

Validating predictors of therapeutic success: A causal inference approach

Peer-reviewed author version

ALONSO ABAD, Ariel; VAN DER ELST, Wim & MOLENBERGHS, Geert (2015)

Validating predictors of therapeutic success: A causal inference approach. In:
Statistical modelling 15(6), p. 619-636.

DOI: 10.1177/1471082X15586286

Handle: <http://hdl.handle.net/1942/20858>

Validating predictors of therapeutic success: A causal inference approach

Ariel Alonso Abad¹, Wim Van der Elst² and Geert Molenberghs^{1,2}

¹I-BioStat, KU Leuven, B-3000 Leuven, Belgium

²I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium

Abstract: In personalized medicine medical decisions, practices and/or products are tailored to the individual patient. The idea is to provide the right patient with the right drug at the right dose at the right time. However, our current lack of ability to predict an individual patient's treatment success for most diseases and conditions is a major challenge to achieve the goal of personalized medicine. In the present work, we argue that many of the techniques often used to evaluate predictors of therapeutic success may not be able to answer the relevant scientific questions and we propose a new validation strategy based on causal inference. The methodology is illustrated using data from a clinical trial in opiate/heroin addiction. The user-friendly R library *EffectTreat* is provided to carry out the necessary calculations.

Key words: causal inference; personalized medicine; prediction of therapeutic success

Received December 2014; revised April 2015; accepted April 2015

1 Introduction

Advances in medical imaging, biomarkers, genetics and computing sciences are paving the way for personalized medical treatments that consider a patient's genetic, anatomical and physiological characteristics. The concept of personalized medicine (PM) is not new. In fact, clinicians have long observed that patients with similar symptoms may have different illnesses, with different causes; and similarly, that medical interventions may work well in some patients with a disease but not in others with apparently the same disease (Abrahams and Silver, 2010).

Personalized medicine may be thought of as tailoring medical treatment to the individual characteristics, needs and preferences of each patient. Therefore, finding good pretreatment predictors of therapeutic success is of utmost importance in PM. Having such a predictor may be particularly relevant when the treatment in question is expensive, invasive and/or may lead to serious adverse events. Indeed, Spear, Heath-Chiozzi and Huff (2001) estimated that the response rates of patients to medications from different therapeutic classes ranged from 80% (analgesics) to 25% (oncology). In addition, an estimated 2.2 million adverse drug reactions occur each year in the

Note: This is a revised version of the article published OnlineFirst on May 21, 2015.

Address for correspondence: Ariel Abad Alonso, I-BioStat, Kapucijnenvoer 35, blok D, bus 7001, 3000 Leuven, Belgium. E-mail: Ariel.AlonsoAbad@kuleuven.be

United States, including more than 100,000 deaths. As a consequence, one would like to be able to assess the probability of treatment success, for a given patient, and weigh it by its risks before a final decision is made.

Banerjee *et al.* (2010) consider such a problem in in-vitro fertilization (IVF) treatments. Nearly 75% of IVF treatments do not result in live births; therefore, providing predictions of success for subsequent IVF treatments should assist a patient with decisions, given the financial, physical and emotional costs of undergoing IVF therapy. Similarly, Shin *et al.* (2013) state that although several psychological and pharmacological treatment options are available for anxiety disorders, not all patients respond well to each option. Therefore, both clinicians and patients would benefit from the identification of objective pretreatment measures that predict which patients will best respond to a given therapy.

Finding appropriate methodological tools for the evaluation of pretreatment predictors of therapeutic success is one of the challenges faced in PM. Most of the currently available validation techniques are based on correlational analysis; methods like linear and logistic regression, discriminant analysis and boosting are often combined with measures of association like odds ratios and Pearson correlations for evaluation purposes (Spielman *et al.*, 1983; van Loendersloot *et al.*, 1992; Banerjee *et al.*, 2010; Masao *et al.*, 2013; Shin *et al.*, 2013; Zeller *et al.*, 2013). In the present work, we argue that studying the association between the putative predictor and the response variable of interest in groups of patients that either receive or did not receive the treatment is not sufficient to answer the relevant scientific question. Actually, in PM one wants to gauge, for a given patient and his associated predictor value, how likely it is that the treatment will be beneficial for him. As stated before, in the following sections, it will be shown that classical correlational analysis does not allow to address the aforementioned question and a new validation strategy, rooted in causal inference, will be proposed for that purpose.

In Section 2, the causal-inference model underlying the validation strategy is introduced. A new validation metric, the so-called predictive causal association (PCA), is proposed in Section 3 and its relationship with the frequently used correlational analysis is discussed in Sections 4 and 6. Some important identifiable issues are discussed in Section 5 and a sensitivity analysis approach is introduced to handle the problem. The case study is presented and analyzed in Section 7 and some final comments are given in Section 8.

2 A causal inference model

In the following, T will denote the most credible outcome to assess therapeutic success, the so-called true endpoint, Z the treatment indicator and S the potential pretreatment predictor. Rubin's model for causal inference assumes the existence, for each patient j , of two potential outcomes for the true endpoint: an outcome T_{0j} that would be observed under the control condition $Z_j = 0$ and an outcome T_{1j} that would be observed under the treatment condition $Z_j = 1$. T_{0j} and T_{1j} are potential

outcomes in that they represent the outcomes of the patient had he received the treatment or control, respectively. The so-called fundamental problem of causal-inference is that, often in practice, only one of T_{0j} and T_{1j} is observed (Holland, 1986). Actually, if we denote by T_j the observed outcome for patient j then, under the *Stable Unit Treatment Value Assumption* (SUTVA), $T_j = Z_j T_{1j} + (1 - Z_j) T_{0j}$. SUTVA basically states that the potential outcomes of an individual are independent of the treatments received by other individuals in the study and that the observed outcome under treatment Z equals the corresponding potential outcome T_Z . Based on the vector of potential outcomes (T_{0j}, T_{1j}) , the individual causal effect of the treatment on a patient can be defined as $\Delta T_j = T_{1j} - T_{0j}$ and the expected causal effect of Z in the population under study as $\beta = E(T_{1j} - T_{0j})$.

Let us now consider for each patient the response vector $\mathbf{Y}_j = (T_{0j}, T_{1j}, S_j)'$. In the following, attention will be restricted to continuous outcomes and it will be further assumed that $\mathbf{Y}_j \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (\mu_{T0}, \mu_{T1}, \mu_S)'$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{T0T0} & \sigma_{T0T1} & \sigma_{T0S} \\ \sigma_{T0T1} & \sigma_{T1T1} & \sigma_{T1S} \\ \sigma_{T0S} & \sigma_{T1S} & \sigma_{SS} \end{pmatrix}.$$

Extensions to other outcome types are possible but algebraically more challenging; they will resort to a numerical approach.

