An efficient algorithm to assess multivariate surrogate endpoints in a causal inference framework

Alvaro J. Flórez[a,b,*], Geert Molenberghs[b,d], Wim Van der Elst[c], Ariel Alonso Abad[d]

[a] *School of statistics, Universidad del Valle, Cali, Colombia*
[b] *Data Science Institute, I-Biostat, Universiteit Hasselt, Diepenbeek, Belgium.*
[c] *Janssen Pharmaceuticals, Beerse, Belgium.*
[d] *I-BioStat, KU Leuven, Leuven, Belgium.*

Abstract

Multivariate surrogate endpoints can improve the efficiency of the drug development process, but their evaluation raises many challenges. Recently, the so-called individual causal association (ICA) has been introduced for validation purposes in the causal-inference paradigm. The ICA is a function of a partially identifiable correlation matrix ($\mathbf{R}$) and, hence, it cannot be estimated without making untestable assumptions. This issue has been addressed via a simulation-based analysis. Essentially, the ICA is assessed across a set of values for the non-identifiable entries in $\mathbf{R}$ that lead to a valid correlation matrix and this has been implemented using a fast algorithm based on partial correlations (PC). Using theoretical arguments and simulations, it is shown that, in spite of its computational efficiency, the PC algorithm may lead to the spurious effect that adding non-informative surrogates, i.e., surrogates that convey no information on the treatment effect on the true endpoint, seemingly reduces the ICA range. To address this, a modified PC algorithm (MPC) is proposed. Based on simulations, it is shown that the MPC algorithm removes this nuisance effect and increases computational efficiency.

Keywords: Causal inference, Matrix completion problem, Multiple surrogate

*The R source code for the simulations of this article and an example are publicly available at github.com/AlvaroFlorez/MPCalgorithm.

*Corresponding author
Email address: `alvaro.florez@uhasselt.be` (Alvaro J. Flórez)

evaluation, Random correlation matrix, Partial correlation, Sensitivity analysis

1. Introduction

In the surrogate endpoint evaluation literature, a true endpoint is defined as the most sensitive and clinically relevant indicator of disease progression and drug response (Biomarkers Definitions Working Group, 2001). Consequently, when a clinical trial aims to determine the effect of a new therapy or treatment, this true endpoint should ideally be used. However, often in practice, its measurement may be expensive or involve medical procedures that are uncomfortable or even dangerous for the patient. In other situations, assessing the drug effect on the true endpoint may require a long follow up time, increasing the duration and cost of the trial. For example, in the current COVID-19 pandemic, evaluating the protective effect of a vaccine candidate may imply the need to follow vaccinated and unvaccinated individuals for a long period of time, in order to compare the infection rates in both groups. In addition, the strict measures taken by authorities around the world may lead to an insufficient number of infections at the end of follow-up, hindering the evaluation of the vaccine candidate. An attractive option in such a situation is to find a less complex "substitute" endpoint that can be measured easier, cheaper, and/or closer in time. Measures of immune response may play such a role in vaccine trials. These replacement outcomes are called surrogate endpoints (Burzykowski et al., 2005; Buyse et al., 2010; Alonso et al., 2016). However, before marketing a drug evaluated using a surrogate endpoint, one needs to show that the surrogate is valid for the true endpoint of interest. For instance, one needs to show that the effect of a vaccine on the surrogate (e.g., measures of immune response) can accurately predict its effect on the true endpoint (protection from infection, hospitalization after infection, intensive care admission, mortality, etc.). Several methods have been proposed for the statistical evaluation of surrogate endpoints, most of them within the so-called causal-inference (Alonso et al., 2015; Van der Elst et al., 2015, 2019) and meta-analytic (Buyse et al., 2000;

Burzykowski et al., 2005; Molenberghs et al., 2010; Buyse et al., 2015) frameworks.

Although in both frameworks most research has focused on univariate surrogate endpoints, it has been argued in recent years that the prediction of the treatment effect on the true endpoint could be considerably improved when multivariate/multiple surrogates are jointly considered. The evaluation of a multivariate and, perhaps high-dimensional surrogate, raises many methodological and computational challenges. For instance, in the causal-inference paradigm, Alonso et al. (2015) proposed assessing univariate surrogate endpoints using the individual causal association (ICA), a metric defined as the association between the individual causal treatment effects on the surrogate and true endpoints. Recently, Van der Elst et al. (2019) extended the ICA to the multivariate setting but, in this scenario, the ICA becomes a function of an only partially identifiable correlation matrix ($\mathbf{R}$) and, hence, it cannot be fully estimated from the data. To deal with this identifiability issue, Alonso et al. (2015) and Van der Elst et al. (2019) proposed a simulation-based analysis. Basically, at each run, the algorithm finds a completion of the non-identifiable $\mathbf{R}$ that is compatible with the data at hand and, based on it, it estimates the ICA. The ICA values obtained from the completed $\mathbf{R}$ matrices allow us to assess the validity of the putative multivariate surrogate across different "plausible" realities compatible with the data available. The range is particularly important because it quantifies the uncertainty associated with the non-identifiability of $\mathbf{R}$.

The algorithm introduced by Van der Elst et al. (2019) to implement the aforementioned simulation-based analysis is computationally costly and can only reasonably handle surrogate endpoints of dimension four or smaller. However, in several applications one may want to consider surrogates of higher dimensions. To address this problem, Flórez et al. (2020) built upon the work of Joe (2006) and Lewandowski et al. (2009), and proposed a new algorithm for carrying out the simulation-based analysis. This method is computationally fast, even in the joint evaluation of a large number of surrogates, and easy to implement in standard statistical software packages (see the `ICA.ContCont.MultS.PC` function in

`Surrogate` R package). In the present work, we show, based on theoretical arguments and simulations, that in the presence of non-informative surrogates, Flórez et al. (2020)'s proposal struggles to completely cover the space of the ICA values compatible with the data. Indeed, even after one million runs, the algorithm produces an artificial shrinkage of the range of the ICA values when non-informative surrogates are jointly evaluated, leading to the misguided belief that adding these outcomes improves surrogacy.

Although at first sight not so harmful from a statistical perspective, substituting a true endpoint by a multivariate proxy that includes both meaningful as well as meaningless surrogates can be very counterproductive. For example, it increments the cost of the trial, puts patients through unnecessary clinical tests, and/or increases the chances of missing data. Therefore, by adding extra information to the matrix completion problem, we introduce an improved version of Flórez et al. (2020)'s algorithm that substantially mitigates the negative effect of adding non-informative surrogates on the range of the ICA values, without losing computational efficiency. The new proposal has the potential of allowing the evaluation of high-dimensional surrogates at an affordable computational cost.

The paper is organized as follows. Section 2 presents a clinical trial with chronic schizophrenic patients. Section 3 introduces the methodology for evaluating multiple surrogates. Section 4 describes the simulation-based analysis for estimating the ICA based on two different algorithms for the matrix completion problem. In Section 5, the effect of adding non-informative surrogates to the surrogacy assessment is illustrated. The modification of Flórez et al. (2020)'s algorithm is presented in Section 6. The algorithms are compared using a simulated scenario and the case study, in Sections 7 and 8, respectively. Finally, Section 9 is reserved for concluding remarks.

