

Estimation of HIV treatment-efficacy by combining structural nested mean models with pharmacokinetic models of antiretroviral drug exposure

LAETITIA COMTÉ, STIJN VANSTEELANDT, RICHARD A. RODE
AND BERNARD VRIJENS*

The aim of treating HIV-1-infected patients is to achieve and maintain suppression of viral load (VL). Achievement of this aim is thwarted by variable adherence to prescribed anti-retroviral drugs. Variable adherence to an antiretroviral regimen creates variability in the patient's internal exposure to the drugs. Structural nested mean models (SNMMs) enabled us to estimate, during the initial phase of treatment, the relationship between variable internal exposure and VL, accounting for measured time-varying confounders and feedback relations using an antiretroviral regimen containing lopinavir/ritonavir (LPV/RTV, LPV/r). Our final SNMM predicts that the short term effect of treatment is modified by the most recent past VL, with higher initial VL's being associated with larger treatment-induced reductions in VL for a given internal exposure to the drugs. Variation in internal exposure to LPV/r in the interquartile interval (P25%–P75%) only slightly affects the overall reduction in VL, supporting the conclusion that the relatively long duration of action of LPV/r lessens the impact on VL of the most frequently recurring intermittent lapses in dosing.

KEYWORDS AND PHRASES: SNMM, HIV, PK.

1. INTRODUCTION

Variable drug exposure resulting from sub-optimal adherence to prescribed therapy creates a major source of statistical noise in drug response (Harter and Peck 1991). This is especially noticeable in the treatment of HIV-1-infected patients, where the importance of punctual medication adherence is vital; see the publications of Vanhove et al. (1996) and Paterson et al. (2000). Moreover, the extent of antiretroviral drug resistance may impact the subsequent virologic response that develops in patients receiving such therapy. In medical practice, HIV-1 VL is routinely measured to monitor virologic response throughout the course of antiretroviral therapy. With the advent of electronic monitoring, drug

dosing histories are more readily available and provide an opportunity to explore the association between drug exposure and both virological response and adverse events. When reliably measured, one can utilize such drug exposure measurements in the analysis, to turn this noise into a source of knowledge (Vrijens and Goetghebeur 1997).

The causal relationship between drug exposure and clinical outcome is however not straightforward as medication non-adherence is often selective. For example, patients who adhere well to their assigned medication often form a selective subgroup of patients who are relatively more healthy, and therefore may not be comparable to patients who do not adhere well (Lee et al. 1991, Goetghebeur and Pocock 1993); one instance of this is the so-called healthy user effect. Analyses of the effects of drug exposure on clinical outcomes must recognize this selection process to prevent confounding bias. If at each time all prognostic factors of the outcome at that time that influence adherence have been measured, then selectivity is said to be explainable, and statistical adjustment possible. However, classical regression methods, adjusting for these time-varying prognostic factors, can be misleading because they do not carefully acknowledge the possibility of feedback relations whereby adherence, prognostic factors and outcome mutually influence each other over time (Robins 1997). This limitation of classical regression models arises because they attempt to parameterize the effects of early and late exposures within a single conditional mean model. Indeed, assessing the effects of late exposures requires adjustment for the entire confounder history; this compromises assessment of the effect of early exposures when the confounder history is (partly) affected by the early exposures for then the corresponding analysis is subject to a post-treatment selection bias.

SNMMs overcome this limitation by separately modeling the effect of the observed drug exposure at each time on future clinical outcomes, conditional on the history of drug exposures and confounders. In this paper, we will adopt such models to assess the relationship among HIV-1-infected patients between the internal drug exposure and the kinetics of VL responses, adjusted for confounding by measured time-varying covariates (e.g. the occurrence of adverse

*Corresponding author.

events (AEs)). To account for pharmacokinetic differences in drug absorption and/or elimination, electronically compiled drug dosing histories will be utilized with individual pharmacokinetic parameters in order to reliably estimate individual temporal patterns of internal drug exposure. In turn these estimated internal drug exposure profiles will be used in the SNMM to quantify the causal relationship between drug exposure and clinical outcome.

When investigating the relationship between drug exposure and VL responses, two distinct phases can be identified. The initial phase reflects treatment initiation in patients naïve to antiretroviral treatment. During this phase the initial decrease in VL that occurs soon after therapy commences, signals the initial success of the antiretroviral therapy. Quantifying the causal relationship between drug exposure and the amount of VL decrease during this initial stage of the treatment imparts unique insight into the treatment process, and will be the focus of the present paper. The subsequent phase of effective antiretroviral therapy consists of the maintenance of viral suppression over many years. During this time many factors which were absent or had a minimal impact during the initial phase can influence the likelihood of continued VL suppression during the second phase. The two phases may vary in length between individual patients. We show that these phases can be separated in the SNMM by incorporating constraints on previous VL, although this requires some adaptation of the estimation procedure. The SNMMs are fitted employing standard statistical software for solving Generalized Estimating Equations (GEE) with independence correlation (Liang and Zeger 1986) (e.g. `proc genmod` in SAS or the library correlated data in S-Plus). Confidence intervals are derived using bias-corrected, accelerated bootstrap analysis (Efron and Tibshirani 1993).

2. DATA

2.1 Study design

A phase III, 96-week randomized (3:2), open-label, multi-centre, multi-country study was conducted to evaluate LPV/r administered once daily (800/200 mg; $n = 115$) versus twice daily (400/100 mg; $n = 75$) in HIV-1-infected antiretroviral-naïve patients. All patients received open label tenofovir disoproxil fumarate (TDF) 300 mg and emtricitabine (FTC) 200 mg, both once daily. Baseline viral resistance was not obtained. For subjects with virological failure after week 12, however, genotype and phenotype resistance was performed and a corresponding baseline sample was submitted for resistance testing. Details of the study design and results have been previously reported (Molina et al. 2007).

To summarize briefly, Molina and colleagues (2007) identified decreasing adherence to LPV/r over time in both groups although a higher rate of decrease was seen in the

twice daily group compared with the once daily group. There was also a significant association between the time to loss of virologic response (TLOVR), defined as confirmed VL rebound of ≥ 50 copies/ml or failure to achieve a VL response of < 50 copies/ml, with adherence.

While the methodology described in this paper can be applied separately to both randomized groups, the focus of the present analysis will involve only the LPV/r once daily group. Of the 115 patients enrolled in the LPV/r once daily group, 6 patients had no dosing history data available to estimate drug exposure and 2 patients had no VL measurements available after baseline. This resulted in a total of 107 evaluable patients for this analysis.

