Ignorance and uncertainty regions as inferential tools in a sensitivity analysis

IGNORANCE AND UNCERTAINTY REGIONS AS INFERENTIAL TOOLS IN A SENSITIVITY ANALYSIS

Stijn Vansteelandt, Els Goetghebeur, Michael G. Kenward and Geert Molenberghs

Ghent University, Harvard School of Public Health, London School of Hygiene and Tropical Medicine, and Limburgs Universitair Centrum

Abstract: It has long been recognised that most standard point estimators lean heavily on untestable assumptions when missing data have occurred. Statisticians have therefore advocated the use of sensitivity analysis, but payed relatively little attention to strategies for summarizing the results from such analyses, which have clear interpretation, verifiable properties and feasible implementation. As a step in this direction, several authors have proposed to shift the focus of inference from point estimators to estimated intervals or regions of ignorance. These regions combine standard point estimates obtained under all possible/plausible missing data models that yield identified parameters of interest. They thus reflect the achievable information from the given data generation structure with its missing data component. The standard framework of inference needs extension to allow for a transparent study of statistical properties of such regions.

In this paper we propose a definition of consistency for the region and introduce the concepts of pointwise, weak and strong coverage for larger regions which acknowledge sampling imprecision in addition to the structural lack of information. The larger regions are called uncertainty regions and quantify an overall level of information by adding imprecision due to sampling error to the estimated region of ignorance. The distinction between ignorance and sampling error is often useful, for instance when sample size considerations are made. The type of coverage required depends on the analysis goal. We provide algorithms for constructing several types of uncertainty regions and derive general relationships between the regions. Based on the estimated uncertainty regions, we show how classical hypothesis tests can be performed without untestable assumptions on the missingness mechanism.

Key words and phrases: Bounds; Identifiability; Incomplete data; Inference; Pattern-mixture Model; Selection Model.

1. Introduction

The problem of missing values has received due attention in the statisti-

cal literature for many years; over the past decade the nature of this work has changed appreciably. Previously, the main concern was the lack of balance induced in data sets by missing values which precluded simple methods of analysis. Recent advances in general statistical methodology and computational developments have greatly reduced this as an issue. The focus has shifted to the nature of inferences that can be drawn legitimately from incomplete data, and how these are bound up with assumptions about the unobserved data. Rubin (1976) and Little and Rubin (1987) provide much of the foundation for this debate. In particular, Rubin delineated those settings in which one could proceed to analyse incomplete data effectively as though they were incomplete by design. This distinction is central to the problem of missing data and rests on the probabilistic relationships between the observed data, the missing data and the random variable representing missingness.

We are concerned here with the situation in which data may be missing in a non-random fashion. That is, conditional on the observed data and covariates, there remains statistical dependence between a data point and the probability that it is missing. Analyses that assume the data are missing by design are then no longer generally valid. The lack of knowledge associated with the missing data now introduces an essential degree of ambiguity into statistical inference. We term this ambiguity 'ignorance' and distinguish it from familiar statistical imprecision, the consequence of random sampling. Our procedure will accommodate this by replacing point estimators by sets of points that estimate intervals or regions of ignorance. Each point in these sets is derived in the usual way from a different plausible model that is compatible with the observed data and yields identified parameters of interest. These sets are quite distinct from confidence regions that represent the statistical imprecision associated with a point estimate. In our approach such measures of sampling error must be added to the region of ignorance to obtain an overall region of uncertainty. These ideas of ignorance and uncertainty were introduced and illustrated in Goetghebeur, Molenberghs and Kenward (1999), Kenward, Molenberghs and Goetghebeur (2001) and Molenberghs, Kenward and Goetghebeur (2001). Related ideas have been formulated and/or used by e.g. Balke and Pearl (1997), Cochran (1977), Horowitz and Manski (2000), Imbens and Manski (2004), Joffe (2001), Nordheim (1984), Robins

(1989) and Scharfstein, Manski and Anthony (2004).

In this paper we develop a formal framework for the study of ignorance and uncertainty, illustrating concepts with a study of HIV prevalence in Kenya (presented in Section 2) where diagnostic test outcomes are incompletely observed. In Section 3 we introduce a formal definition for the region of ignorance. Having replaced conventional point estimators with intervals or regions of estimates, we develop 'classical' frequentist inference to handle standard concepts such as coverage and consistency in these new settings. In Section 4, we define pointwise coverage and consistency of a regional estimator when the target of inference is the unidentified true parameter value. We develop the combination of the estimated region of ignorance with statistical imprecision and derive estimators for the resulting pointwise uncertainty regions. We show how the level of classical hypothesis tests can be protected without untestable assumptions about the missing data mechanism. In Section 5, we define weak and strong coverage, and consistency of a regional estimator when the target of inference is the identified region of ignorance (that contains the true parameter value). We construct uncertainty regions designed to attain a given weak or strong coverage probability. Until Section 6 we impose no restrictions on the observed data law. In Section 6 we discuss the additional challenges that must be met when the observed data model is parametric or semiparametric.

2. Motivation

To motivate the problem setting, consider the following HIV surveillance study described in Verstraeten, Farah, Duchateau and Matu (1998). To evaluate the current situation of the HIV epidemic in Kenya, 787 blood samples were collected as part of the National AIDS Control Programme among pregnant women from rural and urbanised areas near Nairobi in 1996. 52 (699) HIV test results are positive (negative) and coded $Y = 1$ (0). 751 diagnostic test results are observed ($R = 1$) and 36 are missing ($R = 0$). Some sera were hemolysed and therefore produced inconclusive HIV test results; others were not available at the time of diagnostic testing. Under this setting, the observed data $(Y_i R_i, R_i)$ for subjects $i = 1, \ldots, N = 787$ can be regarded as N independent and identically distributed copies of random variables (YR, R).

When the dependence of missingness on the missing outcome is unknown to

the investigator, the pattern-mixture model

$$\begin{aligned} \text{pr}(R=1) &= \nu_0 \qquad (2.1) \\ \text{logit}\{\text{pr}(Y=1|R)\} &= \eta_0 + \gamma^*_{\text{pm}}(1-R) \qquad (2.2) \end{aligned}$$

with γ^*_{pm} known, and the selection model

$$\begin{aligned} \text{pr}(Y=1) &= \beta_0 \\ \text{logit}\{\text{pr}(R=1|Y)\} &= \delta_0 + \gamma_{\text{s}} Y \end{aligned}$$

with γ_{s} known, are nonparametric models for the observed data. With each choice of γ^*_{pm} (γ_{s}) thus corresponds a choice of (ν_0, η_0) ((β_0, δ_0)) that fits the observed data perfectly well so that different choices for γ^*_{pm} (γ_s) cannot be rejected by any statistical test. As a result, HIV risk β_0 cannot be identified from the observed data without unverifiable assumptions (note that $\beta_0 = \text{expit}(\eta_0)\nu_0 + \gamma_{\text{pm}}(1-\nu_0)$ and $\gamma_{\text{pm}} = \text{pr}(Y=1|R=0) = \text{expit}(\eta_0 + \gamma^*_{\text{pm}})$ under the pattern-mixture model defined by restrictions (2.1)-(2.2)). However, β_0 is identified once a value is chosen for γ^*_{pm} (γ_{s}). In line with the missing data literature, parameters like γ^*_{pm} (γ_{s}) that are not identified, but conditional on which the target parameter is identified, will be called 'sensitivity parameters' (see e.g. Molenberghs, Kenward and Goetghebeur 2001).

Since the observed data do not identify the sensitivity parameter, one should be reluctant to analyze the data under a single choice such as $\gamma_{\text{s}} = 0$. For this reason, it has become increasingly common to conduct sensitivity analyses which reveal how estimates for β_0 vary over different values for the sensitivity parameter (see e.g. Copas and Li 1997; Scharfstein, Robins and Rotnitzky 1999; Molenberghs, Kenward and Goetghebeur 2001; Verbeke, Molenberghs, Thijs, Lesaffre and Kenward 2001). Figure 1 shows the varying risk estimates for the Kenyan HIV study. The missing at random (MAR) assumption (Rubin 1976) corresponds to $\gamma_{\text{s}} = 0$ and is itself consistent with a relatively low HIV risk estimate of 0.069. Larger risk estimates would occur if HIV positives were least likely to respond (i.e. $\gamma_{\text{s}} < 0$).

