Available online at www.sciencedirect.com
ScienceDirect
International Journal of Approximate Reasoning
48 (2008) 214–236
INTERNATIONAL JOURNAL OF APPROXIMATE REASONING
www.elsevier.com/locate/ijar

A generic qualitative characterization of independence of causal influence

M.A.J. van Gerven *, P.J.F. Lucas, Th.P. van der Weide

Institute for Computing and Information Sciences, Radboud University Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

Received 23 October 2006; received in revised form 20 August 2007; accepted 22 August 2007
Available online 19 September 2007

Abstract

Independence of causal influence (ICI) offer a high level starting point for the design of Bayesian networks. However, these models are not as widely applied as they could, as their behavior is often not well-understood. One approach is to employ qualitative probabilistic network theory in order to derive a qualitative characterization of ICI models. In this paper we analyze the qualitative properties of ICI models with binary random variables. Qualitative properties are shown to follow from the characteristics of the Boolean function underlying the model. In addition, it is demonstrated that the theory also allows finding constraints on the model parameters given knowledge of the qualitative properties. This high-level qualitative characterization offers a new way of identifying suitable ICI models and may facilitate their exploitation in developing real-world Bayesian networks.

Keywords: Independence of causal influence; Qualitative probabilistic networks; Bayesian networks; Knowledge acquisition

1. Introduction

Since the end of the 1980s, the theory of Bayesian networks has gained considerable attention in the field of artificial intelligence, as it offers a powerful framework for the representation of and reasoning with uncertainty [16]. On the one hand, the theory provides methods for the representation of the dependence and independence information associated with a domain, as well as methods for the representation of the underlying uncertainties. On the other hand, methods to reason with the specified dependencies, independencies and uncertainties are available. The (in)dependence information is specified by means of an acyclic directed graph, whereas the uncertainties are specified by means of a joint probability distribution that respects the independence information specified in the graph. The joint probability distribution is fully determined by a set of local probability distributions, usually in the form of conditional probability tables.

* Corresponding author.
E-mail address: marcelge@cs.ru.nl (M.A.J. van Gerven).

0888-613X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2007.08.012

In designing Bayesian networks, developers try to create acyclic directed graphs that are as sparse as possible, as the size of a conditional probability table is exponential in the number of associated variables. Creating sparse graphs not only saves space, but may also speed up probabilistic inference. Unfortunately, the creation of sparse graphs for a given problem may not always be possible. However, by imposing extra independence assumptions, supplemented by assumptions of functional dependence, it may be possible to reduce the number of conditional probabilities that need to be assessed. The theory of *independence of causal influence* (ICI), also known as *causal independence*, is especially suited for this purpose [17,15,9,4].

ICI theory adopts specific independence assumptions to model the interactions between a set of cause variables and an effect variable; using this approach, the number of parameters that need to be estimated decreases from exponential to linear in the number of variables. The noisy OR model, that expresses that the presence of one or more causes is sufficient to give rise to the occurrence of the effect, is an example of an ICI model that is widely used in practice [8,10,3]. It has been used in the QMR-DT system, which includes knowledge of approximately 600 diseases and approximately 4000 findings [19], the Promedas system, which aims to cover a large diagnostic repertoire of internal medicine [12], and in DIAVAL, an expert system for electrocardiography that uses a generalization of the noisy OR for non-binary random variables [5]. Another, frequently used ICI model is the noisy AND model; it expresses that all causes must be present in order to give rise to the effect. It has, for example, been used to model the joint effect of antibiotics on bacteria causing ventilator-associated pneumonia in patients [14].

The noisy OR and noisy AND models are special cases of ICI models based on Boolean functions since in principle any of the 2^{2^n} possible n-ary Boolean functions can be used to model deterministic interactions between cause and effect variables. Given the favorable properties of ICI models, it is unfortunate that only very few of these are used in practice: only the mentioned noisy OR and noisy AND are popular amongst developers. This is caused by the fact that it is often unclear with what behavior a particular ICI model is endowed when choosing a particular Boolean function. In [13] this problem was addressed by exploiting *qualitative probabilistic network* (QPN) theory to characterize the behavior of ICI models in terms of *influences* and *synergies* [20]. Such a qualitative characterization may then be matched with the behavior that is dictated by the domain, as suggested in Fig. 1. The qualitative pattern associated with a particular ICI model is termed a *qualitative causal pattern*.

The idea that QPN theory might be suitable for analyzing the behavior of ICI models was already recognized by Wellman, who states that: "...*prototypical patterns of systematic interaction might alleviate the burden of specifying qualitative synergies*" and "...*we should expect non-ambiguous synergy results from canonical models because any representation that specifies an n-way influence in terms of $O(n)$ parameters must employ some systematic assumption about interactions*" [20]. However, Lucas [13] offers the first systematic approach to analyzing ICI models in terms of QPN theory. This was done in particular for decomposable ICI models, i.e., ICI models which are characterized in terms of binary functions. There are 16 binary Boolean functions, which can be used to compose a subset of n-ary Boolean function, and which can be classified in terms of presence or absence of the properties of *associativity* and *commutativity*. The previously discussed noisy OR model is based on the Boolean OR, which is both commutative and associative. Although this offers an analysis of a useful subset of Boolean functions, a general characterization of the behavior of Boolean functions is not provided.

The present paper offers a substantial generalization of previously published results as it develops a general theory of qualitative causal patterns. The theory identifies:

(1) The qualitative behavior that holds for a given ICI model.
(2) Properties of ICI models that hold given a qualitative specification.

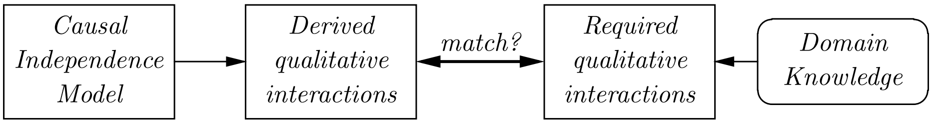

Fig. 1. Comparing the observed qualitative behavior of an ICI model with the desired qualitative behavior as specified by a domain expert.

The theory developed in this paper is useful in Bayesian network design, as it provides a tool for matching desired qualitative behavior of ICI models with the appropriate structural and quantitative parameters. Furthermore, a more widespread use of ICI models in Bayesian networks will facilitate the intelligibility of network behavior, allow the construction of denser networks, and ease the estimation of network parameters.

The structure of this paper is as follows. In Section 2 we review some necessary preliminaries, drawing upon Bayesian network, ICI and QPN theory. Subsequently, we study some general properties of ICI models in Section 3. These properties are then used to identify the qualitative behavior for different Boolean functions in Section 4. Finally, in Section 5 we round off with a discussion of the obtained results.

2. Preliminaries

In this section we will subsequently discuss Bayesian networks, ICI models, the running example of this paper, and QPN theory.

2.1. Bayesian networks

Bayesian networks, also called belief networks or probabilistic networks, were introduced in the 1980s as a framework for probabilistic inference [16]. A Bayesian network representation consists of an acyclic directed graph and an associated joint probability distribution that reflects the independence information specified in the graph.

We make use of the following graph-theoretical concepts in the definition of a Bayesian network. We define a *graph* as a pair $G = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ is a finite set of *vertices* and $\mathbf{E} \subseteq \mathbf{V} \times \mathbf{V}$ is a finite set of *edges*. If for all pairs $(v, v') \in \mathbf{E}$ it holds that $(v', v) \notin \mathbf{E}$, then we say that the graph is *directed*. If, additionally, there is no sequence $v_0, \ldots, v_n$ of distinct vertices such that $(v_{i-1}, v_i) \in \mathbf{E}$ and $v_0 = v_n$, then the graph is an *acyclic* directed graph, or ADG for short. The *parent set* of a vertex $v \in \mathbf{V}$ is defined as the set $\pi_G(v) = \{v' | (v', v) \in \mathbf{E}\}$.

Let P be a joint probability distribution of a set of random variables X and assume that there is a one-to-one correspondence between the vertices in $\mathbf{V}$ and the variables in X. We denote a random variable that corresponds with a vertex $v \in \mathbf{V}$ by X_v and a set of random variables that corresponds with a set of vertices $\mathbf{W} \subseteq \mathbf{V}$ by $X_{\mathbf{W}}$. A *Bayesian network* $\mathscr{B}$ is then defined as a pair $\mathscr{B} = (G, P)$, where $G = (\mathbf{V}, \mathbf{E})$ is an ADG and P is a joint probability distribution that factorizes as

$$P(X_{\mathbf{V}}) = \prod_{v \in \mathbf{V}} P(X_v | X_{\pi_G(v)}).$$

In the following, to simplify notation, we will use vertices $\mathbf{V}$ and random variables in $X_{\mathbf{V}}$ interchangeably, where the interpretation will be clear from the context. In this paper, it is assumed that all random variables are binary. We will use x to denote $X = \top$ (logical truth) and $\bar{x}$ to denote $X = \perp$ (logical falsehood). If the value of variable X is either true of false, but unspecified, then this is indicated by $X = \hat{x}$, or simply by $\hat{x}$.

2.2. ICI models

Independence of causal influence is the notion that causes are independently contributing to the occurrence of an effect through some pattern of interaction, represented as a set of local conditional probability distribu-

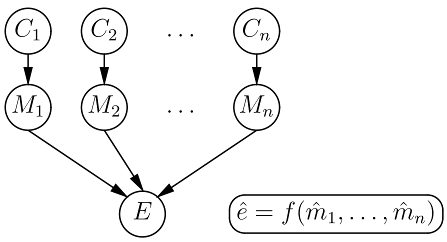

Fig. 2. ICI model.

tions of a Bayesian network [9]. The associated Bayesian network structure is depicted in Fig. 2, where variables C_k indicate cause variables, M_k intermediate variables and E is an effect variable. Let $\mathbb{B} = \{\bot, \top\}$. We use $\mathbf{c} \in \mathbb{B}^n$, possibly with a subscript, to denote an element of $\mathbb{B}^n$ for vectors $\mathbf{C} = (C_1, \ldots, C_n)$; similarly, we use $\mathbf{m} \in \mathbb{B}^n$ for elements of $\mathbf{M} = (M_1, \ldots, M_n)$. These are called *configurations*. To reduce the use of numeric indices, we associate with each cause variable C an intermediate variable M_C. Independence of causal influence is captured by the requirement that an intermediate variable $M_C \in \mathbf{M}$ is dependent of cause variable C and independent of the other cause variables $\mathbf{C} \setminus \{C\}$. According to the independence structure shown in Fig. 2, it holds that:

$$P(e|\mathbf{c}) = \sum_{\mathbf{m}} P(e|\mathbf{m}) P(\mathbf{m}|\mathbf{c}) = \sum_{\mathbf{m}} P(e|\mathbf{m}) \prod_{i=1}^{n} P(\hat{m}_i|\hat{c}_i). \tag{1}$$

An intermediate variable M_C can be interpreted as modulating the contribution of a cause C to the effect E and often specific assumptions are made about this contribution. In this paper, we assume that both *consequentiality* and *accountability* hold. Consequentiality states that the truth of a cause variable increases our belief that the associated intermediate variable is true as well. Formally, we require that $P(m_C|c) > P(m_C|\bar{c})$. Accountability states that the truth of an intermediate variable must imply the truth of its associated cause variable; formally, $P(m_C|\bar{c}) = 0$. The conditional probability distribution $P(E|\mathbf{M})$ used in Eq. (1) is assumed to be deterministic in ICI models, and, thus, can be taken as representing a function $f : \mathbb{B}^n \to \mathbb{B}$, such that $P(e|\mathbf{m}) = 1$ if $f(\mathbf{m}) = e$ and $P(e|\mathbf{m}) = 0$ otherwise. An ICI model is now defined formally as follows:

Definition 1 (*ICI model*). An *ICI model* $\mathscr{C}$ is a tuple $(\mathbf{C}, \mathbf{M}, E, f, \mathbf{P})$, where $\mathbf{C}$ is a set of *cause variables*, $\mathbf{M}$ is a set of *intermediate variables*, E is an *effect variable*, f is an *interaction function* and $\mathbf{P}$ is a set of parameters $\{P(M_C|C) | C \in \mathbf{C}\}$, with $M_C \in \mathbf{M}$, for each $C \in \mathbf{C}$ and vice versa, such that

$$P(e|\mathbf{c}) = \sum_{f(\mathbf{m})=e} \prod_{i=1}^{n} P(\hat{m}_i|\hat{c}_i). \tag{2}$$

By $f(\mathbf{m}) = e$ is denoted the situation where both $f(\mathbf{m}) = e$ and $E = e$ hold. The probability $P(m_C|c)$ will often be abbreviated to $P(m|c)$. In the literature different interpretations of independence of causal influence exist, often taking the form of restrictions on an interaction function f that underlies the model [2,9]. Here, we assume that an interaction function can be any Boolean function $f : \mathbb{B}^n \to \mathbb{B}$.

An ICI model $\mathscr{C} = (\mathbf{C}, \mathbf{M}, E, f, \mathbf{P})$ can act as the basis for the specification of a Bayesian network $\mathscr{B} = (G, P)$, with ADG $G = (\mathbf{V}, \mathbf{E})$, as depicted in Fig. 2, and joint probability distribution P, where G respects all the dependences represented by the joint probability distribution P. The vertices in G are given by

$$\mathbf{V} = \mathbf{C} \cup \mathbf{M} \cup \{E\}$$

such that the sets $\mathbf{C}, \mathbf{M}$ and $\{E\}$ are disjoint, and the arcs in G are given by

$$\mathbf{E} = \{(C, M_C) | C \in \mathbf{C}\} \cup \{(M, E) | M \in \mathbf{M}\}.$$

In addition to the parameters $P(M_C|C)$ and the interaction function f, we also need to specify a prior joint probability distribution $P(\mathbf{C})$ to obtain a complete specification of the Bayesian network $\mathscr{B}$.