The baseline visit included a discussion of the importance of adherence to the drug regimen and of details on the LPV/r-based regimen to which the patient was randomized. Patients were instructed to begin their randomly assigned antiretroviral treatment regimen on the following day. Plasma HIV-RNA levels were measured at baseline and at weeks 4, 8, 16, 24, 32, 40 and 48. Blood samples were collected for measurement of plasma concentrations of LPV and RTV at weeks 4, 8, 16, 24 and 48. Each measurement was obtained 24h after the last dose was taken.

Drug intake of LPV/r between study visits was assessed using the Medication Event Monitoring Systems (MEMS[®]), which records the date and time the patient opens the monitored pill container. For the purpose of this clinical trial, drug intake of TDF and FTC were not recorded using MEMS.

The onset and completion of each gastrointestinal AE were recorded at each clinical visit in terms of the start and the end of each reported event. Finally three baseline covariates were available for this analysis (i.e., gender, age and ethnicity).

2.2 Estimated internal drug exposure

In the treatment of HIV-infected patients, inhibition of viral replication is related to the time-course of drug intake by the patient and is referred to as the ‘external drug exposure’. The drug gains entry into the body either through oral, parenteral or continuous infusion of dosing. Regardless of the route of drug entry into the body, it undergoes distribution via the circulating blood to different parts of the body resulting in a characteristic time course of drug concentration between the extracellular and intracellular fluids. A frequently-used reference concentration is the time-course of the drug’s concentration in plasma, which is considered as defining the drug’s ‘internal exposure’. Typically, following each sequential dose of drug, its concentration in plasma and other body fluids rises for a time, peaks, and then commences to fall, as a variety of processes distribute drug to tissues that metabolize and excrete drug. This continues until the drug concentration in the plasma approaches its nadir value. The pharmacodynamic action of the drug is most

closely related to the prevailing internal exposure, i.e., the prevailing concentration of drug in the plasma. The internal exposure of the drug declines throughout most of the intervals between successive doses because the combined rates of dilution, tissue distribution, metabolism and excretion of the drug and its metabolites exceed the rate of drug entry into the body. The dosing process must, therefore, be repeated at suitable intervals in order to keep the internal exposure of the drug within a therapeutic range that is not too high so as to induce toxic effects and not too low so as to allow drug action to cease.

Between the measures of external and of internal exposures are the processes of absorption, distribution, metabolism, and excretion of the drug commonly abbreviated ‘ADME’. Each of the ADME processes is subject to variation which can contribute to the statistical noise between the patient’s external and internal exposures to the drug. Other sources of variance in drug responses have been described and modeled by Harter and Peck (1991) who highlight the relative magnitudes of the different sources of variance in drug response. One of the major sources of variation in drug response is the patient’s ability to adhere to the prescribed drug dosing regimen. This variation in the external exposure to the drug adds another component of variance to those that arise from the various ADME processes. Thus, internal exposure, with its associated variability, is a logical point of reference upon which to base predictions of the time course of a drug’s therapeutic actions.

For repeated dosing, the estimated plasma drug concentration for a given patient can be recursively estimated by using the MEMS-recorded dosing history as an input to the patient-specific pharmacokinetic model (Vrijens et al. 2005). In particular, the MEMS-recorded dosing history for each patient is used as input to his/her pharmacokinetic model in order to obtain individualized parameter estimates. The output from each patient’s model is the projected time course of the plasma drug concentration throughout the entire study period of time spanned by the MEMS-compiled dosing history. Estimates of the individual pharmacokinetic parameters (absorption rate $k_{a,i}$, elimination rate $k_{e,i}$ and apparent volume of distribution V_i) are derived from a one compartment model which are then combined with the patient’s dosing histories to project a reliable estimation of the internal drug exposure profiles over time (Vrijens et al. 2005). The average internal drug exposure between two consecutive visits can then be estimated by dividing the area under the entire inter-visit estimated internal drug exposure profile by the time (in hours) separating the two visits.

3. METHOD

3.1 Notation

Figure 1 illustrates the temporal sequence of time-varying measurements obtained during the course of this study

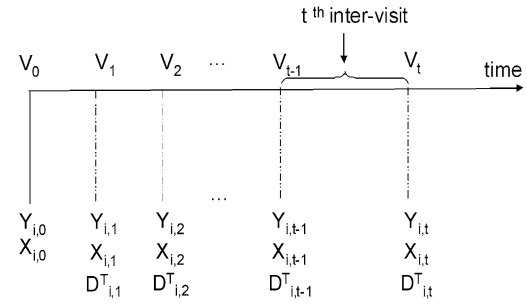


Figure 1. Schematic illustration of the notation used for time-varying measurements during the course of the clinical experiment.

with the corresponding notation used throughout the paper. Here, $Y_{i,t}$ denotes the observed clinical outcome for patient i at visit t . $D_{i,t}^T$ denotes the internal drug exposure or a summary thereof for patient i in the t th inter-visit interval, which is defined as the time between visit $t - 1$ and visit t , $t = 1, \dots, T$. $X_{i,t}$ denotes the vector of confounders of the relationship between drug exposure and clinical outcome (i.e., for all time-varying risk factors of the exposure that are also associated with the outcome), as measured for patient i at visit t or during the t th inter-visit interval. Further, $X_{i,0}$ is used to denote baseline confounders. For any variable $W_{i,t}$, let $\bar{W}_{i,t} \equiv (W_{i,0}, \dots, W_{i,t})$ represent the history of $W_{i,t}$ measured for patient i from baseline until visit t . For each pair of visit times $s, t = 1, \dots, T$ with $s \leq t$, $Y_{i,t}(\bar{D}_{i,s}^T, 0, \dots, 0)$ is used to represent the potential VL outcome that would be observed for patient i at visit t if the internal exposure history were as observed, $\bar{D}_{i,s}^T$, until visit s and zero thereafter. Note that this potential VL outcome may not be observed.

3.2 Structural Nested Mean Model (SNMM)

Classical longitudinal regression methods, like maximum likelihood estimation for linear mixed models, generally cannot be used to assess the effect of internal drug exposure on the clinical outcome of interest, even when regressing on all time-varying confounders that exist between exposure and outcome. The use of these methods can lead to a substantial bias in the exposure-effect estimate whenever the time-varying confounders are themselves affected by earlier exposures. This is so by the same reason that adjustment for post-randomization measurements can be fallible in the analysis of randomized clinical trials (Robins 1997). In this article, we will attempt to make progress through the use of so-called structural nested mean models (Robins 1997).