The individual causal treatment effect ΔT_j is the key variable to assess the impact of the treatment on the patient and, hence, its relationship with the pretreatment predictor S_j will be at the centre of our validation strategy. It is important to point out that when deciding on the appropriateness of treatment for a given patient one always, implicitly or explicitly, needs to consider an alternative intervention. For instance, one may consider if treating a patient with a certain drug may result in a substantial improvement of his condition as compared with leaving him untreated. In such a scenario, the control condition ($Z_j = 0$) would be lack of therapy. The aforementioned distributional assumption leads to

$$\boldsymbol{\psi}_j = \begin{pmatrix} \Delta T_j \\ S_j \end{pmatrix} = \mathbf{A}_\psi \mathbf{Y}_j = \begin{pmatrix} T_{1j} - T_{0j} \\ S_j \end{pmatrix} \sim N(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi), \quad (2.1)$$

where $\boldsymbol{\mu}_\psi = (\beta, \mu_S)$, $\boldsymbol{\Sigma}_\psi = \mathbf{A}_\psi \boldsymbol{\Sigma} \mathbf{A}'_\psi$ and

$$\mathbf{A}_\psi = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad \boldsymbol{\Sigma}_\psi = \begin{pmatrix} \sigma_{T0T0} + \sigma_{T1T1} - 2\sigma_{T0T1} & \sigma_{T1S0} - \sigma_{T0S0} \\ \sigma_{T1S0} - \sigma_{T0S0} & \sigma_{SS} \end{pmatrix}. \quad (2.2)$$

In the next section, a validation method will be introduced based on the stated causal inference model.

3 Predictive causal association

In PM, one is primarily interested in finding a pretreatment predictor S_j that conveys a substantial amount of information about the individual causal treatment effect ΔT_j . In addition, for the normal distribution, the concepts of mutual information and correlation are equivalent and, therefore, it seems logical to assess the adequacy of S_j using the so-called PCA defined as $\rho_\psi = \text{corr}(\Delta T_j, S_j)$. It can be easily shown that

$$\rho_\psi = \frac{\sqrt{\sigma_{T_1T_1}} \rho_{T_1S_0} - \sqrt{\sigma_{T_0T_0}} \rho_{T_0S_0}}{\sqrt{\sigma_{T_0T_0} + \sigma_{T_1T_1} - 2\sqrt{\sigma_{T_0T_0}\sigma_{T_1T_1}} \rho_{T_0T_1}}}, \quad (3.1)$$

where ρ_{XY} denotes the correlation between the variables X and Y . Note further that PCA is also a measure of prediction accuracy, i.e., a measure of how accurately one can predict the individual causal treatment effect on the true endpoint for a given patient, using his pretreatment predictor outcome S_j . In fact, from (2.1) one has

$$\begin{aligned} \Delta T_j | S_j &\sim N [g(S_j), \sigma_{\Delta T}(1 - \rho_\psi^2)], \\ g(S_j) &= \beta + \sqrt{\frac{\sigma_{\Delta T}}{\sigma_{SS}}} \rho_\psi (S_j - \mu_S), \\ \sigma_{\Delta T} &= \sigma_{T_0T_0} + \sigma_{T_1T_1} - 2\sqrt{\sigma_{T_0T_0}\sigma_{T_1T_1}} \rho_{T_0T_1}. \end{aligned}$$

The function $g(S_j)$ can be used to predict the individual causal treatment effect ΔT_j using the pretreatment predictor S_j and the associated prediction mean squared error (PMSE) can be quantified as

$$E \left[\{ \Delta T_j - g(S_j) \}^2 \right] = (1 - \rho_\psi^2) \sigma_{\Delta T}. \quad (3.2)$$

Interestingly, (3.2) also illustrates that the search for a good pretreatment predictor may not be viable in some situations. In fact, the right-hand side of the PMSE is the product of two factors: the first one is a function of the PCA and, hence, it depends on the predictor; the second one, however, can be interpreted as an intrinsic characteristic of the treatment-true-endpoint pair and may have a substantial impact on the PMSE. For instance, if one fixes a value for PMSE, say δ , then $\delta = (1 - \rho_\psi^2) \sigma_{\Delta T}$ and, therefore,

$$\rho_\psi^2 \geq \rho_{min}^2 = 1 - \frac{\delta}{\sigma_{\Delta T}},$$

to achieve the pre-specified level of accuracy. In practice, ρ_{min}^2 could be used to assess the plausibility of finding a good predictor. Indeed, if $\rho_{min}^2 = 0.9$, then one would need to search for a pretreatment predictor that produces a ρ_ψ^2 of at least 90% in order to keep the PMSE smaller than δ . Arguably, such a predictor may be difficult to find.

Let us further assume that $\sigma_{T_0T_0} = \sigma_{T_1T_1} = \sigma_T$, i.e., the variability of the true endpoint is constant across the two treatment conditions, then (3.1) takes the simpler form

$$\rho_\psi = \frac{\rho_{T_1S} - \rho_{T_0S}}{\sqrt{2(1 - \rho_{T_0T_1})}}. \tag{3.3}$$

The previous assumption of homoscedasticity underlies many statistical models like linear regression and analysis of variance. It is also testable using the observable data. In the rest of the manuscript, the homoscedasticity assumption will be used to simplify the algebraic calculations but all conclusions derived will be valid in the most general setting as well.

Some comments are in order. First, note that ρ_{T_1S} and ρ_{T_0S} are the correlations between the potential outcomes for the true endpoint T and the pretreatment predictor S . Under SUTVA, these correlations can be estimated using information from patients who received and did not receive the treatment respectively. As stated in Section 1, these measures or similar ones have been at the core of many evaluation strategies. Second, as (3.3) unequivocally shows, the presence of correlation does not guarantee the validity of the pretreatment predictor. In fact, let us assume, for instance, that $\rho_{T_1S} = \rho_{T_0S} = \gamma > 0$. In such a scenario, and even though the pretreatment predictor is positively correlated with the true endpoint T in groups of patients that received and did not receive the treatment, the PCA equals zero and, consequently, S_j does not convey any information about ΔT_j . Basically, the outcome of the pretreatment predictor is of no help when assessing if the treatment will have a beneficial effect on a given patient. In general, ρ_{T_1S} and ρ_{T_0S} are not sufficient to assess the amount of information the pretreatment predictor conveys on the variable of interest, namely ΔT_j , unless one is willing to assume that the potential outcomes of the true endpoint are uncorrelated, i.e., $\rho_{T_0T_1} = 0$.

