q_2^{x_2}),(1-q_3^{x_3}))$.

8 Multifactor Potential Outcome Models and ICI

In this section we will compute the probability of potential outcome for MFPO treated in Section 3.4 using ICI with a general number of exposures both at levels of an individual and a population.

Let $(\mathcal{S}_i)_{i=1}^k$ be a partitioning of $\{1,\ldots,d\}$. $\omega \in \Omega = \{0,1\}^d$ is decomposed into k disjoint blocks $\omega^{(k)}$ so that

$$\omega^{(i)} = \times_{j\in\mathcal{S}_i}\omega_j, \tag{8.1}$$

and

$$\omega = \left(\omega^{(1)}, \omega^{(2)}, \ldots, \omega^{(k)}\right).$$

In Section 3.4 the completing causes were represented for each individual individual $u \in U$ with values of ξ_i, $i = 1,\ldots,k$ are independent $\mathrm{Be}(\theta_i)$ random variables, respectively. The probability mass function, with $o \in \{0,1\}$, is written as

$$P(\xi_i = o) = \theta_i^o(1-\theta_i)^{1-o},$$

We assume that ξ_is are also independent of the mediators ω.

Let us set $\mathbf{o} = (o_1,\ldots,o_k)$, $\underline{\xi} = (\xi_1,\ldots,\xi_k)$ and thus for any $\mathbf{o} \in \{0,1\}^k$

$$P_{\underline{\xi}}(\mathbf{o}) = P\left(\underline{\xi} = \mathbf{o}\right) = \prod_{i=1}^{k} P(\xi_i = o_i) \tag{8.2}$$

Next we take k Boolean functions β_i on $\{0,1\}^d$, and set

$$\beta_i \wedge \xi_i \equiv \beta_i(\omega^{(i)}) \wedge \xi_i. \quad i = 1,\ldots,k. \tag{8.3}$$

These are seen as Boolean functions on $\{0,1\}^{d+1}$ using the map τ_i in (3.15). We omit τ_i in this section in order not to increase the notational burden any further. Then we recall the multifactor potential outcome function

$$\mathrm{MFPO}_k\left((\beta_l)_{l=1}^k, (\xi_l)_{l=1}^k\right) \equiv \mathrm{OR}_k\left(\beta_1\wedge\xi_1,\ldots,\beta_k\wedge\xi_k\right), \tag{8.4}$$

where OR_k is the disjunction in k Boolean functions.

8.1 Individual Probabilistic Response for MFPO

We plan to compute $P_{\mathrm{MFPO}_k}(D = 1|\underline{x})$ as the probability of potential outcome for an individual $u \in U$. Let the completion pattern of u be $\mathbf{o}(u) = (o_1(u),\ldots,o_k(u))$. We have the individual probabilistic response

$$P_{\mathrm{OR}_k(\beta_1\wedge o_1(u),\ldots,\beta_k\wedge o_k(u))}(D = 1|\underline{x}).$$

But OR_k is a symmetric function of its k binary arguments. Hence it only depends on the number of ones, say, $= l$, in $\mathbf{o}(u)$ and we can write

$$\text{OR}_k\left(\beta_1 \wedge o_{j_1(u)}, \ldots, \beta_k \wedge o_{j_k(u)}\right) = \text{OR}_{l(u)}\left(\beta_{j_1(u)}, \ldots, \beta_{j_l(u)}\right),$$

where $j_1(u), \ldots, j_l(u)$ are the indices of the l ones in $\mathbf{o}(u)$. The last equality follows by a property of symmetric Boolean functions [20, pp. 123-124] and since $\beta_k \wedge 0 = 0$. The structure of $\text{OR}_k\left(\beta_{j_1(u)}, \ldots, \beta_{j_l(u)}\right)$ is, however, analogous to that of Tribes, which we now take advantage of. First, we observe that

$$\text{OR}_l\left(\beta_{j_1(u)}, \ldots, \beta j_l(u)\right) = \text{OR}_{(l-1)}\left(\beta_{j_1(u)}, \ldots, \beta_{j_{l-1}(u)}\right) \vee \beta_{j_l(u)}.$$

Following form here the same rules of computation and using the independencies as in the proof of proposition 7.2 we get as in (7.18)

$$\begin{aligned} P_{\mu_{\underline{x}}}(\text{ OR}_l = 1) &= P_{\mu_{\underline{x}}}(\beta_{j_l(u)} = 1)(1 - P_{\mu_{\underline{x}}}(\text{ OR}_{(l-1)} = 1)) \\ &+ P_{\mu_{\underline{x}}}(\text{ OR}_{(l-1)} = 1). \end{aligned} \tag{8.5}$$

Obviously (8.5) is of the same form as (7.20) and the next proposition follows.

Proposition 8.1: *Take a fixed $u \in U$. Let $j_1(u), \ldots, j_l(u)$ be the indices of the l ones in $\mathbf{o}(u)$. For the noisy $OR_{l(u)}$ that the individual probabilistic response profile is*

$$P_{\mu_{\underline{x}}}(\ OR_{l(u)} = 1) = g_{j_1(u)} \prod_{i=2}^{l}(1 - g_{j_i(u)}) + \sum_{i=2}^{l-1} g_{j_i(u)} \prod_{r=j_i(u)}^{j_l(u)} (1 - g_r) + g_{j_l(u)} \tag{8.6}$$

and for $l = 2$

$$P_{\mu_{\underline{x}}}(\ OR_{2l} = 1) = (1 - g_{j_2(u)})g_{j_1(u)} + g_{j_2(u)}, \tag{8.7}$$

and where for $i = 1, 2, \ldots, l$

$$g_{j_i(u)} = P_{\mu_{\underline{x}}}(\beta_{j_i(u)} = 1). \tag{8.8}$$

8.2 Population Level Probability of Potential Outcome for MFPO

The population U has N individuals u. For each individual u the preceding sections determine the individual probabilistic response as the random variable

$$P_{\text{OR}_k(\beta_1 \wedge \xi_1(u), \ldots, \beta_k \wedge \xi_k(u))}(D = 1 | \underline{x}.)$$

This is a function of $\underline{\xi}(u) = (\xi_1(u), \ldots, \xi_k(u))$, which are r.v.'s generating individual the completion patterns. We assume that $\underline{\xi}(u)$ are N independent and identically distributed

random variables. Then the population level probability of potential outcome is a sample of

$$\bar{P}_{\mathrm{OR}_k}(D=1|\underline{x}) \equiv \frac{1}{N}\sum_{u\in U} P_{\mathrm{OR}_k(\beta_1\wedge\xi_1(u),\ldots,\beta_k\wedge\xi_k(u))}(D=1|\underline{x}). \tag{8.9}$$

The weak law of large numbers entails convergence in probability

$$\bar{P}_{\mathrm{MFPO}_k}(D=1|\underline{x}) \xrightarrow{\mathrm{P}} E_{P_{\underline{\xi}}}\left[P_{\mathrm{OR}_k(\beta_1\wedge\xi_1,\ldots,\beta_k\wedge\xi_k)}(D=1|\underline{x})\right], \quad \text{as } N\to+\infty, \tag{8.10}$$

where $\underline{\xi} \equiv \underline{\xi}(u)$ for all u.

We shall next find an expression for the expectation

$$E_{P_{\underline{\xi}}}\left[P_{\mathrm{OR}_k(\beta_1\wedge\xi_1,\ldots,\beta_k\wedge\xi_k)}(D=1|\underline{x})\right].$$

It turns out again that there is a helpful recursion for $P_{\mathrm{MFPO}a_k}(D=1|\underline{x})$. This is a consequence of decomposability for OR_k, disjoint blocks and the independence assumptions in this section. By the construction above

$$\begin{aligned} &E_{P_{\underline{\xi}}}\left[P_{\mathrm{OR}_k(\beta_1\wedge\xi_1,\ldots,\beta_k\wedge\xi_k)}(D=1|\underline{x})\right] \\ =&\sum_{\mathbf{o}} P_{\underline{\xi}}(\mathbf{o}) P_{\mathrm{OR}_k(\beta_1\wedge o_1,\ldots,\beta_k\wedge o_k)}(D=1|\underline{x}). \end{aligned} \tag{8.11}$$

We consider first a generic term of the sum in the right hand side of (8.11). The decomposability for OR_k gives:

$$\mathrm{OR}_k = \mathrm{OR}_{k-1} \vee \beta_k \wedge o_k(u). \tag{8.12}$$

Thus we get by (5.27)

$$\begin{aligned} &P_{\mathrm{OR}_k}(D=1 \mid \underline{x}) \\ =&P_{\mathrm{OR}_{k-1}\vee\beta_k\wedge o_k(u)}(D=1 \mid \underline{x}) + P_{\mathrm{OR}_{k-1}}(D=1 \mid \underline{x}) \\ +&P_{\beta_k\wedge o_k(u)}(D=1 \mid \underline{x}) - P_{\mathrm{OR}_{k-1}\wedge(\beta_k\wedge o_k(u))}(D=1 \mid \underline{x}). \end{aligned} \tag{8.13}$$

Then by (8.11)

$$\begin{aligned} P_{\mathrm{MFPO}_k}(D=1|\underline{x}) &= \sum_{\mathbf{o}} P_{\underline{\xi}}(\mathbf{o}) P_{\mathrm{OR}_{k-1}}(D=1 \mid \underline{x}) \\ &+ \sum_{\mathbf{o}} P_{\underline{\xi}}(\mathbf{o}) P_{\beta_k\wedge o_k}(D=1 \mid \underline{x}) - \sum_{\mathbf{o}} P_{\underline{\xi}}(\mathbf{o}) P_{\mathrm{OR}_{k-1}\wedge(\beta_k\wedge o_k)}(D=1 \mid \underline{x}). \end{aligned} \tag{8.14}$$

Here the Sum-Product Law (5.15) and (8.2) yield on the first term in the right hand side of (8.14), as $\sum_{o_k} P(\xi_k = o_k) = (1 - \theta_k) + \theta_k$,

$$\begin{aligned}
&\sum_{\mathbf{o}} P_{\underline{\xi}}(\mathbf{o}) P_{\mathrm{OR}_{k-1}}(D = 1 \mid \underline{x}) \\
=&\sum_{o_1} \cdots \sum_{o_{k-1}} \prod_{i=1}^{k-1} P(\xi_i = o_i) P_{\mathrm{OR}_{k-1}}(D = 1 \mid \underline{x}) \cdot \sum_{o_k} P(\xi_k = o_k) \\
=&\sum_{o_1} \cdots \sum_{o_{k-1}} \prod_{i=1}^{k-1} P(\xi_i = o_i) P_{\mathrm{OR}_{k-1}}(D = 1 \mid \underline{x}) \\
=&P_{\mathrm{MFPO}_{k-1}}(D = 1 | \underline{x}).
\end{aligned} \tag{8.15}$$

by re-indexing of the rule in (8.11). Next, for the second term in the right hand side of (8.14)

$$\sum_{\mathbf{o}} P_{\underline{\xi}}(\mathbf{o}) P_{\beta_k \wedge o_k}(D = 1 \mid \underline{x})$$

$$= \sum_{o_k} P(\xi_k = o_k) P_{\beta_k \wedge o_k}(D = 1 \mid \underline{x}) \cdot \sum_{o_1} \cdots \sum_{o_{k-1}} \prod_{i=1}^{k-1} P(\xi_i = o_i)$$

and by the the Sum-Product Law (5.15),

$$\sum_{o_1} \cdots \sum_{o_{k-1}} \prod_{i=1}^{k-1} P(\xi_i = o_i) = \sum_{o_1} P(\xi_1 = o_i) \sum_{o_2} P(\xi_i = o_2) \ldots \sum_{o_{k-1}} P(\xi_{k-1} = o_{k-1}) = 1.$$

Hence

$$\sum_{\mathbf{o}} P_{\underline{\xi}}(\mathbf{o}) P_{\beta_k \wedge o_k}(D = 1 \mid \underline{x}) = \sum_{o_k} P(\xi_k = o_k) P_{\beta_k \wedge o_k}(D = 1 \mid \underline{x}).$$

But $\beta_k(\omega) \wedge 0 = \bot(\omega)$ for all ω, and thus

$$P(\xi_k = 0) P_{\beta_k \wedge 0}(D = 1 \mid \underline{x}) = P(\xi_k = 0) P_{\bot}(D = 1 \mid \underline{x})$$

$$= (1 - \theta_k) \sum_{\omega | \bot(\omega) = 1} \mu(\omega; \underline{x}) = (1 - \theta_k) \cdot P_{\mu_{\underline{x}}}(\emptyset) = 0.$$