Throughout, we will assume that we have available all confounders for the association between exposure at each visit t and the outcome future from visit t onwards. Note that, to satisfy this assumption, the vector $\bar{X}_{i,t-1}$ of con-

founders at time t may need to include the previously measured outcome $Y_{i,t-1}$; e.g. when the exposure at time t is influenced by the previous outcome measurement. Under this so-called ‘no unmeasured confounders’ assumption, the effect of exposure history $\bar{D}_{i,t}^T$ on the outcome $Y_{i,t}$ is identifiable from the observed data at each time t . When the outcome is continuous and lends itself to linear modeling, G-estimation for SNMMs can be used. Here, a SNMM can be viewed as a union model which uses a finite-dimensional (unknown) parameter to describe at each time s , the average effect of taking away the exposure at that time on the outcome at each later time t , conditional on the history of exposure and confounders at time s . In particular, it thus parameterizes the average difference between the observed outcome $Y_{i,t}(\bar{D}_{i,t}^T)$ at each time t and the potential outcome $Y_{i,t}(\bar{D}_{i,t-1}^T, 0)$, which would have been observed for that patient at that time t if that patient had not received the drug in the t th inter-visit interval; this is parameterized for the subpopulation of patients who share the same exposure and confounder history up to visit t . In addition, it parameterizes the average difference between $Y_{i,t}(\bar{D}_{i,t-1}^T, 0)$ and the potential outcome $Y_{i,t}(\bar{D}_{i,t-2}^T, 0, 0)$ that would have been observed for that same patient at time t if he/she were not exposed to the drug after visit $t-2$; this is parameterized for the larger subpopulation of patients who share the same exposure and confounder history up to visit $t-1$. A specific instance of a SNMM is defined by the following simultaneous equations for all t :

$$(1) \quad \begin{cases} E \{Y_{i,t}(\bar{D}_{i,t}^T) - Y_{i,t}(\bar{D}_{i,t-1}^T, 0) | \bar{X}_{i,t-1}, \bar{D}_{i,t}^T\} = \beta_{st} Z_{i,t}^T \\ E \{Y_{i,t}(\bar{D}_{i,t-1}^T, 0) - Y_{i,t}(\bar{D}_{i,t-2}^T, 0, 0) | \bar{X}_{i,t-2}, \bar{D}_{i,t-1}^T\} \\ = \beta_{lg,1} Z_{i,t-1}^T \\ \dots \\ E \{Y_{i,t}(D_{i,1}^T, 0, \dots, 0) - Y_{i,t}(0, \dots, 0) | X_{i,0}, D_{i,1}^T\} \\ = \beta_{lg,t-1} Z_{i,1}^T \end{cases}$$

which assume the short-term and long-term exposure effects, vectors β_{st} and $\beta_{lg,s}$ for $s = 1, \dots, t-1$ respectively, to be constant over the visit times.

Here, $Z_{i,t}^T$ is a known function of the exposure history up to the t th inter-visit interval and of the confounder history up to the $(t-1)$ th visit, and such that $Z_{i,t}^T = 0$ whenever $D_{i,t}^T = 0$ for $1 \leq t \leq T$; for instance, $Z_{i,t}^T = D_{i,t}^T$ or $Z_{i,t}^T$ is the matrix $D_{i,t}^T(1 \times X_{i,t-1})$ when interactions between internal drug exposure and covariates are of interest.

3.3 G-estimation

Robins (1994) developed a semi-parametric efficient estimation method for the parameters indexing a SNMM, which he referred to as G-estimation. We present here a special case which separates the estimation of short-term drug exposure effects from the estimation of long-term effects. Here,

the short-term effect of drug exposure refers to the direct effect of the exposure measured during inter-visit interval t on the immediately following outcome measure $Y_{i,t}$ at time t . It is parameterized by the first equation in (1). The long-term effect of drug exposure refers to the combination of direct effects from exposures preceding inter-visit interval t on the outcome measure $Y_{i,t}$ at time t . It is parameterized by the subsequent equations in (1).

3.3.1 Short-term effects

The estimation of the short-term effect of drug exposure is performed using a two-stage procedure. The first stage involves modeling drug exposure at each time t as a function of drug exposure and confounder history up to time $t-1$. The second stage involves regressing the outcome at each time t on the residuals obtained from the previous model as well as on the confounder and drug exposure history up to time $t-1$. The coefficient corresponding to the residual may then be interpreted as a short-term exposure effect provided that either the first stage or the second stage model is correctly specified. The two stages are developed in more detail below.

In the first stage, the drug exposure $D_{i,t}^T$ is modeled as a function of past exposure $\bar{D}_{i,t-1}^T$ and past confounders $\bar{X}_{i,t-1}$ until visit $t-1$, by means of a conditional mean model:

$$(2) \quad E(D_{i,t}^T | \bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T) = a(\bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T, t; \alpha)$$

for each time t . Here, $a(\bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T, t; \alpha)$ is a known function of the confounder, drug exposure history, time, and a finite-dimensional parameter α . Because model (2) is a so-called sequential conditional mean model which conditions on different covariate sets at different times, GEE with independence working correlation (Vansteelandt 2007, Pepe and Anderson 1994) must be used for estimation (that is, the correlation between repeated observations from the same subject is ignored for estimation); see (Vansteelandt 2007, Chamberlain 1992) for more efficient estimation strategies.

Once model (2) has been fitted, it can be used to calculate $E(Z_{i,t}^T | \bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T)$. When $Z_{i,t}^T$ is not a linear function of $D_{i,t}^T$, then additional expectations must be assessed; e.g. when $Z_{i,t}^T$ is linear in $D_{i,t}^T$ and $D_{i,t}^{T2}$, one needs an additional model for $E(D_{i,t}^{T2} | \bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T)$.