4 PCA versus correlation

In the following, we will study in more detail the relationship between the correlation between both endpoints and PCA. To that effect, let us conceptually think of a population of patients prior to the intervention, i.e, prior to the use of the treatment. We will further assume that the correlation between the pretreatment predictor S and the true endpoint T is non-zero, i.e., $\rho_{T_0S} \neq 0$. The question of interest is whether one can use the outcome S_j to gain information about the potential effect of treatment on a given patient. In other words, one wants to know if a correlate can be considered a good predictor of treatment success. The following lemma characterizes the relationship between ρ_{T_0S} and PCA.

Lemma 1. *Let $\psi_j = (\Delta T_j, S_j)'$ denote the vector containing the individual causal treatment effect on T and the pretreatment predictor S with $\psi_j \sim N(\mu_\psi, \Sigma_\psi)$ and Σ_ψ*

as given in (2.2). If it is further assumed that $\sigma_{T_0S_0} = \sigma_{T_1S_1} = \sigma_T$, then

$$|\rho_\psi - a\rho_{T_0S}| \leq b\sqrt{(1 - \rho_{T_0S}^2)}, \quad (4.1)$$

$$\text{where } a = -\sqrt{\frac{1 - \rho_{T_0T_1}}{2}} \text{ and } b = \sqrt{\frac{1 + \rho_{T_0T_1}}{2}}.$$

For the proof we refer to the Appendix. Lemma 1 reveals some interesting aspects of the relationship between PCA and ρ_{T_0S} . First, note that both concepts are interconnected. Actually, the function $a\rho_{T_0S}$ can be interpreted as an approximation of PCA, with the approximation improving as the correlation ρ_{T_0S} increases. Interestingly, in the limit, when $\rho_{T_0S} = 1$, $\rho_\psi = a$ and even a perfect correlate may not convey a lot of information about the individual causal treatment effect. The problem underlying this counterintuitive result is that the relationship between PCA and ρ_{T_0S} may be largely determined by the correlation between the potential outcomes for the true endpoint ($\rho_{T_0T_1}$). Unfortunately, $\rho_{T_0T_1}$ is unidentifiable from the data and, therefore, it is impossible to characterize PCA using solely the identified correlations ρ_{T_1S} and ρ_{T_0S} .

5 PCA: Some identifiability issues

As already hinted at the end of the previous section, the practical use of PCA is hindered by some important identifiability issues. Indeed, the correlations ρ_{T_0S} and ρ_{T_1S} in (3.3) are identifiable from the data and could be estimated using, for instance, the information from patients assigned to a placebo and experimental group in a clinical trial context. However, the correlation between the potential outcomes for the true endpoint $\rho_{T_0T_1}$ is not identifiable from the data. The reason for this lack of identifiability is the fundamental problem of causal inference, i.e., the fact that the vector (T_0, T_1) is almost never observed in practice.

Two strategies are possible to deal with this identifiability issue. First, one can try to define plausible identifiability conditions based on biological or subject-specific knowledge. For instance, in some settings biological knowledge may suggest that the assumption of independent potential outcomes ($T_0 \perp T_1$) is a sensible one. Although widely applied, such an approach suffers from some serious drawbacks. For instance, such subject-specific knowledge may not always be available and/or, as the fundamental problem of causal inference states, these biologically plausible assumptions can neither be proved nor disproved using data.

A second approach is to implement a sensitivity analysis in which ρ_ψ is estimated across a set of plausible values for the unidentifiable correlation $\rho_{T_0T_1}$. To this end, in a first step, a grid of values $G = \{g_1, g_2, \dots, g_k\}$ is considered for the unidentified correlation between the potential outcomes. Next, several Σ matrices are constructed fixing the identifiable correlations ρ_{T_0S} , ρ_{T_1S} and variances $\sigma_{T_0T_0}$, $\sigma_{T_1T_1}$, σ_{SS} at their

estimated values and considering all the values in G for $\rho_{T_0T_1}$. From all the previous Σ matrices only those that are positive-definite are used in the subsequent step. Finally, ρ_{ψ} is estimated based on these positive-definite matrices. Intuitively, the so-obtained vector ρ_{ψ} quantifies PCA across all plausible ‘realities’, i.e., across those scenarios where the assumptions made for the unidentified correlation are compatible with the observed data. The general behaviour of ρ_{ψ} can subsequently be examined, e.g., by quantifying the variability and the range of its estimates and in this way the sensitivity of the results with respect to the unverifiable assumptions can be assessed.

It is important to point out that the aforementioned strategies are not mutually exclusive. Actually, the sensitivity analysis previously described is flexible enough to allow for a straightforward incorporation of subject-specific knowledge in case it exists. For example, if it is reasonable to assume, based on biological knowledge, that the correlation between the potential outcomes should be positive then the grid $G = \{0, 0.01, 0.02, \dots, 1\}$ could be used for this correlation when carrying out the sensitivity analysis. In the present work, the sensitivity analysis strategy will be adopted for the analysis of the case study.

6 Regression approach

As previously stated, pretreatment predictors of therapeutic success are often evaluated using regression models. In the present setting, where both the true endpoint T and the putative predictor S are continuous and normally distributed, linear regression models are the natural choice. A fundamental difference between the methodology introduced in Section 3 and the regression approach is that, while the former uses individual causal treatment effects as the main building block for the analysis, in the latter the focus is set on the expected causal treatment effects (ECE). Given that the very nature of personalized medicine is to carry out predictions at the level of the individual patient, one may argue that individual causal treatment effects are more fundamental in this area than ECE. However, both approaches are intimately connected and in the following this relationship will be studied in detailed.

When interest is in the validation of predictors of therapeutic success, it is natural to allow for an interaction between S and the treatment indicator, thus, typically, the following model is considered

$$T_j = \beta_0 + \beta_1 Z_j + \beta_2 S_j + \beta_3 Z_j S_j + \varepsilon_i. \quad (6.1)$$

Model (6.1) implies that the expected causal treatment effect varies as a function of the pretreatment predictor S_j , more specifically, $ECE(S_j) = \beta_1 + \beta_3 S_j$ (Gelman and Hill, 2006). Obviously, under model (6.1), the coefficient β_3 fully captures the relationship between ECE and the pretreatment predictor. The following lemma summarizes the relationship between β_3 and PCA (the proof is rather straightforward).