2. A case study in schizophrenia

This is a single-blind multicenter clinical trial in which patients with chronic schizophrenia were treated with a placebo and different doses of antipsychotic drugs (Chouinard et al., 1993; Marder and Meibach, 1994). Particularly, we are using a sub-sample consisting of 170 patients randomly assigned to placebo (85 individuals) and 6 mg/day risperidone treatment (85 individuals) for eight weeks. Patients were evaluated using the Positive and Negative Syndrome Scale (PANSS, Kay et al., 1987). PANSS is a standardized instrument for measuring symptom severity of schizophrenia. Patients are evaluated based on PANSS total score, i.e., the sum of the 30 items each measured in a 7-grade scale. These items describe the symptoms in three dimensions: positive (e.g., hallucination and delusion), negative (e.g., emotional withdrawal and blunted affect), and general psychopathology (e.g., depression and disorientation). However, posterior empirical studies suggest the 30 items can be divided into five factors (Lindenmayer et al., 1995; Van der Gaag et al., 2006; Alonso et al., 2010). In the present work, we will focus on the five factors proposed by Lindenmayer et al. (1995): negative (based on 6 items), positive (based on 5 items), cognitive (based 5 items), excitement (based 4 items), and depression (based on 5 items). The remaining 5 PANSS items are not considered in any of these dimensions. The total PANSS score is the true endpoint of the study whereas the aforementioned five factors can be considered as putative surrogates endpoints: negative (S_1), excitement (S_2), cognitive (S_3), positive (S_4), and depression (S_5).

3. Multiple surrogates assessment

We follow the methodology proposed by Alonso et al. (2015) and Van der Elst et al. (2019) for evaluating univariate and multivariate surrogate endpoints, respectively. Consider the single-trial setting, in which a univariate true endpoint T and a multivariate surrogate endpoint $\mathbf{S} = (S_1, S_2, \ldots, S_p)'$ are collected from

an independent sample of N patients, with all outcomes continuous. Additionally, only two treatments are under evaluation ($Z = 0/1$) in a parallel study design.

Following the so-called Rubin's causal-inference model (Rubin, 1986), let $\mathbf{Y} = (T_0,\ T_1,\ S_{10},\ S_{11},\ldots,\ S_{p0},\ S_{p1})'$ be the $2(p+1)$-dimensional vector of potential outcomes, in which $(T_0, S_{10}, S_{20}, \ldots, S_{p0})$ corresponds to the endpoints under the control condition ($Z = 0$), and $(T_1, S_{11}, S_{21}, \ldots, S_{p1})$ to the endpoints under the treatment condition ($Z = 1$). For simplicity, no subindex has been used to denote the patient. Furthermore, let $\boldsymbol{\Delta} = (\Delta T, \boldsymbol{\Delta}\mathbf{S}')'$ be the individual causal treatment effect, where $\Delta T = T_1 - T_0$ and $\boldsymbol{\Delta}\mathbf{S} = (\Delta S_1, \Delta S_2, \ldots, \Delta S_p)'$ with $\Delta S_k = S_{k1} - S_{k0}$. The so-called fundamental problem of causal inference states that only one of the potential outcomes associated with the true and surrogate endpoints are observed in practice. Therefore, $\boldsymbol{\Delta}$ cannot be estimated from the data (Holland, 1986). Nevertheless, under fairly general identifiability conditions (Rosenbaum and Rubin, 1983), it is possible to obtain consistent estimators of the expected causal treatment effects $E(\boldsymbol{\Delta}) = (\beta, \boldsymbol{\alpha}')'$, where $\beta = E(\Delta T)$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$, with $\alpha_k = E(\Delta S_k)$.

In the surrogacy evaluation context, however, we are more interested in the distribution of $\mathbf{Y}$ rather than in the expected causal effect. Consequently, let us assume that $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{DRD})$, where $\boldsymbol{\mu} = (\mu_{T_0}, \mu_{T_1}, \mu_{S_{10}}, \mu_{S_{11}}, \ldots, \mu_{S_{p0}}, \mu_{S_{p1}})'$, $\mathbf{D}$ is a diagonal matrix with standard deviations $(\sigma_{T_0}, \sigma_{T_1}, \sigma_{S_{10}}, \sigma_{S_{11}}, \ldots, \sigma_{S_{p0}}, \sigma_{S_{p1}})$ along the diagonal, and correlation matrix:

$$
\mathbf{R} = \begin{array}{c} \\ T_0 \\ T_1 \\ S_{10} \\ S_{11} \\ S_{20} \\ S_{21} \\ \vdots \\ S_{p0} \\ S_{p1} \end{array}
\begin{array}{c}
\begin{array}{ccccccccc} T_0 & T_1 & S_{10} & S_{11} & S_{20} & S_{21} & \cdots & S_{p0} & S_{p1} \end{array} \\
\begin{bmatrix}
1 & - & \rho_{T_0S_{10}} & - & \rho_{T_0S_{20}} & - & \cdots & \rho_{T_0S_{p0}} & - \\
 & 1 & - & \rho_{T_1S_{11}} & - & \rho_{T_1S_{21}} & \cdots & - & \rho_{T_1S_{p1}} \\
 & & 1 & - & \rho_{S_{10}S_{20}} & - & \cdots & \rho_{S_{10}S_{p0}} & - \\
 & & & 1 & - & \rho_{S_{11}S_{21}} & \cdots & - & \rho_{S_{11}S_{p1}} \\
 & & & & 1 & - & \cdots & \rho_{S_{20}S_{p0}} & - \\
 & & & & & 1 & \cdots & - & \rho_{S_{21}S_{p1}} \\
 & & & & & & \ddots & \vdots & \vdots \\
 & & & & & & & 1 & - \\
 & & & & & & & & 1
\end{bmatrix}
\end{array},
$$

where $(-)$ indicates non-identifiable entries. The previous distributional

assumptions imply $\boldsymbol{\Delta} \sim N\left(\boldsymbol{\mu}_{\boldsymbol{\Delta}}, \boldsymbol{\Sigma}_{\boldsymbol{\Delta}}\right)$, with $\boldsymbol{\mu}_{\boldsymbol{\Delta}} = (\beta, \boldsymbol{\alpha})'$ and

$$\boldsymbol{\Sigma}_{\boldsymbol{\Delta}} = \begin{pmatrix} \sigma^2_{\Delta T} & \boldsymbol{\Sigma}'_{\boldsymbol{\Delta}\mathbf{S}\Delta T} \\ \boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{S}\Delta T} & \boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{S}} \end{pmatrix},$$

where $\sigma^2_{\Delta T} = \sigma^2_{T_0} + \sigma^2_{T_1} - 2\rho_{T_0T_1}\sqrt{\sigma^2_{T_0}\sigma^2_{T_1}}$ is the variance of ΔT; $\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{S}\Delta T}$ is a p-dimensional vector of covariances between ΔT and $\boldsymbol{\Delta}\mathbf{S}$; and $\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{S}}$ is the $(p \times p)$ variance-covariance matrix of $\boldsymbol{\Delta}\mathbf{S}$.