In the second stage, in order to link the potential outcomes in (1) to the observed outcome, one must rely on the consistency assumption (Robins 1997) that $Y_{i,t}(\bar{D}_{i,t}^T)$ equals the observed outcome $Y_{i,t}$ if patient i has an active drug exposure history equal to $\bar{D}_{i,t}^T$. The causal model corresponding to the short-term effect in the first equation of (1) may then be fitted using a model of the form:

$$(3) \quad E(Y_{i,t} | \bar{X}_{i,t-1}, \bar{D}_{i,t}^T) = b(\bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T, t; \delta) + \beta_{st} \{Z_{i,t}^T - E(Z_{i,t}^T | \bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T)\}$$

where $b(\bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T, t; \delta)$ is a known function of confounder and drug exposure history up to time $t-1$, of time t and of an unknown finite-dimensional parameter δ , and where $E(Z_{i,t}^T | \bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T)$ can be substituted with an estimate based on the first stage. Because model (3) is also a sequential conditional mean model, GEE with independence correlation must be used for estimation. Confidence intervals are based on the bias-corrected, accelerated bootstrap with 10,000 resamples (Efron and Tibshirani 1993).

3.3.2 Long-term effects

The estimation of the long-term effect of drug exposure on the clinical outcome proceeds sequentially by assessing at each time t the effect of drug exposure during the previous inter-visit interval $(t-1, t-2, \dots)$ on the outcome. The first step is described in this section; further sequential steps materialize analogously.

First, the causal effect of past exposure $D_{i,t-1}^T$ on the outcome $Y_{i,t}$ is isolated by estimating its effect on the potential outcome $Y_{i,t}(\bar{D}_{i,t-1}^T, 0)$ that would have been observed at time t if the corresponding patient were not exposed to drug during the t th inter-visit interval. As this potential outcome is not observed, we first predict it by subtracting the causal effect of the observed drug exposure during the t th inter-visit interval from the outcome $Y_{i,t}$, i.e.

$$(4) \quad \hat{Y}_{i,t}(\bar{D}_{i,t-1}^T, 0) = Y_{i,t} - \hat{\beta}_{st} Z_{i,t}^T.$$

Next, the analysis strategy for the short-term effect can be applied to this predicted potential outcome. That is, the long-term effect in the second equation of (1) can be estimated by regressing $\hat{Y}_{i,t}(\bar{D}_{i,t-1}^T, 0)$ on the residuals $Z_{i,t-1}^T - E(Z_{i,t-1}^T | \bar{X}_{i,t-2}, \bar{D}_{i,t-2}^T)$ from the exposure model (2), as well as the drug exposure and confounder history, using GEE with independence correlation:

$$(5) \quad E \left\{ \hat{Y}_{i,t}(\bar{D}_{i,t-1}^T, 0) | \bar{X}_{i,t-2}, \bar{D}_{i,t-1}^T \right\} \\ = c(\bar{X}_{i,t-2}, \bar{D}_{i,t-2}, t; \delta_{lg}) \\ + \beta_{lg,1} \{ Z_{i,t-1}^T - E(Z_{i,t-1}^T | \bar{X}_{i,t-2}, \bar{D}_{i,t-2}^T) \}$$

where $c_1(\bar{X}_{i,t-2}, \bar{D}_{i,t-2}, t; \delta_{lg})$ is a known function of confounder and drug exposure history up to time $t-2$, as well as time t and an unknown finite-dimensional parameter δ_{lg} .

Because the data analysis revealed no significant evidence of a long-term effect $\beta_{lg,1}$ differing from zero, we did not attempt to estimate the long-term effects of exposures administered during earlier intervisit intervals on the VL at time t . Note, however, that if one were willing to assume that the long-term effects of drug exposures administered during all previous intervisit intervals $(t-1, t-2, \dots)$ are the same, then more efficient inferences could be drawn. This would require solving a linear combination of the estimating equations corresponding to each of these long-term effects (Robins 1994).

4. RESULTS

In this section, we will report the results from the application of the SNMM to the data described in Section 2. The clinical outcome ($Y_{i,t}$) is the log10 VL observed after treatment initiation among antiretroviral-naïve patients. Drug exposure ($D_{i,t}^T$) is the average internal drug exposure estimated as described in subsection 2.2 during the t th inter-visit interval. Baseline covariates $X_{i,0}$ include gender, age and ethnicity (Asian; Hispanic; Black; Caucasian and Others). Finally the presence and number of gastro-intestinal AEs during the $(t-1)$ th inter-visit is considered as a time-varying confounder $X_{i,t-1}$ (in addition to the drug exposure and VL history).

The emphasis in this analysis is on the initial phase of antiretroviral treatment during which VL decreases until reaching for the first time the desired limit of quantification for the VL assay (i.e., 50 copies/ml). Because the length of this phase varies for different patients, $X_{i,t-1}$ includes the constraint history $\bar{C}_{i,t-1}$, which encodes the following set of constraints at each time t : $VL_{i,1} > 50 \ \& \ VL_{i,2} > 50 \ \& \dots \ \& \ VL_{i,t-1} > 50$. Integrating these constraints in the model raises further difficulties that will be addressed in Section 4.2.

A total of 348 inter-visit intervals are available across the 107 patients considered in this analysis.

4.1 Result for short-term effects

4.1.1 First stage: Exposure Model

The exposure model (2) relates drug exposure during the t th inter-visit interval to the drug exposure history $\bar{D}_{i,t-1}^T$ up to visit $t-1$, the confounder history $\bar{X}_{i,t-1}$ up to visit $t-1$, and the time t on the subset of patients satisfying the constraint $\bar{C}_{i,t-1}$. Specifically, we consider the following set of equations:

$$(6) \quad a(\bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T, \bar{C}_{i,t-1}, t; \alpha) = \alpha_0 + \alpha_1 D_{i,t-1}^T \\ + \alpha_2 \sum_{k=1}^{t-2} D_{i,k}^T + \alpha_3 Y_{i,t-1} + \alpha_4 Y_{i,t-2} + \alpha_5 AE_{i,t-1} \\ + \alpha_6 AEnum_{i,t-1} + \alpha_7 gender + \sum_{k=8}^{11} \alpha_k ethn_{k-5} \\ + \alpha_{12} age + \sum_{k=13}^{16} \alpha_k I(t = k - 10)$$

with $\sum_{k=1}^{t-2} D_{i,k}^T$ being the sum of past drug exposures estimated on inter-visit intervals preceding the $(t-1)$ th inter-visit interval, $AE_{i,t-1}$ ($AEnum_{i,t-1}$) the presence (number) of adverse events during the $(t-1)$ th inter-visit interval, $gender$ the dummy variable coding 1 for men and 0 for women, $ethn_{k-5}$ with $6 \leq k \leq 9$ being binary (1/0) dummy