Lemma 2. *Let us assume that the causal inference model introduced in Section 2 and the linear regression model given in (6.1) are both valid. If one further assumes that $Z_j \perp S_j$, then $\beta_3 = \sqrt{\frac{\sigma_{\Delta T}}{\sigma_{SS}}} \rho_\psi$.*

Notice that, due to randomization, the assumption of independence $Z_j \perp S_j$ will be often valid for any pretreatment predictor in a clinical trial context. Furthermore, $\rho_\psi = 0$ implies that $\beta_3 = 0$ and, therefore, if the predictor is not valid at the individual level it will not be valid at the population level either. Similarly, if the variances σ_{SS} and $\sigma_{\Delta T}$ are strictly positive, then $\beta_3 = 0$ implies that $\rho_\psi = 0$; and, consequently, if the predictor is not valid at the population level it will not be valid at the individual level. Actually, β_3 and ρ_ψ will always have the same sign and, therefore, the identifiable parameter β_3 can indicate whether the association between the unidentifiable individual causal treatment effect ΔT and the pretreatment predictor S is positive or negative. However, when one moves away from these extreme scenarios ($\beta_3 = 0$ and $\rho_\psi = 0$) important differences between both parameters can emerge. For instance, large values of β_3 may occur when PCA is small if $\sigma_{\Delta T}$ is large relative to σ_{SS} . Consequently, β_3 may not always provide a reliable idea about the magnitude of ρ_ψ .

Note further that model (6.1) implies $E(T_j|Z_j) = \beta_0 + \beta_1 Z_j + \beta_2 \mu_S + \beta_3 Z_j \mu_S$ and, using obvious notation, $\mu_{T0} = \beta_0 + \beta_2 \mu_S$ and $\mu_{T1} = \beta_0 + \beta_1 + (\beta_2 + \beta_3) \mu_S$. Under SUTVA, and using lemma 2, one gets

$$ECE(S_j) = \beta + \sqrt{\frac{\sigma_{\Delta T}}{\sigma_{SS}}} \rho_\psi (S_j - \mu_S) = g(S_j).$$

Thus, both methods provide the same point prediction for ΔT . This is to be expected given that $g(S) = E(\Delta T|S)$, i.e., g describes how the expected causal treatment effect varies as a function of S . However, unlike the regression approach, the causal inference approach allows quantifying the uncertainty of the prediction and offers a direct quantification of the predictive capacity of the pretreatment predictor.

In the following section, the regression and causal inference approaches will be used to evaluate a pretreatment predictor of therapeutic success in patients with opiate/heroin addiction.

7 Case study

A clinical trial in opiate/heroin addiction. The data come from a randomized clinical trial in which the clinical utility of buprenorphine/naloxone (experimental treatment) was compared to clonidine (control treatment) for a short-term (13-day) opiate/heroin detoxification treatment. Before and after the treatment, patients were assessed for relapse, withdrawal symptoms and treatment satisfaction.

Given the emotional, psychological and economical cost of a therapeutic failure, one would like to assess which treatment would likely be more appropriate for a given patient. Along these lines, the Clinical Opiate Withdrawal Scale (COWS) score at screening was evaluated as potential pretreatment predictor of therapeutic success (S). The COWS is an 11-item interviewer-administered questionnaire designed to provide a description of signs and symptoms of opiate withdrawal (e.g., sweating, runny nose, etc). A higher COWS score is indicative for more withdrawal symptoms. The number of days that heroin was used in the 30 days prior to the second follow-up (the second follow-up took place 3 months after the start of the treatment) was used as the true endpoint (T). Data were available for 335 patients, of whom $n = 106$ received the control treatment and $n = 229$ received the experimental treatment.

Missing data. There was a substantial missing data problem in the case study, especially for the true endpoint. Indeed, 231 patients had a missing response for the true endpoint and 6 did not have the score for the pretreatment predictor. Multiple imputation (MI) was used to handle the missing data in all the analyses and 5 imputed datasets were used. In the linear regression approach, models were fitted to each imputed dataset and the results were combined using Rubin's rule. For the causal inference approach, the identifiable variances and covariances were estimated using MI and later passed to the algorithm that carries out the sensitivity analysis.

It is important to point out that MI is valid when the missing generating mechanism is missing at random (MAR) and, hence, the following analyses are based on the MAR assumption.

Data description. The overall treatment effect (treatment difference) on T was $\hat{\beta} = -0.9314$ (*s.e.* = 1.8820, $p = 0.632$). The negative $\hat{\beta}$ indicates that the average number of days that heroin was used post-treatment is smaller in the experimental treatment group ($\hat{\mu}_E = 11.6437$) than in the control treatment group ($\hat{\mu}_C = 12.5751$) – albeit the difference was not significant (2-sided $p = 0.709$).

The correlations between S and T in the experimental and active control treatment groups were both significant with $\hat{\rho}_{T1S} = -0.3367$ ($p = 0.001$) and $\hat{\rho}_{T0S} = -0.3699$ ($p = 0.001$). The negative correlations indicate that patients who have higher $S =$ COWS scores tend to use less heroin in the 30-day interval after the treatment in both treatment conditions. Importantly, ρ_{T1S} and ρ_{T0S} are not significantly different, $p = 0.750$.

Linear regression approach. Table 1 summarizes the results obtained when model (6.1) is fitted to the data ($R^2 = 0.1271$). As it can be seen, the estimated average causal treatment effect is $ECE(S_j) = -1.5504 + 0.0066S_j$. Notice that the point estimate for β_3 was nearly zero and it was not statistically significant. These results together with Lemma 2 hint on a probably small PCA. In the following analysis this unidentifiable parameter will be studied in more detail.

Table 1 Parameter estimates for model (6.1).

	β	s.e.	p
Intercept	26.6627	4.2312	0.001
Z	-1.5504	6.1829	0.8047
S	-0.1677	0.0454	0.0002
Z by S interaction	0.0066	0.0661	0.9208

Source: Authors' own.

Predictive causal association. The sensitivity analysis introduced in Section 5 was carried out using the newly developed R package *EffectTreat* (freely available at <http://cran.r-project.org/>). For conciseness, in the present section no reference to the software is made but in the Supplementary Materials Web Appendix a more comprehensive analysis of the data is provided and the use of the R package *EffectTreat* is explained.

