UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

An Information-Theoretic Approach for the Evaluation of a Time-to -Event Surrogate Endpoints for a Time-to-Event True Endpoint, Based on Causal Inference

Florian Stijven

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :

Prof. dr. Geert MOLENBERGHS
Prof. dr. Ingrid VAN KEILEGOM
Prof. dr. Ariel ALONSO ABAD

2021
2022

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

An Information-Theoretic Approach for the Evaluation of a Time-to -Event Surrogate Endpoints for a Time-to-Event True Endpoint, Based on Causal Inference

Florian Stijven

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :

Prof. dr. Geert MOLENBERGHS

Prof. dr. Ingrid VAN KEILEGOM

Prof. dr. Ariel ALONSO ABAD

Contents

Abstract

The endpoint that is used in a clinical trial is a key factor determining the trial's cost, complexity and duration. Ideally, the clinically most relevant endpoint is used, this is the so-called true endpoint. However, such endpoints often come with important issues. Alternatively, substitute endpoints could be used that address these issues; these are so-called surrogate endpoints. However, a surrogate endpoint should first be evaluated to ensure that it is an appropriate replacement for the true endpoint. Many methods have been proposed for this evaluation process.

In this thesis, the causal-inference approach to the evaluation of surrogate endpoints is extended to the setting with a time-to-event surrogate and true endpoint. The quality of the surrogate endpoint is quantified by the individual causal association (ICA), which is itself based on information-theoretic concepts. Rank-based measures for the ICA are also considered. A flexible model based on a D-vine copula is proposed for the vector of potential outcomes. However, due to the so-called fundamental problem of causal inference, the proposed model is not identifiable. These identifiability issues are tackled by a sensitivity analysis that results in bounds for the ICA. In this sensitivity analysis, the unidentifiable parameters are sampled from the region of the parametric space of the model that is compatible with the observed data. Further, additional assumptions are proposed that restrict this region to obtain tighter bounds for the ICA. The proposed methods are illustrated with an analysis of pooled data from four ovarian cancer trials. This analysis provides convincing evidence that progression-free survival is a good surrogate for overall survival in these four trials.

Introduction

Drug development is a lengthy, complex, and costly process, entrenched with a high degree of uncertainty whether the drug will actually succeed (Pankevich et al., 2014). An important contributing factor is the nature of the endpoint that is used to assess the treatment's efficacy. The *true endpoint* is the best possible indicator for treatment efficacy, or to put it differently, the clinically "most relevant" endpoint (Alonso, Van der Elst, & Meyvisch, 2017). However, true endpoints may have important issues. For instance, the true endpoint might require a long follow-up time such as overall survival (OS) in early cancer types. Amongst other things, a long follow-up time can cause posttreatment confounding and an increased chance of missing data. In addition, the evaluation of treatment efficacy is delayed, which delays a possible market authorization and patient access to a potentially effective treatment. Other issues include the true endpoint being costly to measure (e.g., certain imaging modalities) or having a low incidence (e.g., pregnancy in severe luteinizing hormone deficiency) (Van Der Elst, 2016).

Given the challenges surrounding the true endpoints in some trials, a seemingly attractive strategy is to replace the true endpoint by a "substitute endpoint" in which these issues are not present. Such a substitute endpoint is termed a *surrogate endpoint* (or *surrogate*). The potential of surrogate endpoints is widely recognized by regulatory agencies and medical researchers. Between 2010 and 2012, the FDA approved 45% of new drugs based on a surrogate endpoint (FDA, 2018). The surrogate endpoints that are deemed acceptable by the FDA are very diverse; a sample of those endpoints is given in Table 1 (FDA, 2022). Of course, not just any alternative endpoint can replace the true endpoint. At first sight, a strong association between the surrogate and true endpoint seems to be a sufficient criterion to justify the replacement of the true endpoint. This is however a common misconception that has had serious consequences in the past (Alonso et al., 2017). For example, long-term hormone replacement therapy has been found to lower "bad" cholesterol and raise "good" cholesterol in women; where high levels of "bad" cholesterol increase the cardiovascular risk and high levels of "good" cholesterol decrease the cardiovascular risk. At the same time, this therapy also increased the cardiovas-

Table 1: Example of surrogate endpoints that the FDA considers as acceptable, adopted from FDA (2022).

Disease or Use	Patient Population	Surrogate Endpoint
Anthrax vaccine	Persons at high risk of exposure to anthrax	Anti-protective antigen antibody
Hematological malignancies	Patients with Acute Lymphoblastic Leukemia	Serum asparaginase
Hematological malignancies	Patients with diffuse large B-cell lymphoma	Event-free survival (EFS)
Solid tumors	Patients with nonmetastatic castrate-resistant prostate cancer	Metastasis-free survival

cular risk (for the Women's Health Initiative Investigators et al., 2002). Hence, cholesterol level was not a valid surrogate for cardiovascular risk in that setting. These unfortunate experiences highlight the need for a proper statistical evaluation of surrogate endpoints.

The statistical evaluation of surrogate endpoints is a non-trivial endeavor. Indeed, various methods have been proposed over the last three decades, some of which are now known to be insufficient. Nowadays, the meta-analytic approach is the "gold standard" for the statistical evaluation of surrogate endpoints. However, its use is hindered in some settings by its strong data requirements. Indeed, this approach requires *patient-level* data from *multiple* clinical trials (Alonso et al., 2017).

More recently, Alonso et al. (2015) proposed a new approach for the evaluation of surrogate endpoints, based on causal-inference ideas. This approach is further referred to as the *causal-inference framework*. As opposed to the meta-analytic framework, this framework only requires patient-level data from a single clinical trial. This framework is however not as well-developed as the meta-analytic framework. Indeed, it has only been developed for the gaussian-gaussian and binary-binary setting (Alonso et al., 2015, 2016). In this thesis, the causal-inference framework is extended to the survival-survival setting. In essence, the setting where a time-to-event surrogate is evaluated for a time-to-event true endpoint.

In the first chapter, a brief overview of different surrogate evaluation methods is given. The models proposed in this thesis are based on vine copulas, therefore, the necessary theoretic concepts regarding vine copulas are described in chapter 2. Next, models for the survival-survival setting in the causal-inference framework are proposed in chapter 3. These methods are illustrated in chapter 4 with data from clinical trials in advanced ovarian cancer. Finally, some concluding remarks are formulated.

Chapter 1

Overview of Surrogate Validation Methods

Since Prentice's seminal paper in 1989, the evaluation of surrogate endpoints has received much attention in the statistical literature (Prentice, 1989). The initially proposed methods were grounded in the single-trial setting and are now recognized to be insufficient for a proper evaluation of surrogate endpoints. Some of the subsequent methods are grounded in the multiple trial setting and are now recognized as the gold standard. Other contemporary methods are still grounded in the single-trial setting and are based on causal-inference ideas. A brief overview of different surrogate evaluation methods is given in this chapter. In the last section, the causal-inference framework is introduced. In the remainder of this thesis, the surrogate and true endpoint are respectively denoted by S and T. The treatment is denoted by Z.

1.1 Early Single-Trial Setting Approaches

1.1.1 Prentice's Definition

Prentice (1989, p. 432) defined a surrogate endpoint as "a response variable for which a test of the null hypothesis of no relationship to the treatment groups is also a valid test of the corresponding null hypothesis based on the true endpoint". This is formalized as follows:

$$f(S|Z) = f(S) \Leftrightarrow f(T|Z) = f(T). \tag{1.1}$$

This definition is appealing at first sight. It corresponds to how a surrogate is intended to be used in practice: instead of testing the treatment effect on the true endpoint, the treatment effect is tested on the surrogate endpoint. Indeed, according to Prentice's definition, a treatment effect on the surrogate is a necessary and sufficient condition for a treatment effect on the true endpoint. Prentice (1989) proposed four operational criteria to evaluate whether the definition in Equation 1.1 is fulfilled for a given surrogate and true endpoint. The first two criteria require a treatment effect on both the surrogate and true endpoint.

$$f(S|Z) \neq f(S) \tag{1.2}$$

$$f(T|Z) \neq f(T) \tag{1.3}$$

The third criterion requires the surrogate and true endpoint to be statistically dependent.

$$f(T|S) \neq f(T) \tag{1.4}$$

This entails that the surrogate is prognostic, a condition that any reasonable surrogate should satisfy (Buyse et al., 2000). The fourth criterion requires the treatment to be irrelevant in predicting the true endpoint given the surrogate endpoint.

$$f(T|S, Z) = f(T|S) \tag{1.5}$$

There are several conceptual and practical issues with Prentice's approach. Only some of them are briefly discussed in this thesis, an extensive discussion is provided by Freedman, Graubard, and Schatzkin (1992), Buyse and Molenberghs (1998) and Burzykowski (2001). A fundamental issue is related to the fourth criterion, which requires proving a null hypothesis. Assuming a linear model, $E(T|S,Z) = \beta_0 + \beta_1 \cdot S + \beta_2 \cdot Z$, the fourth criterion requires proving $\beta_2 = 0$. Failing to reject the corresponding null hypothesis does not prove that $\beta_2 = 0$ holds. Another issue pertains to the relation between Prentice's definition and the four operational criteria. The operational criteria are only necessary and sufficient for Prentice's definition in the case of binary endpoints (Buyse & Molenberghs, 1998). Ignoring above issues, the fourth criterion is generally still too restrictive. In fact, it requires that the treatment does not act on the true endpoint, through pathways bypassing the surrogate (Prentice, 1989). Given the complexity of disease-treatment pathways, it is unrealistic that a single endpoint would *fully* capture the effect of the treatment on the true endpoint. Because of these issues, new metrics of surrogacy were proposed in the single trial setting. These are discussed next.

1.1.2 Building upon Prentice's Proposal

Freedman, Graubard, and Schatzkin (1992) proposed an extension of Prentice's approach that moves away from hypothesis testing to estimation. By doing so, Prentice's fourth criterion (Equation 1.5) is relaxed. This is done by estimating the *proportion of treatment effect explained* (PE):

$$PE = \frac{\beta - \beta_S}{\beta} \tag{1.6}$$

where β_S is the treatment effect adjusted for the surrogate endpoint, and β is the unadjusted treatment effect. Prentice's fourth criterion requires that $\beta_S = 0$ or equivalently that $PE = 1$. That indicates that "100% of the treatment effect is explained" by the surrogate endpoint (Freedman, Graubard, & Schatzkin, 1992).

If the treatment effect is not fully captured by the surrogate, Prentice's fourth criterion is not satisfied. However, the surrogate might still be useful if *a large proportion* of the treatment effect is captured by the surrogate. Consequently, a confidence interval can be constructed around the estimate for PE. If the lower limit is sufficiently large, the surrogate is deemed valid. This approach seems attractive at first sight because PE has a direct interpretation, and the conditions required for a surrogate to be valid are more realistic than in Prentice's approach. Nonetheless, the confidence interval for PE is generally too wide to be of much practical value. Another conceptual issue is that the PE is not truly a proportion. Indeed, if the direction of the treatment effect changes after adjusting for the surrogate, PE is greater than one (Buyse & Molenberghs, 1998). A more comprehensive appraisal of the PE is given by Alonso et al. (2017, p. 40)

Later, Buyse and Molenberghs (1998) argued that Prentice's definition as well as the extension of Freedman, Graubard, and Schatzkin (1992) are too limited for a complete evaluation of a surrogate endpoint. Buyse and Molenberghs (1998) therefore proposed two other quantities for evaluating surrogate endpoints in the single-trial setting: (i) the relative effect (RE) and (ii) the adjusted association (AA).

The RE is motivated by the following consideration of what constitutes a good surrogate: "the investigators must be able to predict the effect of treatment on the true endpoint based on the observed effect of treatment on the surrogate" (Buyse & Molenberghs, 1998, p. 1022). The RE is defined as follows:

$$RE = \frac{\beta}{\alpha} \tag{1.7}$$

where α is the effect of treatment on the surrogate, and β is the effect of treatment on the true endpoint (e.g., regression coefficients in a linear model). If $RE = 1$, the surrogate is termed *perfect at the population level*. The AA is defined as the association between S and T after

adjusting for treatment. The exact definition depends on how this association is measured. If the association is maximal (i.e., a deterministic relationship), then the surrogate is termed *perfect at the individual level* (Buyse & Molenberghs, 1998).

If a bivariate normal linear regression model is assumed for S and T, it can be shown that the PE is a composite of RE and AA:

$$PE = \frac{\sigma_T}{\sigma_S} \cdot \frac{AA}{RE} \tag{1.8}$$

with σ_T^2 and σ_S^2 the residual variances of $T|Z$ and $S|Z$, respectively. The use of the RE and AA thus allows for a more detailed assessment than PE (Buyse & Molenberghs, 1998). Still, using the RE to predict the effect of treatment on T rests on the unverifiable assumption that there is a linear regression through the origin. This assumption is unverifiable since only one observation on this line is available. One solution is to ensure replication at the trial level such that the regression line is fitted with several data points. This is done in the meta-analytic approach (Buyse et al., 2000).

1.2 Contemporary Methods

1.2.1 Meta-Analytic Approach

In the meta-analytic approach, surrogacy is evaluated using patient-level data from multiple clinical trials. The adjusted association and relative effect are extended to the *individual-level association* and the *trial-level association*, respectively (Buyse et al., 2000). The individual-level association quantifies the association between the surrogate and true endpoint after adjusting for trial and treatment. The trial-level association quantifies the association between the trial-level treatment effects on the surrogate and the true endpoint (Alonso et al., 2017).

The meta-analytic approach is briefly explained here for Gaussian endpoints. Although this approach is not the subject of this thesis, there are some interesting connections between the meta-analytic approach and the causal-inference framework. Assume we have data from N trials with n_i patients in the i'th trial. S_{ij} and T_{ij} are the surrogate and true endpoint for the j'th patient in the i'th trial. Z_{ij} is the corresponding (binary) treatment indicator. The following linear mixed model is considered:

$$\begin{aligned} S_{ij}|Z_{ij} &= \mu_S + m_{S_i} + (\alpha + a_i)Z_{ij} + \epsilon_{S_{ij}} \\ T_{ij}|Z_{ij} &= \mu_T + m_{T_i} + (\beta + b_i)Z_{ij} + \epsilon_{T_{ij}} \end{aligned} \tag{1.9}$$

where $(\epsilon_{S_{ij}}, \epsilon_{T_{ij}})'$ are mean-zero normally distributed correlated error terms. $(m_{S_i}, m_{T_i}, a_i, b_i)'$ is a random-effects vector that follows a mean-zero normal distribution with an unstructured covariance matrix. The trial specific intercepts are $\mu_S + m_{S_i}$ and $\mu_T + m_{T_i}$ for the surrogate and true endpoint, respectively. The trial specific treatment effects are $\alpha + a_i$ and $\beta + b_i$ for the surrogate and true endpoint, respectively. The individual-level surrogacy is quantified by the coefficient of determination R^2_{indiv} pertaining to the distribution of $\epsilon_{T_{ij}}$ conditional on $\epsilon_{S_{ij}}$. This measure quantifies how well T_{ij} can be predicted based on the observed value of S_{ij} after adjusting for treatment and trial. The trial-level surrogacy is quantified by the coefficient of determination R^2_{trial} pertaining to the distribution of b_i conditional on m_{S_i} and α_i. The latter essentially quantifies how well the treatment effect on the true endpoint can be predicted given the observed treatment effect on the surrogate endpoint (Buyse et al., 2000).

This approach is appealing because it addresses some of the issues in the early single-trial setting approaches of Section 1.1. Moreover, the meta-analytic approach is well-developed for many types of surrogate and true endpoints (Alonso et al., 2017). However, this approach requires patient-level data from multiple clinical trials which are often not available. Although summary-level data are generally more readily available (i.e., treatment effect estimates reported in literature), methods for using both patient and summary-level data are not available.

1.2.2 Principal Stratification

A different approach to evaluating surrogates in the single-trial setting was proposed by Frangakis and Rubin (2002). This approach is based on principal stratification and the notion of posttreatment variables. In the Neyman-Rubin potential outcomes framework, each patient has a four-dimensional vector of potential outcomes $(S_0, T_0, S_1, T_1)'$ with S_k and T_k representing the outcome of the individual under treatment $Z = k$. Under the stable unit treatment value assumption (SUTVA), the following holds

$$\begin{aligned} S^{obs} &= ZS_0 + (1 - Z)S_1 \\ T^{obs} &= ZT_0 + (1 - Z)T_1 \end{aligned} \tag{1.10}$$

where S^{obs} and T^{obs} are the observed outcomes for a patient. SUTVA is generally evident in clinical trials, but requires extra thought in some contexts, e.g., trials in infectious diseases. The four-dimensional vector is, however, never fully observed. Only $(S_0, T_0)'$ or $(S_1, T_1)'$ can be observed for individual patients. Holland (1986) termed this the *fundamental problem of causal-inference*.

Prentice's approach, and extensions thereof, are based on models that adjust the treatment effect for the posttreatment variable S^{obs}, the observed surrogate endpoint. That adjusted treatment effect is based on a comparison of

$$P(T|S^{obs} = s, Z = 0) \quad \text{and} \quad P(T|S^{obs} = s, Z = 1). \tag{1.11}$$

This is not a causal comparison. Under randomized treatment assignment and SUTVA, this comparison can equivalently be written as a comparison of

$$P(T_0|S_0 = s) \quad \text{and} \quad P(T_1|S_1 = s). \tag{1.12}$$

The sets of individuals that are included in the first and second condition of this comparison are generally not the same. Because different sets of individuals are being compared, the corresponding estimands generally do not have a causal interpretation (Frangakis & Rubin, 2002). The more fundamental reason why this comparison is not causal, is that one is conditioning on a *posttreatment* variable. Randomization ensures that treatment assignment and pretreatment variables are independent, but *not* that treatment assignment and posttreatment variables are independent. For a causal comparison after adjustment for posttreatment variables, Frangakis and Rubin (2002) proposed to adjust for principal strata of the posttreatment variables (see further).

Frangakis and Rubin (2002) proposed to evaluate surrogacy based on principal stratification. Principal stratification allows for a causal comparison when adjusting for posttreatment variables. The surrogate endpoint is indeed a posttreatment variable. This approach boils down to controlling for the vector of potential surrogate outcomes $(S_0, S_1)'$ instead of the observed surrogate endpoint S^{obs}. The vector $(S_0, S_1)'$ is independent of the treatment assignment for a randomized experiment, while S^{obs} is generally not. Therefore, a comparison that controls for $(S_0, S_1)'$ is causal, whereas the comparison that controls for S^{obs} is generally not causal. The following definition of surrogacy is proposed for a *principal surrogate* (Frangakis & Rubin, 2002, p. 26).

Definition 1.2.1 (Principal surrogate)**.** *S is a principal surrogate for a comparison of the effect of $Z = 0$ versus $Z = 1$ on T if, for all fixed s, the comparison between the ordered sets*

$$\{T_0 : S_0 = S_1 = s\} \quad \textit{and} \quad \{T_1 : S_0 = S_1 = s\} \tag{1.13}$$

results in equality.