Multivariate individual causal association

In this context, Van der Elst et al. (2019) proposed the so-called squared information coefficient of correlation (Linfoot, 1957; Joe, 1989) to quantify the individual causal association (ICA), i.e.,

$$R^2_H = \frac{\boldsymbol{\Sigma}'_{\boldsymbol{\Delta}\mathbf{S}\Delta T}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\Delta}\mathbf{S}}\boldsymbol{\Sigma}_{\boldsymbol{\Delta}\mathbf{S}\Delta T}}{\sigma^2_{\Delta T}}.$$

The R^2_H ranges over the unit interval $[0, 1]$, and takes the value zero if and only if ΔT and $\boldsymbol{\Delta}\mathbf{S}$ are uncorrelated, i.e., $\rho_{\Delta S_k \Delta T} = 0$, for all $k = 1, \ldots, p$. On the other hand, $R^2_H = 1$ if and only if ΔT is perfectly linearly predictable from $\boldsymbol{\Delta}\mathbf{S}$. Therefore, the R^2_H is an indicator of how well the individual causal treatment effect on T (ΔT) can be predicted by the individual causal treatment effects on the surrogates. Furthermore, it has been recently shown that considering more surrogate endpoints can only improve the capacity to predict ΔT, i.e., $R^2_{H*} \geq R^2_H$ when R^2_{H*} is computed based on $\mathbf{S}'_* = (\mathbf{S}', S_*)'$ (see lemma 2 in Van der Elst et al. (2019)).

4. Estimating R^2_H

As stated before, the R^2_H is a function of a partially identifiable matrix and, hence, it cannot be estimated without making untestable assumptions. To tackle this problem, Van der Elst et al. (2019) proposed a simulation-based analysis in which R^2_H is computed across a broad set of values for the non-identifiable entries in $\mathbf{R}$. Their approach consists of an algorithm that randomly completes the

non-identifiable entries in $\mathbf{R}$ and, after checking the positive definite (PD) condition, calculates the associated R^2_H. The previous algorithm can be interpreted as a procedure that randomly samples the space of all correlation matrices $\mathbf{R}$ compatible with the data at hand. Later, the general behavior of the so-obtained R^2_H values can be examined by using summary measures like the mean, standard deviation, range, and analyzing their frequency distribution. The range is specially important in this setting. Indeed, although the frequency distribution and summary measures may depend on the sampling procedure used to explore the space of all correlation matrices $\mathbf{R}$ compatible with the data, one may expect that if this space is thoroughly sampled, then the range of obtained ICA values should contain the true R^2_H. Notice that the ICA is non-identifiable and, hence, its estimation beyond an uncertainty interval is out of reach, unless one is willing to make untestable assumptions.

Finding an efficient algorithm to carry out the previous simulation-based analysis is particularly challenging in the high-dimensional setting, i.e., when a large number of putative surrogates is available. Two algorithms have now been introduced in the literature to tackle the matrix completion problem and they will be presented in Sections 4.1 and 4.2.

4.1. Rejection-Sampling algorithm

A rejection-sampling (RS) algorithm is the most obvious choice to randomly complete a correlation matrix, i.e., by independently drawing values from a Uniform(-1,1) for each non-identifiable entry of $\mathbf{R}$ and keeping it only if it is PD. Although straightforward, this approach is highly inefficient. As the dimension of the matrix increases, the acceptance rate rapidly goes to zero.

To increase the efficiency of the RS algorithm, Van der Elst et al. (2019) proposed to implement it gradually, based on Sylvester's criterion. One starts completing the upper-left (2×2) sub-matrix, i.e., sampling $\rho_{T_0T_1}$ from a Uniform(-1,1) until the matrix is PD, one then continues to the upper-left (3×3) sub-matrix, i.e., drawing $\rho_{T_1S_{10}}$ from a Uniform(-1,1) until PD is achieved, and so on until the whole matrix is completed. The acceptance rate increases consid-

erably. However, this algorithm constrains the distribution of some unspecified correlations. More specifically, it allows to get wider distributions for the correlations that are fixed first in the procedure (see Figure S4 in the Supplementary Materials).

Although the sampling scheme based on Sylvester's criterion increased computational efficiency, it is still limited to assess surrogacy in a high-dimensional setting. Depending on the fixed entries of $\mathbf{R}$, the method collapses when the number of putative surrogates exceeds 4 or 5.

4.2. Algorithm based on partial correlations

In spite of their potential to increase the efficiency of drug development, multivariate surrogates have received very little attention in the literature. It is plausible that one of the reasons underlying this situation is the increased methodological and numerical complexity associated with their evaluation. The work of Van der Elst et al. (2019) was an important step towards the use of multivariate surrogate endpoints but its numerical limitations seriously hinder its applicability. To achieve computational scalability, Flórez et al. (2020) proposed a new algorithm by sampling partial correlations (PCs) from the Lewandowski-Kurowicka-Joe correlation distribution (Joe, 2006; Lewandowski et al., 2009). The procedure requires to rearrange the $(d \times d)$ matrix $\mathbf{R}$ as:

$$\mathbf{R} = \begin{array}{c} \\ T_0 \\ S_{10} \\ S_{20} \\ \vdots \\ S_{p0} \\ T_1 \\ S_{11} \\ S_{21} \\ \vdots \\ S_{p1} \end{array} \begin{array}{c} \begin{array}{ccccccccccc} T_0 & S_{10} & S_{20} & \dots & S_{p0} & T_1 & S_{11} & S_{12} & \dots & S_{p1} \end{array} \\ \begin{bmatrix} 1 & \rho_{T_0S_{10}} & \rho_{T_0S_{20}} & \cdots & \rho_{T_0S_{p0}} & - & - & - & \cdots & - \\ & 1 & \rho_{S_{10}S_{20}} & \cdots & \rho_{S_{10}S_{p0}} & - & - & - & \cdots & - \\ & & 1 & \cdots & \rho_{S_{20}S_{p0}} & - & - & - & \cdots & - \\ & & & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & 1 & - & - & - & \cdots & - \\ & & & & & 1 & \rho_{T_1S_{11}} & \rho_{T_1S_{21}} & \cdots & \rho_{T_1S_{p1}} \\ & & & & & & 1 & \rho_{S_{11}S_{21}} & \cdots & \rho_{S_{11}S_{p1}} \\ & & & & & & & 1 & \cdots & \rho_{S_{21}S_{p1}} \\ & & & & & & & & \ddots & \vdots \\ & & & & & & & & & 1 \end{bmatrix} \end{array}, \quad (1)$$

with $d = 2(p+1)$. The algorithm completes (1) progressively, starting with the elements lining up closest to the main diagonal in a parallel fashion and following with the subsequent lines in the upper triangle of $\mathbf{R}$, by independently drawing the partial correlation $\rho_{j,j+k|j+1,\ldots,j+k-1}$ $(1 \leq k \leq d-1)$ i.e., the correlation between variables j and $j+k$ after removing the effect of variables $j+1,\ldots,j+k-1$, and using the following equality:

$$\rho_{j,j+k} = \mathbf{r}_1'(j,k)\left\{\mathbf{R}_2(j,k)\right\}^{-1}\mathbf{r}_3(j,k) + \rho_{j,j+k|j+1,\ldots,j+k-1}D_{j,k}, \tag{2}$$

where

$$\mathbf{R}[j:j+k] = \begin{pmatrix} 1 & \mathbf{r}_1'(j,k) & \rho_{j,j+k} \\ \mathbf{r}_1(j,k) & \mathbf{R}_2(j,k) & \mathbf{r}_3(j,k) \\ \rho_{j+k,j} & \mathbf{r}_3'(j,k) & 1 \end{pmatrix},$$

with $\mathbf{r}_1'(j,k) = (\rho_{j,j+1},\ldots,\rho_{j,j+k-1})$, $\mathbf{r}_3'(j,k) = (\rho_{j+k,j+1},\ldots,\rho_{j+k,j+k-1})$, $\mathbf{R}_2(j,k)$ being the middle $(k-1 \times k-1)$ matrix of $\mathbf{R}[j:j+k]$, and:

$$D_{j,k} = \sqrt{\left\{1-\mathbf{r}_1'(j,k)\mathbf{R}_2(j,k)\mathbf{r}_1(j,k)\right\}\left\{1-\mathbf{r}_3'(j,k)\mathbf{R}_2(j,k)\mathbf{r}_3(j,k)\right\}}.$$

to generate $\rho_{j,j+k}$.