We have obtained

$$\begin{aligned}
&P_{\beta_k \wedge o_k}(D = 1 \mid \underline{x}) \\
&= \sum_{o_k} P(\xi_k = o_k) P_{\beta_k \wedge o_k}(D = 1 \mid \underline{x}) = \theta_k \cdot P_{\beta_k \wedge 1}(D = 1 \mid \underline{x}).
\end{aligned} \tag{8.16}$$

By corollary 5.33 we get in the final term in the right hand side of (8.14)

$$P_{\mathrm{OR}_{k-1}\wedge(\beta_k\wedge o_k)}(D=1 \mid \underline{x}) = P_{\mathrm{OR}_{k-1}}(D=1 \mid \underline{x}) \cdot P_{\beta_k\wedge o_k}(D=1 \mid \underline{x}).$$

Then

$$\sum_{\mathbf{o}} P_{\underline{\xi}}(\mathbf{o}) P_{\mathrm{OR}_{k-1}\wedge(\beta_k\wedge o_k)}(D=1 \mid \underline{x}) =$$

$$\sum_{\mathbf{o}} P_{\underline{\xi}}(\mathbf{o}) P_{\mathrm{OR}_{k-1}}(D=1 \mid \underline{x}) \cdot P_{\beta_k\wedge 1}(D=1 \mid \underline{x}),$$

and by the the Sum-Product Law (5.15)

$$= \sum_{o_1} \cdots \sum_{o_{k-1}} \prod_{i=1}^{k-1} P(\xi_i = o_i) P_{\mathrm{OR}_{k-1}}(D=1 \mid \underline{x}) \cdot \sum_{o_k} P(\xi_k = o_k) P_{\beta_k\wedge o_k}(D=1 \mid \underline{x})$$

But the factors in the right hand side have been computed in (8.15) and (8.16) and therefore

$$\begin{aligned} &\sum_{\mathbf{o}} P_{\underline{\xi}}(\mathbf{o}) P_{\mathrm{OR}_{k-1}\wedge(\beta_k\wedge 1)}(D=1 \mid \underline{x}) \\ =&P_{\mathrm{MFPO}_{k-1}}(D=1|\underline{x}) \cdot \theta_k \cdot P_{\beta_k\wedge 1}(D=1 \mid \underline{x}). \end{aligned} \tag{8.17}$$

Lemma 8.2:

$$\begin{aligned} &P_{MFPO_k}(D=1|\underline{x}) \\ =&P_{MFPO_{k-1}}(D=1|\underline{x}) + \theta_k \cdot P_{\beta_k\wedge 1}(D=1 \mid \underline{x}) \\ -&P_{MFPO_{k-1}}(D=1|\underline{x}) \cdot \theta_k \cdot P_{\beta_k\wedge 1}(D=1 \mid \underline{x}) \end{aligned} \tag{8.18}$$

Proof: Substitute (8.15), (8.16) and (8.17) in the right hand side of (8.14). ■

Proposition 8.3: *It holds for the noisy MFPO$_k$, $k > 2$ that*

$$P_{MFPO_k}(D=1|\underline{x}) = g_1 \prod_{j=2}^{k}(1-g_j) + \sum_{j=2}^{k-1} g_j \prod_{i=j+1}^{k}(1-g_i) + g_k \tag{8.19}$$

and for $k = 2$

$$P_{MFPO_2}(D=1|\underline{x}) = (1-g_2)g_1 + g_2, \tag{8.20}$$

and $p_1 = g_1$, where for $i = 1, 2, \ldots, k$

$$g_i = P_{\beta_i}(D=1|\underline{x}) \cdot \theta_i \tag{8.21}$$

Proof: Let us re-arrange the equation in lemma 8.2 as

$$\begin{aligned} &P_{\text{MFPO}_k}(D=1|\underline{x}) \\ =&(1-\theta_k \cdot P_{\beta_k \wedge 1}(D=1 \mid \underline{x}))P_{\text{MFPO}_{k-1}}(D=1|\underline{x}) + \theta_k P_{\beta_k \wedge 1}(D=1 \mid \underline{x}). \end{aligned} \tag{8.22}$$

Let us introduce for convenience of handling the expressions

$$p_k = P_{\text{MFPO}_k}(D=1|\underline{x}), g_k = \theta_k \cdot P_{\beta_k \wedge 1}(D=1 \mid \underline{x}). \tag{8.23}$$

In these terms we have the same recursion as in (8.22), except for a different sequence g_i, where $p_0 = 0$ (= the probability that the empty MFPO$_0$ is true) and hence $p_1 = g_1$. Thus the expressions in (8.19) and (8.20) are as asserted. ■

When we compare proposition 8.1 and (8.8) with proposition 8.3 and (8.21), the only difference is the multiplication of every $P_{\beta_i}(D=1|\underline{x})$ with θ_i, where as in (8.8) the same probabilities of of potential outcome ar multiplied by the bits in respective $\mathbf{o}(u)$. This seems natural.

9 Causal Inference with Causal Effects in Additive Scale at Population Level

The practical significance of the studies above lies at the level of a population, since we can not observe all potential outcomes for an individual. The question is now to determine the population level probability (risk) of the potential outcome $D = 1$ as a function of exposures $\underline{x}$. This has been treated for a special sufficient cause representation in Section 8. In the current setting this requires an expectation with two steps, first with respect to population and then with respect to the probabilities of potential outcomes for individuals.

9.1 Potential population risk

The population U consists of a finite number N of individuals u. Each u has a unique response profile α_i amongst $B = 2^{2^d}$ possible, written as $u \to \alpha_i$. Let

$$\lambda_i \equiv \frac{\sharp\{u \in U : u \to \alpha_i\}}{N}, \quad i = 1, \ldots B. \tag{9.1}$$

be the relative frequency of α_i in the population. Then we denote **the population level probability (risk)** by $E\left[D\right]_{\underline{x}}$ and write it as

$$E\left[D\right]_{\underline{x}} \equiv \frac{1}{N} \sum_{u \in U} E\left[D(u)_{\underline{x}}\right]. \tag{9.2}$$

Then it follows in the right hand side that

$$\frac{1}{N}\sum_{u\in U} E\left[D(u)_{\underline{x}}\right] = \frac{1}{N}\sum_{i=1}^{B}\sum_{u\in U:u\to\alpha_i} E\left[D_{\underline{x}}(u)\right]$$

$$= \frac{1}{N}\sum_{i=1}^{B}\sum_{u\in U:u\to\alpha_i} P\left(D_{\underline{x}}(u)=1 \mid X=\underline{x},u\right)$$

and by consistency and conditional exchangeability in (4.4)

$$= \frac{1}{N}\sum_{i=1}^{B}\sum_{u\in U:u\to\alpha_i} P_{\alpha_i}\left(D \mid \underline{x}\right) = \sum_{i=1}^{B} P_{\alpha_i}\left(D \mid \underline{x}\right)\frac{1}{N}\sum_{u\in U:u\to\alpha_i} 1 = \sum_{i=1}^{B} P_{\alpha_i}\left(D \mid \underline{x}\right)\lambda_j.$$

where we applied (9.1) in the final step. We have thus found

$$E\left[D\right]_{\underline{x}} = \sum_{i=1}^{B} P_{\alpha_i}\left(D=1 \mid \underline{x}\right)\lambda_i \tag{9.3}$$

The issue is, how to make any effective use of this, as the characteristics $(q_i)_{i=1}^{d}$ of the mediators must be seen as unknown and the estimation of λ_i would seem to require the knowledge of each individual's response profile, which involves potential or counterfactual outcomes. We deal with these questions for $d = 2$ by computation of causal contrasts, c.f. [36, pp. 75−−76], [34, 204−205].

9.2 Causal Contrasts

A causal effect is often treated by some form of contrasts between the effects of two or more interventions. In this section we recapitulate, for ease of reference, the interaction contrasts between the effects of two or more interventions. These are counterfactual at the level of an individual, who cannot at the same time be exposed to, e.g., both $X_1 = 1$ and $X_1 = 0$.

We measure first the effect between two exposures X_1 and X_2 on the occurrence of D, by how the association between X_1 and D varies with X_2. First, intervene by $X_2 = 1$ and compute

$$IC_i^{(1)} \equiv \alpha_i(1,1) - \alpha_i(0,1). \tag{9.4}$$

The association between X_1 and D varies with $X_2 = 0$ according to

$$IC_i^{(2)} = \alpha_i(1,0) - \alpha_i(0,0). \tag{9.5}$$

How the association between X_1 and D is varying based on X_2 is then $IC_i^{(1)} - IC_i^{(2)}$, i.e.,

$$IC_i \equiv \alpha_i(1,1) - \alpha_i(1,0) - (\alpha_i(0,1) - \alpha_i(0,0)) \tag{9.6}$$

In the same manner the association between X_2 and D varies under $X_1 = 1$ as

$$IC_i^{(3)} = \alpha_i(1,1) - \alpha_i(1,0).$$

and the association between X_2 and D varies under $X_1 = 0$ as

$$IC_i^{(4)} = \alpha_i(0,1) - \alpha_i(0,0).$$

Arithmetically it holds that

$$IC_i = IC_i^{(3)} - IC_i^{(4)}.$$

For $i = 1, 2, \ldots, 16$ the quantity IC_i in (9.6) is called the **interaction contrast**. It can be computed for any response type $i = 1, 2, \ldots, 16$ by means of Table 1 and this yields that $IC_1 = IC_4 = IC_6 = IC_{11} = IC_{13} = IC_{16} = 0$, which by inspection of Table 1 is seen to correspond to respective Q_{α_i} containing literals of only one of X_1 and X_2 or neither. For the rest of i, $IC_i \neq 0$, and this exhibits some form of joint causal effect of X_1 and X_2 in an additive scale. We continue with a probabilistic extension of the constraint of interaction in (9.6). Here $d = 2$ and by (5.9)

$$\mathbf{p}_{x_1,x_2}^{\alpha_i} \equiv \sum_{(\omega_1,\omega_2)|\alpha_i(\omega_1,\omega_2)=1} \prod_{j=1}^{2} p(\omega_j \mid x_j). \tag{9.7}$$

We emulate the constraint of interaction in (9.6) by setting

$$\text{SIC}_i \equiv \mathbf{p}_{(1,1)}^{\alpha_i} - \mathbf{p}_{(1,0)}^{\alpha_i} - \left(\mathbf{p}_{(0,1)}^{\alpha_i} - \mathbf{p}_{(0,0))}^{\alpha_i}\right). \tag{9.8}$$

This is known as the **additive synergy** of X_1 and X_2 in [21, p. 243]. We find recalling the premise 5.1 that

$$\text{SIC}_1 = \text{SIC}_4 = \text{SIC}_6 = \text{SIC}_{11} = \text{SIC}_{13} = \text{SIC}_{16} = 0, \tag{9.9}$$

which agrees with a result above. Thus we obtain

$$\text{SIC}_2 = \text{SIC}_9 = \text{SIC}_{14} = \text{SIC}_{12} = -(1-q_1)(1-q_2) < 0, \tag{9.10}$$

$$\text{SIC}_{10} = -2(1-q_1)(1-q_2) < 0. \tag{9.11}$$

and

$$\text{SIC}_3 = \text{SIC}_5 = \text{SIC}_8 = \text{SIC}_{15} = (1-q_1)(1-q_2) > 0 \tag{9.12}$$

and

$$\text{SIC}_7 = 2(1-q_1)(1-q_2) > 0. \tag{9.13}$$

These agree in sign $\pm$ with what is found by means of (9.6).