This entails that a causal treatment effect on the true endpoint cannot exist when there is no causal treatment effect on the surrogate endpoint. In other words, a causal treatment effect on the true endpoint can only exist when there is a causal treatment effect on the surrogate (Frangakis & Rubin, 2002). Associative and dissociative effects are also defined in this approach. The *dissociative effect* is a comparison between

$$\{T_0 : S_0 = S_1 = s\} \quad \text{and} \quad \{T_1 : S_0 = S_1 = s\}. \tag{1.14}$$

From Definition 1.2.1, it follows that the above comparison results in an equality for a principal surrogate. The *associative effect* is a comparison between

$$\{T_0 : S_0 \neq S_1\} \quad \text{and} \quad \{T_1 : S_0 \neq S_1\}. \tag{1.15}$$

These effects have an appealing interpretation. If there is a large dissociative effect, then there is a large treatment effect on T for subjects for whom treatment does not affect S. If there is a large associative effect, then there is a large causal treatment effect on T for subjects for whom treatment also affects S. A small dissociative and a large associative effect are desirable properties for a surrogate.

Despite the appealing properties of the principal stratification approach, there is an important limitation. In practice, the vector $(S_0, S_1)'$ is unobservable. Indeed, S_0 and S_1 are never simultaneously observed for the same patient. The comparisons as defined in Equations 1.13-1.15 are thus not identifiable. Several strategies have been proposed to tackle this unidentifiability issue (see also Section 1.4.2). For example, Li, Taylor, and Elliott (2010) reduced the non-identifiability problem with additional (unverifiable) assumptions and by incorporating prior belief in a Bayesian model.

It can further be shown that a principal surrogate is not generally a valid surrogate in Prentice's framework and vice versa (Frangakis & Rubin, 2002). The causal-inference approach to surrogacy as presented in this thesis is based on the principal stratification framework. The same unidentifiability problem arises thus in the causal-inference approach, but it is addressed by a sensitivity analysis (see further).

1.3 Related Causal Frameworks

Joffe and Greene (2009, p. 530) view a surrogate outcome as "an outcome for which knowing the effect of treatment on the surrogate allows prediction of the effect of treatment on the more clinically relevant outcome". They identify two complementary causal paradigms in which surrogates can be evaluated: (i) the causal-association (CA) paradigm and (ii) the causal-effects (CE) paradigm.

1.3.1 Causal-Association Paradigm

In the CA paradigm, "evaluation of a surrogate is based on examination of the association between the effect of the treatment on the putative surrogate and the effect of the treatment on the clinical outcome" (Joffe & Greene, 2009, p. 533). This allows for predicting the treatment effect on the true endpoint based on the observed treatment effect on the surrogate. Both the meta-analytic and the principal stratification approach resort under this paradigm. The former concerns trial-level treatment effects in multiple trials, whereas the latter concerns individual causal effects in a single trial.

Note that these approaches do not model the effect of S on T. Accordingly, confounding of the $S \to T$ relation is not relevant. In fact, only the $Z \to S$ and the $Z \to T$ relations are relevant in the CA paradigm. In randomized trials, these relations are always causal. Thus, no additional assumptions, other than randomized treatment assignment, are required in the CA paradigm.

(a) No confounding (b) Confounding

Figure 1.1: Two possible causal diagrams representing causal relations in a surrogate validation study.

As explained in section 1.2.1, the meta-analytic approach examines the relationship across studies between the treatment effects on the surrogate and true endpoint. Let β_i and α_i denote the treatment effect on respectively T and S in the i'th study. This is represented in the causal graphs of Figure 1.1. The linear regression of β_i on α_i is as follows:

$$E(\beta_i|\alpha_i) = \gamma_0 + \gamma_1 \cdot \alpha_i. \tag{1.16}$$

A good surrogate in the meta-analytic approach satisfies $\gamma_0 \approx 0$ and $\gamma_1 \neq 0$ where simultaneously a large proportion of the variability in β_i is explained by α_i. In a new trial, a near-zero treatment effect on the surrogate then implies a near-zero treatment effect on the true endpoint as well. In addition, the fitted regression line can be used to predict the treatment effect on the true endpoint in a new trial given the observed treatment effect on the surrogate.

This is very similar to the principal surrogacy paradigm. Instead of treating trial-level treatment effects as the building blocks, patient-level treatment effects $T_1 - T_0$ and $S_1 - S_0$ are the primary building blocks in the principal surrogacy paradigm (Joffe & Greene, 2009). The relation between $T_1 - T_0$ and $S_1 - S_0$ is thus of primary interest, e.g., under linearity:

$$E(T_1 - T_0|S_1 - S_0) = \eta_0 + \eta_1 \cdot (S_1 - S_0). \tag{1.17}$$

Principal surrogacy requires there to be no effect on the true endpoint when there is no effect on the surrogate (Definition 1.2.1): $\eta_0 = 0$ in Equation 1.17. Moreover, a non-zero associative effect entails that $\eta_1 \neq 0$. Let us redefine α_i and β_i as the individual causal effects for patient i. Then Figure 1.1 also applies to the principal surrogacy paradigm. This paradigm can thus be seen as the patient-level analog of the meta-analytic approach.

Note that for the meta-analytic as well as the principal surrogacy framework the results are valid under both causal diagram 1.1a and 1.1b. Unmeasured confounding of the $S \to T$ relation is thus not problematic in these frameworks.

1.3.2 Causal-Effects Paradigm

In the CE paradigm, "knowledge of the effects of the treatment on the surrogate and the surrogate on the clinical outcome is used to predict the effect of the treatment on the clinical outcome" (Joffe & Greene, 2009, p. 530). Prentice's approach and an approach based on direct and indirect effects resort under this paradigm.

Joffe and Greene (2009) show that Prentice's approach has ideas in common with the partitioning of the effect of a treatment into direct and indirect effects. The direct effect of treatment with respect to a causal intermediate is that part of the effect that is not mediated by the intermediate, in this case the surrogate. This is represented by $Z \to T$ in Figure 1.1. The indirect effect of treatment is the part that is mediated by the surrogate; this is represented by $Z \to S \to T$ in Figure 1.1.

If there is no direct effect, it seems that statistically controlling for S would block any association between Z and T, i.e., $Z \perp\!\!\!\perp T|S$ which is Prentice's fourth criterion. This is only true if

causal diagram 1.1a holds. In general, this is wrong, which was already shown in Section 1.2.2 by using principal stratification. Two other, more intuitive, reasons why this is wrong in general are given here. First, causal diagram 1.1a is usually not a complete representation of reality. In randomized trials, no factor can influence treatment allocation. There can thus be no common cause for treatment allocation and another variable. However, this does not hold for posttreatment variables such as S. Causal diagram 1.1b is generally a better representation of reality, but possibly still a simplification. In causal diagram 1.1b, U represent variables that confound the relation between S and T. When U is not accounted for in the statistical model, conditioning on S induces a spurious association between Z and T because S is a collider: $Z \rightarrow S \leftarrow U \rightarrow T$. Consequently, Prentice's framework and extensions thereof can only have causal interpretation if causal diagram 1.1a holds. Even so, under this notion of causality, causal effects are defined under manipulations of S which is problematic in its own right, as discussed next.

Second, statistical control is mixed with experimental control. If there is no direct effect of Z on T, then physically holding S fixed at a certain value would mean that Z and T are statistically independent *under that manipulation* (Joffe & Greene, 2009). This is formalized in the notion of direct and indirect effects, see for example Pearl (2009, Chapter 5). The direct and indirect effects defined therein require the notion of manipulating Z and S. This is conceptually hard to justify for S; e.g. what would manipulating time-to-progression mean?

1.3.3 Comparison of the CE and the CA Paradigms

The CE mode of reasoning is mechanistic, it involves a chain of variables in which each variable causally affects the next one(s). This paradigm thus explains the effect of treatment mechanistically and consequently offers an appealing interpretation of the surrogate: the treatment affects T by affecting S. However, causal effects are defined in terms of manipulations of the surrogate. This is often conceptually hard to justify. Moreover, many methods in the CE paradigm require very strong assumptions. The CA approach generally does not require additional assumptions besides randomized treatment allocation. The CA approach is, in comparison with the CE approach, somewhat "black box" in nature because it cannot explain the effect of treatment mechanistically (Joffe & Greene, 2009).

1.4 Causal-Inference Framework

The causal-inference framework is closely connected to the principal surrogacy framework. In these two frameworks, the same notation and assumptions regarding potential outcomes are used, although a different definition for what constitutes a "good" surrogate is used in the causal-inference framework. In the latter, the association between ΔS and ΔT is of primary interest, and a quantification of this association is the primary measure of surrogacy.

1.4.1 Potential Outcomes Framework

In the causal-inference framework, the validation exercise is carried out in a single trial with a well-defined population. The results of the validation exercise hold, strictly speaking, only in this well-defined population. Applying these results in a new trial, with a different population and/or other treatments, would necessarily require a degree of extrapolation. Further, it is assumed that only two treatments are under evaluation ($Z = 0/1$) in a parallel study design. The potential outcomes and underlying assumptions were already introduced in Section 1.2.2 for the principal surrogacy framework.

The individual causal effects are the building blocks of the causal-inference approach (Alonso et al., 2015). They are defined for the true endpoint as $\Delta T = T_1 - T_0$ and analogously for the surrogate endpoint $\Delta S = S_1 - S_0$. The individual causal effects cannot be observed, see also Section 1.2.2, although the *expected* treatment effects, $E(\Delta T)$ and $E(\Delta S)$, are identifiable from

the observed data if $Z \perp\!\!\!\perp (T_0, T_1)$. This latter condition is satisfied for randomized trials. The expected causal treatment effects at the trial level are the primary building blocks in the meta-analytic approach and were previously denoted by β_i and α_i for the i'th trial (e.g., Equation 1.16).

1.4.2 General Approach

Model

The first step in the causal-inference framework is to assume a multivariate model for the vector of potential outcomes:

$$(S_0, T_0, S_1, T_1)' \sim F(\cdot; \boldsymbol{\theta}) \tag{1.18}$$

where $\boldsymbol{\theta}$ is the corresponding parameter vector. Not all elements of $\boldsymbol{\theta}$ are identifiable. Only $(S_0, T_0)'$ or $(S_1, T_1)'$ can be simultaneously observed. Therefore, only the corresponding bivariate distributions

$$(S_0, T_0)' \sim F(S_0, T_0; \boldsymbol{\theta_0}) \quad \text{and} \quad (S_1, T_1)' \sim F(S_1, T_1; \boldsymbol{\theta_1}) \tag{1.19}$$

are identifiable. Let $\boldsymbol{\theta_0}$ and $\boldsymbol{\theta_1}$ denote the corresponding parameter vectors of these bivariate distributions. These parameters can be estimated with the observed data, e.g., via maximum likelihood estimation. Let $\boldsymbol{\theta_n}$ be the set of parameters of $\boldsymbol{\theta}$ not contained in $\boldsymbol{\theta_0}$ or $\boldsymbol{\theta_1}$. Because the elements of $\boldsymbol{\theta_n}$ do not appear in the observable distributions of Equation 1.19, $\boldsymbol{\theta_n}$ is unidentifiable. This unidentifiability issue is further addressed through a sensitivity analysis.

Measures of Surrogacy

From this model, measures of surrogacy are derived that are based on the association between the individual causal effects: ΔT and ΔS. These are termed the *individual causal association* (ICA). In principle, any association measure can be used to quantify this association. The choice for a particular association measure depends on the type of endpoints considered (e.g., binary or continuous). Information theory provides a unifying framework to quantify this association across settings with different types of endpoints (Alonso et al., 2017).

These measures of surrogacy are functions of the joint distribution of $(\Delta S, \Delta T)$. This joint distribution is in turn a function of $F(\cdot; \boldsymbol{\theta})$ in Equation 1.18. Because $\boldsymbol{\theta_n}$ is not identifiable, the measures of surrogacy are also not identifiable. There are three ways to address this identifiability issue: (i) use unverifiable assumptions to make the model identifiable, (ii) use a Bayesian estimation framework with (weakly) informative priors for the unidentifiable parameters, and (iii) implement a sensitivity analysis in which the ICA is computed across a set of plausible values for the unidentifiable parameters $\boldsymbol{\theta_n}$ (Alonso et al., 2015). The sensitivity analysis approach is used in this thesis.

Sensitivity Analysis

The sensitivity analysis consists of a two-step Monte Carlo procedure. Let Γ be the parameter space of $\boldsymbol{\theta}$ in Equation 1.18. In a first step, the region of this parameter space that is compatible with the observed data is determined; this is the so-called Γ_D region where $\Gamma_D \subset \Gamma$. Usually, $\boldsymbol{\theta_0}$ and $\boldsymbol{\theta_1}$ are fixed at their estimated values, $\hat{\boldsymbol{\theta}}_{\boldsymbol{0}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{1}}$. Whereas $\boldsymbol{\theta_n}$ can vary freely. Note that in many cases, the values at which $\boldsymbol{\theta_0}$ and $\boldsymbol{\theta_1}$ are fixed, restrict the region in which $\boldsymbol{\theta_n}$ can vary. Indeed, given $\hat{\boldsymbol{\theta}}_{\boldsymbol{0}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{1}}$, there can be restrictions on $\boldsymbol{\theta_n}$ to ensure that Equation 1.18 still represents a valid distribution: e.g., ensuring a positive definite correlation matrix or cell probabilities in $[0, 1]$.

In the second step, a Monte Carlo approach is implemented to study the behavior of the ICA in this region Γ_D. This is done by sampling the unidentifiable parameters on Γ_D and computing the ICA for each sample of unidentifiable parameters (Alonso, 2018). Each point in Γ_D can be conceptualized as a "world" compatible with ours. The behavior of the ICA on Γ_D thus completely describes the validity of the putative surrogate across all "worlds" compatible with the observed data (Alonso et al., 2016).

One further consideration is the introduction of additional restrictions on Γ_D. Based on the study design or scientific background knowledge, additional restrictions on Γ_D could be justified. For example, Li, Taylor, and Elliott (2010) used the monotonicity assumption to "assist in the identifiability". This assumption states that the treatment cannot worsen a patient's condition, as compared to control. This can be assumed for the surrogate and/or true endpoint. Where a higher value is desirable, the monotonicity assumption is formally written as follows: $S_0 \leq S_1$ and/or $T_0 \leq T_1$. Although these additional restrictions might be justified based on study design aspects or scientific background knowledge, they cannot be empirically verified. They are thus always unverifiable and should therefore be used with caution.

1.4.3 Information Theoretic Concepts

The concept of entropy is a key measure in information theory and quantifies the amount of uncertainty associated with a random variable. Let Y denote a random variable taking values $\{y_1, y_2, ..., y_m\}$ with probability mass function $P(Y = y_i) = p_i$. The entropy of Y is then defined as

$$H(Y) = -E_Y\{\log P(Y)\} = -\sum_{i=1}^{m} p_i \cdot \log p_i. \tag{1.20}$$

The entropy quantifies the amount of uncertainty associated with a random variable. Entropy is non-negative and invariant under bijective transformations. The joint entropy of $(X, Y)'$ is defined similarly as $H(X, Y) = -E_{X,Y}\{\log P(X, Y)\}$. The conditional entropy is also defined similarly as $H(Y|X) = -E_X[E_Y\{\log P(Y|X)\}]$. For continuous outcomes, the concept of entropy is extended to differential entropy. For a continuous random variable Y with density function $f_Y(y)$, the differential entropy $h_d(Y)$ is defined as

$$h_d(Y) = -E\{\log f_Y(Y)\} = -\int f_Y(y) \cdot \log f_Y(y) dy. \tag{1.21}$$

As opposed to entropy, differential entropy can be negative or positive and is coordinate dependent. The joint and conditional entropy are defined analogously as in the discrete case.

A measure of association based on information theory is the mutual information, denoted by $I(X, Y)$. The mutual information between X and Y quantifies the amount of uncertainty in Y expected to be removed if the value of X becomes known: $I(X, Y) = H(Y) - H(Y|X)$ or $I(X, Y) = h_d(Y) - h_d(Y|X)$ for the discrete and continuous case, respectively. The mutual information is symmetric in the sense that it equally well quantifies the amount of uncertainty in X expected to be removed if the value of Y were known: $I(X, Y) = H(X) - H(X|Y)$ or $I(X, Y) = h_d(X) - h_d(X|Y)$. From these definitions, it follows for the continuous case that

$$\begin{aligned} I(X, Y) &= h_d(X) - h_d(X|Y) \\ &= h_d(Y) - h_d(Y|X) \\ &= \int\int f(x, y) \cdot \log\left(\frac{f(x, y)}{f(x)f(y)}\right) dxdy. \end{aligned} \tag{1.22}$$

In the causal-inference framework, the following definition of surrogacy is proposed (Alonso, 2018, p. 3).

Definition 1.4.1 (Surrogacy in the causal-inference framework)**.** *In the single-trial setting, we shall say that S is a good surrogate for T if ΔS conveys a substantial amount of information on ΔT.*

The amount of information shared between ΔT and ΔS is quantified by the mutual information $I(\Delta T, \Delta S)$. It is therefore proposed to assess surrogacy in the causal-inference framework based on the mutual information (Alonso, 2018). The mutual information itself is however difficult to interpret as it lies in $[0, +\infty[$ for continuous endpoints and in $[0, \min\{H(\Delta S), H(\Delta T)\}]$ for discrete endpoints. To obtain a metric of surrogacy with appealing properties, the mutual information is transformed to a metric that lies in $[0, 1]$. This transformation depends on the scale of measurement of both endpoints.

1.4.4 Gaussian-Gaussian Setting

Alonso et al. (2015) addressed the setting where both S and T are continuous. This approach is briefly outlined next.

Model

A multivariate normal distribution is assumed for $(T_0, T_1, S_0, S_1)' \sim N(\boldsymbol{\mu}, \Sigma)$ where

$$\Sigma = \begin{pmatrix} \sigma_{T_0T_0} & \sigma_{T_0T_1} & \sigma_{T_0S_0} & \sigma_{T_0S_1} \\ \sigma_{T_1T_0} & \sigma_{T_1T_1} & \sigma_{T_1S_0} & \sigma_{T_1S_1} \\ \sigma_{S_0T_0} & \sigma_{S_0T_1} & \sigma_{S_0S_0} & \sigma_{S_0S_1} \\ \sigma_{S_1T_0} & \sigma_{S_1T_1} & \sigma_{S_1S_0} & \sigma_{S_1S_1} \end{pmatrix}$$

and $\boldsymbol{\mu} = (\mu_{T_0}, \mu_{T_1}, \mu_{S_0}, \mu_{S_1})'$. Given these distributional assumptions, the following holds for the vector of individual causal effects:

$$\boldsymbol{\Delta} = \boldsymbol{AY} = \begin{pmatrix} T_1 - T_0 \\ S_1 - S_0 \end{pmatrix} \sim N(\boldsymbol{\mu_\Delta}, \Sigma_\Delta) \quad \text{where} \quad \boldsymbol{A} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix} \tag{1.23}$$

with $\Sigma_\Delta = \boldsymbol{A}\Sigma\boldsymbol{A}'$, $\boldsymbol{\mu_\Delta} = (\beta, \alpha)'$ with $\beta = E(\Delta T) = \mu_{T_1} - \mu_{T_0}$ and $\alpha = E(\Delta S) = \mu_{S_1} - \mu_{S_0}$.