If the partial correlations $\rho_{j,j+k|j+1,\ldots,j+k-1}$ are independently drawn from a Beta$[1+1/2(d-1-k), 1+1/2(d-1-k)]$ on $(-1,1)$, then one can show that, marginally, each ρ_{ij} is sampled from a Beta$(d/2, d/2)$ on $(-1,1)$ (Joe, 2006). For more details on its implementation, we refer to Joe (2006) and Flórez et al. (2020).

This procedure is computationally efficient and it allows to complete a high-dimensional correlation matrix relatively fast (Flórez et al., 2020). Theoretically, if the number of runs is sufficiently large, the algorithm should be able to thoroughly sample the entire space of correlation matrices compatible with the data. However, when non-informative surrogates are present, this method struggles to achieve this goal even after a rather large number of runs and, consequently, it may lead to spurious results.

5. PC algorithm and non-informative surrogates

In this section, we explore the performance of the PC algorithm when non-informative components are added to a multivariate surrogate. It will serve as a rationale for proposing a modified algorithm that solves this problem in the next section. By way of illustration, let us consider a three dimensional surrogate $\mathbf{S}' = (S_1, S_2, S_3)$ and the following covariance matrix for the vector of potential outcomes $\mathbf{Y}$:

$$\mathbf{\Sigma}_4 = \begin{array}{c} \\ T_0 \\ T_1 \\ S_{10} \\ S_{11} \\ S_{20} \\ S_{21} \\ S_{30} \\ S_{31} \end{array} \begin{array}{c} \begin{array}{cccccccc} T_0 & T_1 & S_{10} & S_{11} & S_{20} & S_{21} & S_{30} & S_{31} \end{array} \\ \left[\begin{array}{cccc|cccc} 1 & - & 0.95 & - & 0 & - & 0 & - \\ & 1 & - & 0.95 & - & 0 & - & 0 \\ & & 1 & - & 0 & - & 0 & - \\ & & & 1 & - & 0 & - & 0 \\ \hline & & & & 1 & - & 0.80 & - \\ & & & & & 1 & - & 0.80 \\ & & & & & & 1 & - \\ & & & & & & & 1 \end{array}\right] \end{array}. \tag{3}$$

In the previous matrix, the non-identifiable entries are given by "$-$". The potential outcomes of the first component (S_1) are highly correlated with the potential outcomes of the true endpoint and, therefore, S_1 may be a plausible surrogate candidate. On the contrary, S_2 and S_3 are non-informative in the sense that their potential outcomes are uncorrelated with the potential outcomes of T and the potential outcomes of S_1. Intuitively, one would expect that, after applying the PC algorithm, the range of the sampled ICA values obtained from the three-dimensional surrogate should be roughly equal to the one obtained from assessing solely S_1, i.e., only considering the upper-left block of (3). This intuition can be formalized in the following theorem.

Theorem 1. *Let $\mathbf{S}_p$ denote a p-dimensional surrogate candidate and further assume that m components are added to $\mathbf{S}_p$ to create the $p+m$ dimensional outcome $\mathbf{S}'_{p+m} = (\mathbf{S}'_p, \mathbf{S}'_m)$. Let us further assume that the components in $\mathbf{S}_m$ are*

non-informative in the sense that their potential outcomes are uncorrelated with the potential outcomes of T and the potential outcomes of all other components of $\mathbf{S}_p$. Under the previous conditions, and using obvious notation, one has $R^2_{H,p} = R^2_{H,p+m}$.

A proof of Theorem 1 is given in Section A of the Supplementary Materials. This result shows that, as expected, adding non-informative components does not affect the ICA value and, consequently, it should not affect the range emanating from the simulated-based analysis neither. Therefore, a reduction in the range of the ICA after applying the PC algorithm in the previous setting, will clearly contradict the previous theoretical result and indicate a problem with this procedure.

5.1. Evaluating S_1

First, we evaluate the surrogacy of S_1 using the RS algorithm. After obtaining $10,000$ valid correlation matrices, i.e., $10,000$ matrices that were not rejected due to non-positive definiteness, the obtained range for the ICA was virtually identical to $[0, 1]$. Furthermore, the correlation matrix that led to the minimum ICA was:

$$\boldsymbol{\Sigma}_1^{(\min)} = \begin{array}{c} \\ T_0 \\ T_1 \\ S_{10} \\ S_{10} \end{array} \begin{array}{c} \begin{array}{cccc} T_0 & T_1 & S_{10} & S_{11} \end{array} \\ \begin{bmatrix} 1 & 0.90 & 0.95 & 0.96 \\ & 1 & 0.94 & 0.95 \\ & & 1 & 0.97 \\ & & & 1 \end{bmatrix} \end{array}. \tag{4}$$

This matrix illustrates the well-known result that a strong correlation between the true endpoint and the putative surrogate in both treatment groups is not a sufficient condition for surrogacy.

5.2. Evaluating $\mathbf{S}$

The surrogacy of $\mathbf{S}' = (S_1, S_2, S_3)$ was assessed after obtaining $10,000$ valid correlation matrices using the RS algorithm. The ICA values ranged between

0.028 and 0.999. As expected, this range is similar to the one observed when only S_1 was used. Moreover, the minimum ICA value was obtained from the matrix:

$$\boldsymbol{\Sigma}_4^{(\min)} = \begin{array}{c} \\ T_0 \\ T_1 \\ S_{10} \\ S_{11} \\ S_{20} \\ S_{21} \\ S_{30} \\ S_{31} \end{array} \begin{array}{c} \begin{array}{cccccccc} T_0 & T_1 & S_{10} & S_{11} & S_{20} & S_{21} & S_{30} & S_{31} \end{array} \\ \left[\begin{array}{cccc|cccc} 1 & 0.93 & 0.95 & 0.99 & 0 & -0.02 & 0 & -0.05 \\ & 1 & 0.90 & 0.95 & 0.11 & 0 & 0.12 & 0 \\ & & 1 & 0.97 & 0 & 0.13 & 0 & 0.20 \\ & & & 1 & 0.06 & 0 & 0.07 & 0 \\ \hline & & & & 1 & -0.43 & 0.80 & -0.26 \\ & & & & & 1 & -0.38 & 0.80 \\ & & & & & & 1 & -0.38 \\ & & & & & & & 1 \end{array}\right] \end{array}. \quad (5)$$

Note that the upper-left bock matrix of (5) is fairly similar to (4). Furthermore, its off-diagonal block is roughly a zero matrix (all non-identifiable entries are close to zero). Figure 1 shows the relationship between the ICA and four non-identifiable correlations. Low ICA values are mostly obtained when the non-identifiable correlations $\rho_{T_0 S_{31}}$ (see Figure 1a), and $\rho_{S_{11} S_{20}}$ (see Figure 1b), are close to zero; and when the non-identifiable correlation $\rho_{T_0 S_{11}}$ is large (see Figure 1c). On the other hand, small ICA values can be obtained for any value of $\rho_{S_{21} S_{30}}$ between -1 and 1 (see Figure 1d). The same behavior is observed with the rest of the correlations (see Figures S1–S3 of the Supplementary Materials).