In the sequel, we will often use the notation $\mathrm{P}[f]$ to refer to the probability distribution $P(E|\mathbf{c})$. We can alternatively write Eq. (2) in somewhat generalized form as

$$\mathrm{P}[f](e|\mathbf{c}) = \sum_{\mathbf{m}} f(\mathbf{m}) P(\mathbf{m}|\mathbf{c}) = \sum_{\mathbf{m}} f(\mathbf{m}) \prod_{i=1}^{n} P(\hat{m}_i|\hat{c}_i), \tag{3}$$

where we make use of the analogy between Boolean algebra and ordinary arithmetic by interpreting $\bot$ as 0 and $\top$ as 1, i.e., if $f(\mathbf{m}) = \bot$ this is interpreted as $f(\mathbf{m}) = 0$, and as $f(\mathbf{m}) = 1$ otherwise [1]. We will sometimes employ functions f that are not Boolean; even then Eq. (3) still applies, where $\mathrm{P}[f](e|\mathbf{c})$ can be interpreted as the conditional expectation of f given $\mathbf{c}$. If f is a constant and there are no cause variables $\mathbf{C}$ then $\mathrm{P}[f] = f$.

As an example of a realistic ICI model that will be used to illustrate the theory developed in this paper, consider the ICI model shown in Fig. 3 that represents a piece of medical knowledge with respect to the prog-

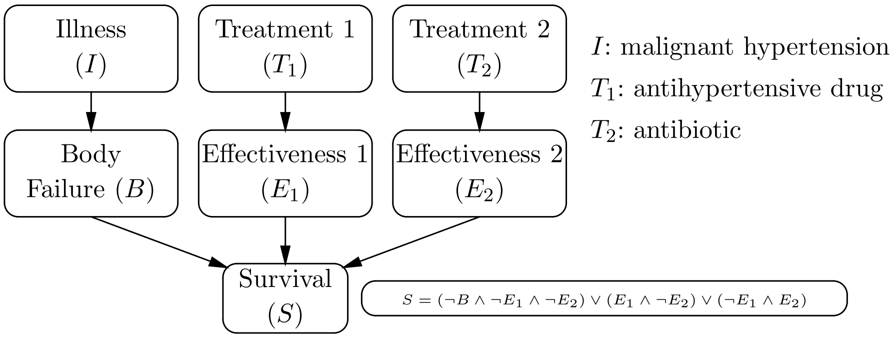

Fig. 3. A prognostic model of survival in serious illness, modeling the interaction between two drugs, expressed as an ICI model.

nosis of a *serious illness* (I), such as malignant hypertension due to chronic kidney infection, infectious hypertension for short, which is handled by two alternative treatments T_1, an antihypertensive drug, and T_2, rifampin (an antibiotic).[1] The seriousness of the infectious hypertension is reflected by the fact that we are interested in the *survival* (S) (e.g., within the next 5 years) of a patient with this illness. The resulting ICI model is shown in Fig. 3. The variable B stands for *body failure* due to the illness, E_1 stands for the *effectiveness* of treatment T_1 and E_2 for the effectiveness of treatment T_2. If body failure occurs and the disease cause is eradicated, it is assumed that the patient will survive. However, if both treatments T_1 and T_2 are effective then the patient will not survive due to the synergistic interaction between the two treatments (rifampin in conjunction with the antihypertensive drug). This can be expressed by means of a Boolean function f defined by the following Boolean expression:

$$S = (\neg B \wedge \neg E_1 \wedge \neg E_2) \vee (E_1 \wedge \neg E_2) \vee (\neg E_1 \wedge E_2) \tag{4}$$

(survival is equivalent to the absence of body failure or eradication of the disease due either treatment T_1 or T_2, but not both). In the sequel, we will use Boolean functions and Boolean expressions interchangeably. The qualitative behavior that arises from this choice should then be in accordance with the domain knowledge as stated above.

According to what has been said above, the Bayesian network model is an example of an ICI model. It will be called the *prognostic model* in the following. Here, the variables I, T_1 and T_2 act as cause variables and B, E_1 and E_2 are the intermediate variables. For example, we have that $B = M_I$.

There are two main tasks in building an ICI model. The first is to determine the underlying interaction function f, in the example a Boolean function that is assumed to model the interaction between the factors Body Failure (B), Effectiveness 1 (E_1) and Effectiveness 2 (E_2) with respect to Survival (S), where S is the effect variable. The second task is to estimate the parameters $P(B|I)$, $P(E_1|T_1)$ and $P(E_2|T_2)$. Notice that just three conditional probabilities need to be estimated, as $P(m_C|\bar{c})$ is assumed to be zero for each cause variable C. Examples of ICI models that model other real-world problems and employ alternative interaction functions can be found in [13].

2.3. Qualitative probabilistic networks

Recall that the aim of the research underlying this paper is to develop a theory that is able to assist Bayesian network developers in quantifying Bayesian networks using qualitative knowledge from a problem domain. Qualitative probabilistic networks are at the core of this theory. We will therefore briefly summarize the theory of qualitative probabilistic networks.

Qualitative probabilistic networks (QPNs) were introduced by Wellman as a qualitative abstraction of ordinary Bayesian networks [20]. The relationships between variables are described by the concepts of *influ-*

[1] The choice of these drugs was inspired by the death of Slobodan Miloševic. It is hypothesized that his death was due to the combined effect of an antihypertensive drug, which was meant to reduce the height of his blood pressure, and the antibiotic rifampin, which counteracted the effect of the antihypertensive drug. Here, we abstract away from the actual course of events.

ences and *synergies*. In the following, let (G, P) be a Bayesian network, let $A, B, C \in \mathbf{X}$ be binary random variables and let (A, C) and (B, C) be arcs in G.

A qualitative influence expresses how the value of one variable influences the probability of observing values for another variable.

Definition 2 (*Qualitative influence*). Let $\mathbf{X}$ denote $\pi_G(C) \setminus \{A\}$. We say that there is a positive qualitative influence of A on C, written as $\delta_{A \to C} = +$, if

$$\delta_{A \to C}(\mathbf{x}) = P(c|a, \mathbf{x}) - P(c|\bar{a}, \mathbf{x}) \geqslant 0,$$

regardless of the configuration $\mathbf{x}$, with a strict inequality for at least one configuration $\mathbf{x}$. *Negative* ($\delta_{A \to C} = -$) and *zero* qualitative influences ($\delta_{A \to C} = 0$) are defined analogously, replacing $\geqslant$ by $\leqslant$ and $=$ respectively. If there are values $\mathbf{x}$ and $\mathbf{x}'$, such that

$$P(c|a, \mathbf{x}) - P(c|\bar{a}, \mathbf{x}) > 0 \quad \text{and} \quad P(c|a, \mathbf{x}') - P(c|\bar{a}, \mathbf{x}') < 0,$$

then we say that the qualitative influence is *non-monotonic*, denoted by $\delta_{A \to C} = \sim$. If none of these cases hold, i.e., when there is incomplete information about the probability distribution, then we say that the qualitative influence is *ambiguous*, written as $\delta_{A \to C} = ?$.

Example 3. In order to illustrate the qualitative concepts we assume for the moment that the exact probabilities associated with the prognostic model are known. We assume $P(b|i) = 0.9$, $P(e_1|t_1) = 0.3$ and $P(e_2|t_2) = 0.6$. Hence, it is very likely that a serious illness gives rise to body failure, as it occurs in 90% of cases, treatment T_1 is effective in 30% of the patients and treatment T_2 is effective in 60% of the patients. What then, we might ask, is the qualitative influence of a serious illness on the survival? This is computed as follows, where the Boolean function f is defined by the Boolean expression (4):

$$\delta_{I \to S}(\{\hat{t}_1, \hat{t}_2\}) = \mathrm{P}[f](s|i, \hat{t}_1, \hat{t}_2) - \mathrm{P}[f](s|\bar{\imath}, \hat{t}_1, \hat{t}_2) = P(\bar{e}_1|\hat{t}_1)P(\bar{e}_2|\hat{t}_2)(P(\bar{b}|i) - P(\bar{b}|\bar{\imath})).$$

It follows that $\delta_{I \to S}(\{t_1, t_2\}) = -0.252$, $\delta_{I \to S}(\{\bar{t}_1, t_2\}) = -0.36$, $\delta_{I \to S}(\{t_1, \bar{t}_2\}) = -0.63$ and $\delta_{I \to S}(\{\bar{t}_1, \bar{t}_2\}) = -0.9$. In accordance with our expectations, serious illness appears to have a negative influence on survival.

An additive synergy expresses how the interaction between two variables influences the probability of observing values for a third variable.

Definition 4 (*Additive synergy*). Let $\mathbf{X}$ denote $\pi_G(C) \setminus \{A, B\}$. We say that there is a *positive additive synergy* of A and B on C, written as $\delta_{(A,B) \to C} = +$, if

$$\delta_{(A,B) \to C}(\mathbf{x}) = P(c|a, b, \mathbf{x}) + P(c|\bar{a}, \bar{b}, \mathbf{x}) - P(c|\bar{a}, b, \mathbf{x}) - P(c|a, \bar{b}, \mathbf{x}) \geqslant 0,$$

regardless of the configuration $\mathbf{x}$, with a strict inequality for at least one configuration $\mathbf{x}$. Negative, zero, non-monotonic and ambiguous additive synergies are defined analogous to qualitative influences.

Example 5. With regard to the prognostic model, we might be interested in the additive synergy between serious illness and treatment T_1 with respect to survival. This is computed as follows, where again we employ Boolean expression (4):

$$\begin{aligned}\delta_{(I,T_1) \to S}(\{\hat{t}_2\}) &= \mathrm{P}[f](s|i, t_1, \hat{t}_2) + \mathrm{P}[f](s|\bar{\imath}, \bar{t}_1, \hat{t}_2) - \mathrm{P}[f](s|\bar{\imath}, t_1, \hat{t}_2) - \mathrm{P}[f](s|i, \bar{t}_1, \hat{t}_2) \\ &= P(\bar{e}_2|\hat{t}_2)(P(\bar{b}|i) - 1)(P(\bar{e}_1|t_1) - 1).\end{aligned}$$

It follows that $\delta_{(I,T_1) \to S}(\{t_2\}) = 0.108$ and $\delta_{(I,T_1) \to S}(\{\bar{t}_2\}) = 0.27$ such that illness I and treatment T_1 have a positive additive synergy with respect to survival.

A product synergy expresses how upon observation of a common child of two vertices, observing the value of one parent vertex influences the probability of observing a value for the other parent vertex. The original definition of a product synergy is as follows [11].

Definition 6 (*Product synergy*). Let $\mathbf{X}$ denote $\pi_G(C) \setminus \{A, B\}$. We say that there is a *positive product synergy* of A and B with regard to the value $\hat{c}$ of variable C, written as $\delta^{\hat{c}}_{(A,B)\to C} = +$, if

$$\delta^{\hat{c}}_{(A,B)\to C}(\mathbf{x}) = P(\hat{c}|a, b, \mathbf{x})P(\hat{c}|\bar{a}, \bar{b}, \mathbf{x}) - P(\hat{c}|\bar{a}, b, \mathbf{x})P(\hat{c}|a, \bar{b}, \mathbf{x}) \geqslant 0,$$

regardless of the configuration $\mathbf{x}$, with a strict inequality for at least one configuration $\mathbf{x}$. It is assumed that the value $\hat{c}$ of variable C is either true or false. Negative, zero, non-monotonic and ambiguous product synergies are again defined analogous to the corresponding types of qualitative influences.

Example 7. With regard to the prognostic model, the product synergy between treatments T_1 and T_2 in the case of survival, is computed as follows:

$$\delta^{s}_{(T_1,T_2)\to S}(\{\hat{\imath}\}) = \mathrm{P}[f](s|\hat{\imath}, t_1, t_2) \cdot \mathrm{P}[f](s|\hat{\imath}, \bar{t}, \bar{t}_2) - \mathrm{P}[f](s|\hat{\imath}, \bar{t}_1, t_2) \cdot \mathrm{P}[f](s|\hat{\imath}, t_1, \bar{t}_2) = -P(e_1|t_1)P(e_2|t_2).$$

It follows that $\delta^{s}_{(T_1,T_2)\to S}(\{\bar{\imath}\}) = \delta^{s}_{(T_1,T_2)\to S}(\{i\}) = 0.18$ such that treatments T_1 and T_2 have a positive product synergy with respect to survival. This positive product synergy arises due to the fact that in the case of survival of a patient, it is more likely that one of both treatments is given. The presence of both T_1 and T_2 and the absence of both T_1 and T_2 will lead to patient death.

The following lemma states that for binary random variables, the product synergy when $C = \bot$ is partially determined by the associated additive synergy.

Lemma 8. *For binary random variables, the product synergy when $C = \bot$ is determined by the product synergy when $C = \top$ and the additive synergy through the following equality*:

$$\delta^{\bar{c}}_{(A,B)\to C}(\mathbf{x}) = \delta^{c}_{(A,B)\to C}(\mathbf{x}) - \delta_{(A,B)\to C}(\mathbf{x}).$$

Proof

$$\begin{aligned}\delta^{\bar{c}}_{(A,B)\to C}(\mathbf{x}) &= P(\bar{c}|\bar{a}, \bar{b}, \mathbf{x})P(\bar{c}|a, b, \mathbf{x}) - P(\bar{c}|a, \bar{b}, \mathbf{x})P(\bar{c}|\bar{a}, b, \mathbf{x}) \\ &= (1 - P(c|\bar{a}, \bar{b}, \mathbf{x}))(1 - P(c|a, b, \mathbf{x})) - (1 - P(c|a, \bar{b}, \mathbf{x}))(1 - P(c|\bar{a}, b, \mathbf{x})) \\ &= (P(c|\bar{a}, \bar{b}, \mathbf{x})P(c|a, b, \mathbf{x}) - P(c|a, \bar{b}, \mathbf{x})P(c|\bar{a}, b, \mathbf{x})) - (P(c|\bar{a}, \bar{b}, \mathbf{x}) + P(c|a, b, \mathbf{x}) \\ &\quad - P(c|a, \bar{b}, \mathbf{x}) - P(c|\bar{a}, b, \mathbf{x})) = \delta^{c}_{(A,B)\to C}(\mathbf{x}) - \delta_{(A,B)\to C}(\mathbf{x}). \qquad \square\end{aligned}$$

Modifications to the definition of a product synergy have been made after the observation that Definition 6 is incomplete when parent vertices in $\mathbf{X}$ are uninstantiated [7,6]. In other words

$$\begin{aligned}&\forall \mathbf{x}[P(\hat{c}|a, b, \mathbf{x})P(\hat{c}|\bar{a}, \bar{b}, \mathbf{x}) - P(\hat{c}|a, \bar{b}, \mathbf{x})P(\hat{c}|\bar{a}, b, \mathbf{x}) \leqslant 0 \\ &\quad \not\Rightarrow P(\hat{c}|a, b)P(\hat{c}|\bar{a}, \bar{b}) - P(\hat{c}|a, \bar{b})P(\hat{c}|\bar{a}, b) \leqslant 0].\end{aligned}$$

This so-called type II product synergy can be formalized in terms of the more intuitive notion of an *intercausal influence* [18].