Figure 1 about here.

While graphical displays like Figure 1 are the most suitable tools in any

sensitivity analysis, they prohibit concise reporting of results, especially when, as usual, many unknown parameters are of interest. One natural and simple strategy for summarizing the results of a sensitivity analysis is to report, besides the usual analysis results obtained under a sole plausible missing data asssumption (e.g. MAR), the range of estimates for β_0 corresponding to a plausible range of values for the sensitivity parameter. We call such range of estimates an Honestly Estimated Ignorance Region (HEIR) for the target parameter because it expresses ignorance due to the missing data. Extreme application of this philosophy has lead to reporting worst case-best case intervals in a number of applications (see e.g. Cochran 1977; Nordheim 1984; Robins 1989; Kooreman 1993; Horowitz and Manski 2000; Balke and Pearl 1997; Molenberghs, Kenward and Goetghebeur 2001). These involve no untestable assumptions about the missing data but have debatable merits because they are often extremely wide. The approach taken here and by others (e.g. Scharfstein, Manski and Anthony 2004) is less extreme because we allow for untestable assumptions (namely that the sensitivity parameter lies within a chosen range) up to a chosen degree in order to obtain narrower and more plausible ranges of estimates. While this procedure is partly subjective, this is inherent to the problem as some untestable assumptions are (usually) unavoidable in any sensitivity analysis (e.g. even graphical displays like Figure 1 can often only be produced for a limited range of values for the sensitivity parameter and their interpretation thus necessarily involves untestable assumptions). Furthermore, reporting estimates that correspond to a range of values instead of a single value for the sensitivity parameter will always be superior, in the sense that it is less sensitive to untestable assumptions. This will be further discussed in Section 7.

Regions of estimates (HEIRs) instead of point estimates have been reported and proved useful in a number of applications. They may be obtained directly from graphical displays like Figure 1 using methods for sensitivity analysis as e.g. described in Scharfstein, Robins and Rotnitzky (1999), or be constructed more rapidly using specialized algorithms or computations (see e.g. Balke and Pearl 1997; Horowitz, Manski, Ponomareva and Stoye 2003; Kooreman 1993; Robins 1989; Vansteelandt and Goetghebeur, 2001). Nonetheless, their frequentist properties have received very little attention so far, with notable exceptions (Horowitz

and Manski, 2001; Imbens and Manski, 2004). The goal of this paper is therefore to examine how one can account for sampling variability on HEIRs and what it takes to be a good estimated region of ignorance. To enable rigorous study, we start by formally defining HEIRs in the next section.

3. Formal setting

Consider a study setting where an $m \times 1$ vector variable $\mathbf{L}_i$ is to be measured on units $i = 1, \ldots, N$, e.g. $\mathbf{L}_i$ may contain a primary outcome and baseline covariates. As the entire vector $\mathbf{L}_i$ may be missing, we observe instead N independent and identically distributed copies $\mathbf{O}_i = (\mathbf{R}_i, \mathbf{L}_{i(R_i)})$ of the observed data vector $\mathbf{O} = (\mathbf{R}, \mathbf{L}_{(R)})$. Here, $\mathbf{R}$ is an $m \times 1$ vector whose tth element, $t = 1, ..., m$, equals 1 if the tth component L_t of $\mathbf{L}$ is observed and 0 otherwise, and $\mathbf{L}_{(R)}$ denotes the observed part of $\mathbf{L}$ (according to the observed response indicator $\mathbf{R}$). We denote the true distribution of the full data $(\mathbf{L}, \mathbf{R})$ by $f_0(\mathbf{L}, \mathbf{R})$.

Suppose for now (and until Section 6) that we impose no restrictions on the full data distribution $f(\mathbf{L}, \mathbf{R})$. Our goal is then to draw inference on a vector functional $\boldsymbol{\beta}_0 = \boldsymbol{\beta}\{f_0(\mathbf{L})\} \in \mathbb{R}^p$ (e.g. the mean) of the true complete data distribution $f_0(\mathbf{L}) = \int f_0(\mathbf{L}, \mathbf{R}) d\mathbf{R}$. This is challenging when there are missing data, because several full data laws $f(\mathbf{L}, \mathbf{R})$ may marginalize to the true observed data law

$$f_0(\mathbf{O}) \quad = \quad \int f_0(\mathbf{L}, \mathbf{R}) d\mathbf{L}_{(1-R)} = \int f(\mathbf{L}, \mathbf{R}) d\mathbf{L}_{(1-R)} \tag{3.1}$$

where $\mathbf{L}_{(1-R)}$ denotes the missing part of $\mathbf{L}$ (according to the observed response indicator $\mathbf{R}$). Different examples of such laws $f(\mathbf{L}, \mathbf{R})$ cannot be distinguished based on realizations from the observed data law. Nevertheless, they may imply different values for the parameter of interest $\boldsymbol{\beta} = \boldsymbol{\beta}\{f(\mathbf{L})\}$, where $f(\mathbf{L}) = \int f(\mathbf{L}, \mathbf{R}) d\mathbf{R}$, in which case the observed data do not identify $\boldsymbol{\beta}_0$.

In view of this identification problem, we follow ideas in Robins (1997) by defining a class $\mathcal{M}(\boldsymbol{\gamma})$ of full data laws, indexed by some vector parameter $\boldsymbol{\gamma}$, to be nonparametric identified (NPI) if for each observed data law $f(\mathbf{O})$, there exists a unique law $f(\mathbf{L}, \mathbf{R}; \boldsymbol{\gamma})$ in the class $\mathcal{M}(\boldsymbol{\gamma})$ such that $f(\mathbf{O})$ is the marginal distribution of $\mathbf{O}$ according to the joint law $f(\mathbf{L}, \mathbf{R}; \boldsymbol{\gamma})$; that is, $f(\mathbf{O}) = \int f(\mathbf{L}, \mathbf{R}; \boldsymbol{\gamma}) d\mathbf{L}_{(1-R)}$. In Section 2 for example, $L = Y$ and each possible value for $\gamma = \gamma_{\text{pm}} \in [0, 1]$ characterizes a single class $\mathcal{M}(\gamma)$ of full data laws defined by restrictions (2.1)-(2.2) for the given γ. For each $\gamma \in [0, 1]$, this class $\mathcal{M}(\gamma)$

contains a unique law that marginalizes to the observed data law. In line with our previous definition, we will call the parameter $\boldsymbol{\gamma}$ indexing the models $\mathcal{M}(\boldsymbol{\gamma})$ a sensitivity parameter.

It follows from the definition that $\boldsymbol{\beta}_0$ is uniquely identified from the observed data law under each NPI model $\mathcal{M}(\boldsymbol{\gamma})$. Furthermore, the observed data cannot distinguish different models $\mathcal{M}(\boldsymbol{\gamma})$ (corresponding to different $\boldsymbol{\gamma}$-values). Suppose however that we have some information about the mechanism leading to the outcomes being missing, that enables us to restrict the class of full data laws to those classes $\mathcal{M}(\boldsymbol{\gamma})$ for which $\boldsymbol{\gamma}$ lives in a chosen set Γ; e.g. to consider the model defined by restrictions (2.1)-(2.2) with $\gamma \in [0, 0.25]$. Then our primary goal is to draw inference for $\boldsymbol{\beta}_0$ under the union model $\mathcal{M}(\Gamma) = \cup_{\gamma\in\Gamma}\mathcal{M}(\gamma)$, assuming that the true value $\boldsymbol{\gamma}_0$ of $\boldsymbol{\gamma}$ lies in Γ.