When S_1 was evaluated using the PC algorithm, the R_H^2 values, obtained from $10,000$ valid correlation matrices, ranged between 0 and 1. However, the joint evaluation of the three surrogates led to a minimum ICA of 0.198, suggesting that adding S_2 and S_3 may improve surrogacy. This seems to contradict the result presented in Theorem 1, i.e., adding non-informative components should have no impact on the ICA. We think this problem is basically due to the marginal distribution used to sample the non-identifiable correlations, which is a Beta$(4, 4)$ on $(-1, 1)$. Indeed, this sampling scheme seems to have problems to sample certain combinations of correlations that produce low ICA values com-

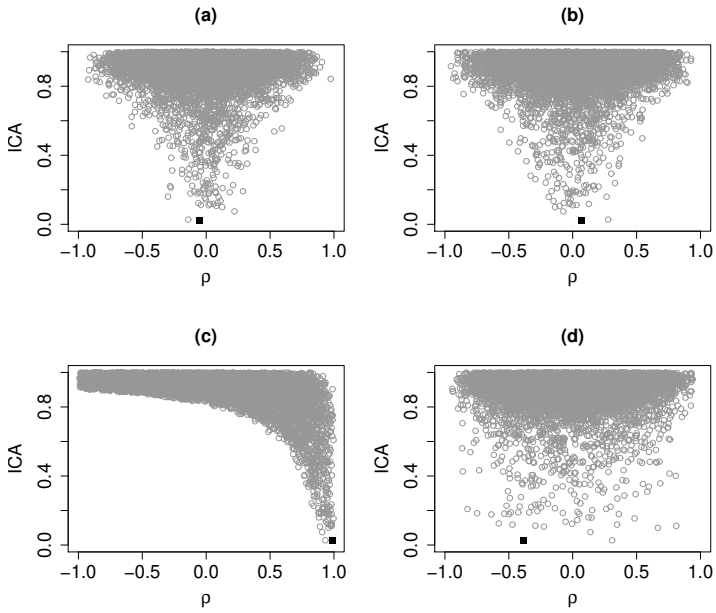

Figure 1: Relationship between the simulated ICA and four non-identifiable correlations simulated using the RS algorithm: (a) $\rho_{T_0 S_{31}}$, (b) $\rho_{S_{11} S_{20}}$, (c) $\rho_{T_0 S_{11}}$, and (d) $\rho_{S_{21} S_{30}}$. The black square indicates the correlation associated to the minimum ICA.

patible with the data. The problem does improve when a very large number of runs is used. However, even after one million runs this issue still persists to a certain degree (see Section C.1 of the Supplementary Materials).

6. Solving the problem: A modified PC algorithm (MPC)

Some remarks are in order. Firstly, note that (1) has a block-diagonal structure. The entries in the first and second block diagonal correspond to the known correlations in the control group $(T_0, S_{10}, S_{20}, \ldots, S_{p0})$, and the treatment arm $(T_1, S_{11}, S_{21}, \ldots, S_{p1})$, respectively; the values in the off-diagonal block consist of non-identifiable correlations between groups. Therefore, any re-indexing that leads to the same block structure is permitted.

Secondly, as the previous example illustrated, the PC algorithm struggles to sample the entire space of correlation matrices compatible with the data when

non-informative surrogates are present. One way to overcome this problem is to fix the non-identifiable correlations between the potential outcomes of T and the non-informative surrogates to zero and conveniently rearranging $\mathbf{R}$. For instance, in setting (3), $\mathbf{R}$ can be re-indexed as:

$$\widetilde{\mathbf{R}} = \begin{array}{c} \\ T_0 \\ S_{10} \\ S_{20} \\ S_{30} \\ S_{21} \\ S_{31} \\ S_{11} \\ T_1 \end{array} \begin{array}{c} \begin{array}{cccccccc} T_0 & S_{10} & S_{20} & S_{30} & S_{21} & S_{31} & S_{11} & T_1 \end{array} \\ \left[\begin{array}{cccccc|cc} 1 & 0.95 & 0 & 0 & 0 & 0 & - & - \\ & 1 & 0 & 0 & 0 & 0 & - & - \\ \hline & & 1 & 0.8 & 0 & 0 & 0 & 0 \\ & & & 1 & 0 & 0 & 0 & 0 \\ & & & & 1 & 0.8 & 0 & 0 \\ & & & & & 1 & 0 & 0 \\ & & & & & & 1 & 0.95 \\ & & & & & & & 1 \end{array}\right] \end{array}. \tag{6}$$

Afterwards the PC algorithm can be implemented in its usual way. Note that fixing some specific correlations to zero and leaving the rest of the non-identifiable entries in the upper-right block of the correlation matrix, as in (6), does not affect the PD constraint or the implementation of the PC algorithm. By including this additional information in the sampling scheme, we are able to explore the parameter space more efficiently, i.e., requiring a much smaller number of runs.

Naturally, in a general setting, we do not know which surrogate, or combination thereof, can be non-informative. Thus, the selection of non-identifiable correlations that are fixed at zero (which depend on the surrogates) can be determined randomly. Alternatively, one may consider all possible combinations of surrogates. However, with several surrogates, this number can be quite large. Therefore, we opt for the following "computationally parsimonious" solution.

MPC algorithm:

1. Randomly select a number r with probability:

$$\pi_r = \frac{\binom{p}{r}}{\sum_{i=0}^{p} \binom{p}{i}}, \text{ for } r = 0, \ldots, p. \tag{7}$$

2. Using equal probabilities, randomly select r surrogates and fixed their non-identifiable correlations at zero.
3. Rearrange the incomplete correlation matrix $\mathbf{R}$ leaving the non-identifiable correlations for the unselected surrogates in the upper-right block of $\mathbf{R}$, as in (6).
4. Implement the PC algorithm in its standard way.
5. Repeat steps (1)–(4) a large number of times, say M.

By steps 1 and 2, each possible combination of r surrogates has the same probability of being selected. Moreover, when $r = 0$, no non-identifiable correlations are set to zero and, hence, the PC algorithm is performed in the usual way. Note further that, in step 1, one can restrict the number of non-identifiable correlations one is willing to fix to zero by assigning the corresponding probabilities to zero (as it will be illustrated in Section 7). Another important practical issue is the selection of the number of runs (M). It is difficult to provide a general rule of thumb for selecting this number that will work in all scenarios. However, some general principles can be followed. For instance, since the dimension of the parameter space increases rapidly as the number of surrogates increases, it is advisable to consider larger values of M when p is large. In general, the larger the value of M the better, but computational feasibility is another important factor one needs to take into account. Therefore, one should use the largest number of runs that is affordable with the computational resources one has.