Table 1. Average log10 VL level and drug exposure at each visit

visit (weeks)	V1 (4)	V2 (8)	V3 (16)	V4 (24)	V5 (32)
number of patients considered	107	96	79	45	21
$E(Y_{i,t-1} C_{i,t-1})$ ($\log_{10}(\text{cp/ml})$)	4.9	2.9	2.5	2.2	2.1
$E(D_{i,t} C_{i,t-1})$ (ng/ml)	7259.4	7408.5	7363.9	7163.4	7626.7

Table 2. Causal estimates with bias-corrected, accelerated bootstrap confidence intervals

		Estimate	95% CI
Short-term effect	effect of internal drug exposure $D_{i,t}^T$	$-3.50 \cdot 10^{-5}$	$[-10.13 \cdot 10^{-5}; -0.33 \cdot 10^{-5}]$
	effect of the interaction between internal drug exposure and $\log_{10}(\text{VL})$ at the previous visit	$-7.65 \cdot 10^{-5}$	$[-18.49 \cdot 10^{-5}; -0.84 \cdot 10^{-5}]$
Long-term effect	effect of internal drug exposure $D_{i,t-1}^T$	$7.50 \cdot 10^{-6}$	$[-3.51 \cdot 10^{-5}; 6.66 \cdot 10^{-5}]$

variables coding Asian in $ethn_1$, Hispanic in $ethn_2$; Black in $ethn_3$ and others than the previous ethnicities or Caucasians in $ethn_4$, age being the age of the patient and $I(t = k - 10)$ with $11 \leq k \leq 14$ encoding visit indicators.

Based on independence GEE parameter estimates, the factors which are most significantly associated with a decrease in the current drug exposure $D_{i,t}^T$ include a decrease in the previous internal exposure $D_{i,t-1}^T$, a decrease in the sum of previous internal exposures estimated on inter-visit intervals prior to the $(t - 1)$ th inter-visit interval, time and ethnicity. To alleviate concerns about model misspecification and model uncertainty, and because G-estimation can endure some over-fitting in the exposure model, this model was not reduced using model selection strategies; see Figure 2 in appendix for a goodness of fit assessment.

4.1.2 Second stage: Causal Model

The second stage of the approach consists of modeling log10 VL as a function of the drug exposure residuals coming from exposure model $Z_{i,t}^T - E(Z_{i,t}^T | \bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T, \bar{C}_{i,t-1})$ where $Z_{i,t}^T = D_{i,t}^T(1, Y_{i,t-1} - E(Y_{i,t-1} | \bar{C}_{i,t-1}))$ as well as the drug exposure and confounder history, but restricted to the subset of patients satisfying constraints $\bar{C}_{i,t}$. The selection of the variables interacting with drug exposure $D_{i,t}^T$ and thus defining $Z_{i,t}$ is based on a backward selection procedure starting from the most complete set of variables including log10 VL at the previous visit $t - 1$, log10 VL at the previous visit $t - 2$, visit indicators, and the baseline covariates age, gender and race. At each step of the selection procedure the least significant interaction was removed from the model based on its p-value and confirmed by its 95% bias corrected accelerated bootstrap confidence interval. Specifically, we let the function $b(\bar{X}_{i,t-1}, \bar{Y}_{i,t-1}, \bar{C}_{i,t-1}, t; \delta)$ of the causal model (3) include the same terms as those used in the exposure model (6), as follows:

$$\begin{aligned}
 E(Y_{i,t} | \bar{X}_{i,t-1}, \bar{D}_{i,t}^T, \bar{C}_{i,t-1}) &= \delta_0 + \delta_1 D_{i,t-1}^T \\
 &+ \delta_2 \sum_{k=1}^{t-2} D_{i,k}^T + \delta_3 \{Y_{i,t-1} - E(Y_{i,t-1} | \bar{C}_{i,t-1})\} \\
 &+ \delta_4 Y_{i,t-2} + \delta_5 AE_{i,t-1} + \delta_6 AEnum_{i,t-1} \\
 &+ \delta_7 gender + \sum_{k=8}^{11} \delta_k ethn_{k-7} + \delta_{12} age \\
 &+ \sum_{k=13}^{16} \delta_k I(t = k - 12) \\
 &+ \beta_{1,st} \{D_{i,t}^T - E(D_{i,t}^T | \bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T, \bar{C}_{i,t-1})\} \\
 &+ \beta_{2,st} \{D_{i,t}^T - E(D_{i,t}^T | \bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T, \bar{C}_{i,t-1})\} \\
 &\{Y_{i,t-1} - E(Y_{i,t-1} | \bar{C}_{i,t-1})\}
 \end{aligned}$$

where $E(Y_{i,t-1} | \bar{C}_{i,t-1})$ is the average log10 VL estimated at each visit $t - 1$ on the subset of patients satisfying constraints; $E(Y_{i,t-1} | \bar{C}_{i,t-1})$ is given in Table 1 for each visit.

The two parameters $\beta_{1,st}$ and $\beta_{2,st}$ are the causal parameters modeling the effect of internal drug exposure $D_{i,t}^T$ and the interaction between internal drug exposure and log10 VL at the previous visit $t - 1$. These were estimated (see Table 2) to be $\hat{\beta}_{1,st} = -3.5 \cdot 10^{-5}$ (95% bias-corrected, accelerated bootstrap confidence interval $[-10.13 \cdot 10^{-5}; -0.33 \cdot 10^{-5}]$) and $\hat{\beta}_{2,st} = -7.65 \cdot 10^{-5}$ (95% bias-corrected, accelerated bootstrap confidence interval $[-18.49 \cdot 10^{-5}; -0.84 \cdot 10^{-5}]$). Quadratic internal drug exposure effect was also included but not kept as non significant. The remaining parameter estimates of model (7) are not shown in Table 2 because they are not directly causally interpretable, but are only included to ensure consistent estimation of the causal parameters $\beta_{1,st}$ and $\beta_{2,st}$ even when the exposure model is misspecified, provided that model (7) is correctly specified; see Figure 3 in appendix for a goodness of fit assessment. Note that by their double robust nature, the reported parameter estimates are also valid when the remaining part of

Table 3. Geometric mean reduction (%) in VL that would arise if the average internal exposure during the preceding inter-visit were reduced from the ρ th% percentile (P_ρ) of the average internal exposure to zero, among patients satisfying the constraints

Difference between observed and average $Log_{10}(VL)$ at previous visit		Inter-visit average internal exposure (ng/ml)				
		P5 (4555)	P25 (6290)	P50: Population median (6940)	P75 (8040)	P95 (11243)
Higher	1	69%	80%	83%	87%	94%
.	0.5	54%	65%	69%	74%	85%
.	0.25	43%	54%	58%	63%	75%
.	0	31%	40%	43%	48%	60%
Lower	-0.25	15%	21%	22%	26%	34%

model (7) is misspecified, provided that the exposure models are correctly specified.