9.3 Population Level Causal Contrasts

Let us first write $E\left[D\right]_{\underline{x}}$ in (9.3) as

$$E\left[D\right]_{x_1,x_2} = \sum_{i=1}^{16} \mathbf{p}_{x_1,x_2}^{\alpha_i} \lambda_i. \tag{9.14}$$

We write the causal constraint at the population level following the discussion above as

$$\text{DR} \equiv E\left[D\right]_{1,1} - E\left[D\right]_{1,0} - \left(E\left[D\right]_{0,1} - E\left[D\right]_{0,0}\right). \tag{9.15}$$

We insert from (9.14) in the right hand side and invoke (9.8) to get

$$\text{DR} = \sum_{i=1}^{16} \left[\mathbf{p}_{(1,1)}^{\alpha_i} - \mathbf{p}_{(1,0)}^{\alpha_i} - \left(\mathbf{p}_{(0,1)}^{\alpha_i} - \mathbf{p}_{(0,0))}^{\alpha_i}\right)\right] \lambda_i = \sum_{i=1}^{16} \text{SIC}_i \cdot \lambda_i. \tag{9.16}$$

We apply (9.9) - (9.13) in the right hand side of (9.8) to find

$$\begin{aligned} \text{DR} \quad = \quad & (1-q_1)(1-q_2)\left[(\lambda_3 + \lambda_5 + \lambda_8 + \lambda_{15} + 2\lambda_7)\right. \\ & \left. - (\lambda_2 + \lambda_9 + \lambda_{12} + \lambda_{15} + 2\lambda_{10})\right]. \end{aligned} \tag{9.17}$$

As one is interested in whether DR $\neq 0$, the value of the product $(1-q_1)(1-q_2) > 0$ needs not to be known. Remarkably, when $(1-q_1)(1-q_2)$ is disregarded, DR in (9.17) is exactly of the same form as what is called **Koopman's interaction constraint** in [12, p.127].

9.4 Causal Inference

At this point we can proceed by the sufficient cause model representations of Sections 3.4, 3.5 and 8. Thus we evoke the complementary component causes in $d = 2$, i.e., for each individual $u \in U$ there is an outcome, the completion pattern, $\mathbf{o}(u) = (o_1(u), \ldots, o_9(u))$ of $\underline{\xi} = (\xi_1, \ldots, \xi_9)$ of independent Be(θ_i) random variables. Here we know in view of the rules in (3.21) that the individual response type α_i of u can be represented for all $\underline{x}$ as

$$\alpha_k(\underline{x}) = \vee_{l=1}^{9}(\beta_l(\underline{x}) \wedge o_l(u)), \quad k = 1, \ldots, 16. \tag{9.18}$$

Let us note that the outcome of all zeroes $\mathbf{o}(u) = (0, \ldots, 0))$ produces α_{16}, the individual response type of total immunity. There are 2^9 different $\mathbf{o} = (o_1, \ldots, o_9)$. We index then as

$$\mathbf{o}^{(i)}, i = 1, \ldots, 2^9.$$

Furthermore, the whole population is partitioned as

$$U = \cup_{i=1}^{2^9}\{u|\mathbf{o}(u) = \mathbf{o}^{(i)}\},$$

where the partitioning sets are disjoint. Let $n_{\mathbf{o}^{(i)}}$ be the number of individuals in U that carry the completion pattern $\mathbf{o}^{(i)}$,

$$n_{\mathbf{o}^{(i)}} = \sharp\{u \in U : \mathbf{o}(u) = \mathbf{o}^{(i)}\}. \tag{9.19}$$

By (9.1)

$$\lambda_k = \frac{1}{N}\sum_{u\in U} I_{\{u\to\alpha_k\}}(u), \tag{9.20}$$

where $I_{\{u\to\alpha_k\}}(u)$ is the indicator function of the event that an invidual has the response profile α_k. This means that $\mathbf{o}(u) = \mathbf{o}^{(i)}$ and $\mathbf{o}^{(i)}$ satisfies (9.18) for all $\underline{x}$. Expressed by indicator functions this means

$$I_{\{u\to\alpha_k\}}(u) = I_{\{\text{for all } \underline{x}:\alpha_k(\underline{x})=\vee_{l=1}^{9}(\beta_l(\underline{x})\wedge o_l^{(i)}\}\cap\{\mathbf{o}(u)=\mathbf{o}^{(i)}\}}(u)$$

$$= I_{\{\text{for all } \underline{x}:\alpha_k(\underline{x})=\vee_{l=1}^{9}(\beta_l(\underline{x})\wedge o_l^{(i)}\}}(u)\cdot I_{\{\mathbf{o}(u)=\mathbf{o}^{(i)}\}}(u)$$

Therefore, in the right hand side of (9.20),

$$\frac{1}{N}\sum_{u\in U} I_{\{u\to\alpha_k\}}(u) = \frac{1}{N}\sum_{u\in U} I_{\{\text{for all } \underline{x}:\alpha_k(\underline{x})=\vee_{l=1}^{9}(\beta_l(\underline{x})\wedge o_l^{(i)})\}}(u)\cdot I_{\{\mathbf{o}(u)=\mathbf{o}^{(i)}\}}(u)$$

$$= \frac{1}{N}\sum_{i=1}^{2^9}\sum_{u:\mathbf{o}(u)=\mathbf{o}^{(i)}} I_{\{\text{for all } \underline{x}:\alpha_k(\underline{x})=\vee_{l=1}^{9}(\beta_l(\underline{x})\wedge o_l^{(i)})\}}(u)\cdot I_{\{\mathbf{o}(u)=\mathbf{o}^{(i)}\}}(u).$$

In fact, the first indicator function inside the sum does not depend on u but on $\mathbf{o}^{(i)}$. This entails

$$= \frac{1}{N}\sum_{i=1}^{2^9} I_{\{\text{for all } \underline{x}:\alpha_k(\underline{x})=\vee_{l=1}^{9}(\beta_l(\underline{x})\wedge o_l^{(i)})\}}(\mathbf{o}^{(i)})\cdot\sum_{u:\mathbf{o}(u)=\mathbf{o}^{(i)}} I_{\{\mathbf{o}(u)=\mathbf{o}^{(i)}\}}(u)$$

$$= \sum_{i=1}^{2^9} I_{\{\text{for all } \underline{x}:\alpha_k(\underline{x})=\vee_{l=1}^{9}(\beta_l(\underline{x})\wedge o_l^{(i)})\}}\cdot\frac{n_{\mathbf{o}^{(i)}}}{N}$$

Hence we have obtained the following formula originally due to Dana Flanders [11, eq. (3), p. 850].

$$\lambda_k = \sum_{i=1}^{2^9} I_{\{\text{for all } \underline{x}:\alpha_k(\underline{x})=\vee_{l=1}^{9}(\beta_l(\underline{x})\wedge o_l^{(i)})\}}(\mathbf{o}^{(i)}) \cdot \frac{n_{\mathbf{o}^{(i)}}}{N}. \tag{9.21}$$

In fact, the derivation above is valid for every $d \geq 2$, assuming that we know a set of functions β_i for the representation (9.18).

Let us now take a look at a case of applying (9.21).

Example 9.1: Let us consider λ_2. This is the relative frequency of the individual response type α_2, i.e., Noisy-Or $(d = 2)$ in a population.

Clearly, data is flat, as data does not as such not permit causality statements without causal assumptions. Here we compute $n_{\mathbf{o}^{(i)}}$ by assuming the P-sufficient causes formulation for α_2.

By Example 6.4 we know that $Q_{\alpha_2}(X) = G_{\beta_6}(X) \bigvee G_{\beta_7}(X) \bigvee G_{\beta_8}(X)$ is the P-sufficient causes formulation for α_2. The completion pattern is $\mathbf{o} = (0, 0, 0, 0, 0, 1, 1, 1, 0)$. The sufficient causes are $G_{\beta_6}(X) = X_1 \bigwedge X_2$, $G_{\beta_7}(X) = X_1 \bigwedge \bar{X}_2$ $G_{\beta_8}(X) = \bar{X}_1 \bigwedge X_2$ as given in (3.20). This means that we need to know the the frequency n_{x_1,x_2} of individuals in the population exposed to $X_1 = 1, X_2 = 1$, $X_1 = 1, X_2 = 0$ or $X_1 = 0, X_2 = 1$ and the incidences i_{x_1,x_2} of D amongst these indivudals (=the frequency of individuals exposed to (x_1, x_2) that actually contracted a certain disease, $D = 1$). Then $n_{\mathbf{o}^{(i)}}$ may be computed from obesrvational data as

$$n_{(0,0,0,0,0,1,1,1,0)} = i_{1,1} \cdot n_{1,1} + i_{1,0} \cdot n_{1,0} + i_{0,1} \cdot n_{0,1}. \tag{9.22}$$

This gives us an estimate of λ_2 by (9.21). ■

10 Linear Risk Model and Fourier Expansions of Probabilities of Potential Outcomes

In the epidemiological literature, see e.g., [3, 38, 42, 44], one refers to

$$P(D = 1 \mid x_1, x_2) = a + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2 \tag{10.1}$$

as the **linear risk model**. If all $P(D = 1 \mid x_1, x_2)$ are small the model coefficients can be estimated in case-control data by replacing the probabilities with the odds [38].

It has been argued, or been pointed out, that the inclusion of the term $b_3 x_1 x_2$ measures interaction in a statistical sense, as there is the need to include this product term for the statistical model to fit the data well, [1].

In this section we show for $d = 2$ that, when $P(D = 1 \mid x_1, x_2) = P_\alpha(D = 1 \mid x_1, x_2)$ for any individual response type α, the expression in (10.1) is in fact the Fourier

expansion of the probabilistic individual response. $P_\alpha(D = 1 \mid x_1, x_2)$ regarded as a function of (x_1, x_2). We start by a quick summary of the Fourier expansions of real functions on the binary hypercube, as found in [4, 23, 26].

10.1 Inner Product, Orthonormality, Fourier Expansions

Let $\underline{x} = (x_j)_{j=1}^d$ be an element of the hypercube $\{0,1\}^d$ and $f(\underline{x})$ be a function with values in $\mathbb{R}$. We shall call $f(\underline{x})$ a generalized (real valued) Boolean function. Let $\mathcal{L}^{\mathbb{R}}$ be the set of such functions, or

$$\mathcal{L}^{\mathbb{R}} = \{f \mid \{0,1\}^d \stackrel{f}{\mapsto} \mathbb{R}\}.$$

The set $\mathcal{L}^{\mathbb{R}}$ is a 2^d dimensional vector space, which is seen by stacking up the 2^d values f into one vector $f = (f_0, \ldots, f_{2^d-1})$ in some fixed order. Then we equip $\mathcal{L}^{\mathbb{R}}$ with the inner product

$$\langle f, g \rangle = \frac{1}{2^d} \sum_{\underline{x} \in \{0,1\}^d} f(\underline{x}) g(\underline{x}), \quad f \in \mathcal{L}^{\mathbb{R}}, g \in \mathcal{L}^{\mathbb{R}}. \tag{10.2}$$

Next $\mathcal{S}$ denotes any subset of $\{1, 2, \ldots, d\}$, and let for each $i \in \{1, 2, \ldots, d\}$

$$w_i = \begin{cases} 1 & i \in \mathcal{S} \\ 0 & i \notin \mathcal{S}. \end{cases}$$

Set $\underline{w} = (w_j)_{j=1}^d$. Given $\underline{w}$ we can determine uniquely the corresponding subset $\mathcal{S}$. Then we introduce the Boolean function $Q_{\mathcal{S}}(\underline{x})$ by

$$Q_{\mathcal{S}}(\underline{x}) \equiv \prod_{j \in \mathcal{S}} (-1)^{x_j} = (-1)^{\sum_{j \in \mathcal{S}} x_j}. \tag{10.3}$$

This shows that we can also write $Q_{\mathcal{S}}(\underline{x}) = f_{\underline{w}}(\underline{x})$ with

$$f_{\underline{w}}(\underline{x}) = (-1)^{\sum_{l=1}^d w_l x_l}. \tag{10.4}$$

We take

$$Q_\emptyset(\underline{x}) \equiv 1. \tag{10.5}$$

It can be shown [26, ch. 1.3] or [4] that the 2^d functions $Q_{\mathcal{S}}$ are an orthonormal basis for the space $\mathcal{L}^{\mathbb{R}}$ with the inner product (10.2). Then we can introduce the Fourier coefficients (also known as Walsh-Hadamard transforms)

$$f_{\mathcal{S}} = \langle Q_{\mathcal{S}}, f \rangle = \frac{1}{2^d} \sum_{\underline{x} \in \{0,1\}^d} Q_{\mathcal{S}}(\underline{x}) f(\underline{x}). \tag{10.6}$$

or

$$\hat{f}_{\underline{w}} = \frac{1}{2^d} \sum_{\underline{x} \in \{0,1\}^d} f_{\underline{w}}(\underline{x}) f(\underline{x}). \tag{10.7}$$

This implies that

$$f(\underline{x}) = \sum_{\mathcal{S}} Q_{\mathcal{S}}(\underline{x}) f_{\mathcal{S}}. \tag{10.8}$$

or

$$f(\underline{x}) = \sum_{\underline{w}} \hat{f}_{\underline{w}}(\underline{x}) \hat{f}_{\underline{w}}. \tag{10.9}$$

This is the **Fourier expansion** of the generalized Boolean function f.