Definition 9 (*Intercausal influence*). Let $\mathbf{X}$ denote $\pi_G(B) \cup \pi_G(C) \setminus \{A\}$. Then a variable A exhibits a *positive intercausal influence* on B with regard to the value $\hat{c}$ if

$$P(b|a, \hat{c}, \mathbf{x}) - P(b|\bar{a}, \hat{c}, \mathbf{x}) \geqslant 0,$$

regardless of the configuration $\mathbf{x}$. Negative, zero, non-monotonic and ambiguous intercausal influences are again defined analogous to the corresponding types of qualitative influences.

For ICI models, intercausal influences describe the dependence between two causes C and C' when the value of the effect variable is observed. We therefore compute $P(c'|c, \hat{e}, \mathbf{c}_2) - P(c'|\bar{c}, \hat{e}, \mathbf{c}_2)$ for all values $\mathbf{c}_2$ of the causes $\mathbf{C}_2 = \mathbf{C} \setminus \{C, C'\}$. Using Bayes' rule we obtain the equal expression:

$$\frac{P(\hat{e}|c, c', \mathbf{c}_2)P(c'|c, \mathbf{c}_2)}{P(\hat{e}|c, \mathbf{c}_2)} - \frac{P(\hat{e}|\bar{c}, c', \mathbf{c}_2)P(c'|\bar{c}, \mathbf{c}_2)}{P(\hat{e}|\bar{c}, \mathbf{c}_2)}. \tag{5}$$

Note that $P(c'|c, \mathbf{c}_2) = P(c'|\bar{c}, \mathbf{c}_2) = P(c')$, as cause variables are independent. This leads to the following expression, whose sign equals that of Formula (5):

$$P(\hat{e}|\bar{c}, \mathbf{c}_2)P(\hat{e}|c, c', \mathbf{c}_2) - P(\hat{e}|c, \mathbf{c}_2)P(\hat{e}|\bar{c}, c', \mathbf{c}_2).$$

By rewriting $P(\hat{e}|\bar{c}, \mathbf{c}_2)$ as $P(\hat{e}|\bar{c}, c', \mathbf{c}_2)P(c') + P(\hat{e}|\bar{c}, \bar{c}', \mathbf{c}_2)P(\bar{c}')$ and $P(\hat{e}|c, \mathbf{c}_2)$ as $P(\hat{e}|c, c', \mathbf{c}_2)P(c')+$ $P(\hat{e}|c, \bar{c}', \mathbf{c}_2)P(\bar{c}')$, we obtain

$$P(\hat{e}|c, c', \mathbf{c}_2)P(\hat{e}|\bar{c}, \bar{c}', \mathbf{c}_2) - P(\hat{e}|\bar{c}, c', \mathbf{c}_2)P(\hat{e}|c, \bar{c}', \mathbf{c}_2),$$

which is the definition of the product synergy, specialized to ICI models. Hence, for ICI models over binary variables the product synergy and intercausal influences are equivalent.

So far, we have assumed that the parameters $P(m_C|c)$ are known when qualitative properties are computed. However, the goal of this paper is to qualitatively characterize ICI models with varying interaction functions. Therefore, we abstract away from the parameters and derive the qualitative properties solely by taking into account the properties of a ICI model's interaction function. In the next section, we infer some general properties of ICI models.

3. Properties of ICI models

In this section, we will investigate general properties of the probability distribution $\mathrm{P}[f]$, where it is assumed that f is a Boolean function.

3.1. General properties

Lemma 10 states that $\mathrm{P}[f]$ is bounded by $f = \bot$ and $f = \top$, which is a basic result due to the first axiom of probability theory.

Lemma 10. $0 = \mathrm{P}[\bot] \leqslant \mathrm{P}[f] \leqslant \mathrm{P}[\top] = 1.$

Lemmas 11–13 show how ICI models may be decomposed by decomposing the Boolean function. We will make use of the analogy between Boolean algebra and ordinary arithmetic by interpreting $\bot$ as 0 and $\top$ as 1 in an arithmetic context [1], in order to allow for a compact notation.

Lemma 11. $\mathrm{P}[\neg f] = 1 - \mathrm{P}[f].$

Proof

$$\mathrm{P}[\neg f](e|\mathbf{c}) = \sum_{\mathbf{m}}(1 - f(\mathbf{m}))P(\mathbf{m}|\mathbf{c}) = \sum_{\mathbf{m}} P(\mathbf{m}|\mathbf{c}) - \sum_{\mathbf{m}} f(\mathbf{m})P(\mathbf{m}|\mathbf{c}) = 1 - \mathrm{P}[f](e|\mathbf{c}). \qquad \square$$

Lemma 12. $\mathrm{P}[f \vee f'] = 1 - \mathrm{P}[\neg f \wedge \neg f'] = \mathrm{P}[f] + \mathrm{P}[f'] - \mathrm{P}[f \wedge f'].$

Proof. As $f \vee f' = \neg(\neg f \wedge \neg f')$, we obtain $\mathrm{P}[\neg(\neg f \wedge \neg f')](e|\mathbf{c}) = 1 - \mathrm{P}[\neg f \wedge \neg f'](e|\mathbf{c})$. It also holds that $\mathrm{P}[f \vee f'] = \mathrm{P}[f \wedge f'] + \mathrm{P}[f \wedge \neg f'] + \mathrm{P}[\neg f \wedge f']$. Since $\mathrm{P}[f] = \mathrm{P}[f \wedge f'] + \mathrm{P}[f \wedge \neg f']$ and $\mathrm{P}[f'] = \mathrm{P}[f' \wedge f]+$ $\mathrm{P}[f' \wedge \neg f]$ we obtain the second equality. $\square$

Lemma 13. *If* $f \wedge f' = \bot$ *then* $\mathrm{P}[f \vee f'] = \mathrm{P}[f \o f'] = \mathrm{P}[f] + \mathrm{P}[f']$, where ø *represents the exclusive OR.*

Proof. The operator $f \o f'$ is defined as $(f \wedge \neg f') \vee (\neg f \wedge f')$. As this is equivalent to $(f \vee f') \wedge \neg(f \wedge f')$, it follows straight from Lemma 12 that if $f \wedge f' = \bot$ then $\mathrm{P}[f \vee f'] = \mathrm{P}[f] + \mathrm{P}[f']$. $\square$

Sometimes, we will add two Boolean functions or compute the difference between two Boolean functions within an ICI model. In that case, Lemma 10 does not hold and the expression is not a proper probability distribution anymore, but *can* be interpreted as a conditional expectation. The following lemma follows directly from the linearity property of conditional expectation.

Lemma 14. $\mathrm{P}[af + bf'] = a\mathrm{P}[f] + b\mathrm{P}[f']$ *for constants* a and b.

$\mathrm{P}[f]$ can be bounded from below and above through the following inequalities.

Corollary 15. $\mathrm{P}[f \wedge f'] \leqslant \mathrm{P}[f] \leqslant \mathrm{P}[f \vee f'] \leqslant \mathrm{P}[f] + \mathrm{P}[f']$.

Proof

$$\begin{aligned}\mathrm{P}[f \wedge f'](e|\mathbf{c}) &= \sum_{\mathbf{m}} f(\mathbf{m})f'(\mathbf{m})P(\mathbf{m}|\mathbf{c}) \leqslant \sum_{\mathbf{m}} f(\mathbf{m})P(\mathbf{m}|\mathbf{c}) = \mathrm{P}[f](e|\mathbf{c}) \\ &\leqslant \sum_{\mathbf{m}}(f(\mathbf{m}) + f'(\mathbf{m}) - f(\mathbf{m})f'(\mathbf{m}))P(\mathbf{m}|\mathbf{c}) = \mathrm{P}[f \vee f'](e|\mathbf{c}) \leqslant \sum_{\mathbf{m}}(f(\mathbf{m}) + f'(\mathbf{m}))P(\mathbf{m}|\mathbf{c}) \\ &= \mathrm{P}[f](e|\mathbf{c}) + \mathrm{P}[f']](e|\mathbf{c}). \qquad \square\end{aligned}$$

3.2. Analytical tools

Next, we introduce a number of analytical tools that will be used in the subsequent sections.

Definition 16 (*Curry*). Let $f : \mathbb{B}^n \to \mathbb{B}$ be a Boolean function. Then, the *curry* of f, denoted by $f_{X_j=\hat{x}_j}$, is defined as the function $f_{X_j=\hat{x}_j} : \mathbb{B}^{n-1} \to \mathbb{B}$, such that $f_{X_j=\hat{x}_j}(\hat{x}_1, \ldots, \hat{x}_{j-1}, \hat{x}_{j+1}, \ldots, \hat{x}_n) = f(\hat{x}_1, \ldots, \hat{x}_{j-1}, \hat{x}_j, \hat{x}_{j+1}, \ldots, \hat{x}_n)$.

Central to the analysis is the notion of a partial order $\leqslant$ on configurations of $\mathbf{C}$ and $\mathbf{M}$.

Definition 17 (*Ordered Boolean n-tuples*). Let $\mathbf{m} = (\hat{m}_1, \ldots, \hat{m}_n), \mathbf{c} = (\hat{c}_1, \ldots, \hat{c}_n) \in \mathbb{B}^n$ be Boolean n-tuples. It holds that $\mathbf{m} \leqslant \mathbf{c}$ iff $\hat{m}_i \leqslant \hat{c}_i$ for all i, $1 \leqslant i \leqslant n$, where $\perp < \top$. The relation $\mathbf{m} < \mathbf{c}$ holds iff $\mathbf{m} \leqslant \mathbf{c}$ and $\mathbf{m} \neq \mathbf{c}$ and the relation $>$ is defined analogously.

Note that for any two tuples $\mathbf{m}$ and $\mathbf{c}$ it holds that either $\mathbf{m} < \mathbf{c}$, $\mathbf{m} > \mathbf{c}$, $\mathbf{m} = \mathbf{c}$ or $\exists C[\hat{m}_C < \hat{c}] \wedge \exists C'[\hat{m}_{C'} > \hat{c}']$. If the latter holds then we say that $\mathbf{m}$ and $\mathbf{c}$ are *incomparable*. By means of this ordering we are in a position to compare configurations $\mathbf{m}$ of the intermediate variables $\mathbf{M}$ with configurations $\mathbf{c}$ of the cause variables $\mathbf{C}$. In other words, we can compare intermediate states with causal states and prove the following lemmas.

Lemma 18. $\mathbf{m} = \mathbf{c} \Rightarrow P(\mathbf{m}|\mathbf{c}) > 0$.

Proof. If $\mathbf{m} = \mathbf{c}$, then

$$P(\mathbf{m}|\mathbf{c}) = \prod_{C \in \mathbf{C}} P(m_C|c)^{\hat{c}} P(\bar{m}_C|\bar{c})^{1-\hat{c}} = \prod_{C \in \mathbf{C}} P(m_C|c)^{\hat{c}} > 0$$

due to the assumptions that $P(m_C|c) > 0$ and $P(m_C|\bar{c}) = 0$. $\square$

Lemma 18 states that the probability that an intermediate state is equal to the causal state is always larger than zero. Hence, the causal state always conveys information about the actual state of the intermediate variables.

Lemma 19. $P(\mathbf{m}|\mathbf{c}) > 0 \Rightarrow \mathbf{c} \geqslant \mathbf{m}$.

Proof. If $\mathbf{c} \ngeqslant \mathbf{m}$ then there is some cause variable $C = \perp$ and $M_C = \top$. Since $P(m_C|\bar{c}) = 0$ it holds that $P(\mathbf{m}_C|\mathbf{c}) = 0$. $\square$

Lemma 19 follows from the notion of accountability and states that the truth of an intermediate variable always implies the truth of its associated cause variable. It is an important lemma, as it essentially shows that we can ignore all configurations $\mathbf{m}$ that are not smaller than or equal, or incomparable, to a given configuration $\mathbf{c}$.

The following lemmas demonstrate how a choice of the parameters influences the value of $P(\mathbf{m}|\mathbf{c})$.

Lemma 20. $\forall C[P(m_C|c) = 1] \Rightarrow \forall \mathbf{m} \neq \mathbf{c}[P(\mathbf{m}|\mathbf{c}) = 0]$ *for arbitrary* $\mathbf{c}$.

Proof. Choose $P(m_C|c) = 1$ for each $C \in \mathbf{C}$. If $\mathbf{m} = \mathbf{c}$, then $P(\mathbf{m}|\mathbf{c}) = 1$, and necessarily $P(\mathbf{m}|\mathbf{c}) = 0$ for $\mathbf{m} \neq \mathbf{c}$. $\square$

Lemma 20 states that if the causal relationship between the causes C the intermediates M_C is deterministic, it is not allowed that the values of causes and intermediate variables differ, which is as expected.

Lemma 21. $\forall C[P(m_C|c) < 1] \Rightarrow \forall \mathbf{m} \leqslant \mathbf{c}[P(\mathbf{m}|\mathbf{c}) > 0]$ *for arbitrary* $\mathbf{c}$.

Proof. Since $\mathbf{m} \leqslant \mathbf{c}$ we have that for each cause variable C such that $M_C = \top$ also $C = \top$ and for each C such that $M_C = \bot$ it is the case that either $C = \bot$ or $C = \top$. Therefore, we may write

$$P(\mathbf{m}|\mathbf{c}) = \prod_{C \in \mathbf{C}} P(m_C|c)^{\hat{m}_C} P(\bar{m}_C|c)^{(1-\hat{m}_C)\hat{c}_C},$$

since $P(\bar{m}_C|\bar{c}) = 1$ by assumption. Since $0 < P(m_C|c) < 1$ by assumption, we have $P(m_C|c) > 0$ and $P(\bar{m}_C|c) > 0$, which proves the proposition. $\square$

Lemma 21 states that if there is an uncertain causal relationship between every cause C and its associated intermediate variable M_C, then it follows that each intermediate state whose true variables form a subset of the true cause variables, has a non-zero probability of occurring.