Because $\boldsymbol{\beta}_0$ is not generally identified from the observed data law under model $\mathcal{M}(\Gamma)$, a whole region of values

$$\begin{aligned} \text{ir}(\boldsymbol{\beta}, \Gamma) \;\; = \;\; \Big\{ & \boldsymbol{\beta}\{f(\mathbf{L})\} \;:\; f(\mathbf{L}) = \int f(\mathbf{L}, \mathbf{R}) d\mathbf{R} \text{ with } f(\mathbf{L}, \mathbf{R}) \in \mathcal{M}(\Gamma) \\ & \text{satisfying } f_0(\mathbf{O}) = \int f(\mathbf{L}, \mathbf{R}) d\mathbf{L}_{(1-R)} \Big\} \end{aligned} \tag{3.2}$$

rather than a single point value for $\boldsymbol{\beta}$, is typically consistent with the observed data law. Extending ideas in Molenberghs, Kenward and Goetghebeur (2001), this region $\text{ir}(\boldsymbol{\beta}, \Gamma)$ will be called the ignorance region for $\boldsymbol{\beta}$. We call an estimator of this set an Honestly Estimated Ignorance Region (HEIR) for $\boldsymbol{\beta}_0$ and view it as an estimate for $\boldsymbol{\beta}_0$ under model $\mathcal{M}(\Gamma)$.

4. Inference for $\boldsymbol{\beta}_0$

In studying the frequentist properties of HEIRs, we first take the viewpoint that the unidentified estimand $\boldsymbol{\beta}_0$ (as opposed to the identified estimand $\text{ir}(\boldsymbol{\beta}, \Gamma)$) is the target of inference under model $\mathcal{M}(\Gamma)$. Our goal is then to construct an appropriate concept of weak consistency for HEIRs and $(1-\alpha)100\%$ uncertainty regions that cover $\boldsymbol{\beta}_0$ with at least $(1-\alpha)100\%$ chance under this model.

4.1. Sampling Variability: Pointwise Coverage

The HEIR inherits variability from the sample of data. This is most easily explored through the parameter $\boldsymbol{\beta}(\boldsymbol{\gamma}) \equiv \boldsymbol{\beta}\{f(\mathbf{L})\}$ where $f(\mathbf{L}) = \int f(\mathbf{L}, \mathbf{R}) d\mathbf{R}$ with $f(\mathbf{L}, \mathbf{R}) \in \mathcal{M}(\boldsymbol{\gamma})$ satisfying $f_0(\mathbf{O}) = \int f(\mathbf{L}, \mathbf{R}) d\mathbf{L}_{(1-R)}$, which is identified under the smaller model $\mathcal{M}(\boldsymbol{\gamma})$. For given $\boldsymbol{\gamma}$, estimates and $(1-\alpha)100\%$ confi-

dence regions for $\boldsymbol{\beta}(\boldsymbol{\gamma})$ under model $\mathcal{M}(\boldsymbol{\gamma})$ can be constructed in the usual way. However, because the true value $\boldsymbol{\gamma}_0$ of $\boldsymbol{\gamma}$ is not identified under model $\mathcal{M}(\Gamma)$, such confidence regions may fail to cover the truth $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\gamma}_0)$ with at least $100(1-\alpha)\%$ chance under the true data-generating model (indeed, only the $\boldsymbol{\gamma}_0$-specific region will). It is hence more meaningful to construct regions that cover $\boldsymbol{\beta}(\boldsymbol{\gamma})$ uniformly over $\boldsymbol{\gamma} \in \Gamma$ under model $\mathcal{M}(\boldsymbol{\gamma})$ with at least $(1-\alpha)100\%$ chance.

Definition 1 *A region $UR_p(\boldsymbol{\beta}, \Gamma)$ is a $(1-\alpha)100\,\%$ pointwise uncertainty region for $\boldsymbol{\beta}_0$ when its pointwise coverage probability, i.e.*

$$\inf_{\gamma \in \Gamma} pr_{\mathcal{M}(\gamma)}\{\boldsymbol{\beta}(\boldsymbol{\gamma}) \in UR_p(\boldsymbol{\beta}, \Gamma)\} \tag{4.1}$$

is at least $(1-\alpha)100\,\%$.

Here the notation $\mathrm{pr}_{\mathcal{M}(\gamma)}(.)$ indicates that probabilities are taken under model $\mathcal{M}(\boldsymbol{\gamma})$. It follows from this definition that $(1-\alpha)100\%$ pointwise uncertainty regions cover the truth $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\gamma}_0)$ with at least $(1-\alpha)100\%$ chance, whatever value $\boldsymbol{\gamma}_0 \in \Gamma$ was used for generating the observed data.

Pointwise uncertainty regions extend confidence regions for identified parameters to partially identified parameters. They retain the well-known link with hypothesis tests: one can test the null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_a : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ at the $\alpha \times 100\%$ significance level by rejecting the null hypothesis when $\boldsymbol{\beta}_0$ is excluded by the $(1-\alpha)100\%$ pointwise uncertainty region $\mathrm{UR}_p(\boldsymbol{\beta}, \Gamma)$. Indeed, under model $\mathcal{M}(\Gamma)$

$$\begin{aligned} \mathrm{pr}_0(\text{reject } \boldsymbol{\beta}_0) &= 1 - \mathrm{pr}_0\{\boldsymbol{\beta}(\boldsymbol{\gamma}_0) \in \mathrm{UR}_p(\boldsymbol{\beta}, \Gamma)\} \\ &\leq 1 - \inf_{\gamma \in \Gamma} \mathrm{pr}_{\mathcal{M}(\gamma)}\{\boldsymbol{\beta}(\boldsymbol{\gamma}) \in \mathrm{UR}_p(\boldsymbol{\beta}, \Gamma)\} \leq \alpha, \end{aligned}$$

where the subscript 0 indicates that probabilities are taken w.r.t. the true observed data law and the last step follows from the definition of pointwise uncertainty regions.

Below, we will show how to construct $(1-\alpha)100\%$ pointwise uncertainty intervals for scalar parameters. To simplify the discussion, let $\boldsymbol{\gamma}_l$ and $\boldsymbol{\gamma}_u$ be values in Γ that correspond to the lower and upper bound of an ignorance interval for β, respectively, so that

$$\mathrm{ir}(\beta, \Gamma) = [\beta_l, \beta_u] = [\beta(\boldsymbol{\gamma}_l), \beta(\boldsymbol{\gamma}_u)]$$

Throughout, suppose that the following assumptions hold.

Assumption 1 *We have available consistent and asymptotically normal (CAN) estimators* $\hat{\beta}_l$ *for* $\beta(\boldsymbol{\gamma}_l)$ *with standard error* $\text{se}(\hat{\beta}_l)$ *under model* $\mathcal{M}(\boldsymbol{\gamma}_l)$ *and* $\hat{\beta}_u$ *for* $\beta(\boldsymbol{\gamma}_u)$ *with standard error* $\text{se}(\hat{\beta}_u)$ *under model* $\mathcal{M}(\boldsymbol{\gamma}_u)$.

Assumption 2 *The values* $\boldsymbol{\gamma}_l$ *and* $\boldsymbol{\gamma}_u$ *in* Γ *that correspond to the lower bound* $\beta_l = \beta(\boldsymbol{\gamma}_l)$ *and upper bound* $\beta_u = \beta(\boldsymbol{\gamma}_u)$*, respectively, are independent of the observed data law.*

Assumption 1 guarantees that CAN estimators for β can be found under models $\mathcal{M}(\boldsymbol{\gamma}_l)$ and $\mathcal{M}(\boldsymbol{\gamma}_u)$. Assumption 2 guarantees that these estimators are CAN for the bounds of the ignorance interval for β_0 with consistent standard errors $\text{se}(\hat{\beta}_l)$ and $\text{se}(\hat{\beta}_u)$, respectively. In Section 6, we give an example where Assumption 2 fails because the values for the sensitivity parameters that correspond to these bounds must be estimated from the observed data. Additional account must then be taken of the sampling variability of these estimated values.