10.2 Detecting Probabilistic Response Profiles by Fourier Expansions; Examples

We describe next a matrix rule for computing the Fourier coefficients as Walsh-Hadamard transforms, c.f., [4, 23, 27].

Let $\underline{w} = (w_l)_{l=1}^d$ be the bits in the conventional binary representation of the integer i, $i \in 0, \ldots, 2^d - 1$ and $\underline{x} = (x_l)_{l=1}^d$ be the bits in the conventional binary representation of the integer j, $j \in 0, \ldots, 2^d - 1$. Let $\text{AND}_d(\underline{w}, \underline{x})$ be the bitwise logical AND. If $H(\text{ AND}_d(\underline{w}, \underline{x}))$ is the number of ones in $\text{AND}_d(\underline{w}, \underline{x})$ (Hamming weight), then

$$h_{ij} = (-1)^{H(\text{ AND}_d(\underline{w}, \underline{x}))}$$

is the text book definition of the element in the ith row and jth column of a Hadamard matrix H_n, where $n = 2^d$. But then it follows

$$h_{ij} = (-1)^{\sum_{l=1}^d w_l x_l}.$$

In this the orthogonality of the functions $f_{\underline{w}}(\underline{x})$ is expressed as

$$\frac{H_n H_n}{2^d} = I_n.$$

Hence holds that

$$H_n^{-1} = \frac{H_n}{2^d}.$$

If $f \in \mathcal{L}^{\mathbb{R}}, f = (f_0, \ldots, f_{2^d-1})^T$, then the Fourier coefficients $\hat{f} = \left(\hat{f}_0, \ldots, \hat{f}_{2^d-1}\right)^T$ in (10.7) are given by the matrix multiplication

$$\hat{f} = H_n f / 2^d \tag{10.10}$$

and the inverse operation (10.9) is

$$f = H_n \hat{f}. \tag{10.11}$$

Let us next return to the discussion in Section 4.2. We have for any $\underline{x} \in \{0,1\}^d$ a number $\mathbf{p}_{\underline{x}} \in [0,1]$ that gives the probability of potential outcome $D = 1$ given the exposures in $\underline{x}$. The probabilistic response profile is

$$\mathbf{p} = \left(\mathbf{p}_{\underline{x}}\right)_{\underline{x} \in \{0,1\}^d} \in [0,1]^{2^d}.$$

The probability of potential outcome $D = 1$ given the exposures in $\underline{x}$ and corresponding to the Boolean function α by means of a qualitative Bayesian network is $\mathbf{p}^{\alpha}_{\underline{x}} \in [0, 1]$. We have $\mathbf{p}^{\alpha} = \left(\mathbf{p}^{\alpha}_{\underline{x}}\right)_{\underline{x}\in\{0,1\}^d}$. Both $\mathbf{p}$ and $\mathbf{p}^{\alpha}$ are in the space $\mathcal{L}^{\mathbb{R}}$. Then we have that $\mathbf{p} \leftrightarrow (\hat{\mathbf{p}}_{\underline{w}})_{\underline{w}\in\{0,1\}^d}$ in the sense of (10.9) or (10.10)

$$\mathbf{p}_{\underline{x}} = \sum_{\underline{w}} \hat{f}_{\underline{w}}(\underline{x})\hat{\mathbf{p}}_{\underline{w}}.$$

and for $\mathbf{p}^{\alpha}$

$$\mathbf{p}^{\alpha}_{\underline{x}} = \sum_{\underline{w}} \hat{f}_{\underline{w}}(\underline{x})\hat{\mathbf{p}}^{\alpha}_{\underline{w}}.$$

Let $\|f\|$ denote the norm on $\mathcal{L}^{\mathbb{R}}$ corresponding to the inner product in (10.2). By the orthogonality of $\hat{f}_{\underline{w}}(\underline{x})$ it follows that

$$\|\mathbf{p} - \mathbf{p}^{\alpha}\|^2 = \sum_{\underline{w}} \left(\hat{\mathbf{p}}_{\underline{w}} - \hat{\mathbf{p}}^{\alpha}_{\underline{w}}\right)^2. \tag{10.12}$$

Example 10.1: We apply Fourier analysis to the nine different probabilities of potential outcome denoted by β_i in Section 3.5. Let thus $d = 2$ and

$$\mathbf{p} = (\mathbf{p}_{11}, \mathbf{p}_{10}, \mathbf{p}_{01}, \mathbf{p}_{00}) = (0.01, 0.02, 0.90, 0.91)\,.$$

This exhibits by definition (4.1) or eq. (4.7) a causal response profile for $\beta_4 = \alpha_{13}$. We use (10.10) to get

$$(\hat{\mathbf{p}}_{\underline{w}}) = (0.46, 0.05, -0.4455, 0).$$

With $q_1 = 0.12$ (q_2 does not matter as X_2 is blocked by X_1) we get

$$\mathbf{p}^{\beta_4} = (0.12, 0.12, 1.0, 1.0).$$

When (10.10) is again applied we get

$$\hat{\mathbf{p}}^{\beta_4} = (0.56, 0, -0.44, 0).$$

For a comparison, with $\alpha_{14} = \beta_8$ with $q_1 = 0.9000$, and $q_2 = 0.1000$ gives $\mathbf{p}^{\beta_8} = (0.8100, 0, 0.9000, 0)$. Then with (10.10)

$$\hat{\mathbf{p}}^{\beta_8} = (0.4275, 0.4275, -0.0225, -0.0225)$$

Hence

$$\|\mathbf{p} - \mathbf{p}^{\beta_4}\|^2 = 0.11$$

and

$$\|\mathbf{p} - \mathbf{p}^{\beta_8}\|^2 = 0.91.$$

Hence, in this comparison $\mathbf{p}^{\beta_4}$ is closer to $\mathbf{p}$, whose response profile β_4. One can check that the rest of β_i are not either in this sense closer to $\mathbf{p}$ than $\mathbf{p}^{\beta_4}$. Of course, this requires tuning q_1 and q_2 by hand in each case so that comparison seems fair. ■

10.3 Examples of Fourier Expansions

Example 10.2: Let for all $\underline{x}$

$$f(\underline{x}) = 1.$$

Then for $\mathcal{S} \neq \emptyset$

$$f_{\mathcal{S}} = \frac{1}{2^d} \sum_{\underline{x} \in \{0,1\}^d} Q_{\mathcal{S}}(\underline{x}) f(\underline{x}) = \frac{1}{2^d} \sum_{\underline{x} \in \{0,1\}^d} Q_{\mathcal{S}}(\underline{x}) Q_{\emptyset}(\underline{x}) = 0,$$

because $\{Q_S(\underline{x})\}_{\mathcal{S}}$ is an orthonormal system, and

$$f_{\emptyset} = \frac{1}{2^d} \sum_{\underline{x} \in \{0,1\}^d} Q_{\emptyset}(\underline{x}) Q_{\emptyset}(\underline{x}) = \frac{1}{2^d} \sum_{\underline{x} \in \{0,1\}^d} 1 = 1,$$

which we, of course, already knew by orthonormality. Hence (10.8) becomes

$$f(\underline{x}) = Q_{\emptyset}(\underline{x}) f_{\emptyset}.$$ ■

Next we point out certain rules of computation that basically cover all that is needed in the sequel for finding Fourier expansions of $P(D = 1|\underline{x})$ regarded as a generalized Boolean function of $\underline{x}$.

Example 10.3:

$$\sum_{x_j=0,1} (-1)^{x_j} = 0. \tag{10.13}$$

$$\sum_{x_j=0,1} (1 - q_j^{x_j})(-1)^{x_j} = q_j - 1, \quad j \in \mathcal{S}, \tag{10.14}$$

$$\sum_{x_j=0,1} q_j^{x_j} (-1)^{x_j} = 1 - q_j, \quad j \in \mathcal{S}, \tag{10.15}$$

$$\sum_{x_j=0,1} q_j^{x_j} = 1 + q_j, \quad j \notin \mathcal{S}, \tag{10.16}$$

$$\sum_{x_j=0,1} (1 - q_j^{x_j}) = 1 - q_j, \quad j \notin \mathcal{S}. \tag{10.17}$$

$$\sum_{x_l=0,1} \sum_{x_j=0,1} (1 - q_l^{x_l}) = 2(1 - q_j), \quad l \neq j. \tag{10.18}$$

$$\sum_{x_l=0,1} \sum_{x_j=0,1} q_l^{x_l} = 2(1 + q_j), \quad l \neq j. \tag{10.19}$$

■

Example 10.4 (Fourier Analysis of General Noisy-Or): In Proposition 7.3 we found for OR_d that

$$P_{\mathrm{OR}_d}(D=1|\underline{x}) = 1 - \prod_{j=1}^{d} q_j^{x_j}. \tag{10.20}$$

We regard $P_{\mathrm{OR}_d}(D=1|\underline{x}) =$ as a generalized Boolean function, i.e.,

$$f_{\mathrm{OR}_d}(\underline{x}) \equiv P_{\mathrm{OR}_d}(D=1|\underline{x}) = 1 - \prod_{j=1}^{d} q_j^{x_j}.$$

We need, in view of the preceding Example 10.2, just to find the coefficients $h_{\mathcal{S}}$ for the auxiliary function

$$h(\underline{x}) = \prod_{j=1}^{d} q_j^{x_j}.$$

By definition we get for any (non-empty) $\mathcal{S}$

$$h_{\mathcal{S}} = \frac{1}{2^d} \sum_{\underline{x}\in\{0,1\}^d} Q_{\mathcal{S}}(\underline{x}) \prod_{j=1}^{d} q_j^{x_j}$$

$$= \frac{1}{2^d} \sum_{\underline{x}\in\{0,1\}^d} \prod_{j\in\mathcal{S}} (-1)^{x_j} \prod_{j=1}^{d} q_j^{x_j} = \frac{1}{2^d} \sum_{\underline{x}\in\{0,1\}^d} \prod_{j\in\mathcal{S}} (-q_j)^{x_j} \prod_{j\notin\mathcal{S}}^{d} q_j^{x_j}.$$

But we get by separability of the summed products and an appropriate application of the Sum-Product law like (5.15)

$$= \frac{1}{2^d} \sum_{x_1,\ldots,x_d} \prod_{j\in\mathcal{S}} (-q_j)^{x_j} \prod_{j\notin\mathcal{S}}^{d} q_j^{x_j} = \prod_{j\in\mathcal{S}} \sum_{x_j=0,1} (-q_j)^{x_j} \cdot \prod_{j\notin\mathcal{S}} \sum_{x_j=0,1} q_j^{x_j}.$$

Hence we have obtained using (10.15) and (10.16),

$$h_{\mathcal{S}} = \frac{1}{2^d} \prod_{j\in\mathcal{S}} (1-q_j) \cdot \prod_{j\notin\mathcal{S}} (1+q_j), \tag{10.21}$$

and

$$h_{\emptyset} = \frac{1}{2^d} \sum_{\underline{x}\in\{0,1\}^d} \prod_{j=1}^{d} q_j^{x_j} = \frac{1}{2^d} \prod_{j=1}^{d} (1+q_j). \tag{10.22}$$

Therefore the Noisy-Or probability is in view of (10.8) expressed by

$$\begin{aligned} &P_{\mathrm{OR}_d}(D=1|\underline{x}) = 1 - \prod_{j=1}^{d} q_j^{x_j} = \\ &1 - \frac{1}{2^d} \prod_{j=1}^{d} (1+q_j) - \frac{1}{2^d} \sum_{\mathcal{S}\neq\emptyset} \prod_{j\in\mathcal{S}} (1-q_j) \prod_{j\notin\mathcal{S}} (1+q_j) \prod_{j\in\mathcal{S}} (-1)^{x_j}. \end{aligned} \tag{10.23}$$

■

We can readily write down explicit expansions for all of the functions $P_{\alpha_i}(D = 1|x_1, x_2)$ for ICI with $d = 2$ by means of (10.13) - (10.19). We give a couple of examples.