As the qualitative behavior of an ICI model is completely determined by its interaction function, in the following we will frequently investigate how these functions behave. This analysis will frequently go beyond pure Boolean functions, as some of the interaction patterns are the result of adding and subtracting Boolean functions. Considerable insight into the interaction patterns is obtained by looking at the function values (positive, negative or zero) of the resulting function for configurations smaller than a given configuration. For this, introduction of a special notation will be convenient, as given in the following definition:

Definition 22 (*Initial non-negative, non-positive function values*). Let $q : \mathbb{B}^m \to W$ be a function, where $W = \{-b, \ldots, 0, \ldots, b\} \subset \mathbb{Z}$, then q is said to have *initial non-negative function values*, denoted by V_q^+, if

$$\exists \mathbf{m}[[q(\mathbf{m}) \in \{1, \ldots, b\}] \wedge \forall \mathbf{m}' < \mathbf{m}[q(\mathbf{m}') \in \{0, \ldots, b\}].$$

Similarly, q is said to have *initial non-positive function values*, denoted by V_q^-, if V_{-q}^+ holds.

Thus, V_q^+ means that the function value of q is *positive* for some value $\mathbf{m}$, and takes non-negative values for any value $\mathbf{m}'$ lower in the ordering $<$. The meaning of V_q^- is analogous.

As an example, consider a function q that indicates quality of life, where the variables 'happiness' and 'beauty', abbreviated to H and B, are used as summary variables. It is defined as follows. With $q(h, b) = 1$ is indicated maximal quality of life; for all $(\hat{h}, \hat{b}) < (h, b)$, for example $(\bar{h}, b) < (h, b)$, unsatisfactory quality of life is quantified by $q(\hat{h}, \hat{b}) = 0$. Thus, for this quality of life function V_q^+ holds whereas V_q^- does not. The properties V_q^+ and V_q^- of a function q will be important tools for the qualitative analysis of ICI models.

4. Qualitative properties of ICI Models

In this section, it is assumed that a Boolean interaction function underlying an ICI model is given; we then identify the signs of qualitative influences (Section 4.1), additive synergies (Section 4.2) and product synergies (Section 4.3). These results can also be used to identify Boolean functions that respect a particular qualitative characterization.

Note that we can assume that the causes are direct parents of E as the intermediate variables are marginalized out of the final computation of $\mathrm{P}[f](e|\mathbf{c})$ (cf. Eq. (2)). For our analysis, we assume some fixed ICI model over a set $\mathbf{C}$ of n cause variables, in which we focus on the interaction between different cause variables C and C' and the effect variable E, where we abbreviate M_C by M and $M_{C'}$ by M'. Throughout this paper we will use $\mathbf{M}_1$ to denote $\mathbf{M} \setminus \{M\}$ and $\mathbf{M}_2$ to denote $\mathbf{M} \setminus \{M, M'\}$. Likewise, we will use $\mathbf{C}_1$ to denote $\mathbf{C} \setminus \{C\}$ and $\mathbf{C}_2$ to denote $\mathbf{C} \setminus \{C, C'\}$.

4.1. Qualitative influences

Let $\delta_{C\to E}[f]$ denote $\delta_{C\to E}$ where f is the interaction function of the corresponding ICI model. A qualitative influence $\delta_{C\to E}[f]$ between a cause C and effect E denotes how the observation of C influences the observation of the effect e. The sign of a qualitative influence for an ICI model mediated by a function f is then determined by the sign of

$$\delta_{C\to E}[f](\mathbf{c}_1) = \mathrm{P}[f](e|c,\mathbf{c}_1) - \mathrm{P}[f](e|\bar{c},\mathbf{c}_1). \tag{6}$$

The analysis of qualitative influences requires that we isolate the contribution of particular cause variables C with respect to the effect E. By writing

$$\begin{aligned}\mathrm{P}[f](e|\hat{c},\mathbf{c}_1) &= \sum_{\mathbf{m}} f(\mathbf{m})P(\mathbf{m}|\mathbf{c}) = P(m|\hat{c})\mathrm{P}[f_m](e|\mathbf{c}_1) + (1-P(m|\hat{c}))\mathrm{P}[f_{\bar{m}}](e|\mathbf{c}_1)\\ &= \mathrm{P}[f_{\bar{m}}](e|\mathbf{c}_1) + P(m|\hat{c})\mathrm{P}[g](e|\mathbf{c}_1),\end{aligned} \tag{7}$$

where g denotes the *difference function* $f_m - f_{\bar{m}}$, we obtain this isolation of C from the remainder of the cause variables. Sometimes, we wish to refer to the variable M over which we vary the interaction function f, and then the notation g_M is used. Note that it holds for the difference function that $g(\mathbf{m}_1) \in \{-1,0,1\}$. If we substitute Eq. (7) into (6) we obtain the following equation for the sign of a qualitative influence in ICI models:

$$\delta_{C\to E}[f](\mathbf{c}_1) = (P(m|c) - P(m|\bar{c}))\cdot \mathrm{P}[g](e|\mathbf{c}_1).$$

Under the assumption that $P(m|c) > P(m|\bar{c})$, which always holds under the assumption of accountability, i.e., $P(m|\bar{c}) = 0$ (cf. Section 2.2), we may write

$$\delta_{C\to E}[f](\mathbf{c}_1) \propto \mathrm{P}[g](e|\mathbf{c}_1). \tag{8}$$

We use Definition 17 and its associated lemmas to derive some properties of qualitative influences in ICI models. We can write

$$\mathrm{P}[g](e|\mathbf{c}_1) = \sum_{\mathbf{m}_1} g(\mathbf{m}_1)P(\mathbf{m}_1|\mathbf{c}_1),$$

where the configuration $\mathbf{m}_1$ ranges over all elements of $\mathbb{B}^{n-1}$. Let these configurations $\mathbf{m}_1$ be represented by $\mathbf{m}_1^i$, for $i = 1,\ldots,2^{n-1}$, and ordered such that if $\mathbf{m}_1^i < \mathbf{m}_1^j$ then $i < j$. The configurations $\mathbf{c}_1$ of $\mathbf{C}_1$ may also be any element of $\mathbb{B}^{n-1}$ and we assume that they are ordered likewise such that $\mathbf{c}_1^i = \mathbf{m}_1^i$ for $i = 1,\ldots,2^{n-1}$. From Lemma 19 it follows that for each configuration $\mathbf{c}_1$:

$$\mathrm{P}[g](e|\mathbf{c}_1) = \sum_{\mathbf{m}_1 \leqslant \mathbf{c}_1} g(\mathbf{m}_1)P(\mathbf{m}_1|\mathbf{c}_1). \tag{9}$$

Therefore, we need only take into account intermediate states that precede a causal state in the ordering. Based on this ordering we derive the properties of qualitative influences in ICI models. We will state these properties compactly in terms of the difference function g.

Proposition 23. $\delta_{C\to E}[f] = 0 \iff g = 0$.

Proof. Using Eq. (9), we prove by induction that if $\mathrm{P}[g](e|\mathbf{c}_1^k) = 0$ then $g(\mathbf{m}_1^k) = 0$, for $k = 1,\ldots,2^{n-1}$.

Basis. Let $k = 1$. Then $\mathrm{P}[g](e|\mathbf{c}_1^k) = g(\mathbf{m}_1^k)\cdot P(\mathbf{m}_1^k|\mathbf{c}_1^k)$. Since $P(\mathbf{m}_1^1|\mathbf{c}_1^1) > 0$ by Lemma 18, it must be the case that $g(\mathbf{m}_1^1) = 0$ if $\mathrm{P}[g](e|\mathbf{c}_1^1) = 0$.

Inductive hypothesis. For $i = 1,\ldots,k$, it holds that from $\mathrm{P}[g](e|\mathbf{c}_1^i) = 0$ it follows that $g(\mathbf{m}_1^i) = 0$, and vice versa.

Induction step. From the inductive hypothesis, it follows that:

$$\mathrm{P}[g](e|\mathbf{c}_1^{k+1}) = \sum_{1\leqslant i\leqslant k+1} g(\mathbf{m}_1^i)P(\mathbf{m}_1^i|\mathbf{c}_1^{k+1}) = g(\mathbf{m}_1^{k+1})P(\mathbf{m}_1^{k+1}|\mathbf{c}_1^{k+1}).$$

As $P(\mathbf{m}_1^{k+1}|\mathbf{c}_1^{k+1}) > 0$ it follows that $g(\mathbf{m}_1^{k+1}) = 0$ if $\mathrm{P}[g](e|\mathbf{c}_1^{k+1}) = 0$, and vice versa. But then $g(\mathbf{m}_1^i) = 0$, for $i = 1,\ldots,2^{n-1}$. $\square$

In order to distinguish the different signs of qualitative influences it is necessary to know when positive and negative contributions are possible in principle. We first state an elementary relationship between positive and negative contributions to the sign of a qualitative influence.

Lemma 24. $\delta_{C\to E}[f](\mathbf{c}_1) > 0 \iff \delta_{C\to E}[\neg f](\mathbf{c}_1) < 0$.

Proof Using the result of Lemma 11, we derive

$$\begin{aligned}\delta_{C\to E}[f](\mathbf{c}_1) > 0 &\iff \mathrm{P}[f_m](e|\mathbf{c}_1) - \mathrm{P}[f_{\bar{m}}](e|\mathbf{c}_1) > 0\\ &\iff (1-\mathrm{P}[f_m](e|\mathbf{c}_1)) - (1-\mathrm{P}[f_{\bar{m}}](e|\mathbf{c}_1)) < 0\\ &\iff \mathrm{P}[\neg f_m](e|\mathbf{c}_1) - \mathrm{P}[\neg f_{\bar{m}}](e|\mathbf{c}_1) < 0\\ &\iff \delta_{C\to E}[\neg f](\mathbf{c}_1) < 0. \qquad \square\end{aligned}$$

Exploring the initial function values of the difference function g, as defined above in Definition 22, yields further insight into the properties of qualitative influences. Note that we use the definition here by taking $b = 1$.

Lemma 25 lists a sufficient condition for observing a positive value of $\delta_{C\to E}[f](\mathbf{c}_1)$.

Lemma 25. *For every ICI model with interaction function f it holds that*

$$V_g^+ \Rightarrow \exists \mathbf{c}_1[\delta_{C\to E}[f](\mathbf{c}_1) > 0].$$

Proof. Recall from Definition 22 that it holds that

$$V_g^+ = \exists \mathbf{m}_1[g(\mathbf{m}_1) = 1 \wedge \forall \mathbf{m}_1' < \mathbf{m}_1[g(\mathbf{m}_1') \in \{0,1\}]].$$

Choosing $\mathbf{c}_1 = \mathbf{m}_1$ we obtain $\mathrm{P}[g](e|\mathbf{c}_1) = \sum_{\mathbf{m}_1' \leqslant \mathbf{c}_1} g(\mathbf{m}_1')P(\mathbf{m}_1'|\mathbf{c}_1)$ according to Eq. (9). Since for each $\mathbf{m}_1' < \mathbf{c}_1$ it holds that $g(\mathbf{m}_1') \in \{0,1\}$ and $g(\mathbf{m}_1) = 1$ with $P(\mathbf{m}_1|\mathbf{c}_1) > 0$ we have proved the lemma. $\square$

We present a similar result for negative values of $\delta_{C\to E}[f](\mathbf{c}_1)$.

Lemma 26. *For every ICI model with interaction function f it holds that*

$$V_g^- \Rightarrow \exists \mathbf{c}_1[\delta_{C\to E}[f](\mathbf{c}_1) < 0].$$

Proof. Recall that $V_g^- = \exists \mathbf{m}_1[g(\mathbf{m}_1) = -1 \wedge \forall \mathbf{m}_1' < \mathbf{m}_1[g(\mathbf{m}_1') \in \{-1,0\}]]$. If we use $\neg f$ in Lemma 25 and the correspondence $\neg f_m(\mathbf{m}_1) - \neg f_{\bar{m}}(\mathbf{m}_1) = 1 \iff g(\mathbf{m}_1) = -1$ then we obtain

$$\exists \mathbf{m}_1[g(\mathbf{m}_1) = -1 \wedge \forall \mathbf{m}_1' < \mathbf{m}_1[g(\mathbf{m}_1') \in \{-1,0\}]] \Rightarrow \exists \mathbf{c}_1[\delta_{C\to E}[\neg f](\mathbf{c}_1) > 0].$$

From Lemma 24 it follows that $\delta_{C\to E}[\neg f](\mathbf{c}_1) > 0 \iff \delta_{C\to E}[\neg\neg f](\mathbf{c}_1) < 0 = \delta_{C\to E}[f](\mathbf{c}_1) < 0$, which proves the proposition. $\square$

The reason why we can find a positive (or negative) value of $\delta_{C\to E}[f](\mathbf{c}_1)$ follows from the fact that we may choose a configuration $\mathbf{c}_1$ that renders all configurations $\mathbf{m}_1$ that are larger than or incomparable with $\mathbf{c}_1$ irrelevant. This is visualized in Fig. 4.

If we consider the functions f_m and $f_{\bar{m}}$ then one of four different situations may arise. First, if neither V_g^+ nor V_g^- hold then the inductive argument of Lemma 23 holds and $\delta_{C\to E}[f] = 0$. Second, if both V_g^+ and V_g^-

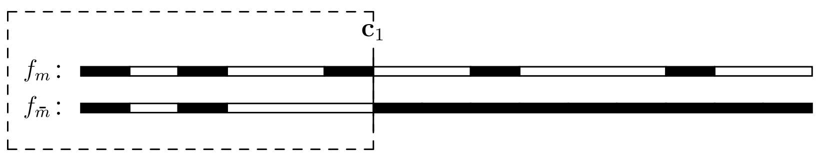

Fig. 4. Illustration of Lemma 25; horizontal bars represent true (black) and false (white) values of $f_m(\mathbf{m}_1)$ and $f_{\bar{m}}(\mathbf{m}_1)$, respectively. We only need to consider configurations $\mathbf{m}_1 \leqslant \mathbf{c}_1$, i.e. the configurations contained within the dashed region.

hold, then we have two incomparable configurations $\mathbf{m}_1$ and $\mathbf{m}'_1$ that render $\delta_{C\to E}[f](\mathbf{c})$ positive and negative, respectively. This leads directly to the following proposition.