Under Assumptions 1 and 2, $(1-\alpha)100\%$ pointwise uncertainty intervals for β_0 can be constructed by adding confidence limits with adjusted critical values to the estimated ignorance limits $\hat{\beta}_l$ and $\hat{\beta}_u$. Thus, with $c_{\alpha^*/2}$ a critical value yet to be derived, we propose $(1-\alpha)100\%$ pointwise uncertainty intervals of the form

$$\text{UR}_p(\beta, \Gamma) = [C_L, C_U] = [\hat{\beta}_l - c_{\alpha^*/2}\text{se}(\hat{\beta}_l), \hat{\beta}_u + c_{\alpha^*/2}\text{se}(\hat{\beta}_u)] \tag{4.2}$$

Next we calculate the critical value $c_{\alpha^*/2}$ needed to attain the desired pointwise coverage level. In the Appendix, we show that under Assumptions 1 and 2, expression (4.2) is an asymptotic $(1-\alpha)100\%$ pointwise uncertainty interval for β_0 if $c_{\alpha^*/2}$ solves the following equation

$$\min\left[\Phi(c_{\alpha^*/2}) - \Phi\left\{-c_{\alpha^*/2} - \frac{\beta_u - \beta_l}{\text{se}(\hat{\beta}_u)}\right\}, \Phi\left\{c_{\alpha^*/2} + \frac{\beta_u - \beta_l}{\text{se}(\hat{\beta}_l)}\right\} - \Phi(-c_{\alpha^*/2})\right] = 1 - \alpha \tag{4.3}$$

where $\Phi(.)$ is the cumulative distribution function of a standard normal variate. We further show that the asymptotic pointwise coverage probability of this interval is the nominal $(1-\alpha)100\%$. Equation (4.3) yields no feasible solution for $c_{\alpha^*/2}$ because it involves unknown functionals of the observed data distribution.

Hence, consistent estimators must be derived for the critical value by replacing β_l, β_u, $\text{se}(\hat{\beta}_l)$ and $\text{se}(\hat{\beta}_u)$ in equation (4.3) by consistent estimators. Resulting Estimated Uncertainty RegiOns will be called EUROs.

In the Appendix, we show that $c_{\alpha^*/2}$ approximates the $(1-\alpha)100\%$ percentile of the standard normal distribution when there is much ignorance about the target parameter and the intended sample size is large. Pointwise uncertainty intervals further enjoy the important property that for monotone mappings $g(.)$, $\text{pr}_0\{\text{ir}(\beta,\Gamma) \subset \text{UR}_s(\beta,\Gamma)\} = \text{pr}_0[g\{\text{ir}(\beta,\Gamma)\} \subset g\{\text{UR}_s(\beta,\Gamma)\}] = \text{pr}_0[\text{ir}\{g(\beta),\Gamma\} \subset g\{\text{UR}_s(\beta,\Gamma)\}]$. Hence, they can be estimated on a transformed scale where, for instance, asymptotic normality is a better approximation and subsequently be backtransformed to the original scale.

In the Kenyan HIV surveillance study, we deduce that β_0 lies between $\beta_l = \beta(\gamma_{pm}=0) = \text{expit}(\eta_0)\nu_0 = \text{E}(YR)$ and $\beta_u = \beta(\gamma_{pm}=1) = \text{expit}(\eta_0)\nu_0 + 1 - \nu_0 = \text{E}(YR+1-R)$. Replacing population values by sampling analogs, we estimate HIV risk between $\hat{\beta}_l = 52/787 = 0.066$ and $\hat{\beta}_u = 88/787 = 0.112$ without assumptions on the missing data mechanism. With $\gamma_{pm,l} = 0$ and $\gamma_{pm,u} = 1$ corresponding to the lower bound β_l and upper bound β_u regardless of the observed data law, Assumption 2 is satisfied. Thus, solving (4.3) with $\hat{\text{se}}(\hat{\beta}_l) = 52 \times 731/787^3$ and $\hat{\text{se}}(\hat{\beta}_u) = 88 \times 699/787^3$ yields $c_{\alpha^*/2} = 1.645$. A 95% pointwise EURO for β_0 is $[0.0515, 0.130]$. Without assumptions on the missing data mechanism, we estimate HIV risk to lie between 5.15% and 13.0% with at least 95% chance. Interestingly, this interval retains the qualitative interpretation of a classical confidence interval but involves no missing data assumptions.

Physicians involved in this study think that 25% is a safe overestimate of the risk of HIV among nonresponders. Assuming they are right, we set $\gamma_{\text{pm}} \in \Gamma = [0, 0.25]$ and estimate β_0 to lie between $52/787 = 0.066$ and $52/787 + (0.25)36/787 = 0.078$. We find a corresponding 95% pointwise EURO from 0.0515 to 0.0924. Better finite-sample approximations are expected by estimating pointwise uncertainty intervals on the logit scale. Solving (4.3) with $\hat{\beta}_l = \text{logit}(52/787)$, $\hat{\beta}_u = \text{logit}(88/787)$, $\hat{\text{se}}(\hat{\beta}_l) = 787/(52 \times 731)$ and $\hat{\text{se}}(\hat{\beta}_u) = 787/(88 \times 699)$ yields $c_{\alpha^*/2} = 1.665$ and a 95% pointwise EURO for $\text{logit}(\beta_0)$ is $[-2.89, -2.25]$. A 95% pointwise EURO for β_0 becomes $[\text{expit}(-2.89), \text{expit}(-2.25)] = [0.0528, 0.0949]$. Thus, under a realistic range of plausible missing data assump-

tions (i.e. provided that $\gamma_{\text{pm}} \in [0, 0.25]$), we estimate that HIV risk lies between 5.28% and 9.49% with at least 95% chance.

4.2. Consistency

To verify whether the HEIR itself is adequate as an estimator of the partially identified parameter $\boldsymbol{\beta}_0$ under model $\mathcal{M}(\Gamma)$, we extend the concept of weak consistency for point estimators to HEIRs. As $\boldsymbol{\gamma}_0$ is not identified under model $\mathcal{M}(\Gamma)$, we require weak convergence of all individual point estimators $\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$ to $\boldsymbol{\beta}(\boldsymbol{\gamma})$ over $\boldsymbol{\gamma} \in \Gamma$.

To formalize this, consider for each $\boldsymbol{\gamma} \in \Gamma$ a sequence of random vectors $\hat{\boldsymbol{\beta}}_1(\boldsymbol{\gamma}), \ldots, \hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma})$ and the theoretical parameter value $\boldsymbol{\beta}(\boldsymbol{\gamma})$.

Definition 2 *We define the HEIR* $\hat{ir}_N(\beta, \Gamma) = \{\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma}); \forall \boldsymbol{\gamma} \in \Gamma\}$ *to be weakly consistent for* $\boldsymbol{\beta}_0$ *if the convergence in probability of* $\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma})$ *to* $\boldsymbol{\beta}(\boldsymbol{\gamma})$ *under model* $\mathcal{M}(\boldsymbol{\gamma})$ *holds for all* $\boldsymbol{\gamma} \in \Gamma$.

HEIRs that are weakly consistent for a parameter $\boldsymbol{\beta}_0$ under model $\mathcal{M}(\Gamma)$ have the desirable property that they cover the truth $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\gamma}_0)$ with arbitrarily large probability as the sample size increases (provided that $\boldsymbol{\gamma}_0 \in \Gamma$).

For example, in the Kenyan HIV surveillance study, HIV risk $\beta(\gamma) = \text{E}(YR) + \gamma\text{E}(1-R)$ where $\gamma = \gamma_{\text{pm}}$. By the weak law of large numbers (Newey and McFadden 1994), $\hat{\beta}(\gamma) = N^{-1}\sum_{i=1}^{N} Y_i R_i + \gamma(1-R_i)$ converges in probability to $\beta(\gamma)$ for all $\gamma \in [0, 1]$. We conclude that the HEIR $[N^{-1}\sum_{i=1}^{N} Y_i R_i, N^{-1}\sum_{i=1}^{N}(Y_i R_i + 1 - R_i)]$ is a weakly consistent estimator for HIV risk β_0 when γ is unrestricted. Likewise, the HEIR $[N^{-1}\sum_{i=1}^{N} Y_i R_i, N^{-1}\sum_{i=1}^{N}\{Y_i R_i + 0.25(1-R_i)\}]$ is a weakly consistent estimator for β_0 when $\gamma \in [0, 0.25]$.