As noted in Sections 3 and 4, the presence of a correlation between the predictor S and the outcome T in both treatment conditions does not necessarily imply that S conveys much information about the individual causal effect ΔT , i.e., it does not necessarily imply that S is a reliable predictor of therapeutic success.

The PCA allows to quantify the predictive capability of S but the identifiability problems discussed in Section 5 hindered its estimation. Essentially, the correlation between the potential outcomes ρ_{T0T1} is unidentifiable and, consequently, ρ_ψ cannot be directly assessed from the data. To handle this problem, the sensitivity analysis introduced in Section 5 was used to estimate ρ_ψ across a set of plausible values for ρ_{T0T1} , i.e., the values in the grid $G = \{-1, -0.99, -0.98, \dots, 1\}$.

Using the values in G for ρ_{T0T1} , $\hat{\sigma}_{T0T0}$, $\hat{\sigma}_{T1T1}$, $\hat{\sigma}_{SS}$, $\hat{\rho}_{T0S}$ and $\hat{\rho}_{T1S}$, a total of 201 Σ matrices were formed of which 175 were positive-definite. Based on the valid matrices, ρ_ψ was estimated. Figure 1 displays the frequency distribution of ρ_ψ . The mean $\rho_\psi = 0.0088$, mode $\rho_\psi = 0.0057$ and median $\rho_\psi = 0.0066$. Further, 95% of the ρ_ψ values were below 0.0198 and ρ_ψ was at most 0.0561. Clearly, the results indicate that in all 'realities' that are compatible with the observed data, ρ_ψ is very small. It can thus be concluded that COWS at screening conveys little information on the individual causal treatment effect and, therefore, the accuracy with which a patient's individual causal treatment effect on T can be predicted based on COWS is small. The results of this analysis are in full agreement with the earlier claim (Section 4) that correlation is not evidence of predictive validity.

Notice that these findings are also in agreement with the results of the regression analysis previously presented. Indeed, the small point estimate and large p value obtained for β_3 were already indicative of a small PCA.

Is there a good pretreatment predictor? The forgoing analysis showed that COWS is not a good pretreatment predictor of the individual causal treatment effect. More generally, given a clinically relevant outcome T and a treatment of interest, one would like to know whether there *exists* a good pretreatment predictor in the first place.

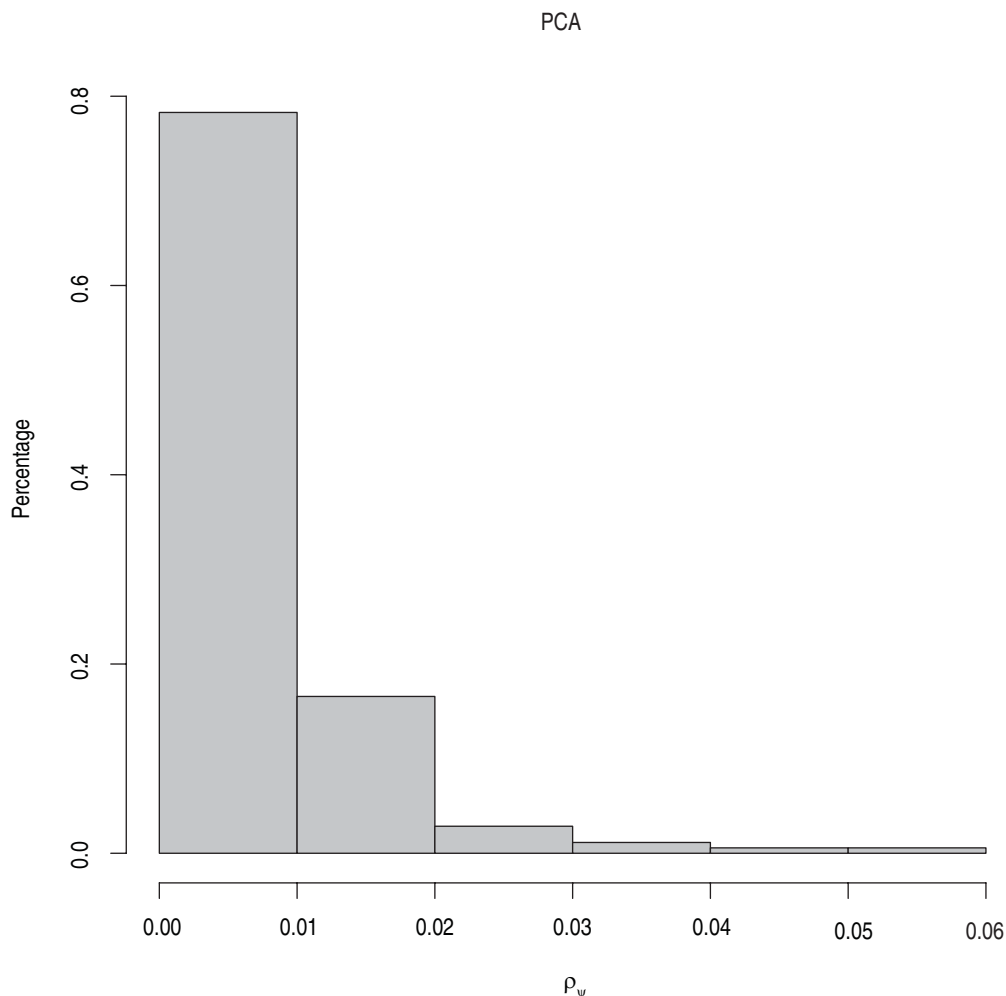


Figure 1 Relative frequencies of ρ_ψ in the opiate/heroin study.

Source: Authors' own.

For instance, let us suppose that one wants to examine the plausibility of finding a pretreatment predictor S that allows for the prediction of ΔT with an average PMSE = 50, i.e., with an average prediction error of about 7 days. Figure 2 shows the distribution of ρ_{min}^2 in such a scenario. About 80% of the ρ_{min}^2 values were above 0.1088, indicating that a candidate S should produce a ρ_ψ of at least $\sqrt{0.1088} = 0.3298$ to achieve the desired level of prediction accuracy. Whether this is a realistic endeavour or not could then be discussed with experts in the field and this information could also help to select potential predictors for future validation studies.