Proposition 27. $V_g^+ \wedge V_g^- \Rightarrow \delta_{C\to E}[f] = \sim$.

Third, if V_g^+ holds and V_g^- does not hold then there is a positive value of $\delta_{C\to E}[f](\mathbf{c}_1)$ for some configuration $\mathbf{c}_1$ of $\mathbf{C}_1$ such that $\delta_{C\to E}[f]$ is either + or ∼. Under a specific condition we can infer that the sign must be positive.

Proposition 28. *If* V_g^+ and $\neg\exists\mathbf{m}_1[g(m_1) = -1]$ *then* $\delta_{C\to E}[f] = +$.

Proof. The proposition follows from the observations that $\delta_{C\to E}[f](\mathbf{c}) > 0$ for some $\mathbf{c}$ and no negative contribution to the sign of the qualitative influence. □

The paper by Lucas [13] includes tables for Boolean functions defined in terms of the 16 binary Boolean functions. We use these results in the following example.

Example 29. For both the AND and the OR operator, we have $\delta_{C\to E}[f] = +$ since for both operators it holds that the difference function $g = f_m - f_{\bar{m}}$ is non-negative and positive for at least one $\mathbf{m}_1$, which implies that the conditions of Proposition 28 hold.

If the conditions of Proposition 28 do not hold then we know for a fact that the sign is ambiguous, since it can be either non-monotonic or positive if the parameters are unknown.

Proposition 30. *If* V_g^+ and $\exists\mathbf{m}_1[g(\mathbf{m}_1) = -1]$ *then* $\delta_{C\to E}[f] = ?$

In order to prove Proposition 30, we need to prove that if V_g^+ holds, and $\exists\mathbf{m}_1[f_m(\mathbf{m}_1) < f_{\bar{m}}(\mathbf{m}_1)]$, then we can find parameters such that $\delta_{C\to E}[f] = \sim$ and other parameters such that $\delta_{C\to E}[f] = +$. The non-monotonic case is easily proven by the following lemma.

Lemma 31. *If* $\exists\mathbf{m}_1[g(\mathbf{m}_1) = 1] \wedge \exists\mathbf{m}'_1[g(\mathbf{m}'_1) = -1]$ *then we can choose parameters such* $\delta_{C\to E}[f] = \sim$.

Proof. From Lemma 20 it follows that we can choose parameters such that $\delta_{C\to E}[f](\mathbf{c}_1) = g(\mathbf{m}_1) = 1$ and $\delta_{C\to E}[f](\mathbf{c}'_1) = g(\mathbf{m}'_1) = -1$. □

It is more complex to prove that we can also find parameters such that $\delta_{C\to E}[f] = +$. The proof is given in Appendix A and relies on the fact that we can always find parameters such that the negative contribution remains smaller than the positive contribution to the sign of the qualitative influence.

Lemma 32. *If* V_g^+ and $\exists\mathbf{m}_1[g(\mathbf{m}_1) = -1]$ *then we can find parameters such that* $\delta_{C\to E}[f] = +$.

Proof. See Appendix A. □

Finally, if V_g^- holds and V_g^+ does not hold then there is a negative value of $d_{C\to E}[f](\mathbf{c}_1)$ for some configuration $\mathbf{c}_1$ of $\mathbf{C}_1$ such that $\delta_{C\to E}[f]$ is either − or ∼. Analogous to positive qualitative influences, under a specific condition we can infer that the sign must be negative. This leads to the following proposition, whose proof is analogous to that of Proposition 28.

Proposition 33. *If* V_g^- and $\neg\exists\mathbf{m}_1[g(\mathbf{m}_1) = 1]$ *then* $\delta_{C\to E}[f] = -$.

Symmetrically to positive qualitative influences, if this condition does not hold then we know for a fact that the sign is ambiguous since it can be either non-monotonic or negative if the parameters are unknown.

Proposition 34. *If* V_g^- and $\exists\mathbf{m}_1[g(\mathbf{m}_1) = 1]$ *then* $\delta_{C\to E}[f] = ?$

The proof that parameters can always be found to generate negative or non-monotonic qualitative influences proceeds in the same way as that for the positive qualitative influences.

In the above, we have shown how properties of the interaction function f influence the qualitative properties of ICI models. It is straightforward to recast properties of the difference function g in terms of properties of the interaction function f due to the identity $g = f_m - f_{\bar{m}}$ as is demonstrated by means of the prognostic model.

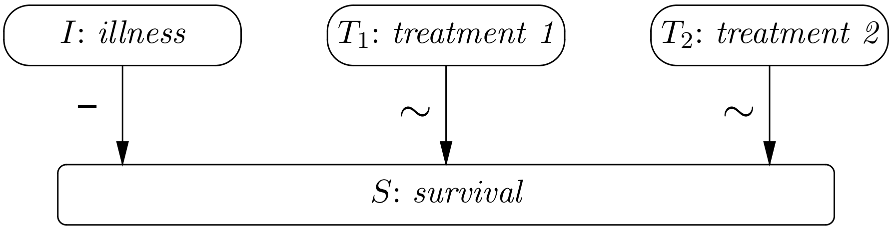

Fig. 5. Qualitative influences with respect to patient survival.

Example 35. We first consider the qualitative influence of I on S. In order to identify the qualitative behavior, we need to investigate the curries f_b and $f_{\bar{b}}$. If we restrict B to $\top$ (i.e., b), we have

$$f_b \equiv (\neg b \wedge \neg E_1 \wedge \neg E_2) \vee (E_1 \wedge \neg E_2) \vee (\neg E_1 \wedge E_2) \equiv (E_1 \wedge \neg E_2) \vee (\neg E_1 \wedge E_2).$$

In a similar vein we can reduce $f_{\bar{b}}$ to $\neg(E_1 \wedge E_2)$. It follows that for g we have[2]

$$\begin{aligned}
g(e_1, e_2) &= f_b(e_1, e_2) - f_{\bar{b}}(e_1, e_2) = (e_1 \wedge \neg e_2) \vee (\neg e_1 \wedge e_2) - \neg(e_1 \wedge e_2) = 0,\\
g(e_1, \bar{e}_2) &= f_b(e_1, \bar{e}_2) - f_{\bar{b}}(e_1, \bar{e}_2) = (e_1 \wedge \neg \bar{e}_2) \vee (\neg e_1 \wedge \bar{e}_2) - \neg(e_1 \wedge \bar{e}_2) = 0,\\
g(\bar{e}_1, e_2) &= f_b(\bar{e}_1, e_2) - f_{\bar{b}}(\bar{e}_1, e_2) = (\bar{e}_1 \wedge \neg e_2) \vee (\neg \bar{e}_1 \wedge e_2) - \neg(\bar{e}_1 \wedge e_2) = 0,\\
g(\bar{e}_1, \bar{e}_2) &= f_b(\bar{e}_1, \bar{e}_2) - f_{\bar{b}}(\bar{e}_1, \bar{e}_2) = (\bar{e}_1 \wedge \neg \bar{e}_2) \vee (\neg \bar{e}_1 \wedge \bar{e}_2) - \neg(\bar{e}_1 \wedge \bar{e}_2) = -1.
\end{aligned}$$

It follows that Proposition 33 holds, such that $\delta_{I \to S}[f] = -$. This negative influence of the serious illness on prognosis is in accordance with the previously stated domain knowledge. We proceed in a similar way for the qualitative influences of T_1 on S and obtain the following results. For the qualitative influence of T_1 on S we have $f_{e_1} \equiv \neg E_2$ and $f_{\bar{e}_1} \equiv (\neg B \wedge \neg E_2) \vee E_2$. It follows that for g we have that $g(i, e_2) = 0$, $g(i, \bar{e}_2) = 1$, $g(\bar{\imath}, e_2) = -1$ and $g(i, e_2) = 0$. As $(i, \bar{e}_2)$ and $(\bar{\imath}, e_2)$ are incomparable and have opposing signs, it follows that $\delta_{T_1 \to S}[f] = \sim$ according to Proposition 27. We remark that $\delta_{T_2 \to S}[f] = \sim$ by symmetry. The qualitative influences are depicted in Fig. 5.

Previously, we have shown how properties of the interaction function f influence the qualitative properties of ICI models. Next, we show that, by means of the propositions and lemmas that have been derived, we can also immediately infer properties of interaction functions that should hold when a qualitative influence is known. First, observe that, based on the lemmas and propositions above

$$\begin{aligned}
&(V_g^+ \wedge \neg \exists \mathbf{m}_1[g(\mathbf{m}_1) = -1]) \vee (V_g^+ \wedge \exists \mathbf{m}_1[g(\mathbf{m}_1) = -1]) \vee\\
&(V_g^- \wedge \neg \exists \mathbf{m}_1[g(\mathbf{m}_1) = 1]) \vee (V_g^- \wedge \exists \mathbf{m}_1[g(\mathbf{m}_1) = 1]) \quad \vee (g = 0)
\end{aligned}$$

covers all possible cases. The first two conjunctions in this disjunction handle the positive qualitative influences (due to Proposition 28 and Lemma 32). The third and fourth conjunctions in this disjunction handle the negative qualitative influences (by symmetry), and the last conjunction is a necessary and sufficient condition for observing a zero qualitative influence (due to Proposition 23). The second and fourth conjunctions are conditions that may lead to non-monotonic qualitative influences, and whose disjunction is equivalent to $\exists \mathbf{m}_1[g(\mathbf{m}_1) = -1] \wedge \exists \mathbf{m}_1[g(\mathbf{m}_1) = 1]$. The properties of interaction functions given a qualitative influence are listed in Table 1.

Example 36. Suppose we knew the qualitative influences but not the underlying interaction function for the prognostic model of Section 2.2. According to Table 1 we have:

$$\begin{aligned}
&\delta_{I \to S}[f] = - \Rightarrow V_{g_B}^-,\\
&\delta_{T_1 \to S}[f] = \sim \Rightarrow \exists \mathbf{m}_1[g_{E_1}(\mathbf{m}_1) = 1] \wedge \exists \mathbf{m}_1'[g_{E_1}(\mathbf{m}_1') = -1],\\
&\delta_{T_2 \to S}[f] = \sim \Rightarrow \exists \mathbf{m}_1[g_{E_2}(\mathbf{m}_1) = 1] \wedge \exists \mathbf{m}_1'[g_{E_2}(\mathbf{m}_1') = -1],
\end{aligned}$$

[2] Recall that in an arithmetic context, we interpret $\top$ as 1 and $\bot$ as 0.

Table 1
Properties of interaction functions given a qualitative influence

Qualitative influence	Property of the interaction function
0	$g = 0$
+	V_g^+
−	V_g^-
∼	$\exists \mathbf{m}_1[g(\mathbf{m}_1) = 1] \wedge \exists \mathbf{m}_1'[g(\mathbf{m}_1') = -1]$

where $g_B = f_b - f_{\bar{b}}, g_{E_1} = f_{e_1} - f_{\bar{e}_1},$ and $g_{E_2} = f_{e_2} - f_{\bar{e}_2}$. The results are indeed properties of the interaction function of the prognostic model, as represented by the Boolean expression (4). The first qualitative influence would, for example, preclude choosing the AND and OR interaction functions, as both do not satisfy property $V_{g_B}^-$.

4.2. Additive synergies

Additive synergies express how two cause variables C and C' from the set of cause variables $\mathbf{C}$ jointly influence the probability of observing the effect E. Recall that the remaining cause variables are denoted by $\mathbf{C}_2 = \mathbf{C} \setminus \{C, C'\}$. Using the general definition of additive synergy from QPN theory, the additive synergy $\delta_{(C,C')\to E}[f]$ between C and C' given interaction function f is determined by

$$\delta_{(C,C')\to E}[f](\mathbf{c}_2) = \mathrm{P}[f](e|c, c', \mathbf{c}_2) + \mathrm{P}[f](e|\bar{c}, \bar{c}', \mathbf{c}_2) - \mathrm{P}[f](e|\bar{c}, c', \mathbf{c}_2) - \mathrm{P}[f](e|c, \bar{c}', \mathbf{c}_2). \tag{10}$$

The analysis requires an isolation of cause variables C and C'. We apply the decomposition (7) twice and obtain:

$$\begin{aligned}\mathrm{P}[f](e|\hat{c}, \hat{c}', \mathbf{c}_2) = {} & P(m|\hat{c})P(m'|\hat{c}')\mathrm{P}[h](e|\mathbf{c}_2) + \mathrm{P}[f_{\bar{m},\bar{m}'}](e|\mathbf{c}_2) + P(m|\hat{c})\mathrm{P}[f_{m,\bar{m}'} - f_{\bar{m},\bar{m}'}](e|\mathbf{c}_2) \\ & + P(m'|\hat{c}')\mathrm{P}[f_{\bar{m},m'} - f_{\bar{m},\bar{m}'}](e|\mathbf{c}_2),\end{aligned} \tag{11}$$

where the function $h : \mathbb{B}^{n-2} \to \{-2, -1, 0, 1, 2\}$ is defined as

$$h(\mathbf{m}_2) = f_{m,m'}(\mathbf{m}_2) + f_{\bar{m},\bar{m}'}(\mathbf{m}_2) - f_{\bar{m},m'}(\mathbf{m}_2) - f_{m,\bar{m}'}(\mathbf{m}_2). \tag{12}$$

The function h is also sometimes indicated by $h_{M,M'}$. By inserting Eq. (11) into (10), we obtain

$$\delta_{(C,C')\to E}[f](\mathbf{c}_2) = (P(m|c) - P(m|\bar{c}))(P(m'|c') - P(m'|\bar{c}'))\mathrm{P}[h](e|\mathbf{c}_2).$$

Under the assumptions that $P(m|c) > P(m|\bar{c})$ and $P(m'|c') > P(m'|\bar{c}')$, which holds under the assumption of accountability, we may write

$$\delta_{(C,C')\to E}[f](\mathbf{c}_2) \propto \mathrm{P}[h](e|\mathbf{c}_2).$$

We take a similar approach as for qualitative influences and use an ordering on configurations of $\mathbf{M}_2$ and $\mathbf{C}_2$ which now range from $\mathbf{m}_1$ to $\mathbf{m}_{2^{n-2}}$ and from $\mathbf{c}_1$ to $\mathbf{c}_{2^{n-2}}$ respectively.