5. Inference for $\mathbf{ir}(\boldsymbol{\beta}, \Gamma)$

The previous notions of pointwise coverage and consistency were designed to measure the validity of HEIRs w.r.t. their ability to estimate the target parameter of interest $\boldsymbol{\beta}_0$. When the ignorance region for $\boldsymbol{\beta}_0$ itself is the primary target of a study, a more natural goal is to construct HEIRs whose distance from the true ignorance region can be made arbitrarily small with arbitrarily large probability and $(1-\alpha)100\%$ uncertainty regions that cover $\text{ir}(\boldsymbol{\beta}, \Gamma)$ with at least $(1-\alpha)100\%$ chance in large samples.

5.1. Sampling Variability: Strong Coverage

A natural strategy for communicating the sampling variability of HEIRs for scalar parameters β is to add the standard $(1-\alpha)100\%$ confidence limits to the estimated ignorance limits; that is:

$$\mathrm{UR}_s(\beta,\Gamma)=[C_L,C_U] \quad = \quad [\hat{\beta}_l - c_{\alpha/2}\mathrm{se}(\hat{\beta}_l), \hat{\beta}_u + c_{\alpha/2}\mathrm{se}(\hat{\beta}_u)] \tag{5.1}$$

where $c_{\alpha/2}$ is the $(1-\alpha/2)100\%$ percentile of the standard normal distribution. This is done in Kenward, Goetghebeur and Molenberghs (2001) and Rosenbaum (1995), for instance. The interval thus constructed covers all $(1-\alpha)100\%$ confidence intervals for $\beta(\boldsymbol{\gamma})$ under $\mathcal{M}(\boldsymbol{\gamma})$ pointwise for all $\boldsymbol{\gamma}\in\Gamma$. In the Appendix, we further show that the resulting interval can be interpreted as the $(1-\alpha)100\%$ strong uncertainty interval $\mathrm{UR}_s(\beta,\Gamma)$, which covers all values in the ignorance region $\mathrm{ir}(\beta,\Gamma)$ simultaneously with $(1-\alpha)100\%$ chance. For vector parameters $\boldsymbol{\beta}$, we define such $(1-\alpha)100\%$ strong uncertainty region as follows.

Definition 3 *A region $UR_s(\boldsymbol{\beta},\Gamma)$ is a $(1-\alpha)100\%$ strong uncertainty region for $\boldsymbol{\beta}_0$ when its strong coverage probability, i.e.*

$$pr_0\{ir(\boldsymbol{\beta},\Gamma)\subset UR_s(\boldsymbol{\beta},\Gamma)\}$$

is at least $(1-\alpha)100\%$ under the true observed data law.

It is immediate from the definition that a $(1-\alpha)100\%$ strong uncertainty region is a conservative $(1-\alpha)100\%$ pointwise uncertainty region.

In the Appendix we show that the strong coverage level of the strong uncertainty interval (5.1) lies between $1-\alpha$ and $1-\alpha/2$ when the estimated confidence limits C_L and C_U satisfy $\mathrm{pr}_0(C_L>\beta_l)=\mathrm{pr}_0(C_U<\beta_u)=\alpha/2$. It equals $1-\alpha$ when, as is generally expected, the true ignorance interval almost never covers the strong uncertainty interval (i.e. $\mathrm{pr}_0\{(C_L>\beta_l)\wedge(C_U<\beta_u)\}=0$). We further show that the strong coverage probability of this interval lies between $1-\alpha$ and $1-\alpha+\alpha^2/4$ when, as expected, $\mathrm{pr}_0(C_L>\beta_l|C_U<\beta_u)\leq\mathrm{pr}_0(C_L>\beta_l)$. For instance, for $\alpha=0.05$ it lies between 95% and 95.0625% (when this property holds).

As with pointwise coverage, strong uncertainty regions can be estimated on a monotonely transformed scale while retaining the original coverage level.

5.1. Sampling Variability: Weak Coverage

While strong uncertainty regions cover all parameter values in the true region of ignorance simultaneously with given probability, one would often be satisfied having covered most of them. Indeed, an estimated region which is expected to cover most parameter values in $\text{ir}(\boldsymbol{\beta}, \Gamma)$ is useful, even when it represents a low strong coverage probability. In view of this, we define the weak coverage probability of an uncertainty region as the expected proportion of overlap between the uncertainty region and the true region of ignorance. This leads to the following definition of $(1-\alpha)100\%$ weak uncertainty regions.

Definition 4 *A region $UR_w(\boldsymbol{\beta}, \Gamma)$ is a $(1-\alpha)100\%$ weak uncertainty region for $\boldsymbol{\beta}_0$ when its weak coverage probability, i.e.*

$$\frac{E_0 \| UR_w(\boldsymbol{\beta}, \Gamma) \cap ir(\boldsymbol{\beta}, \Gamma) \|}{\| ir(\boldsymbol{\beta}, \Gamma) \|}$$

is at least $(1-\alpha)100\%$ under the true observed data law, where $\|A\|$ $(A \subseteq \mathbb{R}^p)$ denotes the volume of A.

This can additionally be interpreted as the probability that a uniform draw from the ignorance region for $\boldsymbol{\beta}_0$ is covered by $\text{UR}_w(\boldsymbol{\beta}, \Gamma)$.

In the Appendix, we show that $100(1-\alpha)\%$ weak uncertainty intervals for scalar β_0 can be constructed following (4.2), where $c_{\alpha^*/2}$ solves the equation

$$\alpha = \frac{\text{se}(\hat{\beta}_l) + \text{se}(\hat{\beta}_u)}{\beta_u - \beta_l} \int_0^{+\infty} z\varphi(z + c_{\alpha^*/2}) dz + \epsilon. \tag{5.2}$$

$\varphi(.)$ denotes the standard normal density function and ϵ is a correction term. In the Appendix, we show that ϵ can be calculated exactly as

$$\begin{aligned} \epsilon \;=\; & \int_{(\beta_u - \beta_l)/\text{se}(\hat{\beta}_u)}^{+\infty} \varphi(z + c_{\alpha^*/2}) dz + \int_{(\beta_u - \beta_l)/\text{se}(\hat{\beta}_l)}^{+\infty} \varphi(z + c_{\alpha^*/2}) dz \\ & - \frac{\text{se}(\hat{\beta}_u)}{\beta_u - \beta_l} \int_{(\beta_u - \beta_l)/\text{se}(\hat{\beta}_u)}^{+\infty} z\varphi(z + c_{\alpha^*/2}) dz - \frac{\text{se}(\hat{\beta}_l)}{\beta_u - \beta_l} \int_{(\beta_u - \beta_l)/\text{se}(\hat{\beta}_l)}^{+\infty} z\varphi(z + c_{\alpha^*/2}) dz \end{aligned} \tag{5.3}$$

which we show to be so small that setting $\epsilon = 0$ will not hamper the accuracy of the calculated critical value. We further show that the weak coverage probability of this interval is the nominal $1-\alpha$.

Equation (5.2) yields no feasible solution for $c_{\alpha^*/2}$ because it involves unknown functionals of the observed data distribution. Hence, consistent estimators must be derived for the critical value by substituting β_l, β_u, $\text{se}(\hat{\beta}_l)$ and $\text{se}(\hat{\beta}_u)$ in equation (5.2) by consistent estimators. When the HEIR is large and its endpoints precisely estimated, it may itself cover more than $(1-\alpha)100\%$ of the true ignorance interval on average. By allowing for negative values of $c_{\alpha^*/2}$, the coverage level is then reached for a weak uncertainty interval which is contained within the HEIR. A drawback with inference for weak uncertainty intervals is that they cannot generally be estimated on a monotonely transformed scale while retaining the original coverage level.