Due to the \log_{10} transformation of the VL, the interpretation of the results of the causal model (given in Table 2) is most easy upon raising all parameter estimates to the power 10. These back-transformed estimates are shown in Table 3, in terms of the geometric mean reduction (in %) in VL that can be causally attributed to the drug. This reduction is a function of the average inter-visit internal exposure on the one hand, and on the other hand, the VL at the previous visit. For example, when a patient initiates an inter-visit period with a VL equal to the population visit average (geometric mean), and when his/her internal exposure during that inter-visit period is equal to the population median value (P50), one can expect that the VL at the visit scheduled at the end of the period is 43% (95%CI [6%–81%]) lower than it would have been if he/she had not taken the drug in the last inter-visit.

4.2 Result for long-term effects

The analysis of long-term effects is complicated by the presence of constraints. For instance, the model for the effect of exposure during the $(t-1)$ th inter-visit interval on the outcome at time t is restricted to those who satisfy $VL_{i,1} > 50 \& \dots \& VL_{i,t-2} > 50$. The analysis of short-term effects focuses on the subset of patients who additionally have $VL_{i,t-1} > 50$. For the latter subgroup, the previous analysis of short-term effects in model (7) enables a prediction on the potential outcome $Y_{i,t}(\bar{D}_{i,t-1}^T, 0)$ as

$$(8) \quad \hat{Y}_{i,t}(\bar{D}_{i,t-1}^T, 0) = Y_{i,t} - \hat{\beta}_{1,st} D_{i,t}^T - \hat{\beta}_{2,st} D_{i,t}^T (Y_{i,t-1} - E(Y_{i,t-1} | \bar{C}_{i,t-1}))$$

In order to have a prediction for the remaining patients, an additional analysis of short-term effects for those patients that satisfy the set of constraints $(\bar{C}_{i,t-2}, -C_{i,t-1}) = (VL_{i,1} > 50 \& \dots \& VL_{i,t-2} > 50 \& VL_{i,t-1} \leq 50)$ must additionally be provided. The principle for obtaining the short-term effect for this subgroup is the same as in the previous study of short-term effects, and is detailed below.

First, an exposure model is fitted on the new subgroup of patients i.e. those achieving VL below the lower limit of quantification of the VL assay (i.e., 50copies/ml) at visit $t-1$:

(9)

$$E(D_{i,t}^T | \bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T, \bar{C}_{i,t-2}, -C_{i,t-1}) = \gamma_0 + \gamma_1 D_{i,t-1}^T + \gamma_2 \sum_{k=1}^{t-2} D_{i,k}^T + \gamma_3 Y_{i,t-2} + \gamma_4 AE_{i,t-1} + \gamma_5 AEnum_{i,t-1} + \gamma_6 gender + \sum_{k=7}^{10} \gamma_k ethn_{k-6} + \gamma_{11} age + \sum_{k=12}^{15} \gamma_k I(t = k - 11)$$

As before, we chose not to simplify this model any further. We subsequently fit the following model:

(10)

$$E(Y_{i,t} | \bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T, \bar{C}_{i,t-2}, -C_{i,t-1}) = \tau_0 + \tau_1 D_{i,t-1}^T + \tau_2 \sum_{k=1}^{t-2} D_{i,k}^T + \tau_3 \{Y_{i,t-2} - E(Y_{i,t-2} | \bar{C}_{i,t-2}, -C_{i,t-1})\} + \tau_4 AE_{i,t-1} + \tau_5 AEnum_{i,t-1} + \tau_6 gender + \sum_{k=7}^{10} \tau_k ethn_{k-6} + \tau_{11} age + \sum_{k=12}^{15} \tau_k I(t = k - 11) + \beta_{1,st2} \{D_{i,t}^T - E(D_{i,t}^T | \bar{X}_{i,t-1}, \bar{D}_{i,t-1}^T, \bar{C}_{i,t-2}, -C_{i,t-1})\}$$

The parameter $\beta_{1,st2}$ was estimated as $4.12 \cdot 10^{-6}$ (95% bias-corrected, accelerated bootstrap confidence interval $[-1.05 \cdot 10^{-5}; 2.06 \cdot 10^{-5}]$) and was not found to be different from zero at the 5% significance level.

Nevertheless, an exposure effect was retained in this model despite its non-significance as the primary interest lies in it. The predicted potential outcome in the corresponding subgroup of patients is now obtained from model (10) by removing the estimated causal short-term effect of internal drug exposure from the observed outcome:

$$(11) \quad \hat{Y}_{i,t}(\bar{D}_{i,t-1}^T, 0) = Y_{i,t} - \hat{\beta}_{1,st2} D_{i,t}^T$$

Now that the potential outcome $Y_{i,t}(\bar{D}_{i,t-1}^T, 0)$ for all patients whose VL has not dropped below the lower limit of quantification of the VL assay (i.e., 50 copies/ml) by visit $t-1$ has been predicted, the long-term effect can be studied. This is done by merging the predicted potential outcomes and evaluating them in the following model:

$$(12) \quad E \left\{ \hat{Y}_{i,t}(\bar{D}_{i,t-1}^T, 0) | \bar{X}_{i,t-2}, \bar{D}_{i,t-1}^T, \bar{C}_{i,t-2} \right\} = \delta_{0,lg} + \delta_{1,lg} D_{i,t-2}^T + \delta_{2,lg} \sum_{k=1}^{t-3} D_{i,k}^T + \delta_{3,lg} \{Y_{i,t-2} - E(Y_{i,t-2} | \bar{C}_{i,t-2})\} + \delta_{4,lg} \{Y_{i,t-3} - E(Y_{i,t-3} | \bar{C}_{i,t-2})\} + \delta_{5,lg} AE_{i,t-2} + \delta_{6,lg} AEnum_{i,t-2} + \delta_{7,lg} gender + \sum_{k=8}^{11} \delta_{k,lg} ethn_{k-7} + \delta_{12,lg} age + \sum_{k=13}^{16} \delta_{k,lg} I(t = k - 12) + \beta_{lg,1} \{D_{i,t-1}^T - E(D_{i,t-1}^T | \bar{X}_{i,t-2}, \bar{D}_{i,t-2}^T, \bar{C}_{i,t-2})\}$$

The parameter $\beta_{1,lg}$ is estimated as $7.5 \cdot 10^{-6}$ (95% bias-corrected, accelerated bootstrap confidence interval $[-3.51 \cdot 10^{-5}; 6.66 \cdot 10^{-5}]$) (see Table 2), suggesting no significant evidence of a relation between the log10 VL and the average internal exposure during the preceding inter-visit interval. As for the short-term effect, quadratic internal drug exposure effect was also included but not kept as non significant. See Figure 4 in appendix for a goodness of fit assessment.