The structure of the expression for qualitative influences and additive synergies is essentially the same, where the only difference is that we sum over 2^{n-2} instead of 2^{n-1} configurations and g is replaced by h. If we consider the proofs of Lemmas 24–32 and Propositions 23–34 in the previous section, then we find that none, with the exception of Lemma 32, are dependent upon these two differences. Due to the analogy between qualitative influences and additive synergies, we state the results in terms of the difference function h without proof.

A necessary and sufficient condition for observing a zero additive synergy is easily found.

Proposition 37. $\delta_{(C,C')\to E}[f] = 0 \iff h = 0.$

Again, interaction functions f and their negations $\neg f$ lead to opposite contributions to the qualitative sign.

Lemma 38. $\delta_{(C,C')\to E}(\mathbf{c}_2) > 0 \iff \delta_{(C,C')\to E}[\neg f](\mathbf{c}_2) < 0.$

We next investigate the implications of function values of the function h, as defined above in Eq. (12), using Definition 22, for the qualitative properties. Here we take $b = 2$. An analysis of positive and negative contributions to the sign of the additive synergy is given by Lemmas 39 and 40.

Lemma 39. *For every ICI model with interaction function f it holds that*

$$V_h^+ \Rightarrow \exists \mathbf{c}_2[\delta_{(C,C')\to E}[f](\mathbf{c}_2) > 0].$$

Lemma 40. *For every ICI model with interaction function f it holds that*

$$V_h^- \Rightarrow \exists \mathbf{c}_2[\delta_{(C,C')\to E}[f](\mathbf{c}_2) < 0].$$

Non-monotonic additive synergies are identified by Proposition 41.

Proposition 41. $V_h^+ \wedge V_h^- \Rightarrow \delta_{(C,C')\to E}[f] = \sim$.

Positive additive synergies are identified by Proposition 42 and ambiguous additive synergies (either non-monotonic or positive signs) are identified by 43. We can always choose parameters such that this ambiguous additive synergy reduces to a non-monotonic or positive additive synergy. The proof is similar to the proof in case of qualitative influences and is omitted here.

Proposition 42. *If* V_h^+ and $\forall \mathbf{m}_2[h(\mathbf{m}_2) \in \{0,1,2\}]$ *then* $\delta_{(C,C')\to E}[f] = +$.

Proposition 43. *If* V_h^+ and $\exists \mathbf{m}_2[h(\mathbf{m}_2) \in \{-2,-1\}]$ *then* $\delta_{(C,C')\to E}[f] = ?$

Symmetric results are obtained for negative additive synergies in Proposition 44, where Proposition 45 identifies ambiguous additive synergies which can be either non-monotonic or negative, depending on the parameters.

Proposition 44. *If* V_h^- and $\forall \mathbf{m}_2[h(\mathbf{m}_2) \in \{-2,-1,0\}]$ *then* $\delta_{(C,C')\to E}[f] = -$.

Proposition 45. *If* V_h^- and $\exists \mathbf{m}_2[h(\mathbf{m}_2) \in \{1,2\}]$ *then* $\delta_{(C,C')\to E}[f] = ?$

We use the results of Lucas [13] to verify some of our results.

Example 46. For the AND operator, we have $\delta_{(C,C')\to E}[f] = +$ since the difference function $h(\mathbf{m}_2) = f_{m,m'}(\mathbf{m}_2) + f_{\bar{m},\bar{m}'}(\mathbf{m}_2) - f_{\bar{m},m'}(\mathbf{m}_2) - f_{m,\bar{m}'}(\mathbf{m}_2)$ must be non-negative and positive for at least one configuration of $\mathbf{m}_2$. On the other hand, for the OR operator we have $\delta_{(C,C')\to E}[f] = -$ since h is non-positive and negative for at least one configuration of $\mathbf{m}_2$.

We can recast properties of the difference function h in terms of properties of the interaction function f as we have the identity $h = f_{m,m'} + f_{\bar{m},\bar{m}'} - f_{\bar{m},m'} - f_{m,\bar{m}'}$. We illustrate the results concerning additive synergies by means of the running example, shown in Fig. 3.

Example 47. With regard to the additive synergy between the treatments T_1 and T_2, we have $f_{e_1,e_2} \equiv \perp$, $f_{\bar{e}_1,\bar{e}_2} \equiv \neg B$ and $f_{\bar{e}_1,e_2} \equiv f_{e_1,\bar{e}_2} \equiv \top$. We then have $h(b) = -2$ and $h(\bar{b}) = -1$ such that $\delta_{(T_1,T_2)\to S}[f] = -$ according to Proposition 44. This agrees with the observation that the administration of one of both treatments is optimal, whereas administration of both treatments yields a suboptimal result. With regard to the additive synergy between I and T_1, we have $f_{b,e_1} \equiv \neg E_2$, $f_{\bar{b},\bar{e}_1} \equiv \top$, $f_{\bar{b},e_1} \equiv \neg E_2$ and $f_{b,\bar{e}_1} \equiv E_2$. We then have that $h(e_2) = 0$ and $h(\bar{e}_2) = 1$ such that $\delta_{(I,T_1)\to S}[f] = +$ according to Proposition 42. We also have $\delta_{(I,T_2)\to S}[f] = +$ by symmetry. This is in agreement with the fact that when a treatment is administered to an ill

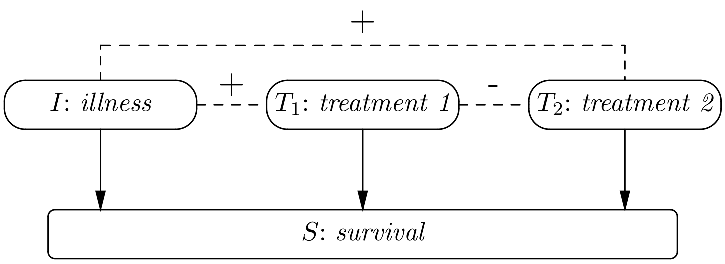

Fig. 6. Additive synergies with respect to patient survival.

person, or when no treatment is administered in the absence of the illness improves survival in comparison to when a non-ill person is treated or when treatment is not given to an ill person. The additive synergies are depicted in Fig. 6.

So far, we have only considered the qualitative behavior of a given interaction function. Again, we infer properties of interaction functions that should hold when an additive synergy is known. These properties are shown in Table 2 and have straightforward derivations due to the correspondence between qualitative influences and additive synergies. An example is again provided by considering the qualitative properties of the prognostic model.

Example 48. Suppose we knew the additive synergies but not the underlying interaction function for the prognostic model. According to Table 2 we have

$$\delta_{(T_1,T_2)\to S}[f] = - \Rightarrow V^-_{h_{E_1,E_2}},$$

where $h_{E_1,E_2} = f_{e_1,e_2} + f_{\bar{e}_1,\bar{e}_2} - f_{\bar{e}_1,e_2} - f_{e_1,\bar{e}_2}$. This is indeed a property of Boolean expression (4) that represents the prognostic model, as may be verified. This constraint would, for example, exclude the AND Boolean function, as it does not satisfy property $V^-_{h_{E_1,E_2}}$.

4.3. Product synergies

Product synergies describe the created dependence between two causes when the value of the effect variable is observed. The sign $\delta^{\hat{e}}_{(C,C')\to E}[f]$ of a product synergy between C and C' with respect to $\hat{e}$ when f is the underlying interaction function, is determined by

$$\delta^{\hat{e}}_{(C,C')\to E}[f](\mathbf{c}_2) = \mathrm{P}[f](\hat{e}|c,c',\mathbf{c}_2)\mathrm{P}[f](\hat{e}|\bar{c},\bar{c}',\mathbf{c}_2) - \mathrm{P}[f](\hat{e}|\bar{c},c',\mathbf{c}_2)\mathrm{P}[f](\hat{e}|c,\bar{c}',\mathbf{c}_2).$$

This can be rewritten for $E = \top$ (presence of the effect has been observed) to

$$\delta^{e}_{(C,C')\to E}[f](\mathbf{c}_2) = P(m|c)P(m'|c')(\mathrm{P}[h](e|\mathbf{c}_2)\mathrm{P}[f_{\bar{m},\bar{m}'}](e|\mathbf{c}_2) - \mathrm{P}[f_{m,\bar{m}'} - f_{\bar{m},\bar{m}'}](e|\mathbf{c}_2)\mathrm{P}[f_{\bar{m},m'} - f_{\bar{m},\bar{m}'}](e|\mathbf{c}_2)),$$

where again $h = f_{m,m'} + f_{\bar{m},\bar{m}'} - f_{\bar{m},m'} - f_{m,\bar{m}'}$. Under our standard assumption of accountability, this yields:

$$\delta^{e}_{(C,C')\to E}[f](\mathbf{c}_2) \propto \mathrm{P}[h](e|\mathbf{c}_2)\mathrm{P}[f_{\bar{m},\bar{m}'}](e|\mathbf{c}_2) - \mathrm{P}[f_{m,\bar{m}'} - f_{\bar{m},\bar{m}'}](e|\mathbf{c}_2)\mathrm{P}[f_{\bar{m},m'} - f_{\bar{m},\bar{m}'}](e|\mathbf{c}_2).$$

This can be alternatively written as

$$\delta^{e}_{(C,C')\to E}[f](\mathbf{c}_2) \propto \mathrm{P}[f_{m,m'}](e|\mathbf{c}_2)\mathrm{P}[f_{\bar{m},\bar{m}'}](e|\mathbf{c}_2) - \mathrm{P}[f_{\bar{m},m'}](e|\mathbf{c}_2)\mathrm{P}[f_{m,\bar{m}'}](e|\mathbf{c}_2).$$

Using the distributive law of arithmetic, we obtain

$$\begin{aligned}\delta^{e}_{(C,C')\to E}[f](\mathbf{c}_2) &\propto \mathrm{P}[f_{m,m'}](e|\mathbf{c}_2)\mathrm{P}[f_{\bar{m},\bar{m}'}](e|\mathbf{c}_2) - \mathrm{P}[f_{\bar{m},m'}](e|\mathbf{c}_2)\mathrm{P}[f_{m,\bar{m}'}](e|\mathbf{c}_2)\\ &= \left(\sum_{\mathbf{m}_2} f_{m,m'}(\mathbf{m}_2)P(\mathbf{m}_2|\mathbf{c}_2)\right)\left(\sum_{\mathbf{m}_2} f_{\bar{m},\bar{m}'}(\mathbf{m}_2)P(\mathbf{m}_2|\mathbf{c}_2)\right)\\ &\quad - \left(\sum_{\mathbf{m}_2} f_{\bar{m},m'}(\mathbf{m}_2)P(\mathbf{m}_2|\mathbf{c}_2)\right)\left(\sum_{\mathbf{m}_2} f_{m,\bar{m}'}(\mathbf{m}_2)P(\mathbf{m}_2|\mathbf{c}_2)\right)\\ &= \sum_{\mathbf{m}_2,\mathbf{m}'_2} r(\mathbf{m}_2,\mathbf{m}'_2)P(\mathbf{m}_2|\mathbf{c}_2)P(\mathbf{m}'_2|\mathbf{c}_2),\end{aligned}$$

Table 2
Properties of interaction functions given an additive synergy

Additive synergy	Property of the interaction function
0	$h = 0$
$+$	V^+_h
$-$	V^-_h
$\sim$	$\exists\mathbf{m}_2[h(\mathbf{m}_2) \in \{1,2\}] \wedge \exists\mathbf{m}'_2[h(\mathbf{m}'_2) \in \{-2,-1\}]$

where the function $r : \mathbb{B}^{n-2} \times \mathbb{B}^{n-2} \rightarrow \{-1, 0, 1\}$ is defined as follows:

$$r(\mathbf{m}_2, \mathbf{m}'_2) = f_{m,m'}(\mathbf{m}_2) f_{\bar{m},\bar{m}'}(\mathbf{m}'_2) - f_{\bar{m},m'}(\mathbf{m}_2) f_{m,\bar{m}'}(\mathbf{m}'_2). \tag{13}$$

We will also sometimes use the notation $r_{M,M'}$. From the expression above, it follows that the behavior of the product synergy is determined by the function r.

It appears that it suffices to carry out the analysis for $E = \top$ (the effect has been observed to be present), as application of the following lemma renders the analysis of the qualitative behavior of the product synergy for $E = \bot$ (absence of the effect has been observed) a straightforward exercise.

Lemma 49. $\delta^{e}_{(C,C')\rightarrow E}[\neg f] = \delta^{\bar{e}}_{(C,C')\rightarrow E}[f]$.

Proof

$$\begin{aligned}\delta^{e}_{(C,C')\rightarrow E}[\neg f](\mathbf{c}_2) &\propto \mathrm{P}[\neg f_{m,m'}](e|\mathbf{c}_2)\mathrm{P}[\neg f_{\bar{m},\bar{m}'}](e|\mathbf{c}_2) - \mathrm{P}[\neg f_{\bar{m},m'}](e|\mathbf{c}_2)\mathrm{P}[f_{m,\bar{m}'}](e|\mathbf{c}_2)\\ &= (1 - \mathrm{P}[f_{m,m'}](e|\mathbf{c}_2))(1 - \mathrm{P}[f_{\bar{m},\bar{m}'}](e|\mathbf{c}_2)) - (1 - \mathrm{P}[f_{\bar{m},m'}](e|\mathbf{c}_2))(1 - \mathrm{P}[f_{m,\bar{m}'}](e|\mathbf{c}_2))\\ &= \mathrm{P}[f_{m,m'}](\bar{e}|\mathbf{c}_2)\mathrm{P}[f_{\bar{m},\bar{m}'}](\bar{e}|\mathbf{c}_2) - \mathrm{P}[f_{\bar{m},m'}](\bar{e}|\mathbf{c}_2)\mathrm{P}[f_{m,\bar{m}'}](\bar{e}|\mathbf{c}_2) \propto \delta^{\bar{e}}_{(C,C')\rightarrow E}[f](\mathbf{c}_2). \qquad \square\end{aligned}$$

Hence, if $\delta^{e}_{(C,C')\rightarrow E}[\neg f](\mathbf{c}_2)$ has a particular sign for configuration $\mathbf{c}_2$ then $\delta^{\bar{e}}_{(C,C')\rightarrow E}[f](\mathbf{c}_2)$ will have the same sign. Therefore, the sign of the product synergy for $E = \top$ with interaction function $\neq f$ will be the same as that for $E = \bot$ with interaction function f. Due to this relationship between the signs of the product synergy for $E = \top$ and $E = \bot$, we will only consider the case where $E = \top$. Recall that by Lemma 8, we have the following interesting relationship between product synergies and additive synergies, which offers an alternative way to compute the product synergy $\delta^{\bar{e}}_{(C,C')\rightarrow E}[f](\mathbf{c}_2)$, based on the associated additive synergy $\delta_{(C,C')\rightarrow E}[f](\mathbf{c}_2)$ and the associated product synergy $\delta^{e}_{(C,C')\rightarrow E}[f](\mathbf{c}_2)$:

$$\delta^{\bar{e}}_{(C,C')\rightarrow E}[f](\mathbf{c}_2) = \delta^{e}_{(C,C')\rightarrow E}[f](\mathbf{c}_2) - \delta_{(C,C')\rightarrow E}[f](\mathbf{c}_2).$$

Lemma 8 is useful for constructing tables of signs for particular Boolean functions, as it saves constructing one of these tables.