For example, in the Kenyan HIV surveillance study, with $\gamma_{\text{pm}} \in [0, 0.25]$, the 95% strong EURO $[0.0487, 0.0950]$ is estimated to cover the true ignorance interval for HIV risk with (at least) 95% chance. We estimated $\hat{\text{pr}}_0\{(C_L > \beta_l) \wedge (C_U < \beta_u)\} = 2.5\ 10^{-16}$, indicating that the nominal coverage level is well approximated by 95%. The 95% weak EURO is $[0.0587, 0.0899]$. It has an expected overlap of 95% with the true ignorance region for HIV risk when $\gamma_{\text{pm}} \in [0, 0.25]$. Since values near the midpoint of the interval are almost always covered, it is considerably smaller than the 95% strong EURO.

5.3. Relationship between the weak and pointwise uncertainty region

To enhance our understanding of the relationships between the different uncertainty regions, we will now prove that a $(1-\alpha)100\%$ pointwise uncertainty region $\text{UR}_p(\boldsymbol{\beta}, \Gamma)$ is a conservative $(1-\alpha)100\%$ weak uncertainty region. Indeed, we know that for all $\boldsymbol{\beta} \in \text{ir}(\boldsymbol{\beta}, \Gamma)$, $\text{pr}_0\{\boldsymbol{\beta} \in \text{UR}_p(\boldsymbol{\beta}, \Gamma)\} \geq 1-\alpha$. Using that the weak coverage probability of $\text{UR}_p(\boldsymbol{\beta}, \Gamma)$ is the probability that a uniform draw $\tilde{\boldsymbol{\beta}}$ from the ignorance region for $\boldsymbol{\beta}$ is covered by $\text{UR}_p(\boldsymbol{\beta}, \Gamma)$, we find

$$\begin{aligned}
\frac{\text{E}_0\|\text{UR}_p(\boldsymbol{\beta}, \Gamma) \cap \text{ir}(\boldsymbol{\beta}, \Gamma)\|}{\|\text{ir}(\boldsymbol{\beta}, \Gamma)\|} &= \text{pr}_{0,\tilde{\beta}}\{\tilde{\boldsymbol{\beta}} \in \text{UR}_p(\boldsymbol{\beta}, \Gamma)\} \\
&= \text{E}_{\tilde{\beta}}[\text{E}_0\{\tilde{\boldsymbol{\beta}} \in \text{UR}_p(\boldsymbol{\beta}, \Gamma)|\tilde{\boldsymbol{\beta}}\}] \\
&\geq \text{E}_{\tilde{\beta}}(1-\alpha|\tilde{\boldsymbol{\beta}}) = 1-\alpha.
\end{aligned}$$

Here, the subscript $\tilde{\boldsymbol{\beta}}$ refers to the uniform sampling distribution over $\text{ir}(\boldsymbol{\beta}, \Gamma)$.

5.4. Asymmetric uncertainty regions

In constructing pointwise, strong and weak uncertainty intervals, we have chosen the same critical values $c_{\alpha^*/2}$ to calculate the lower limit $C_L = \hat{\beta}_l -$

$c_{\alpha^*/2}\text{se}(\hat{\beta}_l)$ and the upper limit $C_U = \hat{\beta}_u + c_{\alpha^*/2}\text{se}(\hat{\beta}_u)$. When $\text{se}(\hat{\beta}_l)$ differs from $\text{se}(\hat{\beta}_u)$ shorter $(1-\alpha)100\%$ uncertainty intervals may sometimes be obtained by allowing for different values $c_{\alpha_l^*/2}$ in $C_L = \hat{\beta}_l - c_{\alpha_l^*/2}\text{se}(\hat{\beta}_l)$ and $c_{\alpha_u^*/2}$ in $C_U = \hat{\beta}_u + c_{\alpha_u^*/2}\text{se}(\hat{\beta}_u)$, respectively. To obtain such intervals, we estimate $c_{\alpha_u^*/2}$ from the observed data for different chosen values of $c_{\alpha_l^*/2}$. For strong uncertainty intervals, for example, we choose $c_{\alpha_u^*/2}$ such that $\text{pr}_0(C_U < \beta_u) = \alpha - \alpha_l^*/2$ when $c_{\alpha_l^*/2}$ is such that $\text{pr}_0(C_L > \beta_l) = \alpha_l^*/2$ (with $\alpha_l^*/2 \leq \alpha^*$). For pointwise and weak uncertainty intervals, a similar strategy is possible along the lines of the Appendix. Having obtained estimates for $c_{\alpha_u^*/2}$ over a range of $c_{\alpha_l^*/2}$-values, we then choose the $(c_{\alpha_l^*/2}, c_{\alpha_u^*/2})$-tuple that minimizes the length $\hat{\beta}_u - \hat{\beta}_l + c_{\alpha_l^*/2}\text{se}(\hat{\beta}_l) + c_{\alpha_u^*/2}\text{se}(\hat{\beta}_u)$ of the uncertainty interval.

5.5. Simulation study

Because our weak and pointwise uncertainty intervals ignore imprecise estimation of the critical value and standard error, we assess their performance in a simulation study. To be able to evaluate the net effects of estimation of the critical value and standard error, we wish to avoid relying on asymptotic approximations and therefore consider inference for the mean of normally distributed observations. We generate ten thousand data sets with sample size 787, standard normally distributed observations for responders and marginal nonresponse probability $\nu_0 = 36/787$. We assume a priori that $\gamma = \text{E}(Y|R=0)$ lies in $\Gamma = [-2, 2]$. For each data set, ignorance and 95% uncertainty intervals for $\beta_0 = \text{E}(Y)$ are estimated for the three coverage definitions. Strong uncertainty limits are estimated via Wald-type confidence limits calculated at the two estimated ignorance limits $\hat{\eta}\hat{\nu} - 2(1-\hat{\nu})$ and $\hat{\eta}\hat{\nu} + 2(1-\hat{\nu})$, where $\eta = \text{E}(Y|R=1)$. Weak and pointwise uncertainty limits are estimated analogously, but with adjusted critical values.

Table 1 about here.

Table 1 gives empirical coverage probabilities, average length of the EURO and (average) adjusted critical values. Strong and pointwise EUROs reach the nominal 95% coverage probability. Some overcoverage is however observed for weak EUROs. This is due to imprecise estimation of the critical values $c_{\alpha^*/2}$. These are highly variable for weak EUROs (mean 0.803 and standard deviation

0.0891), but substantially less so for pointwise EUROs (mean 1.647 and standard deviation 0.0000149). Furthermore, by choosing the exact critical value 0.797 for weak uncertainty intervals, we obtain an estimated coverage of 0.950 (P-value 0.691). The correction term ϵ for weak EUROs is negligible, having a highly skewed distribution with median $-2.053\ 10^{-13}$ (min. $-1.663\ 10^{-6}$, 1st quartile $-5.812\ 10^{-12}$, 3rd quartile $-4.480\ 10^{-15}$, max. $-3.249\ 10^{-28}$).

5.6. Consistency

To verify whether the HEIR is adequate as an estimator of $\mathrm{ir}(\boldsymbol{\beta}, \Gamma)$, we extend the definition of Section 4.2 for weakly consistent HEIRs by defining a HEIR to be weakly consistent for an ignorance region $\mathrm{ir}(\boldsymbol{\beta}, \Gamma)$ under model $\mathcal{M}(\Gamma)$ when its distance to $\mathrm{ir}(\boldsymbol{\beta}, \Gamma)$ can be made arbitrarily small with arbitrarily large probability in large samples.