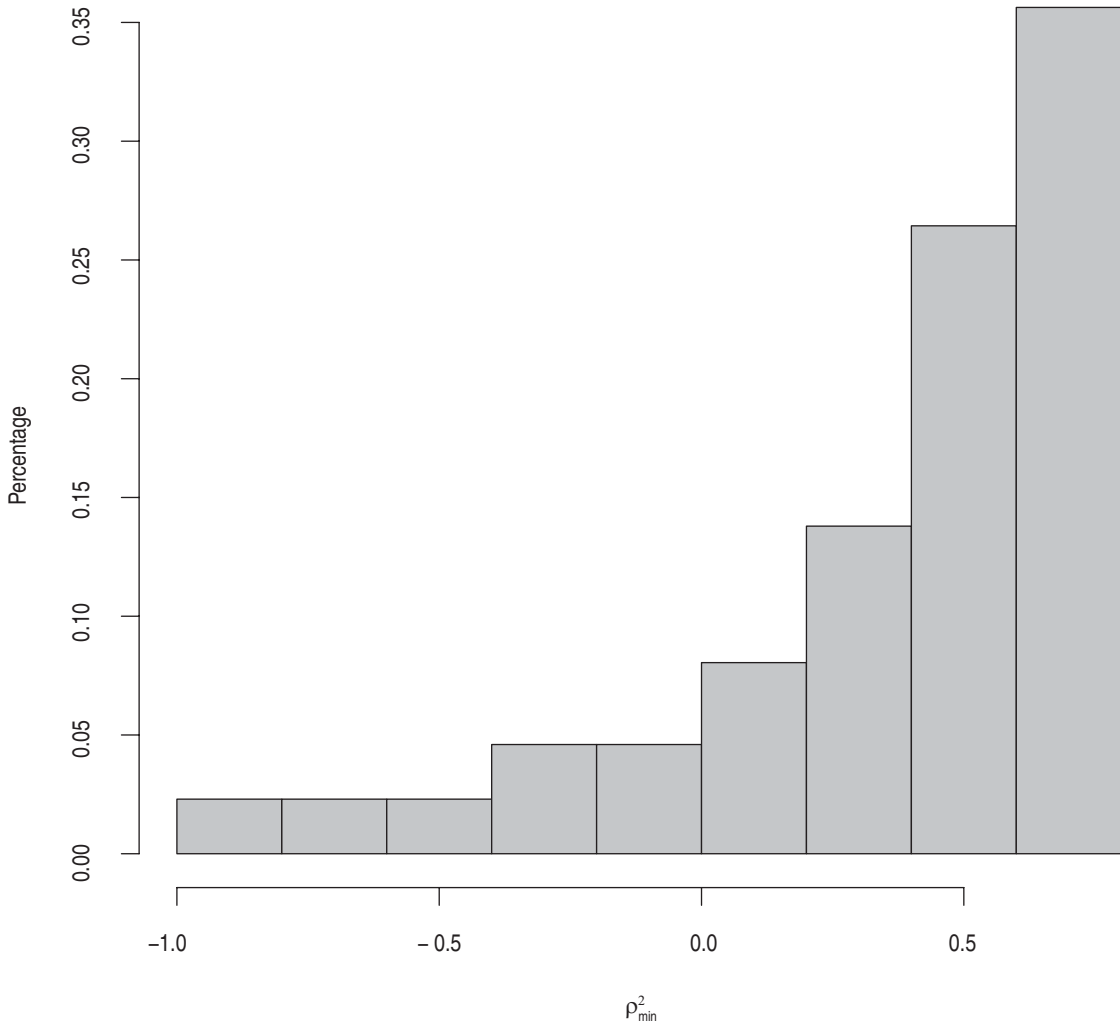


Figure 2 Relative frequencies of ρ_{\min}^2 in the opiate/heroin study using PMSE = 50.
Source: Authors' own.

Predicting ΔT based on S in an individual patient j . In practice, clinicians are interested in the prediction of a patient's individual causal treatment effect (ΔT_j) given the observed S_j . For illustrative purposes, let us assume that a patient scores 60 on COWS, i.e., the patient has a low level of opiate withdrawal symptoms. Using the methodology detailed in Section 3, it is obtained that the expected value of $\Delta T_j | S_j = 60$ equals -1.0409 . Therefore, the patient is expected to have about 1 more heroin-free day in the post-treatment interval with the experimental treatment than with the control.