Example 50. The paper by Lucas [13] includes tables for Boolean functions defined in terms of the 16 binary Boolean functions. Consider the AND operator, $\wedge$; its additive synergy is equal to $\delta_{(C,C')\rightarrow E}[\wedge] = +$, whereas its product synergy for $E = \top$ is equal to $\delta^{e}_{(C,C')\rightarrow E}[\wedge] = 0$. Lemma 8 tells us that the product synergy for $E = \bot$ is equal to $\delta^{\bar{e}}_{(C,C')\rightarrow E}[f] = -$, which is indeed the value for the product synergy for $E = \bot$ in Table 12 in [13].

In the following, we derive sufficient conditions for observing particular qualitative behavior in terms of product synergies.

Proposition 51. $\delta^{e}_{(C,C')\rightarrow E}[f] = 0$ *if it holds that*

$$\forall \mathbf{m}_2, \mathbf{m}'_2[(f_{m,m'}(\mathbf{m}_2) \wedge f_{\bar{m},\bar{m}'}(\mathbf{m}'_2)) \iff (f_{\bar{m},m'}(\mathbf{m}_2) \wedge f_{m,\bar{m}'}(\mathbf{m}'_2))].$$

Proof. Note that if the premise holds, then, according to Definition (13) of the function r, we have that $r(\mathbf{m}_2, \mathbf{m}'_2) = 0$, for each $\mathbf{m}_2, \mathbf{m}'_2$, and thus $\delta^{e}_{(C,C')\rightarrow E}[f] = 0$. $\square$

A special case of this proposition, is the following condition:

$$(f_{m,m'} \equiv \bot \vee f_{\bar{m},\bar{m}'} \equiv \bot) \wedge (f_{m,\bar{m}'} \equiv \bot \vee f_{\bar{m},m'} \equiv \bot),$$

i.e., if at least one Boolean function at both sides of the negation of Definition (13) is equal to falsum, then a zero product synergy results.

We again determine conditions under which $\delta^{e}_{(C,C')\rightarrow E}[f](\mathbf{c}_2)$ is positive or negative. Similar to previous sections, we use the notations V^{+}_{r} and V^{-}_{r}, this time in terms of the function r defined above; for example V^{+}_{r} means that

$$\exists \mathbf{m}_2, \mathbf{m}'_2[[r(\mathbf{m}_2, \mathbf{m}'_2) = 1] \wedge \forall \mathbf{m}''_2 < \mathbf{m}_2, \mathbf{m}'''_2 < \mathbf{m}'_2[r(\mathbf{m}''_2, \mathbf{m}'''_2) \in \{0, 1\}]].$$

Lemma 52. *For every ICI model with interaction function f we have*

$$V_r^+ \Rightarrow \exists \mathbf{c}_2[\delta^e_{(C,C')\to E}[f](\mathbf{c}_2) > 0].$$

Proof. Simply note that if r is initially non-negative, we have a positive $\delta^e_{(C,C')\to E}[f](\mathbf{c}_2)$ for at least one value $\mathbf{c}_2$ by definition. □

An example of a positive value of $\delta^e_{(C,C')\to E}[f](\mathbf{c}_2)$ is demonstrated in Fig. 7.

A similar result holds for negative values of $\delta^e_{(C,C')\to}[f](\mathbf{c}_2)$ and the proof is analogous to that of Lemma 52.

Lemma 53. *For every ICI model with interaction function f we have*

$$V_r^- \Rightarrow \exists \mathbf{c}_2[\delta^e_{(C,C')\to E}[f](\mathbf{c}_2) < 0].$$

The following proposition follows directly from the definition of a non-monotonic product synergy.

Proposition 54. *If both* V_r^+ and V_r^- *hold then* $\delta^e_{(C,C')\to E}[f] = \sim$.

It also follows directly from Lemmas 52 and 53 that if V_r^+ holds and V_r^- does not hold, then the sign of the product synergy is either positive or non-monotonic. Conversely, if V_r^- holds and V_r^+ does not hold, then it follows that the sign of the product synergy is either negative or non-monotonic. The following two propositions identify under which conditions the sign of a product synergy is known to be positive or negative, respectively.

Proposition 55. *If* $\exists \mathbf{m}_2, \mathbf{m}_2'[r(\mathbf{m}_2, \mathbf{m}_2') = 1]$ and $\forall \mathbf{m}_2, \mathbf{m}_2'[r(\mathbf{m}_2, \mathbf{m}_2') \geqslant 0]$ *then it holds that* $\delta^e_{(C,C')\to E}[f] = +$.

Proof. This is just the general case of Lemma 52, where we ensure that the conditions listed for configurations $\mathbf{m}_2'' < \mathbf{m}_2, \mathbf{m}_2''' < \mathbf{m}_2'$ such that $r(\mathbf{m}_2'', \mathbf{m}_2''') \geqslant 0$ not only hold for configurations smaller than $\mathbf{m}_2, \mathbf{m}_2'$, but for all configurations $\mathbf{m}_2'' \neq \mathbf{m}_2, \mathbf{m}_2''' \neq \mathbf{m}_2'$. □

Proposition 56. *If* $\exists \mathbf{m}_2, \mathbf{m}_2'[r(\mathbf{m}_2, \mathbf{m}_2') = -1]$ and $\forall \mathbf{m}_2, \mathbf{m}_2'[r(\mathbf{m}_2, \mathbf{m}_2') \leqslant 0]$ *then it holds that* $\delta^e_{(C,C')\to E}[f] = -$.

Proof. This is the generalized case of Lemma 53. □

The cases that are not covered by the above propositions will be categorized as ambiguous.

Proposition 57. *If none of Propositions 51–56 hold then* $\delta^e_{(C,C')\to E}[f] = ?$

Proposition 57 collects those cases for which no sufficient conditions for observing a particular sign of a product synergy have been derived. In such cases, the sign can still be positive, negative or non-monotonic, depending on the parameters and depending on the structure of the interaction function. It is important to realize that due to Lemma 49, the above results equally hold for the case where $E = \perp$ whenever we replace each occurrence of f by $\neg f$.

We illustrate the results concerning product synergies again by means of the prognostic model, depicted in Fig. 3.

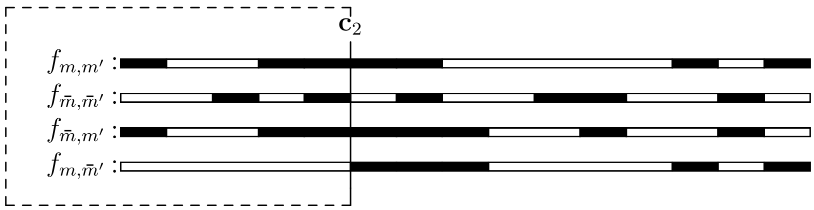

Fig. 7. Illustration of Lemma 52. Similarly to Fig. 4, the horizontal bars represent the truth values of $f_{m,m}(\mathbf{m}_2)$, $f_{\bar{m},\bar{m}}(\mathbf{m}_2)$, $f_{\bar{m},m}(\mathbf{m}_2)$ and $f_{m,\bar{m}}(\mathbf{m}_2)$.

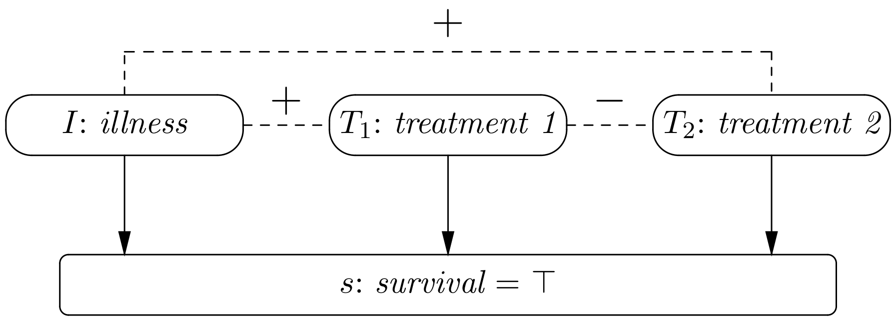

Fig. 8. Product synergies with respect to patient survival.

Example 58. We first focus on the case where we hypothesize that the patient will survive, i.e., $S = \top$. With regard to the product synergy between treatments T_1 and T_2, we have that $f_{e_1,e_2} \equiv \bot$, $f_{\bar{e}_1,\bar{e}_2} \equiv \neg B'$ and $f_{\bar{e}_1,e_2} \equiv f_{e_1,\bar{e}_2} = \top$. Condition 3 of Proposition 56 is satisfied since $r(B,B') = -1$ for each value of B, B', and thus $\delta^s_{(T_1,T_2)\to S}[f] = -$. This agrees with the observation that we expect that one of both treatments was administered given that we observe patient survival. With regard to the product synergy between I and T_1, we have that $f_{b,e_1} \equiv \neg E_2$, $f_{\bar{b},\bar{e}_1} \equiv \top$, $f_{\bar{b},e_1} \equiv \neg E_2$ and $f_{b,\bar{e}_1} \equiv E'_2$. Condition 1 of Proposition 55 is satisfied since $r(\bar{e}_2,\bar{e}'_2) = 1$, whereas $r(E_2,E'_2) = 0$ for any value of E_2, E'_2, with the exception of $E_2 = \bot$ and $E'_2 = \bot$; thus $\delta^s_{(I,T_1)\to S}[f] = +$. Hence, it is likely that treatment T_1 is administered given disease progression and patient survival and that treatment T_1 is not administered given no progression and patient survival. It is less likely that treatment T_1 is administered given no progression and patient survival and that treatment T_1 is not administered given disease progression and patient survival. The same holds for the product synergy between I and T_2 by symmetry. The results are summarized by Fig. 8.

As has been proved in Lemma 49, we can use also the derived propositions for $E = \bot$ by replacing f with $\neg f$. With regard to the product synergy between T_1 and T_2, we have that $\neg f_{e_1,e_2} \equiv \top$, $\neg f_{\bar{e}_1,\bar{e}_2} \equiv B$ and $\neg f_{\bar{e}_1,e_2} \equiv \neg f_{e_1,\bar{e}_2} = \bot$. Condition 3 of Proposition 55 is satisfied, since $r(B,B') = B$, thus $\delta^{\bar{s}}_{(T_1,T_2)\to S}[f] = +$. With regard to the product synergy between I and T_1, we have that $\neg f_{b,e_1} \equiv E_2$, $\neg f_{\bar{b},\bar{e}_1} \equiv \bot$, $\neg f_{\bar{b},e_1} \equiv E_2$ and $\neg f_{b,\bar{e}_1} = \neg E'_2$, thus $r(E_2,E'_2) = -(E_2 \wedge \neg E'_2)$. We classify the product synergy as $\delta^{\bar{s}}_{(I,T_1)\to S}[f] = -$. The same holds for the product synergy between I and T_2 by symmetry. The results are summarized by Fig. 9.

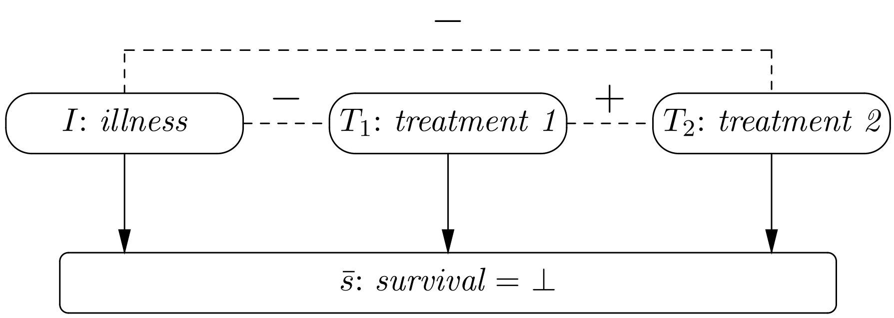

Fig. 9. Product synergies with respect to patient death.

Table 3
Properties of interaction functions given a product synergy for $E = \top$

Product synergy	Property of the interaction function
0	$\neg V_r^+ \wedge \neg V_r^-$
+	$\neg V_r^-$
−	$\neg V_r^+$
∼	$\exists \mathbf{m}_2, \mathbf{m}'_2[r(\mathbf{m}_2,\mathbf{m}'_2) = 1 \wedge \exists \mathbf{m}_2, \mathbf{m}'_2[r(\mathbf{m}_2,\mathbf{m}'_2) = -1]$

Again, we look at the converse analysis from qualitative specification to constraints on interaction functions using the propositions and lemmas that have been derived. Properties of product synergies with the effect observed to be present ($E = \top$) are shown in Table 3 and are derived by negating the properties for opposite signs when $E = \top$. For example, since V_r^+ with $E = \top$ implies that there is a configuration $\mathbf{c}_2$ of cause variables such that $\delta^e_{(C,C')\to E}[f](\mathbf{c}_2) > 0$ (Lemma 52), we know that $\neg V_r^+$ must hold for negative product synergies with $E = \top$. Likewise, $\neg V_r^-$ must hold for positive product synergies with $E = \top$. For the same reason, $\neg V_r^+ \wedge \neg V_r^-$ must hold for zero product synergies with $E = \top$. For non-monotonic product synergies it holds that Propositions 55 and 56 must both be false. Since, according to Proposition 51, it cannot be the case that $\forall \mathbf{m}_2, \mathbf{m}'_2[r(\mathbf{m}_2, \mathbf{m}'_2) = 0]$, it must hold that both $\exists \mathbf{m}_2, \mathbf{m}'_2[r(\mathbf{m}_2, \mathbf{m}'_2) = 1]$ and $\exists \mathbf{m}_2, \mathbf{m}'_2[r(\mathbf{m}_2, \mathbf{m}'_2) = -1]$. Properties of product synergies with $E = \bot$ are obtained using Lemma 49 by replacing the function r with the function

$$\bar{r}(\mathbf{m}_2, \mathbf{m}'_2) = \neg f_{m,m'}(\mathbf{m}_2)\neg f_{\bar{m},\bar{m}'}(\mathbf{m}'_2) - \neg f_{\bar{m},m'}(\mathbf{m}_2)\neg f_{m,\bar{m}'}(\mathbf{m}'_2).$$

In order to demonstrate this converse analysis, we look at the product synergy between treatments T_1 and T_2 of the prognostic model.