Surrogacy Measures

The mutual information $I(\Delta T, \Delta S)$, as defined in Equation 1.22, quantifies the amount of uncertainty that is expected to be removed in ΔT when the value of ΔS becomes known. For the normal distribution, mutual information and Pearson correlation are equivalent. In the normal model, the following relationship holds:

$$I(\Delta T, \Delta S) = -\frac{1}{2}\log\left(1 - \rho_\Delta^2\right) \quad \text{where} \quad \rho_\Delta = \text{corr}(\Delta T, \Delta S). \tag{1.24}$$

The ICA is therefore defined as the Pearson correlation between ΔT and ΔS: $\rho_\Delta = \text{corr}(\Delta T, \Delta S)$. The value for ρ_Δ has a closed form expression in terms of the parameters in Σ.

A generalization of the Pearson correlation measure in the normal model is the so-called squared informational coefficient of correlation (SICC) R_h^2 introduced by Joe (1989) and Linfoot (1957):

$$R_h^2 = 1 - e^{-2I(\Delta T, \Delta S)}. \tag{1.25}$$

For normally distributed outcomes, ρ_Δ^2 and R_h^2 are identical. This new measure is always in the interval $[0, 1]$, is invariant under bijective transformations, and takes value zero if and only if ΔT and ΔS are independent (Alonso, 2018). This measure will also be used in subsequent sections to quantify the ICA in the survival-survival setting.

Sensitivity Analysis

As already alluded to before, not all parameters of the multivariate normal model are identifiable. The covariances $\sigma_{S_0T_1}$, $\sigma_{S_1T_0}$, $\sigma_{T_0T_1}$ and $\sigma_{S_0S_1}$ in Σ are not identifiable from the observed data. Consequently, ρ_Δ is not identifiable from the data. As mentioned in Section 1.4.2, this identifiability issue is tackled by a sensitivity analysis. The identifiable parameters are fixed at their estimated values. The Γ_D region for the unidentifiable correlations is the region in $[-1,1]^4$ such that the resulting covariance matrix, Σ, is positive definite. This entails that the identifiable parameters restrict Γ_D.

The sensitivity analysis implemented in Alonso et al. (2015) proceeds as follows.

1. Select a grid of values $G = \{g_1, g_2, ..., g_k\}$ in $[-1,1]$ for the unidentifiable correlations.
2. The identifiable correlations are fixed at their estimated values. Generate Σ matrices by considering all combinations emanating from G for the unidentifiable correlations.
3. From these generated Σ matrices, only the positive definite ones are retained.
4. For every positive definite matrix, ρ_Δ is computed.

The sequence of values for ρ_Δ that is obtained in this way "quantifies the ICA across all plausible worlds, that is, across those scenarios where the assumptions made for the unidentifiable correlations are compatible with the observed data" (Van der Elst, Molenberghs, & Alonso, 2016, p. 1283). The behavior of ρ_Δ can consequently be used to assess the sensitivity of the results with respect to unverifiable assumptions.

1.4.5 Multivariate Surrogates

Most surrogate evaluation methods allow for considering only one surrogate endpoint. However, given the complex nature of many diseases and the various therapeutic pathways through which a treatment can affect the clinical outcome, it might be unreasonable to expect that only one surrogate can capture the entire treatment effect on the true endpoint. It is therefore expected that multiple surrogates, characterising distinct aspects of the disease-treatment interactions, improve the prediction of the individual causal effect on the true endpoint (Van der Elst et al., 2019).

Van der Elst et al. (2019) extended the Gaussian-Gaussian setting to a setting with a multivariate continuous surrogate endpoint. They found that in some scenarios, the range for the ICA becomes small when multiple surrogate endpoints are considered simultaneously. This is especially the case when the identifiable correlations are strong. In such situations, the "identifiability problem is no longer an issue from a practical perspective as the qualitative conclusion of the analysis is the same in all plausible realities compatible with the identifiable correlations" (Van der Elst et al., 2019, p. 306).

Chapter 2

Vine Copulas

The models proposed in this thesis for the survival-survival setting in the causal-inference framework are based on vine copulas. To make this thesis self-contained, the necessary theoretical concepts regarding vine copulas are described in this chapter. First, bivariate copulas are introduced. These have already been used for surrogate evaluation in the meta-analytic framework for several types of endpoints (Alonso et al., 2016; Burzykowski, 2001). Second, vine copulas are introduced as a flexible way of constructing multivariate copulas with bivariate copulas as building blocks. Finally, a 4-dimensional D-vine copula is proposed on which all further models are based.

2.1 Bivariate Copulas

2.1.1 Definitions and Properties

Copulas are d-dimensional multivariate distribution functions where each variable's marginal distribution is uniform on the interval $[0, 1]$. Copulas are used to describe dependence between random variables, independent of the marginal distributions. Definitions for bivariate copulas are given here, but these can be extended to d-dimensional copulas as well.

Definition 2.1.1 (Bivariate copula)**.** $C : [0, 1]^2 \to [0, 1]$ *is a bivariate copula if*

1. *For every* u, v *in* $[0, 1]$*:*
$$C(u, 0) = C(0, v) = 0$$
2. *For every* u, v *in* $[0, 1]$*:*
$$C(u, 1) = u \quad \textit{and} \quad C(1, v) = v$$
3. *For every* u_1, u_2, v_1, v_2 *in* $[0, 1]$ *such that* $0 \le u_1 \le u_2 \le 1$ *and* $0 \le v_1 \le v_2 \le 1$*:*
$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \ge 0$$

The corresponding *copula density*, c, is obtained by partial differentiation:

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v). \tag{2.1}$$

Sklar's theorem provides the theoretical foundation for the application of copulas (Sklar, 1959). It in essence states that any multivariate distribution function can be written in terms of the marginal distribution functions and a copula that describes the dependence structure. Sklar's theorem for a bivariate copula is given here.

Theorem 2.1.1 (Sklar's theorem)**.** *Let $\boldsymbol{X}$ be a 2-dimensional random vector with joint distribution function F, marginal distribution functions F_1 and F_2, and marginal density functions f_1 and f_2. Then there exists a copula C such that for all x, y in $]-\infty,+\infty[$*

$$F(x,y) = C\{F_1(x), F_2(y)\}. \tag{2.2}$$

If F_1 and F_2 are continuous, then C is unique. The associated density function follows from the copula density and marginal densities:

$$f(x,y) = c\{F_1(x), F_2(y)\} \cdot f_1(x) \cdot f_2(y). \tag{2.3}$$

The converse is also true. Given a copula C and marginals F_1 and F_2, Equation 2.2 defines a joint distribution function with margins F_1 and F_2.

Conditional densities and distributions can be directly derived from the corresponding copula and marginal densities. The proofs are given in Czado (2019, p. 20).

Lemma 2.1.1 (Conditional densities and distribution functions of bivariate distributions in terms of their copula)**.** *The conditional density and distribution function can be rewritten as*

$$f_{x|y}(x|y) = c\{F_1(x), F_2(y)\} \cdot f_2(x) \tag{2.4}$$

and

$$\begin{aligned} F_{x|y}(x|y) &= \frac{\partial}{\partial v} C\{F_1(x), v\}|_{v=F_2(y)} \\ &= \frac{\partial}{\partial F_2(y)} C\{F_1(x), F_2(y)\}. \end{aligned} \tag{2.5}$$

2.1.2 Survival Copula

In survival analysis, it is natural to focus on survival functions instead of distribution functions (Burzykowski, 2001). Let X and Y be continuous random variables with joint distribution function $F(x,y) = C_{XY}\{F_X(x), F_Y(y)\}$. Let $S(x,y) = P(X > x, Y > y)$ be the corresponding joint survival function. Let S_X and S_Y be the marginal survival functions for X and Y, respectively. Then

$$\begin{aligned} S(x,y) &= S_X(x) + S_Y(y) - 1 + F(x,y) \\ &= S_X(x) + S_Y(y) - 1 + C_{XY}\{F_X(x), F_Y(y)\}. \end{aligned}$$

A new copula $\tilde{C}_{XY} : [0,1]^2 \to [0,1]$ is defined as follows

$$\tilde{C}_{XY}(u,v) = u + v - 1 + C_{XY}(1-u, 1-v). \tag{2.6}$$

Let $u = S_X(x)$ and $v = S_Y(y)$, then

$$S(x,y) = \tilde{C}_{XY}\{S_X(x), S_Y(y)\}. \tag{2.7}$$

The new copula $\tilde{C}_{XY} : [0,1]^2 \to [0,1]$ is further referred to as the survival copula.

Lemma 2.1.2 (Density corresponding to the survival copula)**.** *The joint density of (X,Y) is the product of the survival copula density and marginal densities*

$$f_{XY}(x,y) = \tilde{c}_{XY}\{S_X(x), S_Y(y)\} \cdot f_X(x) \cdot f_Y(y) \tag{2.8}$$

where the survival copula density is defined as follows

$$\tilde{c}_{XY}\{S_X(x), S_Y(y)\} = \frac{\partial^2}{\partial u \partial v} \tilde{C}_{XY}(u,v)|_{u=S_X(x), v=S_Y(y)}. \tag{2.9}$$

The proof is given in Appendix A.1.

Survival copulas are convenient in that they can simplify expressions that frequently occurs in survival analysis. For example, if both X and Y are right censored, their likelihood contribution follows from the joint survival function $S(x,y)$ which is directly expressed in terms of the survival copula and marginal survival functions (Equation 2.7).

2.1.3 Rotated Copulas

A given parametric copula, $c(u, v)$, can be further extended by rotating the copula with 90, 180 or 270 degrees.

- 90°: $c_{90}(u, v) = c(1 - v, u)$
- 180°: $c_{180}(u, v) = c(1 - u, 1 - v)$
- 270°: $c_{270}(u, v) = c(v, 1 - u)$

The survival copula is in fact a 180° rotation copula: $\tilde{c}(u, v) = c_{180}(u, v)$. Some parametric copulas can only model positive association, e.g., the Clayton copula. Such parametric copulas can still be used to model negative association by allowing rotations.

2.2 Vine Copulas

Vine copulas constitute a flexible class of copula models. In this class, multivariate copulas are constructed using only bivariate copulas as building blocks. These building blocks are combined to a valid multivariate copula by appropriate conditioning (Czado, 2019). The same notation as in Czado (2019) is further used.

The starting point for vine copulas is the decomposition of a multivariate joint density function into a product of conditional densities. Let $(X_1, ..., X_d)$ be a set of variables with joint distribution function $F_{1,...,d}$ and joint density function $f_{1,...,d}$. The joint density function can be decomposed as follows

$$\begin{aligned} f_{1,...,d}(x_1, ..., x_d) &= f_{d|1,...,d-1}(x_d|x_1, .., x_{d-1}) \cdot f_{1,...,d-1}(x_1, ..., x_{d-1}) \\ &= ... \\ &= \left\{ \prod_{t=2}^{d} f_{t|1,...,t-1}(x_t|x_1, ..., x_{t-1}) \right\} \times f_1(x_1). \end{aligned} \tag{2.10}$$

Here, $f(\cdot|\cdot)$ and $F(\cdot|\cdot)$ refer to conditional density and distribution functions, respectively.

In what follows, we need the notion of *copulas associated with bivariate conditional distributions* (in contrast to bivariate conditional distributions on the copula scale). These are defined as follows (Czado, 2019, p. 88).

Definition 2.2.1 (Copulas associated with bivariate conditional distributions)**.** *Let $(X_1, ..., X_d)$ be a set of random variables.*

- *Let D be a set of indices from $\{1, ..., d\}$ not including i and j. The copula associated with the bivariate conditional distribution $(X_i, X_j)'$ given that $\boldsymbol{X_D} = \boldsymbol{x_D}$ is denoted by $C_{ij;D}(\cdot, \cdot; \boldsymbol{x_D})$.*
- *In contrast, the conditional distribution function of $(U_i, U_j)' = \left(F_{X_i}(X_i), F_{X_j}(X_j)\right)'$ given $\boldsymbol{U_D} = \boldsymbol{u_D}$ is expressed as $C_{ij|D}(\cdot, \cdot; \boldsymbol{u_D})$ with bivariate density function $c_{ij|D}(\cdot, \cdot; \boldsymbol{u_D})$.*
- *For distinct indices i, j and $D = \{i_1, ..., i_k\}$ with $i < j$ and $i_1 < ... < i_k$ we use the abbreviation*

$$c_{i,j;D} = c_{i,j;D}\left\{F_{i|D}(x_i|\boldsymbol{x_D}), F_{j|D}(x_j|\boldsymbol{x_D}); \boldsymbol{x_d}\right\} \tag{2.11}$$

where $F_{i|D}(x_i|\boldsymbol{x_D})$ is the conditional distribution function of X_i given that $\boldsymbol{X_D} = \boldsymbol{x_D}$, and analogously for $F_{j|D}(x_i|\boldsymbol{x_D})$.

In general, the *copula* $C_{ij;D}(\cdot,\cdot;\boldsymbol{x_D})$ is different from the *bivariate distribution function* $C_{ij|D}(\cdot,\cdot;\boldsymbol{x_D})$. Since $C_{ij|D}(\cdot,\cdot;\boldsymbol{x_D})$ is the bivariate distribution function of $U_i, U_j|\boldsymbol{U_D}$, the corresponding margins $U_i|\boldsymbol{U_D}$ and $U_j|\boldsymbol{U_D}$ are generally not uniform. This bivariate distribution function is thus generally not a copula. Note that $C_{ij;D}(\cdot,\cdot;\boldsymbol{x_D})$ is a copula by definition 2.2.1; it thus always has uniform margins.

The notation introduced above allows us to further decompose the joint density function in Equation 2.10 as follows (Czado, 2019, p. 89).

Theorem 2.2.1 (Drawable vine (D-vine) density)**.** *Every joint density $f_{1,\dots,d}$ can be decomposed as*

$$f_{1,\dots,d}(x_1,\dots,x_d) = \left\{\prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{i,i+j;(i+1),\dots,(i+j-1)}\right\}\cdot\left\{\prod_{k=1}^{d} f_k(x_k)\right\} \tag{2.12}$$

where we used the abbreviation introduced in Equation 2.11. The distribution associated with this density decomposition is called a drawable vine (D-vine).

Note that, in the most general case, $c_{i,j;D}$ depends on the conditioning value $\boldsymbol{x_D}$. However, in all that follows, the simplifying assumption is made (Czado, 2019, p. 90). Under this assumption, Equation 2.12 changes from a joint density decomposition to a joint density construction.

Definition 2.2.2 (Simplifying assumption for D-vines)**.** *If*

$$c_{i,j;D}\left\{F_{i|D}(x_i|\boldsymbol{x_D}), F_{j|D}(x_j|\boldsymbol{x_D});\boldsymbol{x_D}\right\} = c_{i,j;D}\left\{F_{i|D}(x_i|\boldsymbol{x_D}), F_{j|D}(x_j|\boldsymbol{x_D})\right\} \tag{2.13}$$

holds for all $\boldsymbol{x_D}$; and i, j and D are chosen to occur in Equation 2.12, then the corresponding D-vine distribution is called simplified.

This means that the copula associated with the bivariate conditional distribution does not depend on the conditioning value(s). The "conditional" dependence structure therefore does not depend on the conditioning value(s), although there remains a dependence on the conditioning value(s) through the univariate distribution functions $F_{i|D}$ and $F_{j|D}$. The simplifying assumption is nicely illustrated in the multivariate Gaussian copula. In a multivariate normal distribution

$$(\boldsymbol{X_1}, \boldsymbol{X_2})' \sim N\left((\boldsymbol{\mu_1}, \boldsymbol{\mu_2})', \Sigma\right) \quad \text{where} \quad \Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix},$$

the distribution of $\boldsymbol{X_1}$ conditional on $\boldsymbol{X_2} = \boldsymbol{a}$ is

$$\boldsymbol{X_1}|\boldsymbol{X_2} = \boldsymbol{a} \sim N\left(\boldsymbol{\mu_1} + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{a} - \boldsymbol{\mu_2}), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right).$$

The covariance matrix of the conditional distribution does not depend on the conditioning value $\boldsymbol{X_2} = \boldsymbol{a}$. Put differently, the dependence structure of the conditional distribution does not depend on the conditioning value $\boldsymbol{X_2} = \boldsymbol{a}$. This confirms that the simplifying assumption is satisfied for the Gaussian copula.

The density of D-vines requires the evaluation of conditional distribution functions. Indeed, the copulas associated with bivariate conditional distributions are the building blocks in the D-vine copula. As can be seen from Equation 2.11, $F_{i|D}(x_i|\boldsymbol{x_D})$ and $F_{j|D}(x_j|\boldsymbol{x_D})$ are required in the construction of $c_{ij;D}$. For the bivariate case, the relation between the bivariate copula and the conditional distribution function is given by Lemma 2.1.1. For extensions to higher dimensions, these conditional distribution functions are obtained through recursion (Czado, 2019, p. 92).

Theorem 2.2.2 (Recursion for conditional distribution functions)**.** *Let X be a random variable and $\boldsymbol{Y}$ be a random vector which have an absolutely continuous joint distribution. Let Y_j be a component of $\boldsymbol{Y}$ and denote the sub-vector of $\boldsymbol{Y}$ with Y_j removed by $\boldsymbol{Y}_{-j}$. In this case, the conditional distribution function of X given $\boldsymbol{Y} = \boldsymbol{y}$, $F_{X|\boldsymbol{y}}(\cdot|\boldsymbol{y})$, satisfies the following recursion*

$$F_{X|\boldsymbol{Y}}(\cdot|\boldsymbol{y}) = \frac{\partial C_{X,Y_j;\boldsymbol{Y}_{-j}}\left(F_{X|\boldsymbol{Y}_{-j}}(x|\boldsymbol{y}_{-j}), F_{Y_j|\boldsymbol{Y}_{-j}}(y_j|\boldsymbol{y}_{-j})\right)}{\partial F_{Y_j|\boldsymbol{Y}_{-j}}(y_j|\boldsymbol{y}_{-j})} \tag{2.14}$$

where $C_{X,Y_j;\boldsymbol{Y}_{-j}}(\cdot,\cdot|\boldsymbol{y}_{-j})$ denotes the copula corresponding to $(X,Y_j)'$ given that $\boldsymbol{Y}_{-j} = \boldsymbol{y}_{-j}$.

2.3 Four-Dimensional D-vine Copula

In this section, a joint model based on a D-vine copula is proposed for $(S_0,T_0,S_1,T_1)'$. Although the D-vine copula is not the only type of vine copula, it is a natural choice to model the vector of potential outcomes, as further explained.