5. DISCUSSION

In this paper, we investigated the effects of LPV/r on the initial decrease in VL in a group of antiretroviral naïve HIV-1 infected patients who were randomized to receive once daily LPV/r dosing within a triple-drug antiretroviral regimen for a period of up to 96 weeks. The presence of time-varying confounders may result in a classical longitudinal regression method yielding possibly severely biased estimates of the exposure-effect when some of these time-varying confounders are themselves affected by earlier exposures (Robins 1997). SNMMs were used to overcome this problem under the assumption that all confounders for the association between exposure at each visit and future outcomes were available.

The parameter estimation strategy adopted in this analysis involved separating, for each visit t , the study of the direct effect of exposure measured on the t th inter-visit on

the directly following VL at visit t (called the short-term effect) and the study of effects of drug exposures preceding the t th inter-visit interval on the VL at visit t (called long-term effect). In both the short- and long-term effects, parameters were estimated using a relatively simple two-stage procedure, which can be applied using standard statistical software. In the first stage, a model was fitted for drug exposure as a function of the preceding drug exposure, time-varying confounders and baseline covariates. In the second stage, a causal model for the effect of drug exposure on VL was fitted by regressing a prediction of the potential exposure-free outcome on residuals obtained from the exposure model as well as previous drug exposure, time-varying confounders and baseline covariates. To acknowledge their sequential nature, parameters in each of the two models were estimated by solving GEE with independence correlation (Vansteelandt 2007, Pepe and Anderson 1994) and confidence intervals were derived from the use of the bias-corrected accelerated bootstrap analysis (Efron and Tibshirani 1993). While our analysis relies on many working models, only some of them need to be correctly specified for consistent estimation. In particular, the obtained estimators of the SNMM have the desirable property of being consistent and asymptotically normally distributed as soon as, at each time, one of the two models (exposure or outcome regression model) is correctly specified (Robins 1997).

As the analysis focuses on the decrease in VL, constraints on previous VLs were imposed and integrated into the SNMM. These constraints have no particular impact on the estimation procedure for the short-term effect, but required further development for the long-term effect. The commonly used framework of marginal structural models (Robins 1999) could not directly be applied in this setting as it would require adaptations 1) to model a continuous exposure variable and 2) to take into account that the relevant set of patients changes over time according to each patient's specific length of initial phase of treatment until viral control is reached.

The causal model for the average concentration on the VL obtained at the end of each inter-visit interval shows a substantial and statistically significant reduction in VL that can be attributed to the internal exposure to LPV that a patient achieves during the initial treatment period until a patient's viral control is achieved. For each visit t , the effect of drug exposures preceding inter-visit interval t on VL measured at visit t were investigated and found to be nonsignificant.

To conclude, our final causal model shows several noteworthy features. In the early phase of treatment in this cohort of antiretroviral-naïve patients for whom viral resistance was not captured for the majority of subjects, when the initial viral load is high, its' subsequent reduction can be as much as 94% within the next inter-visit interval. As depicted in Table 3, past VL has a strong influence on the

short-term effect in the following sense: the higher the previous VL, the greater the reduction created by different levels of internal exposure to LPV/r.

The second significant variable influencing VL response was the patient's variable exposure to LPV/r. The chief source of variability in internal exposure arises from the patient's variable adherence to the prescribed dosing regimen for the prescribed medication in question - LPV/r-based regimen used in this clinical trial. In Table 3, the relatively small changes in mean VL reduction (%) across the patient's variable internal drug exposure to the combination product reflects the product's ability to maintain its therapeutic action in the face of varying levels of patient adherence to the prescribed medicine's recommended dosing regimen.

ACKNOWLEDGEMENTS

We would like to thank Dr John Urquhart for useful discussions and revisions of the manuscript. The second author acknowledges support from IAP research network grant nr. P06/03 from the Belgian government (Belgian Science Policy).

APPENDIX

While at each time, the working models that we adopted are linear regression models and thus in principle testable by the usual goodness-of-fit techniques, we did not include formal tests in the manuscript because the number of tests is too abundant. We instead present residuals versus fitted values plots in this appendix. Note that separate goodness-of-fit assessments are needed at separate times because the models corresponding to different time points condition on different covariate sets.

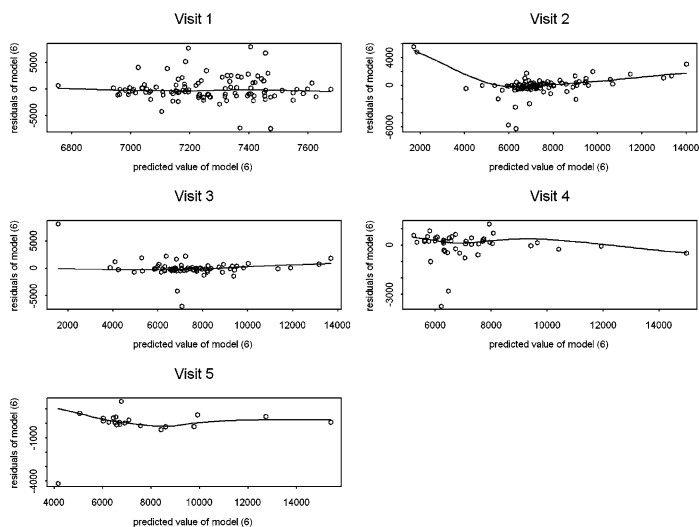


Figure 2. Residuals versus fitted values of exposure model (6).