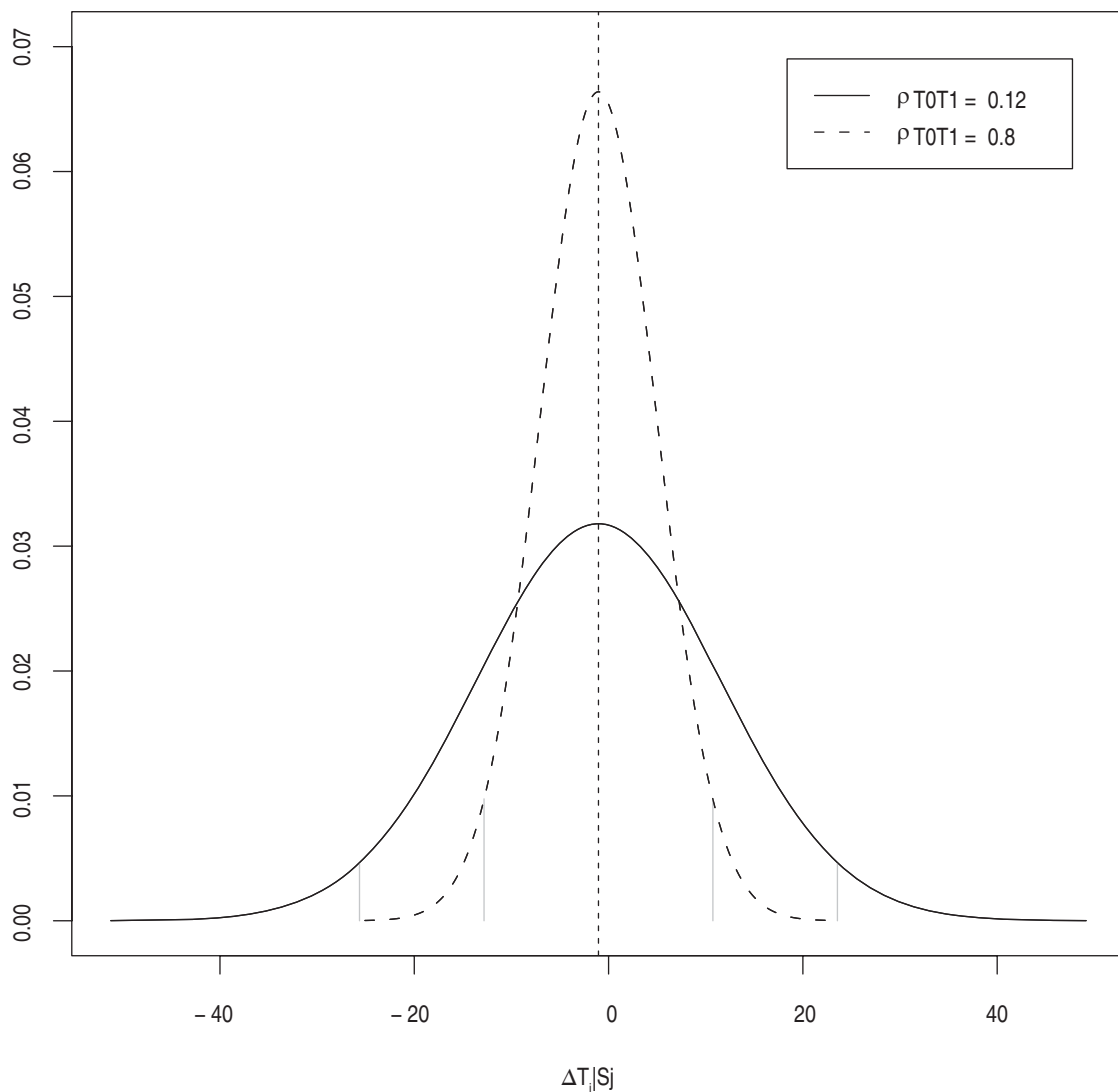


Figure 3 The distributions (and 95% confidence intervals) for $\Delta T_j|S_j = 60$ assuming $\rho_{T0T1} = 0.120$ and $\rho_{T0T1} = 0.800$ in the opiate/heroin study.

Source: Authors' own.

Importantly, the expected value of $\Delta T_j|S_j$ remains constant no matter what assumption is made regarding ρ_{T0T1} and, therefore, the point prediction will be the same across all plausible values of ρ_{T0T1} . However, the variance and 95% support interval obtained from the distribution is actually affected by the value assumed for ρ_{T0T1} . For example, the 95% support interval around the point prediction -1.0409 is rather wide $[-25.6368; 23.5550]$ when it is assumed that ρ_{T0T1} equals its median

value 0.120, whereas it is substantially more narrow $[-12.8194; 10.7376]$ when it is assumed that $\rho_{T_0T_1} = 0.800$. Figure 3 displays the distributions for $\Delta T_j | S_j = 60$ assuming $\rho_{T_0T_1} = 0.120$ and $\rho_{T_0T_1} = 0.800$. Notice that for both $\rho_{T_0T_1}$ the 95% support intervals include 0 and thus no difference between both treatments is to be expected for the patient. Basically, as the low PCA shows, COWS scores do not lead to precise predictions of the individual causal treatment effect. In principle, medical considerations could be used to establish a range of medically plausible values for $\rho_{T_0T_1}$ and carry out the prediction only considering the values in this range.

8 Discussion

Predicting individual causal treatment effects based on pretreatment outcomes is fundamental in PM. It may be argued that the future impact of PM on medical practice largely will depend on our capacity to find and properly validate pretreatment predictors of therapeutic success across different conditions. In the present work, we have argued that classical correlational analysis cannot answer the relevant scientific questions in this area and may be rather misleading. In response, we introduce a new validation strategy to address this problem. The methodology is based on causal inference and allows answering several clinically relevant questions. In addition, an R library is provided to carry out the necessary calculations. There are, however, some important limitations that need to be pointed out. For instance, the current methodology only considers continuous and normally distributed pretreatment predictors and true endpoints. Other type of variables, like binary, are often encountered in medical applications and future research efforts will be directed to generalizing this methodology to other outcome types.

Even though it may seem natural, imputing missing potential outcomes to get direct estimates of $\rho_{T_0T_1}$ is of little value in the present context. In fact, T_0 and T_1 are never simultaneously observed and, consequently, the data at hand do not contain any information about the correlation between the potential outcomes. Therefore, any information about this parameter in the imputed datasets has to come necessarily from the imputation model. So basically, one would need to impute the data using several imputation models that assume different values for $\rho_{T_0T_1}$ in order to avoid bias. Although more laborious, this approach would be equivalent to the sensitivity analysis used in the manuscript and, therefore, MI does not seem to be a better strategy in this scenario.

It is important to point out that the causal inference model used to compute PCA is essentially unidentifiable. However, the distributional assumptions for some marginals, like the bivariate normality of (S_0, T_0) and (S_1, T_1) , can be assessed using the data at hand (details in the Web Appendix). Clearly, the findings of the previous sections are predicated on the assumption that the potential outcomes are continuous and normally distributed. However, many of these results will still be valid for non-normal continuous potential outcomes, although their interpretation will change if the normality assumption is questioned. For instance, the correlation between the individual causal treatment effect and the pretreatment predictor could still

be quantified using ρ_ψ , but these expressions could not be interpreted any longer as the PCA. Indeed, even though ρ_ψ will still be a valid measure of causal correlation, the equivalence between association and correlation will be broken if the normality assumption is dropped. In general, departures from normality will only have a mild effect on the estimation of the parameters of interest but, as already pointed out, the interpretation of the results will become more restricted and limited.

Acknowledgements

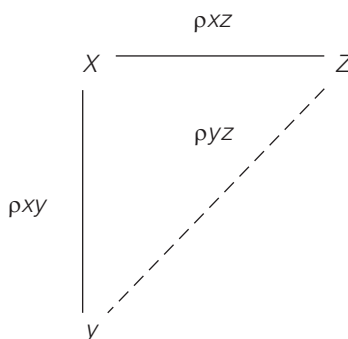
Financial support from the IAP research network #P7/06 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged. The research leading to these results has also received funding from the European Seventh Framework programme [FP7 2007 – 2013] under grant agreement 602552.

Supplementary Materials

An online appendix which details the analysis of the case study using the R package *EffectTreat* is available on the journal’s website.

Appendix

Let us first consider the following diagram describing the association structure between the elements of the vector $\mathbf{Y}' = (x, y, z)$.