2.3.1 Vine Copula Model Formulation

The D-vine copula is a natural way to model the distribution of $(S_0,T_0,S_1,T_1)'$, and to derive the observable bivariate distribution and density functions. The corresponding D-vine copula construction is given in Equation 2.15. The joint density function of $(S_0,T_0,S_1,T_1)'$ is denoted by $f_{1234} = f_{S_0T_0S_1T_1}$. The copula densities for the observable bivariate distributions of $(S_0,T_0)'$ and $(S_1,T_1)'$ are denoted by $c_{12} = c_{S_0T_0}$ and $c_{34} = c_{S_1T_1}$, respectively.

$$\begin{aligned} f_{1234} =& f_1 \cdot f_2 \cdot f_3 \cdot f_4 \\ & \cdot c_{12} \cdot c_{23} \cdot c_{34} \\ & \cdot c_{13;2} \cdot c_{24;3} \\ & \cdot c_{14;23} \end{aligned} \tag{2.15}$$

The full expressions of the individual components of the D-vine construction in Equation 2.15 are given in Table C.1 in Appendix C. For ease of notation, the short-hand notation of Equation 2.15 is further used instead of the expressions in Table C.1.

2.3.2 Comments on this Model

This (D-vine copula) construction of the joint density, f_{1234}, is very appealing in the causal-inference framework for several reasons.

Separation of association from margins. The marginal distributions are specified completely separately from the association structure in this model. The marginal distributions are identifiable, whereas the association structure is only partly identifiable.

This is as opposed to shared frailty proportional-hazards models where the variance of the frailty could be interpreted as a measure of association. In such a model, this variance also depends on violations of the proportional-hazards assumption and more generally misspecifications of the hazard function. Indeed, if there is no association, but the hazard is misspecified, the variance of the frailty distribution can be far from zero.

Flexible association structure. The association structure is very flexible in a D-vine copula. Any copula function can be used for $c_{ij;D}$. The Gaussian copula can also be represented as a D-vine copula. Let c_{1234} be the multivariate copula density for f_{1234}. If all six bivariate copula densities in Equation 2.15 are bivariate Gaussian copula densities, then c_{1234} is a multivariate Gaussian copula density. The D-vine copula is thus a generalization of the multivariate Gaussian copula.

Separation in identifiable and unidentifiable parameters. The vector $(S_0,T_0,S_1,T_1)'$ is never (fully) observed, only $(S_0,T_0)'$ or $(S_1,T_1)'$ can be observed. Therefore, the likelihood emanating from the entire vector of potential outcomes cannot be used directly. Instead, the "observed-data" likelihood emanates from the joint density f_{1234} marginalized over $(S_0,T_0)'$ or $(S_1,T_1)'$. Conveniently, these marginalized densities follow immediately from the D-vine construction

Table 2.1: Observable vectors and their associated distribution functions.

Treatment	Obs. Data	Distribution Function	
$Z = 0$	$(S_0, T_0)'$	$F_{S_0,T_0}(s_0, t_0)$	$= C_{S_0,T_0}\left\{F_{S_0}(s_0), F_{T_0}(t_0)\right\}$
			$= C_{12}$
$Z = 1$	$(S_1, T_1)'$	$F_{S_1,T_1}(s_1, t_1)$	$= C_{S_1,T_1}\left\{F_{S_1}(s_1), F_{T_1}(t_1)\right\}$
			$= C_{34}$

Table 2.2: Observable vectors and their associated density functions.

Treatment	Obs. Data	Density Function	
$Z = 0$	$(S_0, T_0)'$	$f_{S_0,T_0}(s_0, t_0)$	$= c_{S_0,T_0}\left\{F_{S_0}(s_0), F_{T_0}(t_0)\right\} \cdot f_{S_0}(s_0) \cdot f_{T_0}(t_0)$
			$= c_{12} \cdot f_1 \cdot f_2$
$Z = 1$	$(S_1, T_1)'$	$f_{S_1,T_1}(s_1, t_1)$	$= c_{S_1,T_1}\left\{F_{S_1}(s_1), F_{T_1}(t_1)\right\} \cdot f_{S_1}(s_1) \cdot f_{T_1}(t_1)$
			$= c_{34} \cdot f_3 \cdot f_4$

in Equation 2.15. The corresponding marginalized distribution and density functions are given respectively in Table 2.1 and 2.2.

Note that these marginalized densities do not follow immediately for other vine copula constructions such as C-vines. This is problematic because the observed-data likelihood is then determined via the integral of f_{1234} over the unobserved outcomes. Generally, the observed-data likelihood then depends on all components of the vine copula construction. As a consequence, there is then no separation in identifiable and unidentifiable association parameters. In fact, only the D-vine copula model presented here, and variations thereof where the copula for $(S_0, T_0)'$ and $(S_1, T_1)'$ is present in the first tree (second line in Equation 2.15), have this property of separation in identifiable and unidentifiable parameters.

Chapter 3

Models

In this chapter, the models for $(S_0, T_0, S_1, T_1)'$ are introduced in more detail. First, a model that does not take into account time orderings is proposed. Next, a similar model which takes into account possible time orderings is proposed.

3.1 No Time Ordering

The models that do not impose time orderings directly follow from the D-vine copula construction given in Equation 2.15. In this section, the likelihood contributions for observations are derived. This is required to conduct maximum likelihood estimation of the model. Next, different measures of surrogacy in this model are discussed. These quantify the association between the individual causal effects on the surrogate and true endpoint. Finally, a sensitivity analysis is explained to address the lack of identifiability of the joint model.

3.1.1 Likelihood

Observed-data Likelihood

Only $(S_0, T_0)'$ or $(S_1, T_1)'$ are observed with possible (right) censoring. Therefore, the observed-data likelihood emanates from f_{1234} marginalized over $(S_0, T_0)'$ or $(S_1, T_1)'$. The corresponding distribution and density functions follow directly from the components of the D-vine construction, see Table 2.1 and 2.2. Note that further on, survival copulas will be used as explained in Section 2.1.2. This simplifies the expressions for right censored data.

Let the observed vector for patient i be $(s_i, \delta_{s,i}, t_i, \delta_{t,i}, z_i)$. Let s_i and t_i be the, possibly right censored, observed values for the surrogate and true endpoint, respectively. In what follows, independent censoring is always assumed. The corresponding event indicators are $\delta_{s,i}$ and $\delta_{t,i}$ where $\delta_{s,i} = 1$ if the surrogate event is observed and $\delta_{s,i} = 0$ otherwise, analogously for $\delta_{t,i}$. The treatment indicator z_i can only take two values: 0 or 1 for control and experimental treatment, respectively. The possible likelihood contributions for patient i with $z_i = k$ are as follows.

- if $\delta_{s,i} = \delta_{t,i} = 1$:

$$\begin{aligned} L_i &= f_{S_k,T_k}(s_i, t_i) \\ &= c_{S_k,T_k}\left\{S_{S_k}(s_i), S_{T_k}(t_i)\right\} \cdot f_{S_k}(s_i) \cdot f_{T_k}(t_i) \end{aligned} \tag{3.1}$$

  The second equality follows from lemma 2.1.2.

- if $\delta_{s,i} = 1$ and $\delta_{t,i} = 0$:

$$\begin{aligned} L_i &= \int_{t_i}^{\infty} f_{S_k,T_k}(s_i,t)dt \\ &= f_{S_k}(s_i) \cdot \frac{\partial C_{S_k,T_k}}{\partial S_{S_k}} \left\{S_{S_k}(s_i), S_{T_k}(t_i)\right\} \end{aligned} \tag{3.2}$$

  The full derivation is given in Appendix A.2.

- if $\delta_{s,i} = \delta_{t,i} = 0$:

$$\begin{aligned} L_i &= \int_{t_i}^{\infty}\int_{s_i}^{\infty} f_{S_k,T_k}(s,t)dsdt \\ &= S_{S_k,T_k}(s_i,t_i) \\ &= C_{S_k,T_k}\left\{S_{S_k}(s_i), S_{T_k}(t_i)\right\} \end{aligned} \tag{3.3}$$

- if $\delta_{s,i} = 0$ and $\delta_{t,i} = 1$[1]:

$$\begin{aligned} L_i &= \int_{s_i}^{\infty} f_{S_k,T_k}(s,t_i)ds \\ &= f_{T_k}(t_i) \cdot \frac{\partial C_{S_k,T_k}}{\partial S_{T_k}} \left\{S_{S_k}(s_i), S_{T_k}(t_i)\right\} \end{aligned} \tag{3.4}$$

  The full derivation is entirely analogous to the second case where $\delta_{s,i} = 1$ and $\delta_{t,i} = 0$.

In this thesis, four parametric copulas are considered: Gaussian, Clayton, Gumbel and Frank. The expressions for these copulas, copula densities and partial derivatives of copulas are given in Appendix A.3. For the marginal density and survival functions, any (survival) distribution can be used. In what follows, the Royston-Parmar survival model is used as marginal survival function because of its great flexibility.

Identifiability of Parameters

The likelihood contributions for control patients depend on parameters corresponding to f_1, f_2, and c_{12}. Similarly, the likelihood contributions for treated patients depend on parameters corresponding to f_3, f_4, and c_{34}. This is also shown in Table 2.1 and 2.2. Because the likelihood contributions for control patients have no parameters in common with the likelihood contributions for the treated patients, the likelihood can be maximized in the control and treated groups separately.

The parameters corresponding to c_{23}, $c_{13;2}$, $c_{24;3}$, and $c_{14;23}$ do not appear in the likelihood contributions of either group. Indeed, they also do not appear in the observable distribution functions and densities in Table 2.1 and 2.2. Given the observed data, they can thus not be estimated. Hence, they are unidentifiable. This issue is addressed by a sensitivity analysis with respect to the unidentifiable parameters. This is discussed in Section 3.1.3.

The interpretation of these unidentifiable components is summarized in Table 3.1. Note that the simplifying assumption is made (Definition 2.2.2). This means that in Table 3.1, the dependence structure does not depend on the conditioning value(s). Nonetheless, the bivariate conditional distribution generally depends on the conditioning value(s) through the conditional distribution functions (see also the remarks following Definition 2.2.2).

[1]This case implies that the patient remains under observation for the surrogate (e.g. PFS), but not for the true endpoint (e.g. OS). In some study designs, this may be possible. However, this case should generally be regarded with suspicion, and it should be checked whether the study design allows for this.

Table 3.1: Interpretation of the unidentifiable components of the D-vine copula model. The (D-vine) joint density construction is given in Equation 2.15 using short-hand notation. The full expressions for the individual components are given in Table C.1 in Appendix C.

Component	Interpretation
c_{23}	Dependence structure of (T_0, S_1)
$c_{13;2}$	Dependence structure of $(S_0, S_1)\|T_0$
$c_{24;3}$	Dependence structure of $(T_0, T_1)\|S_1$
$c_{14;23}$	Dependence structure of $(S_0, T_1)\|T_0, S_1$

3.1.2 Measures of Surrogacy

Density of Primary Interest

The ICA is based on the association between the individual causal effects on the surrogate and true endpoint, although other definitions for individual causal effects are also possible. From the joint density function, $f_{\Delta S,\Delta T}(\delta_s, \delta_t)$, any association measure can in principle be computed. This joint density function follows from applying the deconvolution formula for differences twice.

Definition 3.1.1 (Deconvolution formula for differences)**.** *The density of $Z = Y - X$ follows from the joint density of $(X, Y)'$ as follows*

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{XY}(x, x + z)dx. \tag{3.5}$$

The joint density function can thus be obtained by calculating following double integral. The construction of $f_{1234} = f_{S_0,T_0,S_1,T_1}$ was discussed in the previous chapter.

$$f_{\Delta S,\Delta T}(\delta_s, \delta_t) = \int_0^{\infty} \int_0^{\infty} f_{1234}(s, t, s + \delta_s, t + \delta_t)dsdt \tag{3.6}$$

This is, however, a complex expression with generally no closed-form solution. Further on, it is discussed how this is solved by numerical approximation. In some special cases, this expression has a closed-form solution, e.g., if the joint distribution of $(S_0, T_0, S_1, T_1)'$ is multivariate normal. This property has been used to derive closed-form expressions for different measures of surrogacy when both endpoints are normally distributed (Alonso et al., 2015; Van der Elst et al., 2021a; Van der Elst et al., 2019).

It is nevertheless straightforward to sample from $f_{\Delta S,\Delta T}(\delta_s, \delta_t)$ by sampling (S_0, T_0, S_1, T_1) from f_{1234} and computing the individual causal effects. The `rvinecopulib` R-package provides flexible functions to sample efficiently from vine copulas (Nagler & Vatter, 2022). Therefore, the ICA is further computed by sampling.

Individual Causal Association: Kendall's τ and Spearman's ρ

Kendall's τ and Spearman's ρ are used to quantify the association between ΔS, and ΔT. They are defined as follows.

Definition 3.1.2 (Spearman's ρ)**.** *Consider three i.i.d. realisations of $(X, Y)'$ denoted by $(X_1, Y_1)'$, $(X_2, Y_2)'$ and $(X_3, Y_3)'$. Spearman's ρ is then defined as*

$$\rho_s = 3\left[P\left\{(X_1 - X_2)(Y_1 - Y_3) > 0\right\} - P\left\{(X_1 - X_2)(Y_1 - Y_3)\right\}\right]. \tag{3.7}$$

Definition 3.1.3 (Kendall's τ)**.** *Consider two i.i.d. realisations of* $(X, Y)'$ *denoted by* $(X_1, Y_1)'$ *and* $(X_2, Y_2)'$*. Kendall's* τ *is then defined as*

$$\tau = P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2)\}. \tag{3.8}$$

Kendall's τ and Spearman's ρ lie in $[-1, 1]$ and are equal to zero when X and Y are independent. Both Kendall's τ and Spearman's ρ are also independent from the marginal distributions of X and Y. These measures thus only depend on the copula C_{XY}.

As noted before, the density of $(\Delta S, \Delta T)'$ is generally intractable to obtain. Therefore, Kendall's τ and Spearman's ρ are computed by a Monte Carlo procedure instead of analytically. In order to compute any measure of association, all parameters of f_{1234} need to be known (or estimated). As explained before, not all parameters are identifiable. This is addressed in a sensitivity analysis by sampling the unidentifiable parameters, discussed in Section 3.1.3. For now, assume that in the following procedure all parameters of f_{1234} are known (or estimated). The procedure to compute Kendall's τ and Spearman's ρ then proceeds as follows:

1. Sample the copula data vector, $(U_{S_0}, U_{T_0}, U_{S_1}, U_{T_1})'$, from the copula C_{1234} of f_{1234} N times. This is implemented in the `rvinecop` function from the `rvinecopula` R-package (Nagler & Vatter, 2022).

2. The copula data from the previous step are transformed to the appropriate scale by the probability integral transform with the corresponding marginal distribution functions:

$$(S_0, T_0, S_1, T_1)' = \left(F_{S_0}^{-1}(U_{S_0}), F_{T_0}^{-1}(U_{T_0}), F_{S_1}^{-1}(U_{S_1}), F_{T_1}^{-1}(U_{T_1})\right)'.$$

   The data thus obtained are i.i.d. samples from f_{1234}.

3. From these sampled vectors, compute the individual causal effects

$$(\Delta S, \Delta T)' = (S_1 - S_0, T_1 - T_0)'.$$

4. Compute the sample estimates for τ and ρ_s. If N is sufficiently large, these estimates approximate the true values.

Individual Causal Association: Information-Theoretic Approach

A third measure for the ICA is based on information theory. It is the squared informational coefficient of correlation (Joe, 1989; Linfoot, 1957), further referred to as R_h^2. This association measure was defined previously in the Gaussian-Gaussian setting, Equation 1.25. As for Kendall's τ and Spearman's ρ, R_h^2 is computed numerically through a slightly different Monte Carlo procedure:

1. Sample the copula data vector, $(U_{S_0}, U_{T_0}, U_{S_1}, U_{T_1})'$, from the copula C_{1234} of f_{1234} N times. This is implemented in the `rvinecop` function from the `rvinecopula` R-package (Nagler & Vatter, 2022).

2. The copula data from the previous step are transformed to the appropriate scale by the probability integral transform with the corresponding marginal distribution functions:

$$(S_0, T_0, S_1, T_1)' = \left(F_{S_0}^{-1}(U_{S_0}), F_{T_0}^{-1}(U_{T_0}), F_{S_1}^{-1}(U_{S_1}), F_{T_1}^{-1}(U_{T_1})\right)'.$$

   The data thus obtained are an i.i.d. sample from f_{1234}.

3. From these sampled vectors, compute the individual causal effects

$$(\Delta S, \Delta T)' = (S_1 - S_0, T_1 - T_0)'.$$

4. Transform the sampled $(\Delta S, \Delta T)'$ to pseudo-copula data (Czado, 2019). The marginal distribution functions are approximated by the empirical distribution functions: $\hat{F}_{\Delta S}$ and $\hat{F}_{\Delta T}$.
$$(U_{\Delta S}, U_{\Delta T})' = \left(\hat{F}_{\Delta S}(\Delta S), \hat{F}_{\Delta T}(\Delta T)\right)'$$
5. Estimate the copula density with a kernel estimator with the `kdecop` function from the `kdecopula` R-package (Nagler, 2018).
6. Given the estimated copula density from the previous step, the mutual information is computed through quasi Monte Carlo integration with the `dep_measures` function from the `kdecopula` R-package (Nagler, 2018).

In principle the transformation from $(\Delta S, \Delta T)'$ to $(U_{\Delta S}, U_{\Delta T})'$ is not necessary. Because efficient functions are available in `R` to compute the mutual information for stochastic vectors with uniform margins, the samples are transformed to uniform scales. The probability integral transform is a monotone transformation and the mutual information is invariant to monotone transformations. Thus, the transformation to pseudo-copula data does not change the mutual information.

Note that three choices influence the accuracy of this procedure. First, a larger N will increase the accuracy. As N increases, the copula density will be more precisely estimated in step 5. Second, the choice of the kernel estimator also influences the accuracy (Nagler, 2018). Third, the number of quasi Monte Carlo samples in step 6 determines how precisely the mutual information is computed for the copula density estimate. In all further analyses presented in this thesis, the number of Monte Carlo and quasi Monte Carlo samples (step 1 and 6, respectively) are set equal to each other.

Accuracy

The accuracy of these procedures to compute Spearman's ρ, Kendall's τ, and R_h^2 is studied through simulations in Appendix D. The "true" value is considered to be the value that is computed with 100.000 (quasi) Monte Carlo samples. The accuracy of the procedure with fewer samples is quantified in terms of the standard deviation based on 50 replications. In consideration of the uncertainty due to the unidentifiability of some parameters, a standard deviation of 0.01 is deemed acceptable. This precision is reached with $N = 2000$ (quasi) Monte Carlo samples.