Example 10.5 (Fourier Expansion of α_9; Noisy Sheffer stroke with $d = 2$):

$$\begin{aligned} P_{\alpha_9}(D = 1|x_1, x_2) = & \frac{1}{4}\left(((1 + q_1)\,(1 - q_2) + 2\,(1 + q_2)\right) \\ & - \frac{1}{4}\left((1 - q_1)(1 - q_2)\right)(-1)^{x_1} \\ & - \frac{1}{4}\left((1 + q_1)(q_2 - 1) + 2(1 - q_2)\right)(-1)^{x_2} \\ & - \frac{1}{4}\,(1 - q_1)\,(1 - q_2)\,(-1)^{x_1}(-1)^{x_2}. \end{aligned} \tag{10.24}$$

■

Example 10.6 (Fourier Expansion of Noisy $\mathbf{Maj_3}$): We use the expression (7.43) for $f(\underline{x}) = P_{Maj_3}(D = 1|\underline{x}_3)$ and get by a mechanical application of the rules (10.13)-(10.19) the Fourier coefficients

$$f_\emptyset = \frac{1}{8}\left[2(1 - q_2)(1 - q_3) + (1 - q_1)(1 + q_2)(1 - q_3) + (1 - q_1)(1 - q_2)(1 + q_3)\right]$$

$$f_{\{1\}} = \frac{1}{8}\left[(q_1 - 1)(1 + q_2)(1 - q_3) + (q_1 - 1)(1 - q_2)(1 + q_3)\right]$$

$$f_{\{2\}} = \frac{1}{8}\left[2(q_2 - 1)(1 - q_3) + (1 - q_1)(1 - q_2)(1 - q_3) + (1 - q_1)(q_2 - 1)(1 + q_3)\right]$$

$$f_{\{3\}} = \frac{1}{8}\left[2(1 - q_2)(q_3 - 1) + (1 - q_1)(1 + q_2)(q_3 - 1) + (1 - q_1)(1 - q_2)(1 - q_3)\right]$$

$$f_{\{1,2\}} = \frac{1}{8}\left[(q_1 - 1)(1 - q_2)(1 - q_3) + (q_1 - 1)(q_2 - 1)(1 + q_3)\right]$$

$$f_{\{1,3\}} = \frac{1}{8}\left[(q_1 - 1)(1 + q_2)(q_3 - 1) + (q_1 - 1)(1 - q_2)(1 - q_3)\right]$$

$$f_{\{2,3\}} = \frac{1}{8}\left[2(q_2 - 1)(q_3 - 1) + (1 - q_1)(1 - q_2)(q_3 - 1) + (1 - q_1)(q_2 - 1)(1 - q_3)\right]$$

$$f_{\{1,2,3\}} = \frac{1}{8}\left[(q_1 - 1)(1 - q_2)(q_3 - 1) + (q_1 - 1)(q_2 - 1)(1 - q_3)\right]$$

Thus

$$
\begin{aligned}
&P_{Maj_3}(D=1|\underline{x}_3) = \\
&\frac{1}{8}\left[2(1-q_2)(1-q_3)+(1-q_1)(1+q_2)(1-q_3)+(1-q_1)(1-q_2)(1+q_3)\right. \\
&+f_{\{1\}}(-1)^{x_1}+f_{\{2\}}(-1)^{x_2}+f_{\{3\}}(-1)^{x_3} \\
&+f_{\{1,2\}}(-1)^{x_1}(-1)^{x_2}+f_{\{1,3\}}(-1)^{x_1}(-1)^{x_3}+f_{\{2,3\}}(-1)^{x_2}(-1)^{x_3} \\
&\left.+f_{\{1,2,3\}}(-1)^{x_1}(-1)^{x_2}(-1)^{x_3}\right].
\end{aligned}
\tag{10.25}
$$

■

Example 10.7 (Fourier Expansion of Noisy Tribes$_{2,2}$): In view of Example 7.1 it holds for any $\underline{x} \in \{0,1\}^4$

$$
P_{\text{Tribes}_{2,2}}(D=1|\underline{x}) = g_1 + g_2 - g_1 g_2, \tag{10.26}
$$

where

$$
g_1 = \prod_{j=1}^{2}(1-q_j^{x_j}), g_2 = \prod_{j=3}^{4}(1-q_j^{x_j}).
$$

First

$$
f_\emptyset = \frac{1}{16}\left[\sum_{i=1}^{7} P_{\text{Tribes}_{2,2}}(D=1|\underline{x}_i)\right],
$$

and we use (7.25) and (7.26) to get

$$
f_\emptyset = \frac{1}{16}\left[4\prod_{j=1}^{2}(1-q_j)+4\prod_{j=3}^{4}(1-q_j)-\prod_{j=1}^{2}(1-q_j)\prod_{j=3}^{4}(1-q_j)\right]. \tag{10.27}
$$

Another mechanical application of the rules (10.13)- (10.19) gives

$$
\begin{aligned}
f_{\{1\}} &= \frac{1}{16}\left[4(q_1-1)(1-q_2)+(q_1-1)(1-q_2)(1-q_3)(1-q_4)\right] \\
f_{\{2\}} &= \frac{1}{16}\left[4(1-q_1)(q_2-1)+(1-q_1)(q_2-1)(1-q_3)(1-q_4)\right] \\
f_{\{3\}} &= \frac{1}{16}\left[4(q_3-1)(1-q_4)+(1-q_1)(1-q_2)(q_3-1)(1-q_4)\right] \\
f_{\{4\}} &= \frac{1}{16}\left[4(q_4-1)(1-q_3)+(1-q_1)(1-q_2)(1-q_3)(q_4-1)\right] \\
f_{\{1,2\}} &= \frac{1}{16}\left[q_1-1)(q_2-1)(4-(1-q_3)(1-q_4)\right] \\
f_{\{1,3\}} &= \frac{1}{16}\left[(q_1-1)(1-q_2)(q_3-1)(1-q_4)\right]
\end{aligned}
$$

$$
\begin{aligned}
f_{\{1,4\}} &= \frac{1}{16}\left[(q_1-1)(1-q_2)(1-q_3)(q_4-1)\right] \\
f_{\{2,3\}} &= \frac{1}{16}\left[(1-q_1)(q_2-1)(q_3-1)(1-q_4)\right] \\
f_{\{2,4\}} &= \frac{1}{16}\left[(1-q_1)(q_2-1)(1-q_3)(q_4-1)\right] \\
f_{\{3,4\}} &= \frac{1}{16}\left[(q_4-1)(q_3-1)(4-(1-q_1)(1-q_2)\right] \\
f_{\{1,2,3\}} &= \frac{1}{16}\left[(q_4-1)(q_3-1)(4-(q_3-1)(1-q_4)\right] \\
f_{\{1,3,4\}} &= \frac{1}{16}\left[(q_1-1)(1-q_2)(q_3-1)(q_4-1)\right] \\
f_{\{2,3,4\}} &= \frac{1}{16}\left[(1-q_1)(q_2-1)(q_3-1)(q_4-1)\right] \\
f_{\{1,2,3,4\}} &= \frac{1}{16}\left[(q_1-1)(q_2-1)(q_3-1)(q_4-1)\right]
\end{aligned}
$$

■

10.4 The Linear Risk Model Revisited

We now return to the discussion in Section 10 above, or put in one formula, we return to

$$P(D=1 \mid x_1, x_2) = a + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2 \tag{10.28}$$

We point out now that this is the the general form of the exact Fourier series of $P_{\alpha_i}(D = 1|x_1, x_2)$, $i = 1, 2, \ldots, 16$. We begin with α_2. For $d = 2$ we obtain via (10.23) the expression

$$
\begin{aligned}
P_{\alpha_2}(D=1|x_1,x_2) \quad = \quad & 1 - \frac{1}{4}\prod_{j=1}^{2}(1+q_j) \\
& -\frac{1}{4}(1-q_1)(1+q_2)(-1)^{x_1} - \frac{1}{4}(1+q_1)(1-q_2)(-1)^{x_2} \\
& -\frac{1}{4}(1-q_1)(1-q_2)(-1)^{x_1}(-1)^{x_2}.
\end{aligned} \tag{10.29}
$$

Let us simplify writing by setting

$$\bar{a} = 1 - \frac{1}{4}\prod_{j=1}^{2}(1+q_j), b = \frac{1}{2}(1-q_1)(1+q_2),$$

$$c = \frac{1}{2}(1+q_1)(1-q_2), d = -(1-q_1)(1-q_2)$$

and by

$$u_i = (-1) \cdot (-1)^{x_i}, i = 1, 2, \tag{10.30}$$

so that $u_i \in \{-1, 1\}$. Then (10.29) becomes

$$P_{\alpha_2}(D = 1|x_1, x_2) = \bar{a} + \frac{1}{2}bu_1 + \frac{1}{2}cu_2 + \frac{1}{4}du_1u_2. \tag{10.31}$$

Next we introduce $o_i \in \{0, 1\}$ by

$$u_i = 2o_i - 1. \tag{10.32}$$

When we substitute these expressions for u_i in (10.31) we obtain in the right hand side the sum

$$\bar{a} - 1/2(b + c - d/2) + (b - d/2)o_1 + (c - d/2)o_2 + do_1 \cdot o_2$$

We agree on new assignments

$$a \equiv \bar{a} - 1/2(b + c - d/2), b_1 \equiv (b - d/2), b_2 \equiv (c - d/2), b_3 \equiv d.$$

Here $a > 0$, $b_1 > 0$, $b_2 > 0$ and $b_3 < 0$, if $0 < 1 - q_i < 1$, $i = 1, 2$. Hence we see that $P_{\alpha_2}(D = 1|x_1, x_2)$ is of the form

$$P_{\alpha_2}(D = 1|x_1, x_2) = a + b_1o_1 + b_2o_2 + b_3o_1 \cdot o_2.$$

Let us note that by (10.30) and (10.32) $x_i = 0 \leftrightarrow u_i = 0$ and $x_i = 1 \leftrightarrow u_i = 1$ and we may write

$$P_{\alpha_2}(D = 1|x_1, x_2) = a + b_1x_1 + b_2x_2 + b_3x_1 \cdot x_2. \tag{10.33}$$

Hence we have shown for that the linear model (10.28) is a built-ine property of $P_{\alpha_2}(D = 1|x_1, x_2)$ as its Fourier expansion. By comparison with (9.10) we find also that $b_3 = \text{SIC}_2(= d)$. The role of b_3 in (9.17) is also significant.

A glance at Example 10.5 shows that $P_{\alpha_9}(D = 1|x_1, x_2)$ can be represented by means of a linear model in the manner shown above. Hence we have the following claim.

Proposition 10.1: *For each α_i, $i = 1, \ldots, 16$ we have a linear model*

$$P_{\alpha_i}(D = 1|x_1, x_2) = a^{(i)} + b_1^{(i)}x_1 + b_2^{(i)}x_2 + b_3^{(i)}x_1 \cdot x_2, \tag{10.34}$$

which is its Fourier expansion as a function of $(x_1, x_2) \in \{0, 1\}^2$. We have in addition $b_3^{(i)} = SIC_i$.

Proof: $P_{\alpha_1}(D = 1|x_1, x_2) = a^{(1)} = 1$ and $P_{\alpha_{16}}(D = 1|x_1, x_2) = a^{(16)} = 0$. For the rest of α_is find

$$f_{\mathcal{S}}^{(i)} = \frac{1}{4} \sum_{(x_1,x_2)\in\{0,1\}^2} Q_{\mathcal{S}}(\underline{x})P_{\alpha_i}(D = 1|x_1, x_2) \tag{10.35}$$

by using the rules in Example 10.3. This implies that

$$P_{\alpha_i}(D = 1|x_1, x_2) = \sum_{\mathcal{S}} Q_{\mathcal{S}}(x_1, x_2) f_{\mathcal{S}}^{(i)}, \tag{10.36}$$

which is re-written in the same way as the case α_2 to get (10.34). The details are omitted.■

Equation 10.34 has previously been proposed as a model [42, 44], but without the formal connection with the model. Hence, it can be argued that for binary variables the true model is linear. This has implications for non-saturated models, since a logistic model then can lead to different results than the linear model [38].

11 Concluding Remarks

In this paper we have further developed the model for probabilistic potential outcomes and probabilistic sufficient causes by combining and generalizing results from [11, 31]. As well as exploring the model's connection with Boolean functions in the form of BCF and Fourier expansion. We presented an equation (8.3) to calculate the probability for a potential outcome with d variables based on this model.

References

[1] A. Ahlbom and L. Alfredsson. Interaction: a word with two meanings creates confusion. *European Journal of Epidemiology*, 20(7):563–564, 2005.