To formalize this, consider the sequence of random sets $\hat{\mathrm{ir}}_1(\boldsymbol{\beta}, \Gamma), \ldots, \hat{\mathrm{ir}}_N(\boldsymbol{\beta}, \Gamma)$ and the theoretical region of ignorance $\mathrm{ir}(\boldsymbol{\beta}, \Gamma)$. Define $\|\mathrm{ir}(\boldsymbol{\beta}, \Gamma) - \hat{\mathrm{ir}}_N(\boldsymbol{\beta}, \Gamma)\|$ as the maximum distance between the true $(\mathrm{ir}(\boldsymbol{\beta}, \Gamma))$ and estimated $(\hat{\mathrm{ir}}_N(\boldsymbol{\beta}, \Gamma))$ region of ignorance for $\boldsymbol{\beta}_0$, i.e.

$$\|\mathrm{ir}(\boldsymbol{\beta}, \Gamma) - \hat{\mathrm{ir}}_N(\boldsymbol{\beta}, \Gamma)\| = \max \left(\sup_{\beta_N \in \hat{\mathrm{ir}}_N(\beta, \Gamma)} \inf_{\tilde{\beta} \in \mathrm{ir}(\beta, \Gamma)} \|\boldsymbol{\beta}_N - \tilde{\boldsymbol{\beta}}\|, \sup_{\tilde{\beta} \in \mathrm{ir}(\beta, \Gamma)} \inf_{\beta_N \in \hat{\mathrm{ir}}_N(\beta, \Gamma)} \|\boldsymbol{\beta}_N - \tilde{\boldsymbol{\beta}}\| \right)$$

where $\|\boldsymbol{\beta}_N - \tilde{\boldsymbol{\beta}}\|$ is the Euclidian distance between $\boldsymbol{\beta}_N$ and $\tilde{\boldsymbol{\beta}}$. The above distance is known as the Hausdorff metric over the metric space $2^{\mathbb{R}^p}$ of all subsets of $\mathbb{R}^p$.

Using general results on stochastic convergence in metric spaces (van der Vaart 1998), we come to the following definition.

Definition 5 *We define the HEIR $\hat{ir}_N(\boldsymbol{\beta}, \Gamma)$ weakly consistent for $ir(\boldsymbol{\beta}, \Gamma)$ if*

$$(\forall \epsilon, \delta > 0)(\exists N_0(\epsilon, \delta))\, (N > N_0(\epsilon, \delta) \Rightarrow pr_0\{\|\hat{ir}_N(\boldsymbol{\beta}, \Gamma) - ir(\boldsymbol{\beta}, \Gamma)\| < \delta\} > 1 - \epsilon)\,.$$

The following theorem, which is proved in Appendix 2, gives simple sufficient rules for verifying this property.

Theorem 1 *Define the HEIR $\hat{ir}_N(\boldsymbol{\beta}, \Gamma) = \{\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma}); \forall \boldsymbol{\gamma} \in \Gamma\}$ as a set of individual point estimators $\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma})$, and the true region of ignorance $ir(\boldsymbol{\beta}, \Gamma) = \{\boldsymbol{\beta}(\boldsymbol{\gamma}); \forall \boldsymbol{\gamma} \in \Gamma\}$ as the set of corresponding estimands $\boldsymbol{\beta}(\boldsymbol{\gamma})$. Then, $\hat{ir}_N(\boldsymbol{\beta}, \Gamma)$ is a weakly consistent estimator for $ir(\boldsymbol{\beta}, \Gamma)$ under model $\mathcal{M}(\Gamma)$ when the point estimators $\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma})$ in $\hat{ir}_N(\boldsymbol{\beta}, \Gamma)$ are weakly consistent estimators for $\boldsymbol{\beta}(\boldsymbol{\gamma})$ uniformly*

over all $\boldsymbol{\gamma} \in \Gamma$*; that is*

$$(\forall \epsilon, \delta > 0)(\exists N_0(\epsilon, \delta)) \left(N > N_0(\epsilon, \delta) \Rightarrow pr_{\mathcal{M}(\gamma)}\{\sup_{\gamma \in \Gamma} \|\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma}) - \boldsymbol{\beta}(\boldsymbol{\gamma})\| < \delta\} > 1 - \epsilon \right)$$

where $\|\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma}) - \boldsymbol{\beta}(\boldsymbol{\gamma})\|$ *is the Euclidian distance between* $\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma})$ *and* $\boldsymbol{\beta}(\boldsymbol{\gamma})$*.*

For example, because in the Kenyan HIV surveillance study, $YR + \gamma(1-R)$ is continuous at each $\gamma \in [0,1]$ w.p.1 and $YR+\gamma(1-R)$ is bounded above by 1, it follows from the uniform law of large numbers (Newey and McFadden 1994) that $\hat{\beta}(\gamma) = N^{-1}\sum_{i=1}^{N} Y_i R_i + \gamma(1-R_i)$ converges in probability to $\beta(\gamma)$ uniformly over all $\gamma \in [0,1]$. We conclude that the HEIR $[N^{-1}\sum_{i=1}^{N} Y_i R_i, N^{-1}\sum_{i=1}^{N}(Y_i R_i + 1 - R_i)]$ is a weakly consistent estimator for $\text{ir}(\beta, \Gamma)$ under $\mathcal{M}(\Gamma)$ when $\Gamma = [0,1]$. Likewise, the HEIR $[N^{-1}\sum_{i=1}^{N} Y_i R_i, N^{-1}\sum_{i=1}^{N}\{Y_i R_i + 0.25(1-R_i)\}]$ is a weakly consistent estimator for $\text{ir}(\beta, \Gamma)$ under $\mathcal{M}(\Gamma)$ when $\Gamma = [0, 0.25]$. Note that the reverse of Theorem 1 is not true.

6. Parametric and semiparametric models

So far, we have conducted inference for specific functionals of the complete data law by constructing ignorance and uncertainty regions under a family of NPI models. Such families are e.g. constructed in Robins (1997) and Scharfstein, Robins and Rotnitzky (1999). In practice, parametric restrictions on the full data distribution are often necessary, for instance when we are interested in low dimensional models for the complete data or when we are forced to impose dimension reducing modelling restrictions due to the curse of dimensionality. In this section, we therefore discuss whether meaningful ignorance and uncertainty regions can still be defined when the full data law is required to satisfy the restrictions of some parametric or semiparametric model $\mathcal{M}^*$.

Suppose that we are interested in assessing the effect of age X (in years) on HIV risk in the Kenyan surveillance study through model $\mathcal{M}^*$ defined by

$$\text{logit}\{\text{pr}(Y = 1|X)\} = \eta_0 + \beta_0 X \tag{6.1}$$

$$\text{logit}\{\text{pr}(R = 1|Y, X)\} = \alpha_0^{(1)} + \alpha_0^{(2)} X + \gamma Y. \tag{6.2}$$

Because model $\mathcal{M}^*$ imposes restrictions on the observed data law, there will generally be few (often at most one) full data laws that marginalize to the observed data law and satisfy the model restrictions (6.1)-(6.2). As a result, the

dependence γ of missingness on the missing outcome may become identified, in which case the central identifiability problem of this paper disappears. The fact that γ can be identified despite the missing data, is due to the (semi)parametric restrictions that model $\mathcal{M}^*$ imposes. This is undesirable as one would rarely have sufficient information to know, before seeing the observed data, whether the model restrictions (6.1)-(6.2) hold (see also the discussion of Diggle and Kenward 1994; Little and Rubin 1987; Scharfstein, Robins and Rotnitzky 1999). By the same token, it would usually be unreasonable to restrict attention to those full data laws that satisfy the intersection model $\mathcal{M}^* \cap \mathcal{M}(\Gamma)$ by defining the ignorance region for $\boldsymbol{\beta}$ to be

$$\begin{aligned} \text{ir}(\boldsymbol{\beta}, \Gamma) \;=\; \Big\{ \boldsymbol{\beta}\{f(\mathbf{L})\} \;:\; f(\mathbf{L}) = \int f(\mathbf{L},\mathbf{R}) d\mathbf{R} \text{ with } f(\mathbf{L},\mathbf{R}) \in \mathcal{M}^* \cap \mathcal{M}(\Gamma) \\ \text{satisfying } f_0(\mathbf{O}) = \int f(\mathbf{L},\mathbf{R}) d\mathbf{L}_{(1-R)} \Big\} \end{aligned}$$

where $\mathcal{M}(\Gamma)$ is defined as before as a class of NPI models for the full data in the absence of the restrictions of model $\mathcal{M}^*$. The ignorance region for $\boldsymbol{\beta}_0$, and hence also the strong uncertainty region, is therefore generally no longer well defined when the full data model $\mathcal{M}^*$ places restrictions on the observed data law.