3.1.3 Sensitivity Analysis

The different measures for the ICA, as defined above, can be computed if all parameters in f_{1234} are known or estimated. However, not all parameters are identifiable. This is resolved by fixing the identifiable parameters at their estimated values and sampling the unidentifiable parameters from a certain distribution that is compatible with the observed data. For each sample of unidentifiable parameters, all parameters of f_{1234} are "known", and the ICA can thus be computed. First, it is discussed how the unidentifiable vine copula parameters can be sampled. Second, it is discussed how additional assumptions regarding unidentifiable parameters can be incorporated into the sensitivity analysis.

Vine Copula Parameters

A distribution has to be specified to sample the unidentifiable copula parameters from. If the unidentifiable parameters are defined on an interval with finite limits a and b, then a straightforward choice is to sample the parameters from $U(a, b)$. This is for example the case for Gaussian copulas where the correlation parameters lie in $[-1, 1]$. For other copulas, one or two of the

limits might be infinite. A different approach is needed there. The approach that is presented here, is applicable for 1 parameter copulas where all copulas of the D-vine decomposition are of the same parametric form. This can easily be extended to the case where the copulas of the D-vine decomposition are not all of the same parametric form.

Sampling vine copula parameters. The unidentifiable association parameters, in terms of Spearman's ρ, are sampled from a uniform distribution: $\rho_{\cdot;\cdot} \sim U(-1, 1)$. The sampled parameters on Spearman's ρ scale are then transformed to the scale of the original copula parameter $\theta_{\cdot;\cdot}$. If the unidentifiable copulas allow for positive and negative associations, this approach suffices. However, some copulas only allow for positive association, e.g., the Clayton copula where ρ_s is restricted to $[0, 1]$. This entails that we implicitly assume all unidentifiable associations to be positive in the sensitivity analysis. This is not warranted in general and frequently contradictory to the goal of the sensitivity analysis. The sensitivity analysis is intended to explore the (range of) values for the ICA that are compatible with the observed data. Copulas such as the Clayton copula can still be used in the sensitivity analysis with a slight modification. If the Clayton copula is rotated by 90 or 270 degrees, ρ_s is restricted to $[-1, 0]$ (see also Section 2.1.3). In addition to sampling the unidentifiable copula parameters from $\rho_{\cdot;\cdot} \sim U(0, 1)$, rotation parameters are sampled from a uniform discrete distribution with four elements $(0, 90, 180, 270)$.

This approach generates samples for the unidentifiable parameters that are compatible with the (estimated) identifiable parameters. One may wish to incorporate additional assumptions in the sensitivity analysis. We could impose restrictions on the unidentifiable copula parameters. However, these parameters have an obscure interpretation as all of them, except one, relate to copulas associated with *conditional* bivariate distributions. Next, it is discussed how "interpretable" restrictions can be imposed on the marginal association structure.

Additional Assumptions

An attractive approach for incorporating assumptions on unidentifiable association parameters is to consider the marginal unidentifiable association parameters, instead of the unidentifiable (conditional) copula parameters. However, the joint density is defined in terms of (conditional) copula parameters, and the relation with the marginal association parameters is in general not clear. As before, an approach based on sampling is proposed here.

For each vector of sampled unidentifiable parameters f_{1234} is known. The marginal Kendall's τ's are then computed by sampling. Next, only the sampled sets of unidentifiable parameters are retained that satisfy the restrictions on the marginal association parameters, i.e., the marginal Kendall's τ's. Two sets of restrictions are proposed.

Monotonicity. All (unidentifiable) marginal associations are positive. Hence, every sample where at least one of the marginal Kendall's τ's is negative, is discarded. This condition is written formally as follows:

$$\min(\tau_{S_0,S_1}, \tau_{T_0,T_1}, \tau_{S_0,T_1}, \tau_{S_1,T_0}) > 0 \tag{3.9}$$

where $\tau_{\cdot,\cdot}$ refers to the marginal Kendall's τ between the respective potential outcomes.

Weaker Cross-Association. This restriction says that the association between potential outcomes across treatment groups is weaker than within treatment groups. This condition is written formally as follows:

$$\min(\tau_{S_0,T_0}, \tau_{S_1,T_1}) > \max(\tau_{S_0,T_1}, \tau_{S_1,T_0}). \tag{3.10}$$

These assumptions are discussed extensively in Section 4.2.2 in relation to the case study.

3.2 Time Ordering

The models presented above do not incorporate time orderings. This is not very problematic if the probability mass of the (fitted) joint distribution is small in the region that does not adhere to these time orderings. This is not guaranteed by the model proposed in the previous section. Therefore, a model is needed that explicitly takes into account the time orderings in the data. The approach proposed here is appropriate for a surrogate-true endpoint pair with the following time orderings: $S_0 \leq T_0$, $S_1 \leq T_1$ where $P(S_0 = T_0) > 0$ and $P(S_1 = T_1) > 0$. PFS is such a surrogate. PFS is always smaller than or equal to OS. Moreover, there is a proportion of patients that die without progression such that $P(S_0 = T_0) > 0$ and $P(S_1 = T_1) > 0$.

The general approach in this section is to model the surrogate indirectly. A joint model for time-to-progression (TTP) and OS is proposed where TTP is dependently censored by OS. From this joint model for TTP and OS, the joint distribution of PFS and OS is derived. Based on the distribution of PFS and OS, the measures of surrogacy are computed. Of course, this also applies to surrogates other than PFS and TTP that are defined similarly.

3.2.1 Semi-Competing Risks

The joint modelling of variables such as TTP and OS has been extensively described and discussed in literature. This data structure has been termed *semi-competing risks*, though it has been described earlier by an illness-death model (Fine, Jiang, & Chappell, 2001).

Let S be the time to the surrogate endpoint[2], T the time to the true endpoint[3], and C the time to independent censoring. As before, Z is the binary treatment indicator where $Z = 0/1$. In what follows, independent censoring is always assumed.

In the semi-competing risks framework, the observed data consist of $(X_i, \delta_i^X, Y_i, \delta_i^Y, Z_i)$ for patient $i = 1, ..., n$. The observed data are a random sample from $(X, \delta^X, Y, \delta^Y, Z)$ where $X = \min(S, T, C)$, $\delta^X = I(X = S)$, $Y = \min(T, C)$ and $\delta^Y = I(Y = T)$. The particularity of the semi-competing risks framework is an asymmetry which is not present in the competing risks framework. Indeed, S is *dependently* censored by the minimum of T and C, whereas T is *independently* censored by C. This entails that the joint distribution of (S, T) is only observable on the upper wedge, i.e., the region where $S \leq T$. The notation in terms of X and Y is, however, not further used.

3.2.2 Likelihood

Observable Likelihood

Let the observed vector for patient i be $(s_i, \delta_{s,i}, t_i, \delta_{t,i}, z_i)$, where the elements are defined as in Section 3.1.1. The only difference is that S is now dependently censored by T. In essence, S now refers TTP instead of PFS. This entails that if the patient dies before the surrogate event is observed, $\delta_{s,i} = 0$; whereas, in the previous section, the following would hold in that case: $\delta_{s,i} = 1$.

Note that in this framework, a latent value for S is assumed if $S > T$. As noted before, the joint distribution is only observable for $S \leq T$. The fitted joint distribution for $S > T$ is thus unverifiable. However, this is not problematic if only "verifiable quantities" are of interest, more specifically, if only quantities derived from the fitted joint distribution which are verifiable with the observed data are used.

Let T_k be the time to the potential true endpoint event (e.g., death) for $Z = k$, S_k the time to the potential surrogate event (e.g., progression), and $S_k^* = \min(S_k, T_k)$ the potential time to the composite event (e.g., PFS). Under independent censoring, the marginal survival function

[2] This is often referred to as the non-terminal event in the semi-competing risks literature.

[3] This is often referred to as the terminal event in the semi-competing risks literature.

for T_k can be derived from the fitted joint model. In addition, the marginal survival function for S_k^* is observable and can be derived from the joint distribution of (S_k, T_k) as follows:

$$\begin{aligned} P(S_k^* > x) &= P(\min(S_k, T_k) > x) \\ &= P(S_k > x, T_k > x). \end{aligned} \tag{3.11}$$

Finally, the joint survival function of (S_k^*, T_k) is also observable.

$$\begin{aligned} P(S_k^* > x, T_k > t) &= P(\min(S_k, T_k) > x, T_k > t) \\ &= P(S_k > x, T_k > x, T_k > t) \\ &= P(S_k > x, T_k > \max(x, t)). \end{aligned} \tag{3.12}$$

Consistent with the definitions for PFS, TTP and OS, let $S^* = PFS$, $S = TTP$ and $T = OS$. Then it is clear that the model is fitted using TTP and OS, but the joint distribution for (PFS, OS) can be derived from this model, as shown by Equation 3.12. Moreover, this latter joint distribution is used in deriving surrogacy measures. The latent variable specification is not problematic for these surrogacy analyses as the joint survival function in Equation 3.12 is observable.

Identifiability of Parameters

Note that in this framework, the bivariate distributions for (PFS, OS) that derive from the fitted models do not rely on unverifiable assumptions, even though latent variables are used. It would be a mistake, however, to directly interpret the fitted association measures between TTP and OS. For example, one should not interpret the Kendall's τ derived from the estimated copula parameters of c_{12} and c_{34} in Equation 2.15. Indeed, such association measures rely on the assumption that the specification of the model is correct for $S > T$. This is unverifiable, and S is not well-defined for $S > T$. Lee et al. (2015) set S equal to ∞ in that case to emphasize that it is not well-defined. Using that notation would complicate the model specification, but would not alter the observable part of the model. That notation is therefore not used.

As before, the association between potential outcomes *across* treatment groups still relies on unverifiable assumptions. This is again addressed by a sensitivity analysis. The same comments as in Section 3.1.1 still hold here.

3.2.3 Measures of Surrogacy

All further comments made in Section 3.1.2 and 3.1.3 also apply for the models with time orderings. The only difference is that the sampling procedure is slightly different as detailed next.

Sampling

One can sample from f_{1234} to obtain a sample of $(S_0^*, T_0, S_1^*, T_1)'$ as follows:

1. Sample the copula data vector, $(U_{S_0}, U_{T_0}, U_{S_1}, U_{T_1})'$, from the copula C_{1234} of f_{1234}. This is implemented in the `rvinecop` function from the `rvinecopula` R-package (Nagler & Vatter, 2022).

2. These copula data from the previous step are transformed to the appropriate scale by the probability integral transform with the corresponding marginal distribution functions:

$$(S_0, T_0, S_1, T_1)' = \left(F_{S_0}^{-1}(U_{S_0}), F_{T_0}^{-1}(U_{T_0}), F_{S_1}^{-1}(U_{S_1}), F_{T_1}^{-1}(U_{T_1})\right)'.$$

   The data thus obtained is an i.i.d. sample from f_{1234}.

3. Compute

$$(S_0^*, T_0, S_1^*, T_1)' = (\min(S_0, T_0), T_0, \min(S_1, T_1), T_1)' .$$

Note that this model formulation is somewhat like random-effects models. Such models are constructed by assuming a latent random-effect distribution which is in itself not observable. However, the "consequences" of the random effects are observable as they lead to an observable marginal distribution which is the target of inference. In our approach, a latent value for S_k is assumed if $S_k > T_k$ which is not observable. However, the "consequences" are observable as it leads to an observable distribution for $(S_k^*, T_k)'$, which is the target of inference. In addition, simulating data from a random-effects model is similar to simulating data from our model. One can sample from a random-effects model by first sampling the random effects and then sampling from the conditional distributions given the random effects. This yields a sample from the observable marginal distribution. In our approach, we first sample $(S_0, T_0, S_1, T_1)'$. Then we derive $(S_0^*, T_0, S_1^*, T_1)'$. Note that the second step is deterministic in our approach, but not when sampling from a random-effects model.

3.2.4 Sensitivity Analysis

The sensitivity analysis proceeds in the same way as for the model without time orderings. However, extra care is needed with respect to the additional assumptions. In Section 3.1.3, assumptions were proposed for the marginal association parameters in terms of Kendall's τ. For the model with time orderings, these marginal association parameters are defined in terms of S_k^* and T_k and are computed by sampling as described above. These parameters are *not* the same as the Kendall's τ parameters corresponding to c_{12} and c_{34}. Indeed, the latter are defined on the (latent) scale of S_k, but not on the scale of S_k^*.

3.3 Further Remarks

The models without time orderings are simpler than the models of this section. It is however difficult to compare the fit of models of this section with models without time orderings. The response variable is defined differently, so the maximized likelihoods cannot be compared across these two types of models. In any case, if there exist time orderings in the data, models that take these restrictions into account are more appropriate. The model with time orderings could still be preferred because of its relative simplicity. Moreover, even if the model of this section provides a much better fit, the conclusions with regards to surrogacy might be the same as for the simpler model. A very compelling reason to prefer the models with time orderings, is that additional restrictions (i.e., the time orderings) can reduce the uncertainty in the sensitivity analysis.

Chapter 4

Case Study: Advanced Ovarian Cancer

The methods introduced in the previous sections are applied to a data set, further referred to as the *ovarian cancer data*.

4.1 Data Description

The ovarian cancer data combine the data that were collected in four double-blind randomized clinical trials in advanced ovarian cancer (Omura et al., 1991). In these trials, the objective was to examine the efficacy of cyclophosphamide plus cisplatin (CP) versus cyclophosphamide plus adriamycin plus cisplatin (CAP) to treat advanced ovarian cancer. The four individual trials failed to show a significant effect of CAP on survival. However, a meta-analysis of the pooled data showed a significant survival benefit. These data have been used previously in the meta-analytic framework, where center was used as unit of analysis (Buyse et al., 2000). In what follows, all data are pooled and the hierarchy originating from the four trials is ignored.

The ovarian cancer data contain the PFS and OS for 1192 patients. These data are freely available through the `Surrogate` R-package (Van der Elst et al., 2021b). One patient is excluded from further analysis because the recorded PFS is larger than the recorded OS. All further analyses are thus based on 1191 observations.

4.2 Surrogacy Analysis

4.2.1 Model Fitting

The four copula models described previously (Gaussian, Frank, Gumbel, Clayton) are fitted to these data through maximizing the likelihood as detailed in Section 3.1.1. For illustrative purposes, both the models with and without time orderings are considered, although the former should be preferred because it makes full use of the data structure. The corresponding marginal distribution functions are modelled with a Royston-Parmar model with three internal knots. More details on this survival model are given in Appendix B. The fitted models are summarized in Table 4.1 by the corresponding maximized log-likelihoods together with the Kendall's τ association measures corresponding to the fitted copulas, c_{12} and c_{34}. The best fitting models, with and without time orderings, are the Clayton copula models. The further surrogacy analysis is therefore based on those two fitted models.

The goodness-of-fit is checked for both Clayton copula models. In Appendix E.1, the KM estimates are plotted with the corresponding model-based estimates of the survival functions. The model-based estimates match the KM estimates closely which indicates that the model fit is good. Moreover, the estimated strength of association is similar across different choices for the identifiable copulas (τ_{12} and τ_{34} in Table 4.1).

Table 4.1: Fitted models with margins based on the Royston-Parmar survival model with 3 internal knots. τ_{12} and τ_{34} are the Kendall's τ measures corresponding to the fitted identifiable copulas, c_{12} and c_{34}. Note that the identifiable copulas are survival copulas. $l(\cdot;\hat{\boldsymbol{\theta}})$: maximized log-likelihood.

Model		$l(\cdot;\hat{\boldsymbol{\theta}})$	τ_{12}	τ_{34}
Ordering	Copula			
No Ordering	Gaussian	-3110.87	0.77	0.83
	Clayton	-3063.41	0.79	0.86
	Frank	-3067.00	0.80	0.85
	Gumbel	-3068.44	0.79	0.84
Ordering	Gaussian	-3223.76	0.72	0.79
	Clayton	-3150.92	0.75	0.81
	Frank	-3171.03	0.75	0.85
	Gumbel	-3209.78	0.74	0.80

Table 4.2: Results of the sensitivity analysis for R_h^2 in the ovarian cancer data. Every row is based on n replications. M: monotonicity assumption, W-CA: weaker cross-association assumption.

Ordering	Assumptions	n	Range of R_h^2	$[p_1, p_{99}]$	median
Ordering	-	5000	[0.745, 0.992]	[0.842, 0.991]	0.981
	M	2234	[0.745, 0.991]	[0.817, 0.990]	0.969
	W-CA	4785	[0.813, 0.992]	[0.882, 0.991]	0.981
	M + W-CA	2107	[0.813, 0.991]	[0.857, 0.990]	0.971
No Ordering	-	5000	[0.662, 0.992]	[0.782, 0.990]	0.975
	M	2210	[0.662, 0.990]	[0.744, 0.987]	0.958
	W-CA	4661	[0.701, 0.992]	[0.833, 0.990]	0.975
	M + W-CA	2037	[0.701, 0.990]	[0.797, 0.987]	0.961

4.2.2 Sensitivity Analysis

The results of the sensitivity analyses are presented here for R_h^2. Only Clayton copulas as unobservable copulas are further considered, with rotations of 0, 90, 180 and 270 degrees. In Table 4.2, the results of the sensitivity analyses are given for all sets of assumptions and restrictions discussed previously. The corresponding frequency distributions are visualized in Figure 4.1. Irrespective of whether time orderings are taken into account and any additional assumptions, the results of the sensitivity analyses indicate that R_h^2 is large in all scenarios compatible with the observed data.

Similar results are obtained when the ICA is quantified with Spearman's ρ or Kendall's τ, as shown in Appendix E.2. However, the lower limits for the ICA are considerably smaller for Spearman's ρ and Kendall's τ. These are different measures of association, so differences are expected, although the magnitude of these differences warrants some further consideration. When the follow-up stops at about 14 years, a considerable number of patients have not experienced any event (see plots in Appendix E.1). The fitted survival functions beyond the end of the study are thus merely based on extrapolation. Moreover, the fitted survival functions become very flat after about 10 years. According to the fitted models, a considerable proportion of patients will therefore have a very large time-to-event. This also implies that very large individual causal effects are possible. This is problematic as it is mostly based on extrapolation beyond

the end of the study. The rank-based measures of association are not sensitive to these very large values, but the R_h^2 is possibly much more sensitive. The results for R_h^2 should thus be interpreted with extra care and ideally together with the rank-based measures.

The results of this case study also show that it is beneficial to take the time orderings explicitly into account. Indeed, looking at Table 4.2, one can see that the range is consistently shifted upwards when time orderings are taken into account, comparing the ranges under corresponding assumptions. This equivalently holds for the percentiles and medians in Table 4.2.

The additional assumptions can also have a considerable influence on the results. Assuming monotonicity has practically no effect on the results. While assuming weaker cross-association, with or without monotonicity, restricts the ranges considerably. These interpretations can also be derived from the histograms in Figure 4.1, though it should be noted that these histograms cannot be interpreted as posterior distributions. They rather represent the range of values for R_h^2 that are compatible with the observed data.