Example 59. Suppose we knew the product synergies but not the underlying interaction function for the prognostic model. For the product synergy between treatment T_1 and T_2 with the effect assumed to be present ($E = \top$), we have

$$\delta^e_{(T_1,T_2)\to S}[f] = - \Rightarrow \neg V^+_{r_{E_1,E_2}},$$

whereas its product synergy for the effect assumed to be absent ($E = \bot$) is given by

$$\delta^{\bar{e}}_{(T_1,T_2)\to S}[f] = + \Rightarrow \neg V^-_{\bar{r}_{E_1,E_2}}.$$

Note that here we use the complementary function $\bar{r}_{E_1,E_2}$. Again, it may be verified that these are properties of the Boolean expression (4) that underlies the prognostic model. For example, these properties are not satisfied by the AND function, which, therefore, cannot be selected as a basis for a prognostic model that satisfies the given qualitative constraints.

5. Discussion

This paper offers a detailed analysis of ICI models that employ Boolean interaction functions. In contrast to previous work, [13], the present paper offers a characterization of ICI models based on Boolean functions in general, and it can, thus, be used as a foundation for the analysis of any of such ICI models. It was shown that QPN theory can be applied to these models in order to characterize model behavior in terms of influences and synergies. By making use of difference functions and an order on Boolean tuples we were able to derive both the conditions under which positive, negative, zero, non-monotonic and ambiguous signs for qualitative influences, additive synergies and product synergies are observed and the constraints these signs impose on the underlying interaction functions.

The theory developed in this paper allows one to identify whether a particular ICI model with a chosen interaction function can fulfill the specified qualitative properties in principle. This is a useful development since without the theory one would need to estimate the conditional probabilities $P(\hat{m}'\hat{c})$ for each of the causes and exhaustively compute the influences and synergies for the model as in Section 2.3. By means of the theory, the qualitative behavior can be read off directly from the underlying interaction function.

The developed theory can also be employed for placing direct constraints on the structure of the underlying interaction function *given* a qualitative specification in terms of influences and synergies, as demonstrated by Tables 1–3. These results can also be used to generate the set of interaction functions that respect the constraints which facilitates the selection of a suitable interaction function for problems that can be represented as ICI models. Even though there exist a superexponential number of interaction functions, it may still be possible to identify a small set of interaction function that respects a (partial) qualitative specification. This results from the fact that most interaction functions are characterized by ambiguous influences and synergies whereas this is not to be expected for most real world models.

In conclusion, we believe that the theory can aid in Bayesian network construction, where the prognostic model served as an example to illustrate the usefulness of the theory in practice. If the ICI assumptions hold then the appropriateness of an interaction function can be determined without the need to specify the parameters in advance and properties of the interaction function can be derived from a qualitative specification.

Appendix A

Lemma 32. *If V_g^+ and $\exists \mathbf{m}_1[g(\mathbf{m}_1) = -1]$ then we can find parameters such that $\delta_{C\to E}[f] = +$.*

Proof. It suffices to prove that $\forall \mathbf{c}[\delta_{C\to E}[f](\mathbf{c}) \geqslant 0]$ for some choice of the parameters. We know that there must be some configuration $\mathbf{m}_1$ with $g(\mathbf{m}_1) = 1$ and for all configurations $\mathbf{m}_1'' < \mathbf{m}_1$ it holds that $g(\mathbf{m}_1'') \in \{0, 1\}$. We assume that $\forall \mathbf{m}_1'' < \mathbf{m}_1[g(\mathbf{m}_1'') = 0]$ and $\forall \mathbf{m}_1' > \mathbf{m}_1[g(\mathbf{m}_1') = -1]$, which minimizes $\mathrm{P}[g](e|\mathbf{c}_1)$. The incomparable configurations must be either zero or positive (otherwise a non-monotonic qualitative influence is implied) such that these cannot contribute negatively. We therefore obtain

$$\mathrm{P}[g](e|\mathbf{c}_1) \geqslant \prod_{C\in\mathbf{C}_1} P(m_C|\hat{c})^{\hat{m}_C} P(\bar{m}_C|\hat{c})^{1-\hat{m}_C} - \sum_{\mathbf{m}_1'>\mathbf{m}_1} P(\mathbf{m}_1'|\mathbf{c}_1). \tag{A.1}$$

By choosing $\mathrm{P}(m_C|c) = 1$ for each C such that $M_C = \top$, we obtain

$$\mathrm{P}[g](e|\mathbf{c}_1) \geqslant \prod_{C\in\mathbf{C}_1} P(\bar{m}_C|\hat{c})^{1-\hat{m}_C} - \sum_{\mathbf{m}_1'>\mathbf{m}_1} \prod_{C\in\mathbf{C}_1} P(m_C|\hat{c})^{(1-\hat{m}_C)\hat{m}_C'} P(\bar{m}_C|\hat{c})^{(1-\hat{m}_C)(1-\hat{m}_C')}$$

due to the fact that if $M_C = \bot$ then $M_C' = \top$ or $M_C' = \bot$. Given that for each $\mathbf{m}_1'$ there must exist at least one cause $C_{u(\mathbf{m}_1')}$ with $u : \mathbb{B}^{n-1} \to \{1, \ldots, n\}$, such that $M_{u(\mathbf{m}_1')}' = \top$ and $M_{u(\mathbf{m}_1')} = \bot$, we obtain

$$\mathrm{P}[g](e|\mathbf{c}_1) \geqslant \prod_{C\in\mathbf{C}_1} P(\bar{m}_C|\hat{c})^{1-\hat{m}_C} - \sum_{\mathbf{m}_1'>\mathbf{m}_1} P(m_{u(\mathbf{m}_1')}|\hat{c}_{u(\mathbf{m}')}).$$

By distinguishing present and absent causes, we may write

$$\mathrm{P}[g](e|\mathbf{c}_1) \geqslant \prod_{C\in\mathbf{C}_1} P(\bar{m}_C|c)^{(1-\hat{m}_C)\hat{c}} - \sum_{\mathbf{m}_1'>\mathbf{m}_1} P(m_{u(\mathbf{m}_1')}|c_{u(\mathbf{m}_1')})^{\hat{c}_{u(\mathbf{m}_1')}} \cdot 0^{1-\hat{c}_{u(\mathbf{m}_1')}}.$$

A key step is to distinguish $\mathbf{C}_1$ into $\mathbf{C}_a = \{C|C \in \mathbf{C}_1, \forall \mathbf{m}_1' > \mathbf{m}_1[C \neq C_{u(\mathbf{m}_1')}]\}$ and $\mathbf{C}_b = \{C|C \in \mathbf{C}_1, \exists \mathbf{m}_1' > \mathbf{m}_1[C = C_{u(\mathbf{m}_1')}]\}$, such that

$$\mathrm{P}[g](e|\mathbf{c}_1) \geqslant \prod_{C\in\mathbf{C}_a} P(\bar{m}_C|c)^{(1-\hat{m}_C)\hat{c}} \prod_{C'\in\mathbf{C}_b} P(\bar{m}_{C'}|c')^{(1-\hat{m}_{C'})\hat{c}'} - \sum_{\mathbf{m}_1'>\mathbf{m}_1} P(m_{u(\mathbf{m}_1')}|c_{u(\mathbf{m}_1')})^{\hat{c}_{u(\mathbf{m}_1')}} \cdot 0^{1-\hat{c}_{u(\mathbf{m}_1')}}.$$

By choosing $P(\bar{m}_C|c) = q$ for each $C \in \mathbf{C}_a$ such that $M_C = \bot$ and choosing $P(m_{u(\mathbf{m}_1')}|c_{u(\mathbf{m}_1')}) = p$ for all $\mathbf{m}_1' > \mathbf{m}_1$, we obtain

$$\mathrm{P}[g](e|\mathbf{c}_1) \geqslant \prod_{C\in\mathbf{C}_a} q^{(1-\hat{m}_C)\hat{c}} \prod_{C'\in\mathbf{C}_b} (1-p)^{(1-\hat{m}_{C'})\hat{c}'} - \sum_{\mathbf{m}_1'>\mathbf{m}_1} p^{\hat{c}_{u(\mathbf{m}_1')}} \cdot 0^{1-\hat{c}_{u(\mathbf{m}_1')}}.$$

Let w be the cardinality of $\{\mathbf{m}_1'|\mathbf{m}_1' \in \mathbb{B}^{n-1}, \mathbf{m}_1' > \mathbf{m}_1\}$. As there are at most $n-1$ cause variables in $\mathbf{C}_1$, we obtain:

$$\mathrm{P}[g](e|\mathbf{c}_1) \geqslant q^n(1-p)^n - wp,$$

where w is the cardinality of $\{\mathbf{m}_1'|\mathbf{m}_1' \in \mathbb{B}^{n-1}, \mathbf{m}_1' > \mathbf{m}_1\}$. It follows from Bernoulli's inequality that $\mathrm{P}[g](e|\mathbf{c}_1) \geqslant q^n(1-np) - wp$, such that by choosing $p < \frac{q^n}{q^n n+w}$, we have ensured that $\mathrm{P}[g](e|\mathbf{c}_1) \geqslant 0$. As there must be at least one configuration of $\mathbf{C}_1$ for which $\mathrm{P}[g](e|\mathbf{c}_1) \neq 0$, we have proved the proposition. □

References

[1] G. Birkhoff, S. Mac Lane, A Survey of Modern Algebra, fifth ed., A.K. Peters, 1997.

[2] F.G. Cozman, Axiomatizing noisy-OR, in: R. López de Mántaras, L. Saitta (Eds.), European Conference on Artificial Intelligence, IOS Press, Amsterdam, 2004, pp. 979–980.

[3] F.J. Díez, Parameter adjustment in Bayes networks. The generalized noisy OR-gate, in: Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufman, San Francisco, CA, 1993, pp. 99–105.
[4] F.J. Díez, M.J. Druzdzel, Canonical Probabilistic Models for Knowledge Engineering, Technical Report CISIAD-06-01, UNED, Madrid, Spain, 2006.
[5] F.J. Díez, J. Mira, E. Iturralde, S. Zubillaga, DIAVAL: a Bayesian expert system for electrocardiography, Int. Artif. Intell. Med. 10 (1997) 59–73.
[6] M.J. Druzdzel, M. Henrion, Efficient reasoning in qualitative probabilistic networks, in: R. Fikes, W. Lehnert (Eds.), Proceedings of the Eleventh National Conference on Artificial Intelligence, AAAI Press, Menlo Park, CA, 1993, pp. 548–553.
[7] M.J. Druzdzel, M. Henrion, Intercausal reasoning with uninstantiated ancestor nodes, in: D.E. Heckerman, A. Mamdani (Eds.), Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Mateo, CA, 1993, pp. 317–325.
[8] I.J. Good, Good Thinking: the Foundations of Probability and its Applications, University of Minnesota Press, Mineapolis, MN, 1983.
[9] D.E. Heckerman, J. Breese, Causal independence for probability assessment and inference using Bayesian networks, IEEE Trans. Syst. Man Cybern. 26 (1996) 826–831.
[10] M. Henrion, Some practical issues in constructing belief networks, in: J.F. Lemmer, T. Levitt, L.N. Kanal (Eds.), Proceedings of the Third Conference on Uncertainty in Artificial Intelligence, Elsevier Science, New York, NY, 1989, pp. 161–173.
[11] M. Henrion, M.J. Druzdzel, Qualitative propagation and scenario-based approaches to explanation in probabilistic reasoning, in: P.P. Bonissone, M. Henrion, L.N. Kanal, J.F. Lemmer (Eds.), Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence, Elsevier Science, New York, NY, 1991, pp. 17–32.
[12] H.J. Kappen, J.P. Neijt Promedas, A Probabilistic Decision Support System for Medical Diagnosis, Technical Report, Stichting Neurale Netwerken, Nijmegen, The Netherlands, 2002.
[13] P.J.F. Lucas, Bayesian network modelling by qualitative patterns, Artif. Intell. 163 (2005) 233–263.
[14] P.J.F. Lucas, N.C. de Bruijn, K. Schurink, A. Hoepelman, A probabilistic and decision-theoretic approach to the management of infectious disease at the ICU, Artif. Intell. Med. 19 (2000) 251–279.
[15] J. Pearl, Fusion, propagation and structuring in belief networks, Artif. Intell. 29 (1986) 241–288.
[16] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, second ed., Morgan Kaufman, San Francisco, CA, 1988.
[17] Y. Peng, J.A. Reggia, Plausibility of diagnostic hypotheses, in: Proceedings of the Fifth National Conference on AI (AAAI-86), Philadelphia, PA, 1986, pp. 140–145.
[18] S. Renooij, Qualitative Approaches to Quantifying Probabilistic Networks, PhD Thesis, University of Utrecht, Utrecht, The Netherlands, 2001.
[19] M.A. Shwe, B. Middleton, D.E. Heckerman, M. Henrion, E.J. Horvitz, H.P. Lehmann, G.F. Cooper, Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base, Methods Inform. Med. 30 (4) (1991) 241–255.
[20] M.P. Wellman, Fundamental concepts of qualitative probabilistic networks, Artif. Intell. 44 (1990) 257–303.