[2] S. M. Aji and R. J. McEliece. The generalized distributive law. *IEEE Transactions on Information Theory*, 46(2):325–343, 2000.

[3] T. Andersson, L. Alfredsson, H. Källberg, S. Zdravkovic, and A. Ahlbom. Calculating measures of biological interaction. *European Journal of Epidemiology*, 20(7):575–579, 2005.

[4] A. Bernasconi. Mathematical techniques for the analysis of boolean functions. *PhD thesis*, 1998.

[5] F. M. Brown. *Boolean reasoning: the logic of Boolean equations*. Springer Science & Business Media, 2012.

[6] N. Cartwright. Causal laws and effective strategies. *Noûs*, pages 419–437, 1979.

[7] C. A. Dawid et al. Fundamentals of statistical causality. *Research Report 279*, 2007.

[8] A. B. de González and D. R. Cox. Interpretation of interaction: A review. *The Annals of Applied Statistics*, 1(2):371–385, 2007.

[9] F. J. Dıez and M. J. Druzdzel. Canonical probabilistic models for knowledge engineering. *UNED, Madrid, Spain, Technical Report CISIAD-06-01*, 2006.

[10] D. J. Finney. *Probit analysis: a statistical treatment of the sigmoid response curve*. Cambridge university press, 1952.

[11] W. D. Flanders. On the relationship of sufficient component cause models with potential outcome (counterfactual) models. *European Journal of Epidemiology*, 21(12):847–853, 2006.

[12] S. Greenland and C. Poole. Invariants and noninvariants in the concept of interdependent effects. *Scandinavian Journal of Work, Environment & Health*, pages 125–129, 1988.

[13] P. Gustavsson, F. Nyberg, G. Pershagen, P. Schéele, R. Jakobsson, and N. Plato. Low-dose exposure to asbestos and lung cancer: dose-response relations and interaction with smoking in a population-based case-referent study in stockholm, sweden. *American Journal of Epidemiology*, 155(11):1016–1022, 2002.

[14] M. A. Harrison. *Introduction to switching and automata theory*. McGraw-Hill (New York), 1965.

[15] D. Heckerman and J. S. Breese. Causal independence for probability assessment and inference using bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 26(6):826–831, 1996.

[16] P. W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

[17] C. Howson and P. Urbach. *Scientific reasoning: the Bayesian approach*. Open Court Publishing, 2006.

[18] P. K. Kearns, H. A. Casey, and J. P. Leach. Hypothesis: Multiple sclerosis is caused by three-hits, strictly in order, in genetically susceptible persons. *Multiple Sclerosis and Related Disorders*, 24:157–174, 2018.

[19] T. J. Koski and J. Noble. A review of bayesian networks and structure learning. *Mathematica Applicanda*, 40(1), 2012.

[20] S. C. Lee. *Modern switching theory and digital design*, volume 441. Prentice-Hall Englewood Cliffs, NJ, 1978.

[21] P. J. Lucas. Bayesian network modelling through qualitative patterns. *Artificial Intelligence*, 163(2):233–263, 2005.

[22] J. L. Mackie. Causes and conditions. *American Philosophical Quarterly*, 2(4):245–264, 1965.

[23] Y. Mansour. Learning boolean functions via the fourier transform. In *Theoretical advances in neural computation and learning*, pages 391–424. Springer, 1994.

[24] O. S. Miettinen. Causal and preventive interdependence: elementary principles. *Scandinavian Journal of Work, Environment & Health*, pages 159–168, 1982.

[25] L. R. Novick and P. W. Cheng. Assessing interactive causal influence. *Psychological Review*, 111(2):455, 2004.

[26] R. O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[27] J. Ott and R. A. Kronmal. Some classification procedures for multivariate binary data using orthogonal functions. *Journal of the American Statistical Association*, 71(354):391–399, 1976.

[28] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

[29] J. Pearl, M. Glymour, and N. P. Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.

[30] D. L. Poole and A. K. Mackworth. *Artificial Intelligence: foundations of computational agents*. Cambridge University Press, 2010.

[31] R. R. Ramsahai. Probabilistic causality and detecting collections of interdependence patterns. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):705–723, 2013.

[32] D. H. Rehkopf, M. M. Glymour, and T. L. Osypuk. The consistency assumption for causal inference in social epidemiology: When a rose is not a rose. *Current Epidemiology Reports*, 3(1):63–71, 2016.

[33] B. Reusch. On the generation of prime implicants. Technical report, Cornell University, 1976.

[34] K. J. Rothman. *Epidemiology: an introduction*. Oxford University Press, 2012.

[35] K. J. Rothman and S. Greenland. Causation and causal inference in epidemiology. *American Journal of Public Health*, 95(S1):S144–S150, 2005.

[36] K. J. Rothman, S. Greenland, and T. L. Lash. *Modern epidemiology*. Lippincott Williams & Wilkins, 2008.

[37] D. B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688, 1974.

[38] A. Skrondal. Interaction as departure from additivity in case-control studies: a cautionary note. *American Journal of Epidemiology*, 158(3):251–258, 2003.

[39] J. Splawa-Neyman, D. M. Dabrowska, T. P. Speed, et al. On the application of probability theory to agricultural experiments. *Statistical Science*, 5(4):465–472, 1990. Reprint.

[40] P. Suppes. *A probabilistic theory of causality*. North-Holland Publishing Company Amsterdam, 1970.

[41] T. J. VanderWeele. *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press, 2016.

[42] T. J. VanderWeele and T. S. Richardson. General theory for interactions in sufficient cause models with dichotomous exposures. *Annals of Statistics*, 40(4):2128, 2012.

[43] T. J. VanderWeele and J. M. Robins. The identification of synergism in the sufficient-component-cause framework. *Epidemiology*, 18(3):329–339, 2007.

[44] T. J. VanderWeele and J. M. Robins. Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika*, 95(1):49–61, 2008.

[45] T. J. VanderWeele and J. M. Robins. Stochastic counterfactuals and stochastic sufficient causes. *Statistica Sinica*, 22(1):379, 2012.

[46] J. Vomlel. Noisy-or classifier. *International Journal of Intelligent Systems*, 21(3):381–398, 2006.

[47] K. M. Wogau. On historical explanation 1. *Theoria*, 28(3):213–233, 1962.

[48] Z. Yu and M. J. van der Laan. Construction of counterfactuals and the g-computation formula. *U.C. Berkeley Division of Biostatistics Working Paper Series*, 2002.

[49] N. L. Zhang and D. Poole. Exploiting causal independence in bayesian network inference. *Journal of Artificial Intelligence Research*, 5:301–328, 1996.

Paper C

Measures of Additive Interaction and Effect Direction

Measures of Additive Interaction and Effect Direction

by

Daniel Berglund, Helga Westerlind, and Timo Koski

Abstract

Measures for additive interaction are defined using risk ratios. These ratios need to be modeled so that all combinations of the exposures are harmful, as the scale between protective and harmful factors differs. This remodeling is referred to as recoding. Previously, recoding has been thought of as random. In this paper, we will examine and discuss the impact of recoding in studies with small effect sizes, such as genome wide association studies, and the impact recoding has on significance testing.

Keywords: Additive Interaction; Sufficient Cause; Logistic Regression; Linear Odds; RERI; Effect Direction

1 Introduction

The risk of some diseases can be effected by the presence of a combination of exposures that together have a large impact on the risk, but where the exposures have smaller marginal effects [1, 10, 17]. Such combinations can be studied by the estimation of additive interaction, and there is large number of such methods [1, 2, 7–9, 24, 29].

However, the measures used to estimate additive interaction are based on risk ratios, and the ratios reference group needs to be set so that all exposures are harmful to prevent errors in the estimation [14]. This choice of reference group is not only based on the exposure's effect in the general population, but also conditioned on the other exposures present in the study.

In previous papers [9, 11, 14, 24] when estimating the additive interaction the reference group has not been viewed as random. However, in in a study if a factor is harmful or protective is affected by uncertainty. Since it is also the conditional effect, and not the effect from the exposure in general, it can be unknown and have high uncertainty. In this paper we will examine the impact of the randomness of the effect directions on the estimation of additive interaction, although its' estimate is conditional on the reference group.

We show that the randomness of the choice of reference group for the ratios could cause errors in the estimation of the confidence interval for additive interaction, essentially the confidence interval becomes conditioned on the reference group. However, the interactions interpretation remains unchanged if the reference is adjusted for.

This paper is organized as follows: In Section 2 we will summarize the background on potential outcomes, sufficient causes and additive interaction. The effect directions are defined and explained in Section 3. The measures of additive interaction and the effect on them from considering effect directions as random are described in Section 4. And finally in Section 5 a concluding discussion.

2 Potential Outcomes and Sufficient Causes

Consider a potential outcome model [6, 19, 25] with d binary exposures, $X_1, \ldots, X_d$, and let D be some binary outcome. U is the sample space of individuals, and let u be an individual in the population. Also let Z be a set of covariates that are included to adjust for confounding, the covariates are not required to be binary. We also set the negation of some exposure X as $\bar{X} = 1 - X$.

Definition 2.1: *Potential outcome*
For an individual u let $D_x(u)$ be the potential outcome if the binary exposures are set to $X = x$ ■

Definition 2.2: *Observed outcome*
For an individual u let $D(u)$ be the observed outcome for the individual ■

We also let

$$p_x = P(D = 1|X = x). \tag{2.1}$$

The inference about the potential outcomes is based on the observed outcomes since we can not observe the potential outcomes. For us to be able to use the observed outcomes for inferences about the potential outcomes we make the usual assumptions, consistency and conditional exchangeability.

Assumption 2.1: *Consistency*
For an individual u with exposures set to $X = x$ it holds that

$$D_x(u) = D(u). \tag{2.2}$$

■

Assumption 2.2: *Conditional exchangeability*
Given a set of exposures E and a set of covariates Z the potential outcome D_e is independent of expsoure level(e) of E. Formally,

$$D_e \perp\!\!\!\perp E|Z. \tag{2.3}$$

■

The consistency assumption says that an individuals observed outcome is the same as the potential outcome with the individuals exposures. The assumption is for instance broken if the interventions effect is not the same as the original effect. E.g., the difference between income from work or as a social program [16].

The conditional exchangeability assumption says that given the strata of the covariates Z then the potential outcome D_e is independent of the exposure level of E. I.e., the confounding effect from Z has been adjusted for. It is also called by other names ignorable treatment assignment, no unmeasured confounding or exogeneity[22].

2.1 Sufficient Causes and Additive Interaction

The sufficient cause framework is model for causality where for an outcome a set of binary events form a sufficient cause if for some part of the population the presence of all the events in an individual causes the outcome [25]. A set of sufficient causes forms the sufficient cause model. Each sufficient cause contains several component events and the necessity of all components being present for the outcome to occur based on that particular sufficient cause means there is interaction between the components.

Sufficient cause synergism between a set of exposures means that the exposures are all in present in the same sufficient cause. By estimating additive interaction it can be possible to detect the presence of sufficient cause synergism. However, due to masking effects the conditions on the additive interaction are sufficient, not necessarily [20].

For two exposures the interaction contrast (IC) is defined as [17]

$$IC = p_{11} - p_{10} - p_{01} + p_{00}. \tag{2.4}$$

If we form all possible combinations of the potential outcomes for the two exposures we get all possible responses that an individual in the population could have. This also corresponds to all possible boolean functions with four inputs, they are shown in Table 1. Some of these response patterns imply additive interaction, and are marked with either $+$ for superadditive interaction, or $-$ for subadditive interaction. In general, for d number of exposures there are 2^{2^d} possible response patterns.

3 Effect Direction

A statement that an exposure is harmful or protective can have several different implications, is it the effect from only that exposure, or in some population, or context? Most of the time a statement that an exposure is harmful(e.g., smoking increases risk for lung cancer) referred to the exposure alone in the general population. However it could also refer to the effect in some specific subpopulation(e.g., lung cancer risk for smoking among asbestos workers). As we will see later for additive interaction it is important that we distinguish between the two different types. We start with defining formally what the *effect direction* and *conditional effect direction* are.