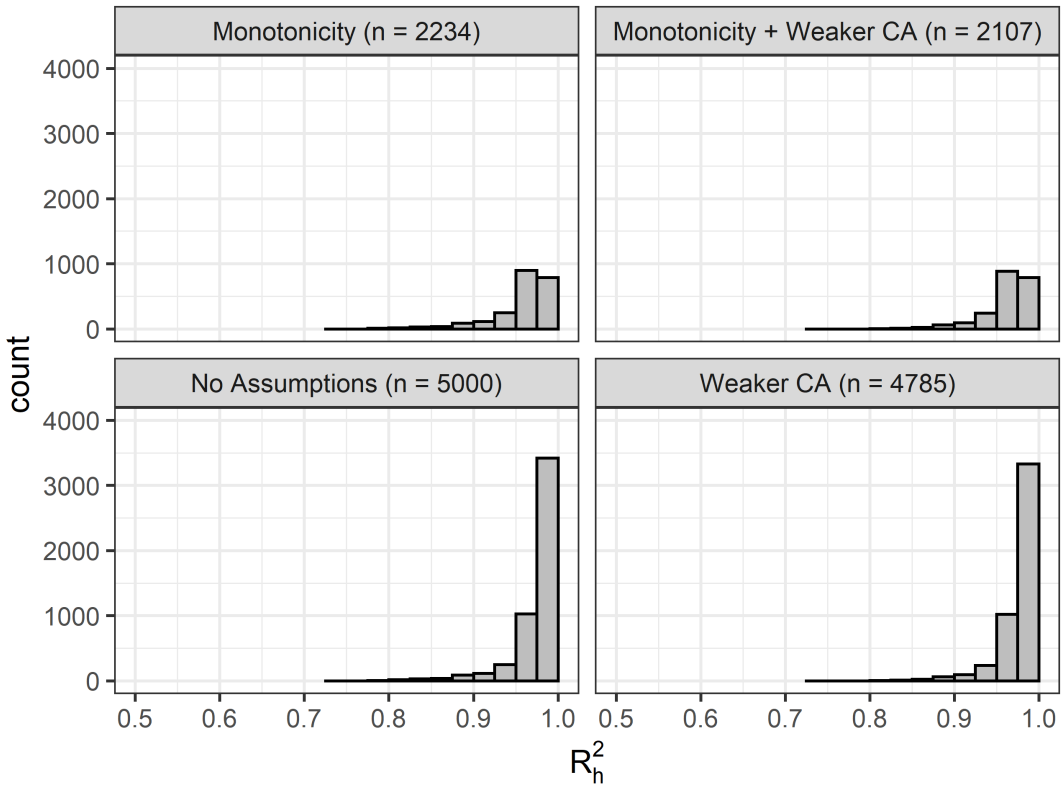
It is judged that both the monotonicity and weaker cross-association assumptions are realistic in this case study, as explained next.

Monotonicity in this case study means, roughly speaking, that patients that do better than average on CP ($Z = 0$), are also expected to do better than average on CAP ($Z = 1$) for both PFS and OS. There are certainly situations where the monotonicity assumption would not be well-justified. For example, consider a trial where chemotherapy is compared with immunotherapy. Also assume that there exists a genetic marker that is associated with prognosis; if the marker is present, prognosis is worse. In addition, if this marker is present, the immunotherapy will be very effective, otherwise the immunotherapy is not effective. The effectiveness of chemotherapy does not depend on any marker. In this case, a patient with a worse than average outcome on chemotherapy is more likely to have this marker. Consequently, this patient will tend to do better than average on immunotherapy. There is thus a negative association in this example, violating the monotonicity assumption.

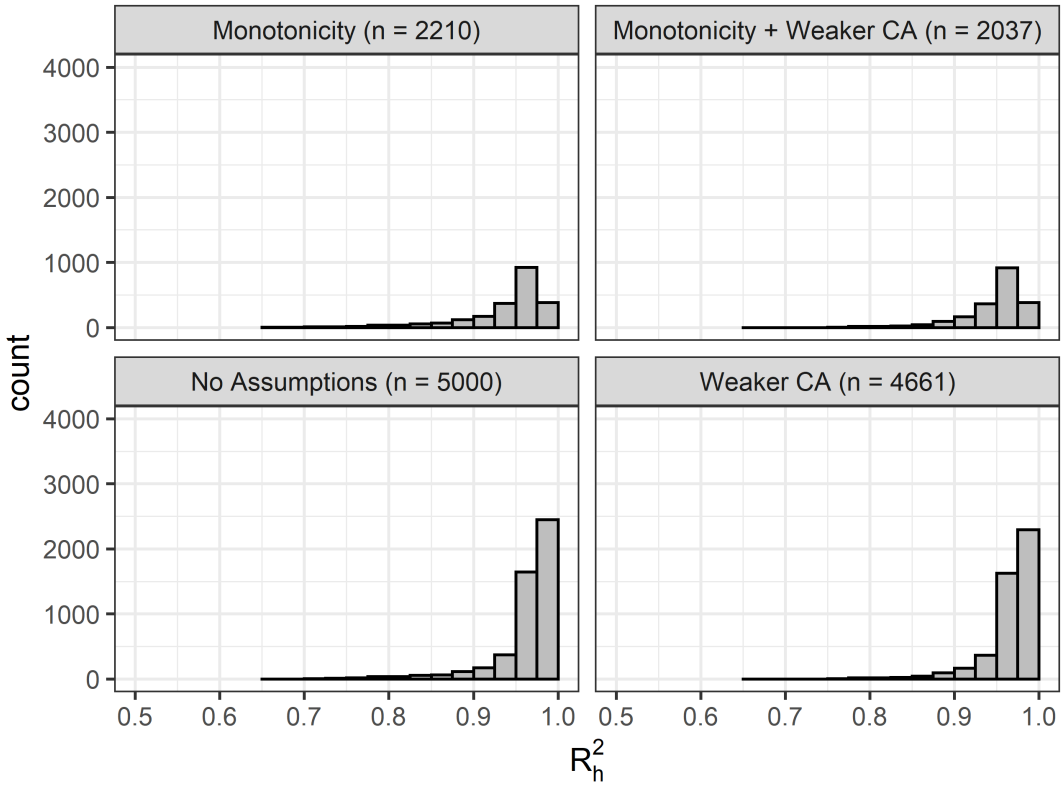
Weaker cross-association is a less clear-cut assumption than monotonicity. It can however be justified more generally than monotonicity. We assume that a potential outcome is determined by two components in addition to inherent variability: (i) prognostic factors and (ii) *treatment-specific* predictive factors. The prognostic factors induce a positive association between potential outcomes *across* and *within* treatments. Indeed, prognostic factors are expected to affect *all* potential outcomes in the same direction. They could be represented as a shared-frailty term for the four potential outcomes. Treatment-specific predictive factors only affect the potential outcomes under the same treatment. As in the immunotherapy example above, in particular disease (sub)types, some treatments may work better. This component causes a dilution of the association across treatments, but not within treatments. This component could be represented as a treatment specific shared-frailty term. Hence, if this component is present, one would expect the weaker cross-association assumption to hold.

Justifying these assumptions should be done in agreement with field experts. The justification of these assumptions for the ovarian cancer data was not done together with field experts and should as such be read.

Under these assumptions and taking time orderings into account, the fourth row of Table 4.2 contains the most relevant results. There, the R_h^2 is bounded between 0.813 and 0.991. The corresponding Spearman's ρ is bounded between 0.545 and 0.996, and the corresponding Kendall's τ is bounded between 0.389 and 0.956 (Table E.2 and E.1 in Appendix E). These results indicate that the ICA is strong across "all realities" compatible with the observed data.



(a) With time orderings



(b) Without time orderings

Figure 4.1: Frequency distribution of R_h^2 in the sensitivity analysis for different assumptions regarding the unidentifiable associations. Each histogram contains n replications.

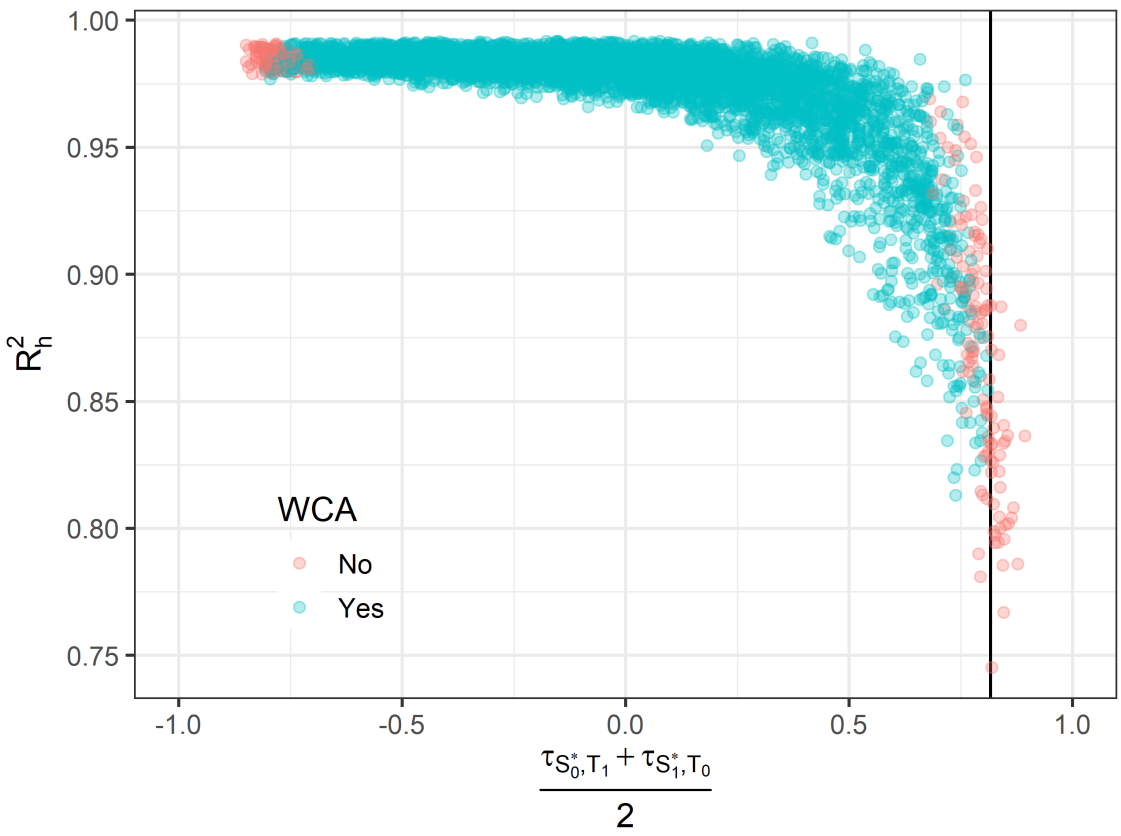


Figure 4.2: Plot of $n = 5000$ replications of the sensitivity analysis with time orderings taken into account without additional assumptions. The vertical line is at $\tau = 0.816$ which is the minimum of $\tau_{S_0^*,T_0}$ and $\tau_{S_1^*,T_1}$. Note that $\tau_{S_0^*,T_0}$ and $\tau_{S_1^*,T_1}$ are *not* the values as reported in Table 4.1. Rather, these values represent the association between S_k^* and T_k as detailed in Section 3.2.3. The points corresponding to replications where the weaker cross-association assumption is satisfied, are colored blue.

4.2.3 Further Exploration

In the previous subsection, an all-or-nothing approach is followed regarding additional assumptions in the sensitivity analysis. Alternatively, one can make no additional assumptions, but explore how the ICA varies with varying values of unidentifiable quantities. In this subsection, we further explore how the ICA varies with the *average cross-association*

$$\tau_{CA} = \frac{\tau_{S_0^*,T_1} + \tau_{S_1^*,T_0}}{2}. \tag{4.1}$$

for the model that takes the time orderings into account. In Figure 4.2, $n = 5000$ replications of the sensitivity analysis without additional assumptions are plotted. The average cross-association is plotted against the R_h^2. There is a relationship between both. Indeed, for a negative average cross-association, R_h^2 tends to be very strong. Conversely, the R_h^2 tends to be weaker for a positive cross-association. The vertical line indicates the minimum of $\tau_{S_0^*,T_0}$ and $\tau_{S_1^*,T_1}$. Under the weaker cross-association assumption, the average cross-association should be smaller than this value. In fact, the weaker cross-association assumption is slightly more restrictive than this because *both* τ_{S_0,T_1} and τ_{S_1,T_0} should be smaller than $\min(\tau_{S_0,T_0}, \tau_{S_1,T_1})$, not only the average τ_{CA}.

Using this plot, one could judge which values are plausible for the ICA. This introduces an extra degree of flexibility at the cost of increased subjectivity. Figure 4.2 also shows why inter-

preting the shapes of the histograms in Figure 4.1 might be misleading. One might judge values of τ_{CA} close to zero as very unlikely, but still possible. Thus, the corresponding R_h^2 values should be included in the results of the sensitivity analysis. However, as is clear from Figure 4.2, most replications lie in that region. This does not mean that such values are more likely than those in the sparser regions of Figure 4.2. Hence, these histograms should not be interpreted as reflecting the "posterior uncertainty".

From Figure 4.2, one can also deduce that for τ_{CA} close to 1, the R_h^2 might become too small for the surrogate to be useful. However, bounds obtained this way are of little relevance if they are based on underlying associations that are impossible. In most cases, even relatively weak assumptions on τ_{CA} could provide useful bounds. In this line of thought, one might prespecify a minimum value for the ICA for the surrogate to be deemed valid. Then, the restrictions needed on τ_{CA} to reach this lower bound can be determined. One can then judge how plausible it is that the putative surrogate is a valid surrogate, based on the plausibility of the "required" restrictions thus found.

4.2.4 Number of Replications

The data were reanalysed in a sensitivity analysis with $n = 100,000$ replications to examine how close the previous 5000 replications would be to the lower and upper bounds. The reanalysis only considered Spearman's ρ because of computational limitations. These additional results are reported in Table E.1 in Appendix E. The upper bounds for 5000 replications are very close to those for 100,000 replications, whereas there are some differences between the respective lower bounds. The lower bound for 5000 replications (no assumptions or monotonicity) is 0.447, while the respective lower bound for 100.000 replications is 0.379. However, when the additional weaker cross-association assumption is made, the difference is minimal. Indeed, the lower bound for 5000 replications (weaker cross-association, with or without monotonicity) is 0.545, while the respective lower bound for 100.000 replications is 0.528

4.3 Conclusion

There is convincing evidence that PFS is a valid surrogate for OS for the treatment of advanced ovarian cancer with CP and CAP. Assuming monotonicity and weaker cross-association, the individual causal association is strong across all scenarios compatible with the observed data, regardless of the measure used to quantify the association.

Discussion and Conclusion

Discussion

Assumptions Underlying Proposed Methodology

The explicit and implicit assumptions underlying the causal-inference framework and methods presented in this thesis, are discussed in some detail here.

In the definition of the potential outcomes, SUTVA is assumed. This is a standard assumption underlying methods that use Rubin's causal model. This assumption consists of two parts: (i) *no interference* and (ii) *no hidden variations of treatments* (Imbens & Rubin, 2015). The *no interference* part of SUTVA requires that the treatment applied to one patient does not affect the outcome for other patients. In most clinical trials, this holds. But there are some settings where this might be problematic, e.g., large vaccine trials where vaccination of one patient indirectly protects other patients. The *no hidden variations of treatments* part of SUTVA requires that an individual receiving a specific treatment level cannot receive different forms of that treatment. In clinical trials, the treatment formulation and administration are strictly controlled, so this part of SUTVA generally holds in clinical trials. Without SUTVA, the potential outcomes are not well defined, and the causal-inference approach cannot be applied as presented in this thesis.

The causal-inference approach also implicitly assumes a *non-zero variability* in the individual causal effects in the population. At first sight, this implicit assumption may seem trivial. Nonetheless, some statistical methods "rely" on a *zero variability* in the individual causal effects. For example, the sharp null hypothesis of no treatment effect states that for each patient in the experiment both values of the potential outcomes are identical. Although this sharp null hypothesis is for many applications too strong, it is important to realize that a non-zero variability in the individual causal effects a realistic, but not a trivial assumption. In principle, in a trial with a zero average treatment effect, there can still be a non-zero variability in the individual causal effects if both treatments have distinct effects which cancel out *on average*. The methods presented in this thesis can thus still be applied in such "negative" trials. However, if both treatments have *exactly* the same effect on each patient, then the sharp null hypothesis of no treatment effect is satisfied. In that case, the methods presented in this thesis are no longer valid as there is no variability in the individual causal effects.

The implicit assumption of a non-zero variability in the individual causal effects in the causal-inference framework has an analog in the meta-analytic framework. Indeed, it is required in the latter framework that there is variability in the *trial-level* treatment effects. Without variability in the trial-level treatment effects on the surrogate, α_i, a prediction function $\hat{\beta}_0 = f(\alpha_0)$ cannot be estimated (for values other than α_i).

The D-vine density is in general a density decomposition. Any joint density can be decomposed in this way. Using D-vines thus does not necessarily constitute an assumption. However, in the models presented in this thesis, the simplifying assumption for D-vines is made. In that case, the D-vine constitutes a joint density *construction*, but not in general a joint density *decomposition*. The sensitivity analysis is thus restricted in the sense that the measure of surrogacy is explored across joint densities compatible with the observed data where the simplifying assumption holds. Moreover, in the case study, only Clayton copulas (and rotations thereof) were

considered as the unidentifiable copulas. Arguably, this set of densities is still broad enough for the results of the sensitivity analysis to be useful.

Limitations

In this thesis, the ICA was quantified using three measures (R_h^2, Spearman's ρ and Kendall's τ) with a focus on R_h^2. It is for any of these measures difficult to provide general guidelines regarding the magnitude that is sufficient for a surrogate to be valid. The analysis presented in this thesis can be considered as a quantitative component complementary to other clinical and biological arguments in evaluating a surrogate.

Evaluating a surrogate in the single-trial setting comes with inherent limitations. Indeed, the results of the validation exercise are, strictly speaking, only valid in the well-defined population of the trial in which the validation exercise is carried out. This is of limited use as the goal of evaluating surrogates is usually to justify their use in new clinical trials. To justify such an extrapolation, clinical and biological arguments are required. This could be problematic because the therapies used in new trials are frequently not yet well-studied, and unexpected things can happen.

The ICA in the causal-inference framework is an "individual-level analog" of the trial-level surrogacy in the meta-analytic framework, but they still represent distinct measures of surrogacy that can lead to different conclusions. Trial-level surrogacy quantifies how well the trial-level treatment effect on the true endpoint can be predicted in a new trial based on the observed treatment effect on the surrogate endpoint. This formulation corresponds exactly to the goal of surrogate endpoints. In the causal-inference framework, trial-level treatment effects are replaced with individual causal effects. The surrogacy measure in the causal-inference framework thus quantifies how well the individual causal effect on the true endpoint can be predicted based on the (non-observable) individual causal effect on the surrogate endpoint. This is not directly relevant to the use of surrogates in practice. It is only relevant if one believes that good prediction at the level of individual causal effects extends to trial-level effects. In principle, accurate predictions can be possible in one level, but not the other.