The MPC algorithm was used to complete matrix (3) $M = 70,000$ times, assigning probabilities as (7). This means that each combination of $(n - r)$ surrogates was solely evaluated (by fixing the non-identifiable correlations of the other r surrogates to zero) on average $10,000$ times. As displayed in Figure 2, the three methods produce similar frequency distributions for the ICA. However, the ones obtained from the RS and MPC algorithms have heavier left-tails allowing for lower minimum ICA values. These results show that the MPC algorithm was able to solve the under-sampling problem of the PC algorithm without sacrificing computational efficiency or drastically affecting the frequency distribution of the

ICA.

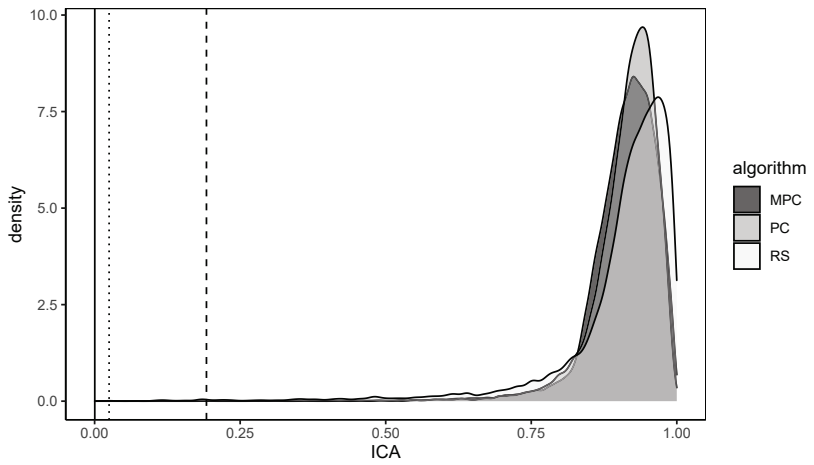

Figure 2: Densities of the ICA computed using the RS, PC and MPC algorithm. The vertical lines are the minimum ICA computed by the RS (dotted line), PC (dashed line), and MPC (solid line) algorithms.

Moreover, in Section C.1 of the Supplementary Materials, we performed an additional simulation study to show that, in the presence of non-informative surrogates, the MPC algorithm can indeed find low ICA values compatible with the data with a substantially smaller number of runs. In fact, when implemented with the covariance matrix (3) the MPC algorithm basically solved the problem after 50,000 runs, whereas after one million runs of the PC algorithm the problem was still present (even though it was attenuated). Regarding computation time, this means that the MPC algorithm was able to solve in 30 seconds a problem that the PC algorithm could not solve after 13 minutes. Notice that the covariance matrix (3) represents a rather low dimensional surrogate with only 3 components and the situation will get more involved when more surrogate are used.

It is important to point out that, in these simulations, it is assumed that the identifiable entries are estimated with a negligible error. In Section C.2 of the Supplementary Materials, we perform a simulation study to explore the effect of

the sample size in the estimation of the ICA using the PC and MPC algorithms. Additionally, in Section C.3 of the Supplementary Materials, we compare both algorithms in a setting with only informative surrogates, i.e., using only non-zero non-identifiable correlations. As expected, in the absence of non-informative surrogates, both methods provide somewhat the same results.

The MPC algorithm has been implemented in the function `ICA.ContCont.MultS.MPC` of the `Surrogate` R package. In the accompanying help file an example of its use is discussed in detail. We refer the interested reader to this file and to the Supplementary Materials for more information.

7. MPC versus PC algorithm: A more realistic comparison

To compare both algorithms in a realistic setting, we considered a p-dimensional surrogate $\mathbf{S}_p = (S_1, S_2, \ldots, S_p)$, in which only the first two components were informative (S_1, S_2), while the other $(S_3, S_4, \ldots, S_p)$ were non-informative. Furthermore, the ICA was assessed based on both the PC and MPC algorithms. The data-generating mechanism is given by $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{DRD})$ where $\boldsymbol{\mu} = \mathbf{0}$,

$$\mathbf{R} = \begin{array}{c} \\ T_0 \\ T_1 \\ S_{10} \\ S_{11} \\ S_{20} \\ S_{21} \\ S_{30} \\ S_{31} \\ \vdots \\ S_{p0} \\ S_{p1} \end{array} \begin{array}{c} \begin{array}{cccccccccccc} T_0 & T_1 & S_{10} & S_{11} & S_{20} & S_{21} & S_{30} & S_{31} & \cdots & S_{p0} & S_{p1} \end{array} \\ \left[\begin{array}{cccccc|ccccc} 1 & -0.2 & 0.9^* & -0.5 & -0.6^* & 0.5 & 0^* & 0 & \ldots & 0^* & 0 \\ & 1 & -0.5 & 0.7^* & 0.4 & -0.5^* & 0 & 0^* & \ldots & 0 & 0^* \\ & & 1 & -0.7 & -0.5^* & 0.7 & 0^* & 0 & \ldots & 0^* & 0 \\ & & & 1 & 0.6 & -0.4^* & 0 & 0^* & \ldots & 0 & 0^* \\ & & & & 1 & -0.2 & 0^* & 0 & \ldots & 0^* & 0 \\ & & & & & 1 & 0 & 0^* & \ldots & 0 & 0^* \\ \hline & & & & & & 1 & 0.5 & \ldots & 0.5^* & 0.5 \\ & & & & & & & 1 & \ldots & 0.5 & 0.5^* \\ & & & & & & & & \ddots & \vdots & \vdots \\ & & & & & & & & & 1 & 0.5 \\ & & & & & & & & & & 1 \end{array}\right] \end{array}, \quad (8)$$

and $\mathbf{D}$ is a diagonal matrix with $(5, 2, 4, 1, 4, 1, \ldots, 4, 1)$ along the diagonal. The identifiable correlations in (8) are marked with a * symbol and the dimension of $\mathbf{S}_p$ was fixed at $p = \{1, 2, 3, 5, 10, 15\}$. Based on Theorem 1, this setting

leads to $R_H^2 = 0.928$. This value is obtained when the two informative surrogates (S_1, S_2) are considered, i.e., when using the upper-left block of (8) and one should not expect any noticeable differences in the ICA values after adding more surrogates.

Based on the aforementioned distributional assumption a data set of size $N = 200$ was generated and the identifiable parameters were estimated. Afterwards, the ICA range was calculated using the RS, PC and MPC algorithms. It is important to point out that the RS algorithm was only able to handle up to 3 surrogates.

Note that in the scenarios with $p = 1$ and $p = 2$, one only considers a single (S_1) or a couple of informative surrogates (S_1, S_2). But as more surrogates are added, one needs to ensure that each individual combination of $(p - r)$ surrogates is evaluated a sufficient number of times. Therefore it is necessary to appropriately increase the number of iterations M. If the number of surrogates in the first step of the MPC algorithm (r) is selected using the probabilities (7), then to obtain $10,000$ valid correlation matrices for each subset of r surrogates, will require around $310,000$, 10^7, and 3.28×10^8 runs to assess 5, 10, and 15 surrogates, respectively. Therefore, to achieve computational feasibility, we decided to fix $r = \{0, p-1, p-2\}$ when evaluating more than five surrogates. It means we are considering only three situations: (1) assessing surrogacy without fixing any correlation to zero, (2) setting to zero the inestimable correlations associated to $p-2$ surrogates (i.e., sampling the correlations of two surrogates), and (3) equating to zero the non-identifiable correlations associated to $p - 1$ surrogates (i.e., sampling the correlations of one surrogate). Consequently, we run each algorithm 10, 30, 70, 310, 560, and 1200 thousand times to assess 1, 2, 3, 5, 10, and 15 surrogates, respectively.