Response Type	D_{11}	D_{01}	D_{10}	D_{00}
1	1	1	1	1
2^-	1	1	1	0
3^+	1	1	0	1
4	1	1	0	0
5^+	1	0	1	1
6	1	0	1	0
7^+	1	0	0	1
8^+	1	0	0	0
9^-	0	1	1	1
10^-	0	1	1	0
11	0	1	0	1
12^-	0	1	0	0
13	0	0	1	1
14^-	0	0	1	0
15^+	0	0	0	1
16	0	0	0	0

Table 1: Response patterns with two binary exposures. The patterns marked with $+$ corresponds to superadditive interaction and the ones marked with $-$ subadditive interaction.

Definition 3.1: *Direction of effect*
A set of exposures, x, is a harmful exposure and has risk direction of effect if $p_x > p_{\bar{x}}$. If $p_x < p_{\bar{x}}$ then the effect is protective. ■

Definition 3.2: *Conditional direction of effect*
A set of exposures, x, is a risk or protective exposure condtional on a set of exposures c, $c \not\subset e$, if $p_{x,c} > p_{\bar{x},c}$ respectively $p_{x,c} < p_{\bar{x},c}$ ■

Definition 3.3: *Direction of interaction effect*
The interaction effect for set of exposures, x, is superadditive and has risk direction of effect if $IC > 0$, and subadditive and protective direction if $IC < 0$. ■

Definition 3.4: *Conditional direction of interaction effect*
The interaction effect for set of exposures, x, is superadditive and has risk direction of effect conditional on a set of exposures c if $IC|c > 0$, and sub additive and protective direction if $IC|c < 0$. ■

The conditional effect direction is related to the sufficient cause model, it is if the exposure in some particular pie is the exposure itself or its negation. However, the effect direction is not directly related, since its a combination of conditional effects,

$$P(p_x > p_{\bar{x}}) = \sum_A P(p_x > p_{\bar{x}}|A)P(A) = \sum_A P(p_{x,A} > p_{\bar{x},A})P(A). \quad (3.1)$$

The magnitude of an effect from an exposure can vary over different strata. However, assuming that there is no interaction between the effect of the covariate and the exposures then it does not impact the effect directions. If the assumption can not be made then, just as for the interaction itself, the interaction's effect direction will depend on the value of the covariate.

Theorem 3.1: *Assuming no additive interaction between the exposures and the covariates then the effect direction and conditional effect direction is the same in all strata.*

Proof: This follows from that the effect of the strata on the outcome is the same both with and without the exposure, i.e. if c is the effect from the strata on the probability then the condition for the effect direction,

$$p_x + c > p_{ref} + c. \quad (3.2)$$

is independent of c. ■

3.1 Equivalence Classes

The possible individuals in the population can be can be divided into equivalence classes based on invariance to the conditional effect direction. For two exposures Greenland and Poole [4] grouped the possible individuals, shown in Table 1, into equivalence classes, shown in Table 2. These classes are invariant to changing if one or more exposures are harmful or protective. For each class all the members of that class have the same type of interaction independent of the conditional effect directions of the exposures.

C_M and C_S is referred to as definite interdependence and implies that there is co-action between the exposures. Co-action means that there is synergism between the exposures or their negation, e.g., for two exposures the sufficient cause includes XY, $\bar{X}Y$, $X\bar{Y}$ or $\bar{X}\bar{Y}$ [26]. The class of competing types, C_T, implies either competition between the separate sufficient causes of the exposures or co action for $\bar{D}$. For instance for type 2 it is either competing between the two separate sufficient causes, X and Y, or synergism between $\bar{X}$ and $\bar{Y}$ for $\bar{D}$ [23]. The combination of C_M, C_S and C_T are referred to as causal interdependence and are all the response patterns that imply sufficient cause interaction between any combination of the exposures or their complements for the outcome or its' complement [22].

Class	Response types	Description
C_D	1	Doomed
C_X	6, 11	X causal
C_Y	4, 13	Y causal
C_M	7, 10	Mutual antagonism
C_S	8, 12, 14, 15	Synergistic causation under recoding
C_T	2, 3, 5, 9	Competing/Synergistic prevention
C_I	16	Immune

Table 2: Greenland and Poole's equivalence classes

4 Measures of Interaction and Estimation

In practice instead of calculating IC it is common to use other measures based on relative risks (RR). Since RR can not be estimated from case-control data they are often approximated by using ORs, however, for the approximation to work either the rare disease assumption needs to hold, or certain study designs are needed [12]. Note that the rare disease assumption is that the risk in all *strata* studied is small, not only the prevalence in the general population. A rare disease in the general population is not necessarily rare in some specific strata. The relative risks and odds ratios for the exposures $X = x$ with the reference group $r = \{r_1, r_2\}$ are defined as,

$$\text{RR}_x^r = \frac{p_x}{p_r}, \tag{4.1}$$

and

$$\text{OR}_x^r = \frac{\frac{p_x}{1-p_x}}{\frac{p_r}{1-p_r}}. \tag{4.2}$$

4.1 Models

The ratios are commonly estimated by using logistic regression [17, 24],

$$\ln\left(\frac{p_{x,y}}{1-p_{x,y}}\right) = \alpha + \beta_1 x + \beta_2 y + \beta_3 xy + \gamma_z z. \tag{4.3}$$

An alternative model proposed by Skrondal [18] is the linear odds model,

$$\frac{p_{x,y}}{1-p_{x,y}} = a + b_1 x + b_2 y + b_3 xy + g_z z. \tag{4.4}$$

Note the absence of coefficients corresponding to interaction between the exposures and the covariates, i.e., they are not fully saturated. The logistic model is a multiplicative

model, while the linear odds model is additive [3, 18]. However, since we are interested in additive differences this causes some issues for the logistic model. Under some conditions the logistic model can be used to make inferences about the underlying true additive model [3], but there can be issues if covariates are included.

The logistic model is multiplicative which means that $\beta_3 = 0$ does not have to imply no additive interaction, in fact it implies no multiplicative interaction. This means that if the interaction terms for the interaction between the exposures and covariates are not included then it is implied that there is additive interaction between the exposures and the covariates [5, 18]. The means that the interaction estimated from the logistic model can be incorrect [18].

The linear odds model do not have this issue, but it can return negative odds and fail to converge [18, 27, 28]. Its' maximum likelihood estimators can also have problems when used with continuous covariates [27].

4.2 Measures of Interaction

Using the RRs IC can then be expressed as the relative excces ratio due to interaction (RERI) [17],

$$\text{RERI} = \text{RR}_{11} - \text{RR}_{10} - \text{RR}_{01} + 1. \tag{4.5}$$

The interaction can also be measured using proportional measures [17] which traditionally are defined as

$$\text{SI}_{trad} = \frac{\text{RR}_{11} - 1}{(\text{RR}_{10} - 1) + (\text{RR}_{01} - 1)}, \tag{4.6}$$

$$\text{AP}_{trad} = \frac{\text{RERI}}{\text{RR}_{11}}, \tag{4.7}$$

and

$$\text{AP}^*_{trad} = \frac{\text{RERI}}{\text{RR}_{11} - 1}. \tag{4.8}$$

The synergy index measures how much more the the effect in the doubly exposed group exceeds 1 compared to how much more the exposures separately exceeds 1 together [24].

The interpretations of the attributable proportions are similar, they both measures the proportion of interaction in the group with both exposures. However, AP is the proportion of the *disease* in the doubly exposed group that is due to the interaction, while AP^* is the *effect* that is due to interaction in the doubly exposed group [21]. Ap^* more closely follows the intuitive interpretation one would expect as demonstrated in the following example.

Example 4.1: Let $\text{RR}_{11} = 2$, $\text{RR}_{10} = 1$, $\text{RR}_{01} = 1$. Then $\text{AP}_{trad} = \frac{1}{2} = 50\%$, while $\text{AP}^*_{trad} = \frac{1}{1} = 100\%$. One would expect 100% from RR_{trad} since the $\text{RR}_{10} = 1$, $\text{RR}_{01} = 1$, but this is not case since AP is the proportion of the disease in the doubly exposed group, and not the effect. ■

AP^*_{trad} is also better than AP_{trad} in relation to the issue with covariates and confounding described earlier in the model section [18]. With the linear odds model only AP^*_{trad} and SI_{trad} does not depend on the covariates. For the logistic model all the measures are independent of the covariates since the ORs from the logistic model are independent of the covariates, but this can mean that the measures are incorrect [18].

However, there is a problem with the traditional proportional measures when the interaction is subadditive, they become negative and lack interpretation [8].

Example 4.2: Suppose $RR_{11} = 1.1$, $RR_{10} = 1.2$, and $RR_{01} = 1.3$. Then RERI $= -0.4$, $AP_{trad} = -0.36$, and $AP^*_{trad} = -4$. ■

We have no interpretations for these values since we are no longer observing a superadditive interaction. There is no additional effect or additional individuals affected by the disease because of the interaction. In fact, the interaction has the opposite effect; the effect is lower due to the interaction, and some individuals are protected from the disease because of it [8].

If there is no additive interaction then $RR_{11} = RR_{10} + RR_{01} - 1$. We can use this and instead of comparing the effect of the doubly exposed group against RERI to see how much *more* interaction there is, we can compare RERI against what the effect would be if there was no interaction (i.e., then it would hold that: $RR_{11} = RR_{10} + RR_{01} - 1$) to see how much *less* interaction there is. Which leads to a different version of AP and AP^* [8]:

$$\text{AP} = \frac{\text{RERI}}{\max(\text{RR}_{11}, \text{RR}_{10} + \text{RR}_{01} - 1)} \tag{4.9}$$

$$\text{AP}^* = \frac{\text{RERI}}{\max(\text{RR}_{11}, \text{RR}_{10} + \text{RR}_{01} - 1) - 1} \tag{4.10}$$

For subadditive interaction both AP and AP^* are between 0 and -1 [8]. They are negative to reflect the protective nature of the interaction and so that they can be combined with their counterpart into the interval $[-1, 1]$. The more the interaction reduces the effect compare to what the effect would have been if there was interaction the closer the measures are to -1. With subadditive interaction the interpretations of AP is the proportion of the *disease* in the doubly exposed group that is protected by the interaction, while the interpretations of AP^* is the proportion of the *effect* on the additive scale in the doubly exposed group that is removed due to the interaction.

Example 4.3: Using the same numbers as in the previous example, $RR_{11} = 1.1$, $RR_{10} = 1.2$, and $RR_{01} = 1.3$. This leads to $AP = \frac{-0.4}{1.5} = -0.267$ and $AP^* = \frac{-0.4}{0.5} = -0.8$. The interpretations is then that 26.7% of the individuals in the doubly exposed group that would have had the disease if there was no interaction are healthy. We could also recalculate it as how many more would be sick if there was not an interaction; In this case $\frac{0.267}{1-0.267} = 36.4\%$ more individuals would have been sick if there was no interaction. For AP^{*-} the interpretations is that 80% of the *effect* is removed because of the interaction in the doubly

exposed group, it can also be interpreted as the effect being $100 - 80 = 20\%$ of what it would have been if there was no interaction on the additive scale. ■

The synergy index was not modified by Ola et. al. in [8], but we propose that it should also be modified in a similar manner in order to have the same interpretation for both subadditive and superadditive interaction.

$$\text{SI} = \frac{\max(\text{RR}_{11}, \text{RR}_{10} + \text{RR}_{01} - 1) - 1}{\min(\text{RR}_{11}, \text{RR}_{10} + \text{RR}_{01} - 1) - 1} \tag{4.11}$$

However, based on this definition we can show that SI and AP* are expressions of the same property.

Theorem 4.1: *It holds that:*

$$\text{AP}^* = 1 - \frac{1}{\text{SI}} \tag{4.12}$$

Proof:

$$\begin{aligned} \text{AP}^* =& \frac{\text{RR}_{11} - \text{RR}_{10} - \text{RR}_{01} + 1}{\max(\text{RR}_{11}, \text{RR}_{10} + \text{RR}_{01} - 1) - 1} \\ =& 1 - \frac{\min(\text{RR}_{11}, \text{RR}_{10} + \text{RR}_{01} - 1) - 1}{\max(\text{RR}_{11}, \text{RR}_{10} + \text{RR}_{01} - 1) - 1} \\ =& 1 - \frac{1}{\text{SI}} \end{aligned} \tag{4.13}$$

■

AP* is the better measure since it has an easier and more intuitive interpretations than the synergy index, and unlike the synergy index, it does not have issues with approaching infinity when $\text{RR}_{11} - 1 = 0$ or $\text{RR}_{10} + \text{RR}_{01} - 2 = 0$.