As the discussion above indicates, the causal-inference framework should not be regarded as a substitute for the meta-analytic framework, but rather a complementary framework. Indeed, the meta-analytic framework is considered as the gold standard for evaluating surrogate endpoints. Still, in the early stages of drug development, limited data are available, possibly only from a single trial. In principle, the meta-analytic approach could still be considered by using e.g. centers as units of analysis, instead of trials. Such alternative clustering units are, however, not always available. In those cases, the causal-inference framework can be considered while still acknowledging its limitations.

Future Research

The causal-inference framework to surrogacy evaluation is an approach that only relatively recently gained more attention. It is therefore not as well-developed as the well-established meta-analytic approach which was proposed more than 20 years ago. Consequently, there are still many possibilities for future research in the causal-inference framework.

Many types of surrogate and true endpoint combinations have not yet been addressed in the causal-inference framework. Some relevant combinations are briefly discussed here. Tumor response is a putative surrogate for OS. Tumor response can be binary, ordinal, or continuous. Arguably, a continuous surrogate is more promising as it is more informative than binary or ordinal endpoints. The methods presented in this thesis can be easily adapted for the continuous-survival setting. Indeed, one could use marginals for the surrogate endpoints that are defined on the entire real line instead of marginals based on survival models that are only defined for $S_k > 0$.

Another possible direction of future research is to extend the proposed methods to multivariate surrogates. These surrogates can be both of a continuous or time-to-event type. The D-vine copula model can be readily extended to 6 dimensions for 2 surrogates, 8 dimensions for 3 surrogates and so on. However, Kendall's τ and Spearman's ρ do not readily extend to measure the association between ΔT and $(\Delta S_1, ..., \Delta S_k)'$. The SICC directly extends to measure this association, although this would be computationally prohibitive if the same numerical methods as in this thesis are used.

Conclusion

In this thesis, the causal-inference framework for validating surrogate endpoints was extended to the setting with time-to-event surrogate and time-to-event true endpoints. These methods were applied to the ovarian cancer data. The analysis found convincing evidence that PFS is a good surrogate for OS in advanced ovarian cancer patients when the efficacy of cyclophosphamide plus cisplatin is compared with cyclophosphamide plus cisplatin plus adriamycin.

R-code

The R-code is available from github.com/florianstijven/Master-Thesis. All analyses presented in this thesis can be replicated with the available R-code.

Data

The ovarian cancer data are freely available in the `Surrogate` R package (Van der Elst et al., 2021b).

Bibliography

Alonso, A. (2018). An information-theoretic approach for the evaluation of surrogate endpoints, 1–7. https://doi.org/https://doi.org/10.1002/9781118445112.stat08157

Alonso, A., Bigirumurame, T., Burzykowski, T., Buyse, M., Molenberghs, G., Muchene, L., Perualila, N. J., Shkedy, Z., & Van der Elst, W. (2017). *Applied surrogate endpoint evaluation methods with sas and r*. CRC Press.

Alonso, A., Van der Elst, W., & Meyvisch, P. (2017). Assessing a surrogate predictive value: A causal inference approach. *Statistics in Medicine*, *36*(7), 1083–1098.

Alonso, A., Van der Elst, W., Molenberghs, G., Buyse, M., & Burzykowski, T. (2015). On the relationship between the causal-inference and meta-analytic paradigms for the validation of surrogate endpoints. *Biometrics*, *71*(1), 15–24.

Alonso, A., Van der Elst, W., Molenberghs, G., Buyse, M., & Burzykowski, T. (2016). An information-theoretic approach for the evaluation of surrogate endpoints based on causal inference. *Biometrics*, *72*(3), 669–677.

Burzykowski, T. (2001). *Statistical evaluation methodology for surrogate endpoints in clinical studies* (Doctoral dissertation). UHasselt.

Buyse, M., & Molenberghs, G. (1998). Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics*, 1014–1029.

Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D., & Geys, H. (2000). The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*, *1*(1), 49–67.

Czado, C. (2019). Analyzing dependent data with vine copulas. *Lecture Notes in Statistics, Springer*.

FDA. (2018). *Surrogate endpoint resources for drug and biologic development*. Retrieved May 3, 2022, from https://www.fda.gov/drugs/development-resources/surrogate-endpoint-resources-drug-and-biologic-development

FDA. (2022). *Table of surrogate endpoints that were the basis of drug approval or licensure*. Retrieved May 3, 2022, from https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure

Fine, J. P., Jiang, H., & Chappell, R. (2001). On semi-competing risks data. *Biometrika*, *88*(4), 907–919.

for the Women's Health Initiative Investigators, W. G. et al. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the women's health initiative randomized controlled trial. *Jama*, *288*(3), 321–333.

Frangakis, C. E., & Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, *58*(1), 21–29.

Freedman, L. S., Graubard, B. I., & Schatzkin, A. (1992). Statistical validation of intermediate endpoints for chronic diseases. *Statistics in medicine*, *11*(2), 167–178.

Fu, H., Wang, Y., Liu, J., Kulkarni, P. M., & Melemed, A. S. (2013). Joint modeling of progression-free survival and overall survival by a bayesian normal induced copula estimation model. *Statistics in medicine*, *32*(2), 240–254.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, *81*(396), 945–960.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Jackson, C. (2016). flexsurv: A platform for parametric survival modeling in R. *Journal of Statistical Software*, *70*(8), 1–33. https://doi.org/10.18637/jss.v070.i08

Joe, H. (1989). Relative entropy measures of multivariate dependence. *Journal of the American Statistical Association*, *84*(405), 157–164.

Joffe, M. M., & Greene, T. (2009). Related causal frameworks for surrogate outcomes. *Biometrics*, *65*(2), 530–538.

Lee, K. H., Haneuse, S., Schrag, D., & Dominici, F. (2015). Bayesian semiparametric analysis of semicompeting risks data: Investigating hospital readmission after a pancreatic cancer diagnosis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *64*(2), 253–273.

Li, Y., Taylor, J. M., & Elliott, M. R. (2010). A bayesian approach to surrogacy assessment using principal stratification in clinical trials. *Biometrics*, *66*(2), 523–531.

Linfoot, E. H. (1957). An informational measure of correlation. *Information and control*, *1*(1), 85–89.

Nagler, T. (2018). kdecopula: An R package for the kernel estimation of bivariate copula densities. *Journal of Statistical Software*, *84*(7), 1–22. https://doi.org/10.18637/jss.v084.i07

Nagler, T., & Vatter, T. (2022). *Rvinecopulib: High performance algorithms for vine copula modeling* [R package version 0.6.1.1.3].

Omura, G., Buyse, M., Marsoni, S., Bertelsen, K., Conte, P., Jakobsen, A., & Vermorken, J. (1991). Cyclophosphamide plus cisplatin versus cyclophosphomide, doxorubicin, and cisplatin chemotherapy of ovarian carcinoma: A meta-analysis. *Journal of Clinical Oncology*, *9*(9), 1668–1674.

Pankevich, D. E., Altevogt, B. M., Dunlop, J., Gage, F. H., & Hyman, S. E. (2014). Improving and accelerating drug development for nervous system disorders. *Neuron*, *84*(3), 546–553.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics surveys*, *3*, 96–146.

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in medicine*, *8*(4), 431–440.

Royston, P., & Parmar, M. K. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in medicine*, *21*(15), 2175–2197.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, *8*, 229–231.

Sorrell, L., Wei, Y., Wojtyś, M., & Rowe, P. (2022). Estimating the correlation between semi-competing risk survival endpoints. *Biometrical Journal*, *64*(1), 131–145.

Van Der Elst, W. (2016). *Validation of surrogate endpoints from multiple randomized clinical trials with a failure-time true endpoint* (Doctoral dissertation). Limburgs Universitair Centrum.

Van der Elst, W., Abad, A. A., Coppenolle, H., Meyvisch, P., & Molenberghs, G. (2021a). The individual-level surrogate threshold effect in a causal-inference setting with normally distributed endpoints. *Pharmaceutical Statistics*.

Van der Elst, W., Alonso, A., Geys, H., Meyvisch, P., Bijnens, L., Sengupta, R., & Molenberghs, G. (2019). Univariate versus multivariate surrogates in the single-trial setting. *Statistics in Biopharmaceutical Research*.

Van der Elst, W., Meyvisch, P., Poveda, A. F., Alonso, A., Ensor, H. M., & Molenberghs, C. J. W. G. (2021b). *Surrogate: Evaluation of surrogate endpoints in clinical trials* [R package version 2.0].

Van der Elst, W., Molenberghs, G., & Alonso, A. (2016). Exploring the relationship between the causal-inference and meta-analytic paradigms for the evaluation of surrogate endpoints. *Statistics in medicine*, *35*(8), 1281–1298.

Appendix A

Proofs and Derivations

A.1 Density Corresponding to the Survival Copula

Lemma A.1.1 (Density corresponding to the survival copula)**.** *The joint density of* $(X, Y)'$ *is the product of the survival copula density and marginal densities.*

$$f_{XY}(x, y) = \tilde{c}_{XY}\{S_X(x), S_Y(y)\} \cdot f_X(x) \cdot f_Y(y) \tag{A.1}$$

where the survival copula density is defined as follows

$$\tilde{c}_{XY}\{S_X(x), S_Y(y)\} = \frac{\partial^2}{\partial u \partial v} \tilde{C}_{XY}(u, v)|_{u=S_X(x), v=S_Y(y)}. \tag{A.2}$$

Proof. The copula density c_{XY} follows from the “ordinary" copula C_{XY}

$$\begin{aligned} c_{XY}(1-u, 1-v) &= \frac{\partial^2}{\partial(1-u)\partial(1-v)} C_{XY}(1-u, 1-v) \\ &= \frac{\partial^2}{\partial(1-u)\partial(1-v)} \tilde{C}_{XY}(u, v) - u - v + 1 \\ &= \frac{\partial^2}{\partial u \partial v} \tilde{C}_{XY}(u, v) - u - v + 1 \\ &= \frac{\partial^2}{\partial u \partial v} \tilde{C}_{XY}(u, v) \\ &= \tilde{c}_{XY}(u, v). \end{aligned} \tag{A.3}$$

The joint density of $(X, Y)'$ follows from the “ordinary" copula density and corresponding marginal densities:

$$f_{XY}(x, y) = c_{XY}\{F_X(x), F_Y(y)\} \cdot f_X(x) \cdot f_Y(y). \tag{A.4}$$

Let $u = S_X(x)$ and $v = S_Y(y)$, then the following holds by Equation A.3

$$c_{XY}\{F_X(x), F_Y(y)\} = \tilde{c}_{XY}\{S_X(x), S_Y(y)\} \tag{A.5}$$

and thus replacing the ordinary copula density with the survival copula density in Equation A.4 gives the following result

$$f_{XY}(x, y) = \tilde{c}_{XY}\{S_X(x), S_Y(y)\} \cdot f_X(x) \cdot f_Y(y). \tag{A.6}$$

□

A.2 Likelihood Contribution for Partly Censored Data

The full derivation for the likelihood contribution of observations where only one of the two endpoints are censored, is given here. Let $\delta_{s,i} = 1$ and $\delta_{t,i} = 0$ for patient i. The likelihood contribution for such observations is derived as follows.

$$\begin{aligned}
L_i &= \int_{t_i}^{\infty} f_{S_k,T_k}(s_i,t)dt \\
&= \int_{t_i}^{\infty} c_{S_k,T_k}\left\{S_{S_k}(s_i), S_{T_k}(t_i)\right\} \cdot f_{S_k}(s_i) \cdot f_{T_k}(t_i)dt \\
&= \int_{t_i}^{\infty} \frac{\partial^2 C_{S_k,T_k}}{\partial s \partial t}\left\{S_{S_k}(s_i), S_{T_k}(t_i)\right\} dt \\
&= \frac{\partial}{\partial s}\int_{t_i}^{\infty} \frac{\partial C_{S_k,T_k}}{\partial t}\left\{S_{S_k}(s_i), S_{T_k}(t_i)\right\} dt \\
&= -\frac{\partial}{\partial s} C_{S_k,T_k}\left\{S_{S_k}(s_i), S_{T_k}(t_i)\right\} \\
&= -\frac{\partial}{\partial S_{S_k}} C_{S_k,T_k}\left\{S_{S_k}(s_i), S_{T_k}(t_i)\right\} \cdot \frac{dS_{S_k}}{ds} \\
&= f_{S_k}(s_i) \cdot \frac{\partial C_{S_k,T_k}}{\partial S_{S_k}}\left\{S_{S_k}(s_i), S_{T_k}(t_i)\right\}
\end{aligned} \tag{A.7}$$

The second equality follows from lemma 2.1.2. In the other steps, the rules of differentiation and integration are applied. The use of the survival copula results in convenient expressions for the observed likelihood for right censored data.

A.3 Copulas

The entire derivations for the partial derivatives are given in Sorrell et al. (2022, Supplementary Information) for the Clayton, Gumbel, and Frank copula. The derivations for the Gaussian copula are given in Fu et al. (2013).

A.3.1 Clayton

The Clayton copula is given by

$$C(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\delta}}. \tag{A.8}$$

The first derivatives of the Clayton copula are as follows:

$$\frac{\partial C(u,v)}{\partial u} = \frac{C(u,v)^{\theta+1}}{u^{\theta+1}} \tag{A.9}$$

and

$$\frac{\partial C(u,v)}{\partial u} = \frac{C(u,v)^{\theta+1}}{v^{\theta+1}}. \tag{A.10}$$

The second derivative of the Clayton copula is

$$\frac{\partial^2 C(u,v)}{\partial u \partial v} = \frac{(\theta+1)C(u,v)^{2\theta+1}}{u^{\theta+1}v^{\theta+1}}. \tag{A.11}$$

A.3.2 Gumbel

The Gumbel copula is given by

$$C(u,v) = \exp\left[-\left\{(-\log u)^{\theta} + (-\log v)^{\theta}\right\}\right]. \tag{A.12}$$

The first derivatives of the Gumbel copula are as follows

$$\frac{\partial C(u,v)}{\partial u} = \frac{C(u,v)(-\log u)^{\theta-1}}{u(-\log C(u,v))^{\theta-1}} \tag{A.13}$$

and

$$\frac{\partial C(u,v)}{\partial v} = \frac{C(u,v)(-\log v)^{\theta-1}}{v(-\log C(u,v))^{\theta-1}}. \tag{A.14}$$

The second derivative of the Gumbel copula is

$$\frac{\partial^2 C(u,v)}{\partial u \partial v} = \frac{C(u,v)(-\log u)^{\theta-1}(-\log v)^{\theta-1}(\theta - 1 - \log C(u,v))}{uv(-\log C(u,v))^{2\theta-1}}. \tag{A.15}$$

A.3.3 Frank

The Frank copula is given by

$$C(u,v) = -\frac{1}{\theta}\log\left[\frac{1}{1-e^{-\theta}}\left\{(1-e^{-\theta}) - (1-e^{-\theta u})(1-e^{-\theta v})\right\}\right]. \tag{A.16}$$

The first derivatives of the Frank copula are as follows

$$\frac{\partial C(u,v)}{\partial u} = \frac{1-e^{\theta C(u,v)}}{1-e^{\theta u}} \tag{A.17}$$

and

$$\frac{\partial C(u,v)}{\partial u} = \frac{1-e^{\theta C(u,v)}}{1-e^{\theta v}}. \tag{A.18}$$

The second derivative of the Frank copula is

$$\frac{\partial^2 C(u,v)}{\partial u \partial v} = \frac{\theta e^{\theta C(u,v)}(e^{\theta C(u,v)} - 1)}{(e^{\theta u} - 1)(e^{\theta v} - 1)}. \tag{A.19}$$

A.3.4 Gaussian

The Gaussian copula is given by

$$\Psi\left[\Phi^{-1}(u), \Phi^{-1}(v)\right]. \tag{A.20}$$

The first derivatives of the Gaussian copula are as follows

$$\frac{\partial C(u,v)}{\partial u} = \frac{\dot{\Psi}_u\left[\Phi^{-1}(u), \Phi^{-1}(v)\right]}{\phi\left\{\Phi^{-1}(u)\right\}} \tag{A.21}$$

and

$$\frac{\partial C(u,v)}{\partial u} = \frac{\dot{\Psi}_v\left[\Phi^{-1}(u), \Phi^{-1}(v)\right]}{\phi\left\{\Phi^{-1}(u)\right\}} \tag{A.22}$$

where

$$\begin{aligned}\dot{\Psi}_u\left[\Phi^{-1}(u),\Phi^{-1}(v)\right] &= \frac{\partial\Psi(u,v|\rho)}{\partial\Phi^{-1}(u)} \\ &= \phi\left\{\Phi^{-1}(u)\right\}\int_{\Phi^{-1}(v)}^{\infty}\phi(x|\Phi^{-1}(u),\rho)dx\end{aligned} \tag{A.23}$$

and

$$\begin{aligned}\dot{\Psi}_v\left[\Phi^{-1}(u),\Phi^{-1}(v)\right] &= \frac{\partial\Psi(u,v|\rho)}{\partial\Phi^{-1}(v)} \\ &= \phi\left\{\Phi^{-1}(v)\right\}\int_{\Phi^{-1}(u)}^{\infty}\phi(x|\Phi^{-1}(v),\rho)dx.\end{aligned} \tag{A.24}$$

The second derivative of the Gaussian copula is

$$\frac{\partial^2 C(u,v)}{\partial u\partial v} = \frac{\psi\left\{\Phi^{-1}(u),\Phi^{-1}(v)|\rho\right\}}{\phi\left\{\Phi^{-1}(u)\right\}\left\{\Phi^{-1}(v)\right\}}. \tag{A.25}$$

where $\phi(.)$ is the standard normal pdf and $\psi(.|\rho)$ is the standard bivariate normal pdf with correlation coefficient as ρ.

Appendix B

Royston-Parmar Model

Royston and Parmar (2002) extended the Weibull proportional hazards model and log-logistic proportional odds model to a flexible parametric modelling procedure. In this thesis, only the first is considered and explained here. The hazard at time t, $h(t;\boldsymbol{x})$, in a Weibull proportional hazards model (PH) is

$$h(t;\boldsymbol{x}) = h_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x}) \tag{B.1}$$

where $h_0(t) = \lambda\gamma t^{\gamma-1}$ is the baseline hazard function. γ and λ are the shape and scale parameters of the corresponding Weibull distribution. The log-cumulative hazard of this model is

$$\log H(t;\boldsymbol{x}) = \boldsymbol{\beta}'\boldsymbol{x_i} + \log\lambda + \gamma\log t. \tag{B.2}$$

Let $\eta = \boldsymbol{\beta}'\boldsymbol{x_i}$, $\gamma_0 = \log\lambda$, $\gamma_1 = \gamma$ and $z = \log t$. The log-cumulative hazard can thus be rewritten as

$$\log H(t;\boldsymbol{x}) = \gamma_0 + \gamma_1 \cdot z + \eta. \tag{B.3}$$

For the Weibull proportional-hazards model, the log-cumulative hazard is thus linear in $z = \log t$.