A summary of the results is presented in Figure 3. The graph shows the range and median of simulated R_H^2's based on the RS, PC, and MPC algorithms. Regarding the median, all procedures led to similar results: a noticeable increment when one moves from S_1 to (S_1, S_2), along with a negligible growth as non-informative outcomes are added. Up to three surrogates, there are no

noticeable differences between the PC and MPC algorithms. Nevertheless, the RS algorithm display the lowest minimum. A different pattern emerges when analyzing five or more surrogates. In fact, the PC algorithm exhibits a considerably narrowed range when p is equal to 10 or 15. Actually, the minimum R_H^2 passed from 0.45 with $p = 2$ (considering the "useful" outcomes) to 0.65 with $p = 15$ after adding 13 non-informative outcomes. This result may lead to the deceptive conclusion that the added outcomes improve surrogacy. In addition, no noticiable reduction in the range was observed with the MPC algorithm. The minimum ICA with $p = 15$ was 0.53.

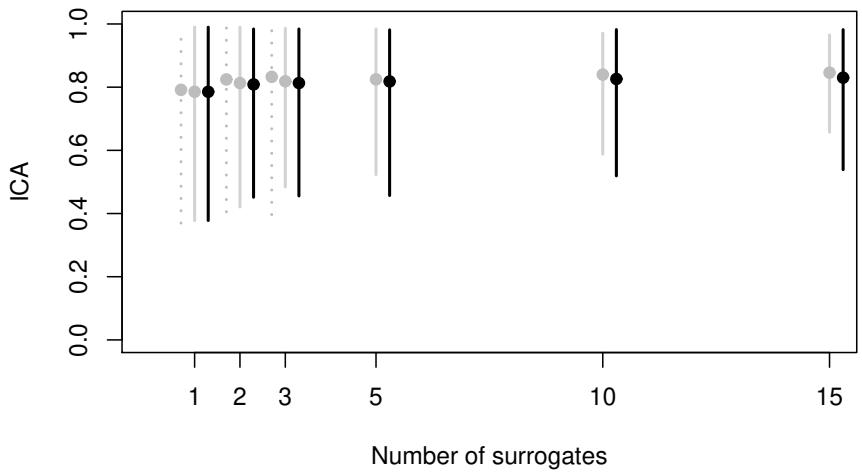

Figure 3: Range of the ICA computed using the RS algoritm (gray dashed line), PC algorithm (gray solid line), and MPC algorithm (black solid line) for $p = \{1, 2, 3, 5, 10, 15\}$ surrogates. The dots indicate the median.

Several additional data sets were generated and fairy similar findings were always obtained. In all cases, the PC algorithm produced an artificially narrowed range whereas the MPC led to correct conclusions. Although the differences in the median were negligible between both methods when only a few surrogates were considered, these increase as the number of surrogates got larger. Similarly,

differences in the ICA densities obtained by the PC and MPC algorithms were also more noticeable as more surrogates were jointly evaluated (see Figure S6 of the Supplementary Materials).

Importantly, regarding computation time, the MPC algorithm was more efficient than the PC. Indeed, when sampling 1000 correlation matrices to evaluate 5, 10 and 15 surrogates, the former took on average 4.2, 5.6 and 9.8 seconds, respectively while the latter took roughly 5, 9 and 20 seconds, respectively. These times were obtained with a laptop computer with an Intel(R) Core(TM) i5-6200U CPU 2.30GHz processor and 16GB of RAM memory.

8. Data Analyses

For the schizophrenia data the partially observed correlation matrix is:

$$\mathbf{R} = \begin{array}{c} \\ T_0 \\ T_1 \\ S_{10} \\ S_{11} \\ S_{20} \\ S_{21} \\ S_{30} \\ S_{31} \\ S_{40} \\ S_{41} \\ S_{50} \\ S_{51} \end{array} \begin{array}{c} \begin{array}{cccccccccccc} T_0 & T_1 & S_{10} & S_{11} & S_{20} & S_{21} & S_{30} & S_{31} & S_{40} & S_{41} & S_{50} & S_{51} \end{array} \\ \begin{bmatrix} 1 & - & 0.69 & - & 0.69 & - & 0.75 & - & 0.75 & - & 0.68 & \\ & 1 & - & 0.72 & - & 0.69 & - & 0.82 & - & 0.81 & - & 0.71 \\ & & 1 & - & 0.21 & - & 0.49 & - & 0.27 & - & 0.36 & - \\ & & & 1 & - & 0.25 & - & 0.53 & - & 0.36 & - & 0.32 \\ & & & & 1 & - & 0.32 & - & 0.60 & - & 0.40 & - \\ & & & & & 1 & - & 0.50 & - & 0.57 & - & 0.60 \\ & & & & & & 1 & - & 0.45 & - & 0.31 & - \\ & & & & & & & 1 & - & 0.61 & - & 0.39 \\ & & & & & & & & 1 & - & 0.48 & - \\ & & & & & & & & & 1 & - & 0.61 \\ & & & & & & & & & & 1 & - \\ & & & & & & & & & & & 1 \end{bmatrix} \end{array}.$$

The estimable correlations are positive and relatively high. Particularly, the PANSS score (T) is strongly correlated with cognitive (S_3) and positive (S_4) scales in both arms. The correlations between the surrogates range between roughly 0.2 and 0.6. We proceed calculating the ICA for all possible combinations of surrogates using the RS, PC and MPC algorithms. Table 1 shows the

best combination of surrogates (based on the maximum median of the ICA) for each subset.

Table 1: PANSS data. ICA range and median computed by M (in thousand) runs of the RS, PC, and MPC algorithms of the best set of surrogates. * The RS algorithm was run $M = 10.000$ times.

		RS algorithm*			PC algorithm			MPC algorithm		
Surrogates	M	min	median	max	min	median	max	min	median	max
S_4	20	0.000	0.642	0.997	0.000	0.640	0.999	0.000	0.641	0.998
(S_3, S_4)	60	0.001	0.836	0.998	0.003	0.828	0.998	0.008	0.821	0.999
(S_1, S_3, S_4)	140	0.058	0.930	0.999	0.154	0.924	0.998	0.148	0.919	0.999
(S_1, S_3, S_4, S_5)	300	-	-	-	0.512	0.963	0.999	0.561	0.960	0.999
$(S_1, \ldots, S_5)$	620	-	-	-	0.892	0.992	1.000	0.933	0.992	1.000

When evaluating each surrogate independently, the positive factor (S_4) provides the highest median R^2_H. Nevertheless, the range covers the whole parameter space. Similar results are observed when we jointly evaluate the best subsets of up to three surrogates. When four or five surrogates are assessed simultaneously, the ICA range narrows considerably. Although the PC algorithm leads to a lower minimum than the MPC algorithm in some simulations, theses differences are negligible. Furthermore, the minimum R^2_H based on the MPC algorithm is found when none of the non-identifiable correlations are fixed to zero, suggesting that all the surrogates are informative. This is in complete agreement with the clinical expectations and confirms that, when non-informative surrogates are absent, the PC and MPC algorithm will produce comparable results.