In what follows it will be assumed that

$$\mathbf{Y} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\rho), \quad \text{where} \quad \boldsymbol{\Sigma}_\rho = \begin{pmatrix} 1 & \rho_{xy} & \rho_{xz} \\ \rho_{xy} & 1 & \rho_{yz} \\ \rho_{xz} & \rho_{yz} & 1 \end{pmatrix}.$$

Notice that, without loss of generality, it has been assumed that the components of Y have mean zero and variance one. This can be achieved by standardization and will have no impact on statements regarding the association between these components. The question of interest is, given ρ_{xy} and ρ_{xz} , what can be said about the value of ρ_{yz} . The previous distributional assumptions imply that

$$y = \rho_{xy}x + \varepsilon_y, \quad \varepsilon_y \sim N(0, 1 - \rho_{xy}^2),$$

$$z = \rho_{xz}x + \varepsilon_z, \quad \varepsilon_z \sim N(0, 1 - \rho_{xz}^2),$$

with $\varepsilon_y, \varepsilon_z \perp x$. Moreover, $\rho_{yz} = \text{cov}(y, z) = \text{cov}(\rho_{xy}x + \varepsilon_y, \rho_{xz}x + \varepsilon_z) = \rho_{xy}\rho_{xz} + \text{cov}(\varepsilon_y, \varepsilon_z)$.

Note further that $\text{cov}(\varepsilon_y, \varepsilon_z) = \rho(\varepsilon_y, \varepsilon_z)\sqrt{(1 - \rho_{xy}^2)(1 - \rho_{xz}^2)}$ and, therefore,

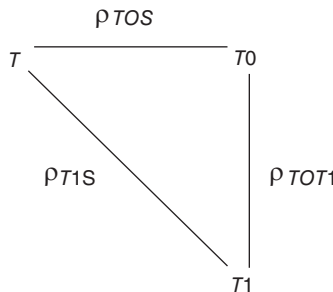
$$-\sqrt{(1 - \rho_{xy}^2)(1 - \rho_{xz}^2)} \leq \text{cov}(\varepsilon_y, \varepsilon_z) \leq \sqrt{(1 - \rho_{xy}^2)(1 - \rho_{xz}^2)},$$

and the previous inequality leads to

$$\rho_{xy}\rho_{xz} - \sqrt{(1 - \rho_{xy}^2)(1 - \rho_{xz}^2)} \leq \rho_{yz} \leq \rho_{xy}\rho_{xz} + \sqrt{(1 - \rho_{xy}^2)(1 - \rho_{xz}^2)}. \quad (\text{A-1})$$

Proof of Lemma 1

Based on the result introduced in Appendix and the next diagram one has



$$\begin{aligned} \rho_{ST0}\rho_{T0T1} - \sqrt{(1 - \rho_{ST0}^2)(1 - \rho_{T0T1}^2)} &\leq \rho_{T1S} \leq \rho_{ST0}\rho_{T0T1} \\ &+ \sqrt{(1 - \rho_{ST0}^2)(1 - \rho_{T0T1}^2)}, \\ -\rho_{T0S}(1 - \rho_{T0T1}) - \sqrt{(1 - \rho_{ST0}^2)(1 - \rho_{T0T1}^2)} &\leq \rho_{T1S} - \rho_{T0S} \leq -\rho_{T0S}(1 - \rho_{T0T1}) \\ &+ \sqrt{(1 - \rho_{ST0}^2)(1 - \rho_{T0T1}^2)} \\ -\rho_{T0S}\sqrt{\frac{1 - \rho_{T0T1}}{2}} - \sqrt{(1 - \rho_{ST0}^2)\left(\frac{1 + \rho_{T0T1}}{2}\right)} &\leq \rho_{\psi} \leq -\rho_{T0S}\sqrt{\frac{1 - \rho_{T0T1}}{2}} \\ &+ \sqrt{(1 - \rho_{ST0}^2)\left(\frac{1 + \rho_{T0T1}}{2}\right)}. \end{aligned}$$

References

- Abrahams, E. and Silver, M. (2010). The history of personalized medicine. In Gordon, E., and Koslow, S. (eds), *Integrative Neuroscience and Personalized Medicine*, pages 3–16. New York, USA: Oxford University Press.
- Banerjee, P., Choi, B., Shahine, L.K., Jun, S.H., O’Leary, K., Lathi, R.B., Westphal, L.M., Wong, W.H., Yao, M.W. (2010). Deep phenotyping to predict live birth outcomes in in vitro fertilization. *Proc Natl Acad Sci USA*, **107**, 13570–13575.
- Gelman, A., and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York, USA: Cambridge University Press.
- Holland, Paul W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, **81**, 945–960.
- Masao, H., Takayoshi, S., Tetsuro, S., Akito, S., Rika, H., Kuniaki, A., Tatsuya, Y., Yoshio, S., Taro, Y., Hikari, O., Kazuhisa, M., Mikiko, N., Eishiro, M., and Shuichi, K. (2014). Hepatic interferon-stimulated genes are differentially regulated in the liver of chronic hepatitis C patients with different interleukin-28B genotypes. *Hepatology*, **59**, 828–838. doi: 10.1002/hep.26788
- Miwa, T., Hayter, A.J., and Kuriki, S. (2003). The evaluation of general non-centred orthant probabilities. *Journal of The Royal Statistical Society Series B*, **65**, 223–234.
- Roest, J., van Heusden, A.M., Mous, H., Zeilmaker, G.H., and Verhoeff, A. (1996). The ovarian response as a predictor for successful in vitro fertilization treatment after the age of 40 years. *Fertil Steril*, **66**, 969–973.
- Shin, L.M., Davis, F.C., Van Elzaker, M.B., Dahlgren, M.K., and Dubois, S.J. (2013). Neuroimaging predictors of treatment response in anxiety disorders. *Biology of Mood & Anxiety Disorders*, **3**, 3–15.
- Spear, B.B., Heath-Chiozzi, M., and Huff, J. (2001). Clinical application of pharmacogenetics. *TRENDS in Molecular Medicine*, **7**, 201–204.
- Spielman, S.R., Schwartz, J.S., McCarthy, D.M., Horowitz, L.N., Greenspan, A.M., Sadowski, L.M., Josephson, M.E., and Waxman, H.L. (1983). Predictors of the success or failure of medical therapy in patients with chronic recurrent sustained ventricular tachycardia: a discriminant analysis. *J Am Coll Cardiol.*, **1**, 401–408.

- van Loendersloot L.L., van Wely, M., Limpens, J., Bossuyt, P.M.M., Repping, S., and van der Veen, F. (2010). Predictive factors in in vitro fertilization (IVF): A systematic review and meta-analysis. *Human Reproduction Update*, **16**, 577–589. doi: 10.1093/humupd/dmq015
- Zeller, D., Reiners, K., Brauning, S., and Buttmann, M. (2013). Central motor conduction time may predict response to fampridine in multiple sclerosis patients. *J Neurol Neurosurg Psychiatry*, **85**, 707–709. doi: 10.1136/jnnp-2013-306860