If the distribution of $T|\boldsymbol{X} = \boldsymbol{x}$ departs from the Weibull distribution, then $\log H(t;\boldsymbol{x})$ will be related to x by a non-linear function $s + \eta = s(z;\boldsymbol{\gamma}) + \eta$ where $s = \gamma_0 + \gamma_1 \cdot z$ in the conventional Weibull PH model. The survival, density and hazard functions of this extended model are

$$S(t;\boldsymbol{\gamma},\eta) = \exp\{-\exp(s+\eta)\} \tag{B.4}$$

$$f(t;\boldsymbol{\gamma},\eta) = \exp\{s - \exp(s)\} \cdot \frac{ds}{dt} \tag{B.5}$$

$$h(t;\boldsymbol{\gamma},\eta) = \exp(s) \cdot \frac{ds}{dt}. \tag{B.6}$$

The approach in the Royston-Parmar model is to model $s(z;\boldsymbol{\gamma})$ as a natural cubic spline function. This entails that the baseline log-cumulative hazard is modelled as a natural cubic splines function of log time.

Natural cubic splines are defined as cubic splines constrained to be linear beyond the boundary knots k_{min} and k_{max}. These boundary knots are placed at the extreme uncensored log survival times in the Royston-Parmar model. In addition, m distinct internal knots $k_1 < ... < k_m$, with $k_{min} < k_1$ and $k_m < k_{max}$, are specified. These are specified based on the centiles of the distribution of uncensored log survival times as shown in B.1. Given the boundary and internal knots the natural cubic spline is written as

$$s(z;\boldsymbol{\gamma}) = \gamma_0 + \gamma_1 z + \gamma_2 v_1(z) + ... + \gamma_{m+1} v_m(z) \tag{B.7}$$

where $v_j(z)$ is the j'th basis function. This basic function is defined as follows

$$v_j(z) = (z-k_j)_+^3 - \lambda_j(z-k_{min})_+^3 - (1-\lambda_j)(z-k_{max})_+^3 \tag{B.8}$$

Table B.1: Placement of internal knots in the Royston-Parmar model. (Adapted from (Royston & Parmar, 2002))

m	Centiles		
1	50		
2	33	67	
3	25	50	75

where

$$\lambda_j = \frac{k_{max} - k_j}{k_{max} - k_{min}} \quad \text{and} \quad (z-a)_+ = \max(0, z-a). \tag{B.9}$$

Algebraic details on the natural cubic splines can be found in Appendix B of Royston and Parmar (2002). The Royston-Parmar model can be fitted with maximum likelihood. The likelihood contributions of observations follow directly from Equation B.4 and B.5 for censored and uncensored observations, respectively. Note that $\frac{ds}{dx} = \frac{ds(x;\gamma)}{dx}$ from Equation B.5 is easily obtained

$$\begin{aligned} \frac{ds}{dx} &= \gamma_1 + \sum_{j=2}^{m} \gamma_j \frac{dv_j(x)}{dx} \\ &= \gamma_1 + \sum_{j=2}^{m} \gamma_j [3(x-k_j)_+^2 - 3\lambda_j (x-k_{min})_+^2 - 3(1-\lambda_j)(x-k_{max})_+^2]. \end{aligned} \tag{B.10}$$

This model can be fitted with the `flexsurvspline` function from the `flexsurv` R-package (Jackson, 2016). In addition, this package provides additional useful functions for the Royston-Parmar model such as functions to compute the density and survival function.

As detailed above, the Royston-Parmar model reduces to the Weibull distribution when there are zero internal knots. However, there are different parameterisations of the Weibull distribution available; the connections between the Weibull parameters for different parameterisations and the Royston-Parmar parameters are given next.

Parameterisation 1

The Weibull distribution is parameterised as in the above derivations.

$$f(x;\lambda,\gamma) = \lambda\gamma x^{\gamma-1} e^{-\lambda x^\gamma} \quad \text{and} \quad h(x;\lambda,\gamma) = \lambda\gamma t^{\gamma-1} \tag{B.11}$$

The relation with the Royston-Parmar model parameters is as follows:

$$\lambda = e^{\gamma_0} \quad \text{and} \quad \gamma = \gamma_1. \tag{B.12}$$

Parameterisation 2

The Weibull distribution is parameterised as in `dweibull` from the R `stats` package.

$$f(x;\lambda,k) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} \quad \text{and} \quad h(x;\lambda,k) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} \tag{B.13}$$

The relation with the Royston-Parmar model parameters is as follows:

$$\lambda = e^{-\frac{\gamma_0}{\gamma_1}} \quad \text{and} \quad k = \gamma_1. \tag{B.14}$$

Appendix C

Vine Copulas

Table C.1: Full expressions of the components of the D-vine construction in Equation 2.15.

Component	Expression
f_1	$= f_{S_0}(s_0)$
f_2	$= f_{T_0}(t_0)$
f_3	$= f_{S_1}(s_1)$
f_4	$= f_{T_1}(t_1)$
c_{12}	$= c_{S_0,T_0}\left\{F_{S_0}(s_0), F_{T_0}(t_0)\right\}$
c_{23}	$= c_{T_0,S_1}\left\{F_{T_0}(t_0), F_{S_1}(s_1)\right\}$
c_{34}	$= c_{S_1,T_1}\left\{F_{S_1}(s_1), F_{T_1}(t_1)\right\}$
$c_{13;2}$	$= c_{S_0,S_1;T_0}\left\{F_{S_0\|T_0}(s_0\|T_0 = t_0), F_{S_1\|T_0}(s_1\|T_0 = t_0)\right\}$
$c_{24;3}$	$= c_{T_0,T_1;S_1}\left\{F_{T_0\|S_1}(t_0\|S_1 = s_1), F_{T_1\|S_1}(t_1\|S_1 = s_1)\right\}$
$c_{14;23}$	$= c_{S_0,T_1;T_0,S_1}\left\{F_{S_0\|T_0,S_1}(s_0\|T_0 = t_0, S_1 = s_1), F_{T_1\|T_0=t_0,S_1=s_1}(t_1\|T_0 = t_0, S_1 = s_1)\right\}$

Appendix D

Monte Carlo Integration

A simulation study is conducted to asses the accuracy of the Monte Carlo integration procedure to compute Kendall's τ, Spearman's ρ and R_h^2. The parameters for the marginal distributions are given in Table D.1. Clayton copulas are used as identifiable copulas. The corresponding association parameter is fixed at $\delta = 6$, which corresponds to $\tau = 0.75$. As unidentifiable copulas, four options are considered. For each option, all four unidentifiable copulas are assumed to be one of the following parametric copulas: Gaussian, Clayton, Frank or Gumbel. The Spearman's ρ values for corresponding copulas across these four options are set equal to each other. These are given in the second column of Table D.2. The corresponding copula parameters are given in the third to sixth columns. No time orderings are assumed.

The results are given in Table D.3. The number of Monte Carlo samples from f_{1234} and quasi Monte Carlo samples in the `dep_measures` function are set equal to each other, and further denoted by N. All measures are computed 50 times under the same settings to examine how variable the computations are. The values in this table are the means across these 50 computations, and the corresponding standard deviations. Only the computation for $N = 100.000$ is not repeated; this can be considered as the true value.

The standard deviations in Table D.3 show that the Monte Carlo procedure increases in precision as N increases. If a standard deviation of 0.01 is deemed acceptable, then $N = 2000$ results in a sufficient precision. In consideration of the uncertainty due to the unidentifiability of some parameters, this precision is deemed acceptable.

The computed measures of surrogacy are all very close to each other. This indicates that the specific choice for the parametric form of the unidentifiable copulas doesn't influence the measures of surrogacy much, if the Spearman's ρ values of the corresponding copulas are the same.

Table D.1: Parameters for the marginal distributions of the simulation study. λ and k are the scale and shape parameters of the Weibull distribution, respectively.

Outcome	λ	k	$E(Y)$
S_0	0.368	2	0.326
S_1	0.472	2	0.419
T_0	0.607	2	0.538
T_1	0.687	2	0.609

Table D.2: The unidentifiable copula parameters are fixed at these values. *: 90 degree rotation copula.

Copula	ρ_s	Copula Parameter: $\theta_{..;.}$			
		Gaussian	Clayton	Frank	Gumbel
c_{23}	-0.266	-0.277	0.436*	-1.650	1.218*
$c_{13;2}$	0.372	0.387	0.683	2.399	1.343
$c_{24;3}$	0.572	0.591	1.375	4.164	1.692
$c_{14;23}$	0.908	0.916	5.916	12.896	3.895

Table D.3: Results of the simulations to assess the accuracy of the Monte Carlo integration. N is the number of MC samples from f_{1234}, and also the number of quasi MC samples to compute R_h^2 with the `dep_measures` function. The approximated value is the mean over the 50 simulations, and SD is the corresponding standard deviation.

Unid. Cop.	N	Approximated value (SD)		
		ρ_s	τ	R_h^2
Gaussian	500	0.864 (0.016)	0.688 (0.018)	0.808 (0.019)
	1000	0.863 (0.010)	0.685 (0.011)	0.804 (0.011)
	2000	0.864 (0.007)	0.685 (0.008)	0.803 (0.008)
	5000	0.865 (0.005)	0.687 (0.006)	0.804 (0.006)
	100.000	0.866 (NA)	0.689 (NA)	0.803 (NA)
Clayton	500	0.869 (0.014)	0.692 (0.016)	0.818 (0.017)
	1000	0.869 (0.009)	0.691 (0.010)	0.815 (0.010)
	2000	0.870 (0.006)	0.691 (0.008)	0.815 (0.008)
	5000	0.871 (0.004)	0.693 (0.005)	0.817 (0.005)
	100.000	0.873 (NA)	0.694 (NA)	0.815 (NA)
Frank	500	0.866 (0.015)	0.688 (0.018)	0.807 (0.02)
	1000	0.866 (0.009)	0.687 (0.010)	0.804 (0.010)
	2000	0.866 (0.007)	0.687 (0.008)	0.803 (0.008)
	5000	0.868 (0.005)	0.688 (0.006)	0.805 (0.005)
	100.000	0.868 (NA)	0.689 (NA)	0.803 (NA)
Gumbel	500	0.863 (0.016)	0.686 (0.019)	0.806 (0.019)
	1000	0.862 (0.01)	0.683 (0.011)	0.803 (0.010)
	2000	0.862 (0.008)	0.683 (0.009)	0.803 (0.009)
	5000	0.863 (0.005)	0.685 (0.006)	0.803 (0.006)
	100.000	0.865 (NA)	0.686 (NA)	0.802 (NA)

Appendix E

Additional Results Ovarian Cancer Data

E.1 Goodness of Fit

The KM estimates are plotted with the corresponding model-based estimates of the survival functions in Figure E.1. The model based estimates follow the KM estimates closely, which indicates that the model fit is good.

For the model without time-orderings, the marginal survival functions as plotted in Figure E.1b, follow directly from the fitted model. This also holds for the marginal survival functions for OS in the model with time orderings, Figure E.1a. However, in the model with time orderings, the marginal survival functions for PFS follow from the fitted model by Equation 3.11.

E.2 Results for Other Surrogacy Measures

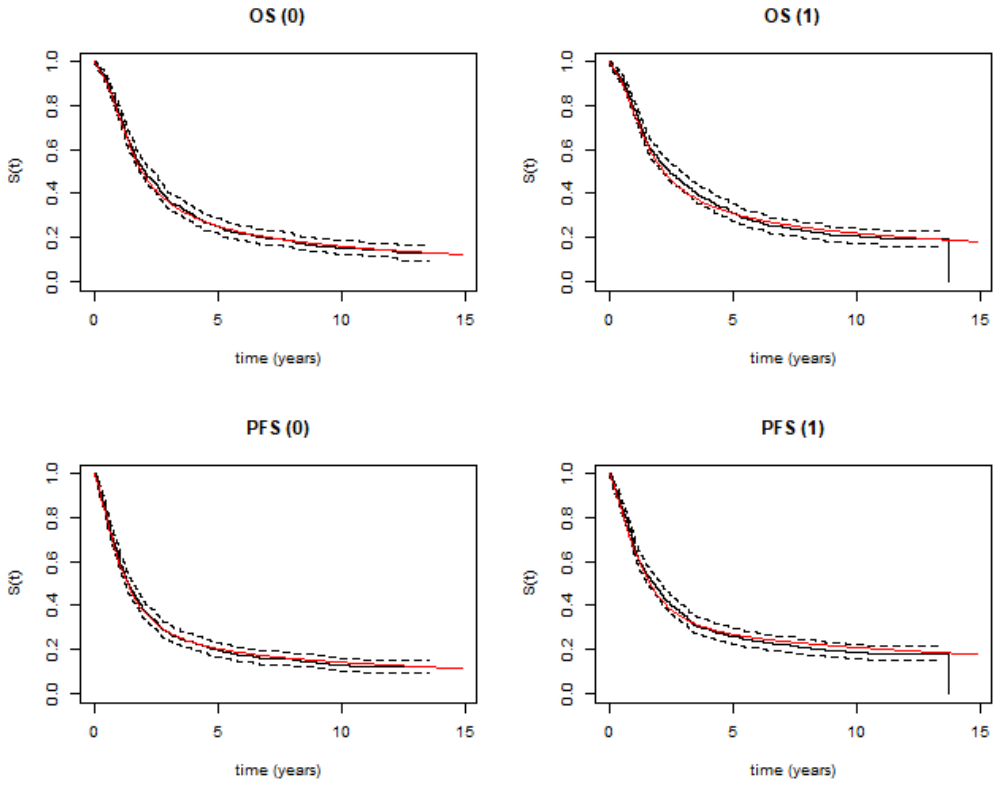
In the main text, only the results for R_h^2 are reported. The same results for Spearman's ρ and Kendall's τ are reported here in Table E.1 and E.2, respectively. Additionally, the results of the sensitivity analysis with 100.000 replications, instead of 5000, are reported in Table E.1.

Table E.1: Results of the sensitivity analysis for the ovarian cancer data for Spearman's ρ. Every row is based on n replications. M: monotonicity assumption, W-CA: weaker cross-association assumption.

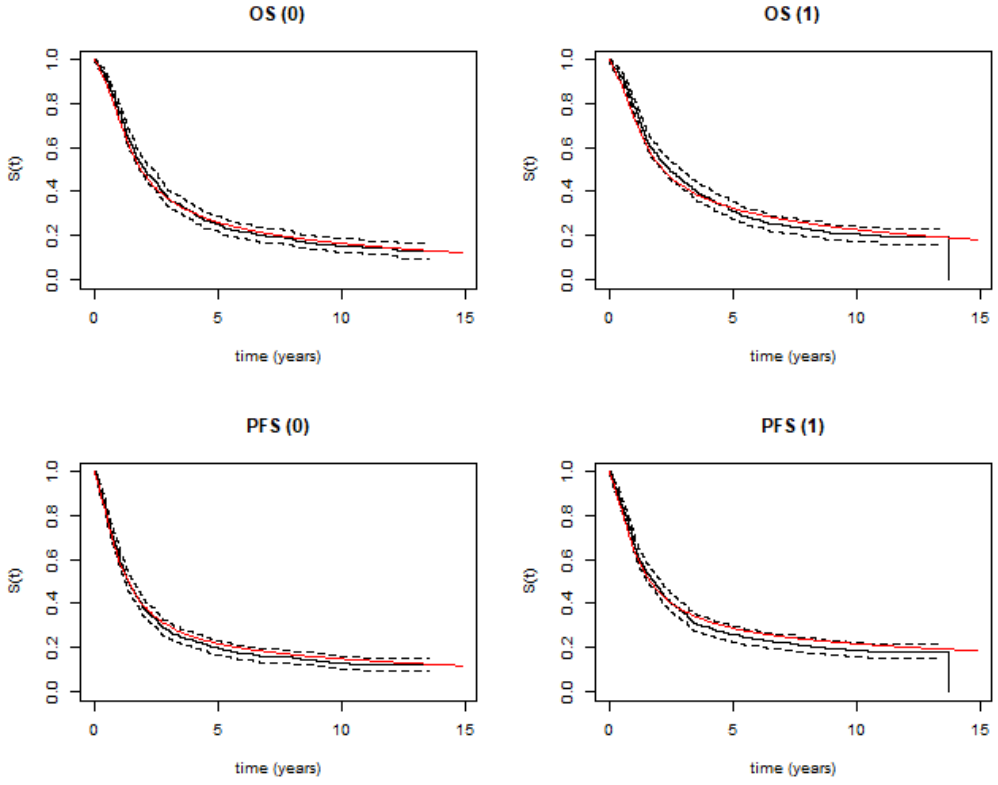
Ordering	Assumptions	n	Range of ρ_s	$[p_1, p_{99}]$	median
Ordering	-	5000	[0.447, 0.998]	[0.673, 0.995]	0.978
	M	2234	[0.447, 0.996]	[0.587, 0.992]	0.955
	W-CA	4785	[0.545, 0.998]	[0.759, 0.995]	0.978
	M + W-CA	2107	[0.545, 0.996]	[0.711, 0.992]	0.957
No Ordering	-	5000	[0.296, 0.997]	[0.627, 0.994]	0.972
	M	2210	[0.296, 0.993]	[0.555, 0.987]	0.931
	W-CA	4661	[0.605, 0.997]	[0.754, 0.994]	0.972
	M + W-CA	2037	[0.605, 0.993]	[0.712, 0.987]	0.936
Ordering	-	100.000	[0.379, 0.999]	[0.685, 0.995]	0.978
	M	43.673	[0.379, 0.998]	[0.612, 0.992]	0.954
	W-CA	95.926	[0.528, 0.999]	[0.763, 0.995]	0.978
	M + W-CA	41.581	[0.528, 0.998]	[0.715, 0.992]	0.956

Table E.2: Results of the sensitivity analysis for the ovarian cancer data for Kendall's τ. Every row is based on n replications. M: monotonicity assumption, W-CA: weaker cross-association assumption.

Ordering	Assumptions	n	Range of τ	$[p_1, p_{99}]$	median
Ordering	-	5000	[0.305, 0.972]	[0.521, 0.952]	0.901
	M	2234	[0.305, 0.956]	[0.433, 0.941]	0.854
	W-CA	4785	[0.389, 0.972]	[0.607, 0.953]	0.902
	M + W-CA	2107	[0.389, 0.956]	[0.563, 0.942]	0.858
No Ordering	-	5000	[0.151, 0.959]	[0.468, 0.942]	0.887
	M	2210	[0.151, 0.941]	[0.401, 0.927]	0.821
	W-CA	4661	[0.452, 0.959]	[0.599, 0.942]	0.887
	M + W-CA	2037	[0.452, 0.941]	[0.557, 0.928]	0.828



(a) With time orderings



(b) Without time orderings

Figure E.1: Goodness of fit for the fitted Clayton copula models. The Kaplan-Meier estimates of the survival functions for PFS and OS are given as solid black lines with the pointwise 95% confidence interval as dashed black lines. The model-based survival functions are shown by solid red lines.