Protective Exposures and Interaction

Independently of how the ratios are estimated there is a problem if one or more of the ratios are below 1, i.e., the exposure have a protective effect. A preventive ratio is in the range $\{0, 1\}$, while a risk ratio is in the range $\{1, \infty\}$. This means that they are not directly comparable on an additive scale [14]. For instance $\text{RR} = 2$ is equivalent to $\text{RR} = 0.5$ for a preventive exposure, but $2 - 0.5 \neq 0$. This problem can be adjusted for by changing the reference group to the group with lowest risk so that all ratios are above 1. However, this is not equivalent to that all exposures have harmful effect direction, it also involves the conditional effect direction [14]. That all ratios are above 1 is equivalent to that all three of the following combinations are harmful: $\{\bar{r_1}, \bar{r_2}\}$, $\{\bar{r_1} \mid r_2\}$, and $\{\bar{r_2} \mid r_1\}$.

No actual recoding of the data and recalculation of the model needs to be performed. The coefficients for the recoded model can be calculated from the first model's estimation,

since the estimation of the model's parameters is unaffected. The conditions for the different reference groups are also the same for both the logistic model, and the linear model using their respective coefficients, and are shown in Table 3.

Parameter \ Reference group	00	10	01	11
θ_1	≥ 0	< 0	$> \theta_2$	$> \theta_1 + \theta_2 + \theta_3$
θ_2	≥ 0	$> \theta_1$	< 0	$> \theta_1 + \theta_2 + \theta_3$
$\theta_1 + \theta_2 + \theta_3$	≥ 0	$> \theta_1$	$> \theta_2$	< 0

Table 3: Parameter conditions for the different reference groups for two exposures

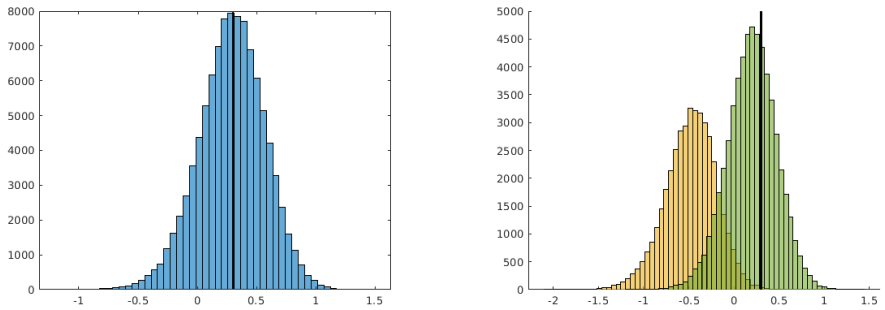
It is often known if an exposure is harmful or protective in the general population. However, the exposures conditional effect direction is commonly not known which causes an issue with the choice of reference group. This is especially problematic if an exposure's true conditional effect is small, i.e., a ratio close to 1, then there is large uncertainty of that exposure's conditional effect direction. In previous methods the reference group has been set using the ratios estimated from the data [13–15, 22, 24]. However, these methods do not account for the randomness of the effect directions.

The point estimation of the interaction is unaffected, though, the confidence interval and, depending on the model, the hypothesis testing can be incorrect. When assuming that the measure is distributed around the estimate of the measure without accounting for the randomness of the recoding parts of the distribution is based on calculating the interaction with protective exposures, since the distribution for the ratios includes the volume where the ratio is below one, even if the ratios distribution is centered around a value above one.

Example 4.4: 100 000 cohorts were simulated with 10 000 individuals in each. The true probabilities for the outcome were set as $p_{11} = 0.07$, $p_{10} = 0.06$, $p_{01} = 0.031$, $p_{00} = 0.03$ so there is high uncertainty in the cohorts if the second exposure is harmful or protective. For each cohort RERI was calculated using either the true reference group, i.e., the second exposure is harmful, or the reference group as estimated from the cohort's data. The histograms for the estimated RERI are shown in Figure 1.

In Figure 1b there are two different distributions; The left distribution is when the second exposure is estimated as protective, and the right distribution when the exposure is estimated as harmful. The high uncertainty about the second exposure's conditional effect direction means we do not know which distribution is the true one. Note that the right distribution is not centered on the true value of RERI, even though its reference group is the true reference, because the underlying distributions for the RRs are truncated, and the truncation skews the distribution to the left in this case. ■

However, the effect of the uncertainty of reference group is smaller than it might seem from the example. The left peak in Figure 1b is estimated with viewing the second exposure as protective so the interpretation of RERI has to account for that. Because of the



(a) Classically assumed distribution for RERI (b) True distribution for RERI including the randomness of the reference group

Figure 1: Histograms from simulating cohorts and estimating RERI with and without accounting for the reference group. Vertical black line is RERI calculated from the true probabilities.

equivalence classes both of them have the same interpretation. There can be problem for the estimation of the confidence interval for the magnitude of the interaction.

The different RERI values, with different reference groups, can no be directly compared without introducing errors. We illustrate this with the following example.

Example 4.5: Assume that there is uncertainty of the conditional effect direction of the second exposure just as in the previous example. Then we have that,

$$RERI^{00} = \frac{p_{11} - p_{10} - p_{01} + p_{00}}{p_{00}}, \tag{4.14}$$

and

$$RERI^{01} = \frac{p_{10} - p_{11} - p_{00} + p_{01}}{p_{01}}. \tag{4.15}$$

Hence,

$$p_{00}RERI^{00} = -p_{01}RERI^{01}. \tag{4.16}$$

The two RERIs have different sign, but also different scaling (p_{00}, p_{01}) which can not be estimated in case-control studies. However, the uncertainty of the choice of reference group, 00 or 01, is directly tied to the difference between p_{00} and p_{01}. So with high uncertainty the error made can be small, but this also depends on the variance of p_{00} and p_{01}.

However, when accounting for the effect directions for some interpretations the different RERI values will still lead to the same interpretation. For instance superadditive interaction for one is equivalent to subadditive in the other. ■

5 Discussion

In this paper we have defined the direction of effects, and clarified its' connection with the estimation of measures of additive interaction that are defined using risk ratios. Conditional effect directions can be highly uncertain, which can cause errors in the estimated variance of the measure. However, the point estimate is unchanged. For the linear odds model the hypothesis testing for the presence of additive interaction is also unchanged, since the test is done by testing $b_3 = 0$. However, any other hypothesis with the linear odds model is affected. For the logistic model there could be an effect on the tests, since no additive interaction does not directly correspond to $\beta_3 = 0$.

The confidence intervals for additive interaction will be different, since the interaction measures are incorrect for some parts of the distribution of the model coefficients. These different areas can not be combined due to the difference in scaling as illustrated in Example 4.5. However, the interpretation is likely the same.

Hypothesis tests of other interpretations than direction/sign, or presence, of additive interaction can be effected depending on what the hypothesis is. For example, uncertainty about exposure A's conditional effect when conditioned on an exposure B can impact the testing of sufficient cause synergism between A and B. However, it is possible to adjust for this by keeping track of the different reference groups. If A conditioned on B is protective, then synergism between A and B when A's conditional effect is harmful is equivalent to synergism between $\bar{A}$ and B, since if A has conditional protective effect then $\bar{A}|B$ is harmful. A test for synergism between A and B is then a test of both synergism between A and B when A is conditionally harmful, and synergism between $\bar{A}$ and B when A is conditionally protective. However, the test does not need to account for this as long as the type of hypothesis tested is the same as an equivalence class, or a combination of classes.

References

[1] T. Andersson, L. Alfredsson, H. Källberg, S. Zdravkovic, and A. Ahlbom. Calculating measures of biological interaction. *European Journal of Epidemiology*, 20(7):575–579, 2005.

[2] S. F. Assmann, D. W. Hosmer, S. Lemeshow, and K. A. Mundt. Confidence intervals for measures of interaction. *Epidemiology*, pages 286–290, 1996.

[3] S. Greenland. Additive risk versus additive relative risk models. *Epidemiology*, pages 32–36, 1993.

[4] S. Greenland and C. Poole. Invariants and noninvariants in the concept of interdependent effects. *Scandinavian Journal of Work, Environment & Health*, pages 125–129, 1988.

[5] S. Greenland, J. M. Robins, and J. Pearl. Confounding and collapsibility in causal inference. *Statistical Science*, pages 29–46, 1999.

[6] P. W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

[7] D. W. Hosmer and S. Lemeshow. Confidence interval estimation of interaction. *Epidemiology*, 3(5):452–456, 1992.

[8] O. Hössjer, I. Kockum, L. Alfredsson, A. K. Hedström, T. Olsson, and M. Lekman. A general framework for and new normalization of attributable proportion. *Epidemiologic Methods*, 6(1), 2017.

[9] H. Källberg, A. Ahlbom, and L. Alfredsson. Calculating measures of biological interaction using R. *European Journal of Epidemiology*, 21(8):571–573, 2006.

[10] H. Källberg, B. Ding, L. Padyukov, C. Bengtsson, J. Rönnelid, L. Klareskog, L. Alfredsson, E. S. Group, et al. Smoking is a major preventable risk factor for rheumatoid arthritis: estimations of risks after various exposures to cigarette smoke. *Annals of the Rheumatic Diseases*, 70(3):508–511, 2011.

[11] M. J. Knol, I. van der Tweel, D. E. Grobbee, M. E. Numans, and M. I. Geerlings. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. *International Journal of Epidemiology*, 36(5):1111–1118, 2007.

[12] M. J. Knol, J. P. Vandenbroucke, P. Scott, and M. Egger. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *American Journal of Epidemiology*, 168(9):1073–1081, 2008.

[13] M. J. Knol and T. J. VanderWeele. Recoding preventive exposures to get valid measures of interaction on an additive scale. *European Journal of Epidemiology*, 26(10):825–826, 2011.

[14] M. J. Knol, T. J. VanderWeele, R. H. Groenwold, O. H. Klungel, M. M. Rovers, and D. E. Grobbee. Estimating measures of interaction on an additive scale for preventive exposures. *European Journal of Epidemiology*, 26(6):433–438, 2011.

[15] M. B. Mathur and T. J. VanderWeele. R function for additive interaction measures. *Epidemiology (Cambridge, Mass.)*, 29(1):e5–e6, 2018.

[16] D. H. Rehkopf, M. M. Glymour, and T. L. Osypuk. The consistency assumption for causal inference in social epidemiology: When a rose is not a rose. *Current Epidemiology Reports*, 3(1):63–71, 2016.

[17] K. J. Rothman, S. Greenland, and T. L. Lash. *Modern epidemiology*. Lippincott Williams & Wilkins, 2008.

[18] A. Skrondal. Interaction as departure from additivity in case-control studies: a cautionary note. *American Journal of Epidemiology*, 158(3):251–258, 2003.

[19] J. Splawa-Neyman, D. M. Dabrowska, T. P. Speed, et al. On the application of probability theory to agricultural experiments. *Statistical Science*, 5(4):465–472, 1990. Reprint.

[20] T. J. VanderWeele. Sufficient cause interactions and statistical interactions. *Epidemiology*, 20(1):6–13, 2009.

[21] T. J. VanderWeele. Reconsidering the denominator of the attributable proportion for interaction. *European Journal of Epidemiology*, 28(10):779–784, 2013.

[22] T. J. VanderWeele. *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press, 2016.

[23] T. J. VanderWeele and M. J. Knol. Remarks on antagonism. *American Journal of Epidemiology*, 173(10):1140–1147, 2011.

[24] T. J. VanderWeele and M. J. Knol. A tutorial on interaction. *Epidemiologic Methods*, 3(1):33–72, 2014.

[25] T. J. VanderWeele and T. S. Richardson. General theory for interactions in sufficient cause models with dichotomous exposures. *Annals of Statistics*, 40(4):2128, 2012.

[26] T. J. VanderWeele and J. M. Robins. The identification of synergism in the sufficient-component-cause framework. *Epidemiology*, 18(3):329–339, 2007.

[27] S. Wachholder. Binomial regression in GLIM: Estimating risk ratios and risk differences. *American Journal of Epidemiology*, 123(1):174–184, 1986.

[28] L. N. Yelland, A. B. Salter, and P. Ryan. Relative risk estimation in randomized controlled trials: a comparison of methods for independent observations. *The International Journal of Biostatistics*, 7(1):1–31, 2011.

[29] G. Y. Zou. On the estimation of additive interaction by use of the four-by-two table and beyond. *American Journal of Epidemiology*, 168(2):212–224, 2008.