9. Discussion

The simulation-based analysis proposed to assess the ICA, is certainly an appealing methodology that allows to quantify the uncertainly emanating from its non-identifiability. The methodology is based on random draws of a partially identifiable correlation matrix $\mathbf{R}$. Currently, this can be done considerably fast using an algorithm based on partial correlations (Flórez et al., 2020).

However, this computational efficiency comes at a cost. Indeed, in presence of non-informative surrogates, the PC algorithm fails to properly sample some regions of the space of correlation matrices compatible with the data. Based on some theoretical elements and simulations, we showed that this leads to a spurious reduction of the uncertainty about the ICA as non-informative surrogates are added. It should be point out that the problem is attenuated when the number of runs is increased. However, our simulations show that in a relatively simple setting with 3 surrogates, even after a million runs the problem is still present. This is a critical drawback of the algorithm when evaluating multivariate surrogate endpoints.

We proposed a modified PC algorithm that includes extra information in the sampling scheme to tackle this problem. It is shown via simulations that the MPC algorithm completely solves the issues of non-informative surrogates, and it also increases computational efficiency. We believe that the MPC algorithm could make the evaluation of high-dimensional surrogates computationally affordable. This may be particularly useful in research areas like probiotics or bioinformatics.

As previously shown, the MPC algorithm allows to efficiently explore the PD space of correlation matrices compatible with the data and, as a consequence, it correctly assesses the ICA range. However, by design, the MPC algorithm produces zero-inflated marginal distributions for the unspecified correlations. Given that our focus is on the range of the plausible ICA values and not on the non-identifiable correlations' distributional properties, this issue is irrelevant in our context. If the focus lies on sampling from different joint distributions for the non-identifiable correlations then more complex procedures are required. For instance, Kurowicka (2014) thoughtfully analyzed this problem when the correlation matrix exhibits a chordal sparsity pattern and we refer the interested reader to this work for further information.

Finally, we point out that a number of extensions are worth exploring. For instance, copula models (Nelsen, 2007) have been successfully applied within the meta-analytic approach to jointly model survival outcomes (Emura et al.,

2019, 2021). The use of these models in the causal inference framework is not evident due to the unidentifiability issues. However, they do have the potential to extend the proposed methodology to no-normal variables. This will be the objective of future research.

Funding

Financial support from the IAP research network #P7=06 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged. Alvaro J. Flórez acknowledges funding from the European Seventh Framework programme FP7 2007-2013 under grant agreement Nr. 602552.

Supplementary Materials

Additional supporting information can be found online.

References

Alonso, A., Bigirumurame, T., Burzykowski, T., Buyse, M., Molenberghs, G., Muchene, L., Perualila, N., Shkedy, Z., Van der Elst, W., 2016. Applied Surrogate Endpoint Evaluation with SAS and R. Chapman & Hall/CRC, Boca Ratón.

Alonso, A., Laenen, A., Molenberghs, G., Geys, H., Vangeneugden, T., 2010. A unified approach to multi-item reliability. Biometrics 66, 1061–1068.

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, 15–24.

Biomarkers Definitions Working Group, 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clinical Pharmacology & Therapeutics 69, 89–95.

Burzykowski, T., Molenberghs, G., Buyse, M., 2005. The Evaluation of Surrogate Endpoints. Springer-Verlag GMBH.

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

Buyse, M., Molenberghs, G., Paoletti, X., Oba, K., Alonso, A., der Elst, W.V., Burzykowski, T., 2015. Statistical evaluation of surrogate endpoints with examples from cancer clinical trials. Biometrical Journal 58, 104–132.

Buyse, M., Sargent, D.J., Grothey, A., Matheson, A., de Gramont, A., 2010. Biomarkers and surrogate end points—the challenge of statistical validation. Nature Reviews Clinical Oncology 7, 309–317.

Chouinard, G., Jones, B., Remington, G., Bloom, D., Addington, D., MacEwan, G.W., Labelle, A., Beauclair, L., Arnott, W., 1993. A canadian multicenter placebo-controlled study of fixed doses of risperidone and haloperidol in the treatment of chronic schizophrenic patients. Journal of Clinical Psychopharmacology 13, 25–40.

Emura, T., Matsui, S., Rondeau, V., 2019. Survival Analysis with Correlated Endpoints. Springer, Singapore.

Emura, T., Sofeu, C.L., Rondeau, V., 2021. Conditional copula models for correlated survival endpoints: Individual patient data meta-analysis of randomized controlled trials. Statistical Methods in Medical Research 30.

Flórez, A.J., Alonso Abad, A., Molenberghs, G., Van Der Elst, W., 2020. Generating random correlation matrices with fixed values: An application to the evaluation of multivariate surrogate endpoints. Computational Statistics & Data Analysis 142, 1–10.

Holland, P.W., 1986. Statistics and causal inference. Journal of the American Statistical Association 81, 945–960.

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

Joe, H., 2006. Generating random correlation matrices based on partial correlations. Journal of Multivariate Analysis 97, 2177–189.

Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophrenia Bulletin 13, 261–276.

Kurowicka, D., 2014. Joint density of correlations in the correlation matrix with chordal sparsity patterns. Journal of Multivariate Analysis 129, 160–170.

Lewandowski, D., Kurowicka, D., Joe, H., 2009. Generating random correlation matrices based on vines and extended onion method. Journal of Multivariate Analysis 100, 1989–2001.

Lindenmayer, J.P., Bernstein-Hyman, R., Grochowski, S., Bark, N., 1995. Psychopathology of schizophrenia: Initial validation of a 5-factor model. Psychopathology 28, 22–31.

Linfoot, E., 1957. An informational measure of correlation. Information and Control 1, 85–89.

Marder, S.R., Meibach, R.C., 1994. Risperidone in the treatment of schizophrenia. American Journal of Psychiatry 151, 825–835.

Molenberghs, G., Burzykowski, T., Alonso, A., Assam, P., Tilahun, A., Buyse, M., 2010. A unified framework for the evaluation of surrogate endpoints in mental-health clinical trials. Statistical Methods in Medical Research 19, 205–236.

Nelsen, R.B., 2007. An Introduction to Copulas. Springer, New York.

Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.

Rubin, D.B., 1986. Statistics and causal inference: Comment: which ifs have causal answers. Journal of the American Statistical Association 81, 961–962.

Van der Elst, W., Alonso, A., Geys, H., Meyvisch, P., Bijnens, L., Sengupta, R., Molenberghs, G., 2019. Univariate versus multivariate surrogate endpoints in the single-trial setting. Statistics in Biopharmaceutical Research 11, 301–311.

Van der Elst, W., Molenberghs, G., Alonso, A., 2015. Exploring the relationship between the causal-inference and meta-analytic paradigms for the evaluation of surrogate endpoints. Statistics in Medicine 35, 1281–1298.

Van der Gaag, M., Hoffman, T., Remijsen, M., Hijman, R., Dehaan, L., van Meijel, B., van Harten, P., Valmaggia, L., de Hert, M., Cuijpers, A., 2006. The five-factor model of the positive and negative syndrome scale II: A ten-fold cross-validation of a revised model. Schizophrenia Research 85, 280–287.