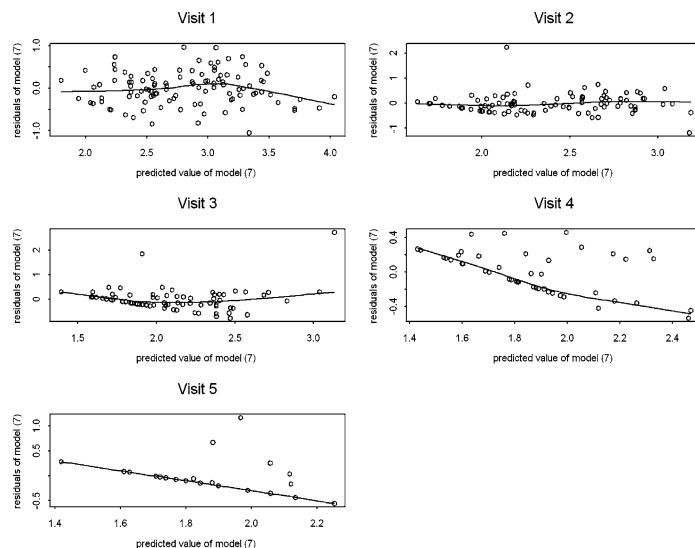


Figure 3. Residuals versus fitted values of causal model (7) estimating short term effect of exposure.

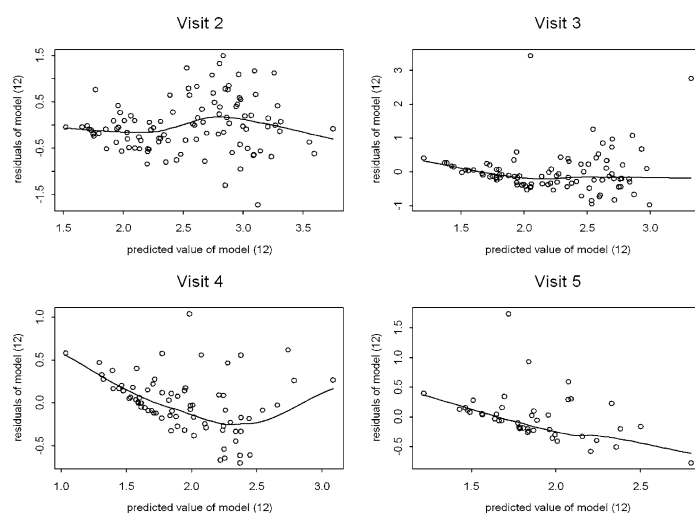


Figure 4. Residuals versus fitted values of causal model (12) estimating long term effect of exposure.

Received 26 November 2010

REFERENCES

- CHAMBERLAIN, G. (1992). Sequential moment restrictions in panel data - comment. *Journal of Business & Economic Statistics* **10** 20.
- EFRON, B. AND TIBSHIRANI, R.J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, San Francisco. [MR1270903](#)
- GOETGHEBEUR, E. AND POCKOCK, S. (1993). Statistical issues in allowing for noncompliance and withdrawal. *Drug Inform J* **27** 837-845.
- HARTER J.G. AND PECK C.C. (1991). Chronobiology. Suggestions for integrating it into drug development. *Ann N Y Acad Sci* **618** 563-571.
- LEE, J., ELLENBERG, J., HIRTZ, D. AND NELSON, K. (1991). Analysis of clinical trials by treatment actually received: is it really an option? *Statist Med* **10** 1595-1605.

- LIANG, K.-Y. AND ZEGER, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. [MR0836430](#)
- MOLINA, J.M., PODSADECKI, T.J., JOHNSON, M.A., WILKIN, A., DOMINGO, P., MYERS, R., HAIRRELL, J.M., RODE, R.A., KING, M.S. AND HANNA, G.J. (2007). A lopinavir/ritonavir-based once-daily regimen results in better compliance and is non-inferior to a twice-daily regimen through 96 weeks. *AIDS Res Human Retroviruses* **223** 1505–1514.
- PATERSON, D.L., SWINDELLS, S., MOHR, J., BRESTER, M., VERGIS, E.M., SQUIER, C., WAGENER, M.M. AND SINGH, N. (2000). Adherence to protease inhibitor therapy and outcomes in patients with HIV infection. *Ann Intern Med.* **133** 21–30.
- PEPE, M. AND ANDERSON, G.L. (1994). A cautionary note on inference for marginal regression-models with longitudinal data and general correlated response data. *Commun Stat Simul Comput* **23** 939–951.
- ROBINS, J.M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Commun Stat Theory Meth* **23** 2379–2412. [MR1293185](#)
- ROBINS, J.M. (1997). Causal Inference from Complex Longitudinal Data. *Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics (120)*, M. Berkane, Editor. NY: Springer Verlag 1997 (pages 69–117). [MR1601279](#)
- ROBINS, J.M. (1999). Marginal structural models versus structural nested models as tools for causal inference. *Statistical Models in Epidemiology: The Environment and Clinical Trials* **116** 95–134. [MR1731682](#)
- VANHOVE, G.F., SCHAPIRO, J.M., WINTERS, M.A., MERIGAN, T.C. AND BLASCHKE, T.F. (1996). Patient compliance and drug failure in protease inhibitor monotherapy. *Ann Intern Med.* **276**(24) 1955–1956.
- VANSTEELENDT, S. (2007). On confounding, prediction and efficiency in the analysis of longitudinal and cross-sectional clustered data. *Scand J Statist* **34** 478–498. [MR2368794](#)
- VRIJENS, B. AND GOETGHEBEUR, E. (1997). Comparing compliance patterns between randomized treatments. *Control Clin Trials* **18**(3) 187–203.
- VRIJENS, B., TOUSSET, E., RODE, R., BERTZ, R., MAYER, S. AND URQUHART, J. (2005). Successful projection of the time course of drug concentration in plasma during a 1-year period from electronically compiled dosing-time data used as input to individually parameterized pharmacokinetic models. *J Clin Pharmacol* **45**(4) 461–467.
- Laetitia Comté
University of Liège
Liège, Belgium
Chaussée Colonel Joset 43
4630 Soumagne, Belgium
E-mail address: comte.laetitia@gmail.com
- Stijn Vansteelandt
Department of Applied Mathematics and Computer Science
Ghent University
Ghent, Belgium
E-mail address: stijn.vansteelandt@ugent.be
- Richard A. Rode
Abbott Laboratories
Dept. 436, Bldg. AP9A-2
100 Abbott Park Road
Abbott Park, IL 60064-6124, USA
E-mail address: Richard.A.Rode@abbott.com
- Bernard Vrijens
AARDEX Group
Rue des Cyclistes Frontière 24
4600 Visé, Belgium
E-mail address: bernard.vrijens@aardegroupp.com