It remains useful, however, to repeat the analysis for different fixed values $\boldsymbol{\gamma}$ of the sensitivity parameter in the set Γ as if model $\mathcal{M}^* \cap \mathcal{M}(\boldsymbol{\gamma})$ were true; e.g. to conduct inference for β_0 indexing model (6.1) for different values of γ in model (6.2). We continue to call the resulting range of estimates a HEIR. Pointwise uncertainty intervals can be constructed as before and stay meaningful provided that, for each $\boldsymbol{\gamma}$, probabilities in (4.1) are taken w.r.t. model $\mathcal{M}^* \cap \mathcal{M}(\boldsymbol{\gamma})$. Indeed, when $\gamma_0 \in \Gamma$ the resulting intervals will cover the truth $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\gamma}_0)$ with at least $(1-\alpha)100\%$ chance provided that model $\mathcal{M}^* \cap \mathcal{M}(\boldsymbol{\gamma}_0)$ holds. Likewise, it remains meaningful to define a HEIR weakly consistent for $\boldsymbol{\beta}_0$ when the convergence of $\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma})$ to $\boldsymbol{\beta}(\boldsymbol{\gamma})$ under model $\mathcal{M}^* \cap \mathcal{M}(\boldsymbol{\gamma})$ holds for all $\boldsymbol{\gamma} \in \Gamma$. Indeed, such HEIR will cover the truth $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\gamma}_0)$ with arbitrarily large probability as the sample size increases provided that model $\mathcal{M}^* \cap \mathcal{M}(\boldsymbol{\gamma}_0)$ holds with $\gamma_0 \in \Gamma$.

For example, consider the full data model defined by (6.1) with unrestricted response model. Figure 2 (left) shows HEIRs and 95% EUROs for age-specific HIV risk under this model. HEIRs were obtained via the IDE algorithm (Vansteelandt and Goetghebeur 2001) which yields point estimates on the boundary of

the HEIR. Without assumptions on the missing data, we find HIV risk estimates between 6% and 12% over a reasonably wide age range. Following Vansteelandt and Goetghebeur (2001), the boundary of the HEIR was constructed by defining a sensitivity parameter $\boldsymbol{\gamma}$ and estimating from the observed data the 'extreme' values of that parameter that yield estimates on the boundary of the HEIR. The standard errors of the ignorance limits, which were used to construct EUROs in Figure 2 (left), were calculated supposing that these 'extreme' values are fixed. Assumption 2 failed in this example because the values of $\boldsymbol{\gamma}$ that yielded estimates on the boundary of the HEIR, vary over different samples and hence depend on the observed data law. The reported EUROs ignore this extra variability. Figure 2 (right) therefore examines 500 bootstrap resamples, but reveals no undercoverage/overcoverage of our intervals. The 95% pointwise EUROs are the most meaningful here because they are known to cover the true age-specific HIV risk with at least 95% chance, regardless of the missing data mechanism, when model (6.1) holds.

Figure 2 about here.

7. Discussion

This article proposes a formalism for sensitivity analysis when (nonignorably) missing data have occurred. The usual parameter of interest is estimated by a region of estimates instead of a single point. Imprecision on estimated (ignorance) regions is incorporated via uncertainty regions which cover the truth with given probability. These broaden classical confidence regions to acknowledge additional model uncertainty due to the missing data. As such, our formalism can be viewed as a frequentist alternative to Bayesian approaches for sensitivity analysis (see e.g. Scharfstein, Daniels and Robins 2003). Our approach chooses not to average out the extremes, which is especially important where the possibility of high risks must be confronted. The nonparametric Bayesian approach of Scharfstein, Daniels and Robins (2003) is attractive and useful when there are strong scientific beliefs about the degree of selection bias, which can be expressed in a prior distribution $f(\gamma)$ for γ. In a similar spirit, our weak uncertainty regions express where the estimand can be expected when each value in the ignorance region is a priori considered equally plausible. Ultimately, more general prior knowledge

could be incorporated in our frequentist framework by redefining a $(1-\alpha)100\%$ weak uncertainty region for $\boldsymbol{\beta}_0$ to be a region whose coverage probability, i.e.

$$\int \mathrm{pr}_{\mathcal{M}(\gamma)} \left\{\beta\left(\gamma\right) \in \mathrm{UR}_w(\boldsymbol{\beta}, \Gamma)\right\} f\left(\gamma\right) d\gamma$$

is at least $(1-\alpha)100\%$ under the true observed data law. Such regions enjoy the desirable property that they are guaranteed not to use the observed data to gather additional information about the sensitivity parameter. In agreement with Scharfstein, Robins and Rotnitzky (1999), using such information would be undesirable as it can only come from the postulated model assumptions, which are usually made for convenience. The Bayesian approach does not generally enjoy this property. We plan to report on this alternative development elsewhere.

Detailed consideration of γ-specific point estimates and confidence intervals (as in Figure 1 and Scharfstein, Robins and Rotnitzky (1999)) remains most informative and is therefore highly valuable at the analysis stage. The methods in this paper form by no means an attempt to be competitive. They aim instead to summarize the detailed information that results from a sensitivity analysis, with appropriate account of sampling variability, and thus to make the results of a sensitivity analysis feasible for practical reporting. Such procedure has proved especially valuable in settings where the simultaneous impact of several sensitivity parameters is studied (Vansteelandt and Goetghebeur 2005). In general, we believe that a worthwhile summary strategy is to report, besides the usual analysis under a sole plausible missing data assumption (e.g. MAR), a HEIR and 95% pointwise uncertainty interval for the target parameter corresponding to one or several credible ranges of values for the sensitivity parameter that were selected with the help of subject-matter experts' insight. We believe this will help decision-makers to actually use the results of sensitivity analyses in practice, because they can interpret and use 95% *pointwise* uncertainty intervals like confidence intervals for point parameters, also with (semi)parametric observed data models. Hypothesis tests derived from the pointwise EURO are valuable for significance testing. Equivalence can be concluded when the pointwise EURO is contained in a chosen equivalence range. The methods are generally applicable and easy to implement with simple S-Plus programs that can be obtained from the first author.

Additional challenges must be met to adjust for imprecise estimation of the

standard errors and critical values. In a small simulation study, ignoring this imprecision yielded slight overcoverage for weak uncertainty intervals, but not for strong and pointwise intervals. Further improvements are also possible for estimation of the standard errors of the ignorance limits. In many examples, the sensitivity parameters can be chosen such that over repeat samples the same limiting values generate the bounds of the HEIR (i.e. such that Assumption 2 holds). In that case, these standard errors can be calculated in the usual way, conditional on those values for the sensitivity parameter. If this is not possible, our methods will yield approximate results that have shown good performance in bootstrap simulations. Alternatively, a bootstrap procedure itself may be used. For strong uncertainty intervals, this approach was taken by Horowitz and Manski (2000), but is expensive in terms of computation time and did not yield improved results for our example. For pointwise uncertainty intervals, mathematically equivalent estimators were published by Imbens and Manski (2004) while this work was under review.

The need to select the range of plausible values Γ for the sensitivity parameter is not an inherent drawback of our method, but typical of most meaningful sensitivity analyses. Including implausible γ-values may not only broaden the ignorance region unnecessarily, but also introduce implausible values. Furthermore, even a relatively narrow range of carefully chosen full data models may be able to convey sufficient caution. In the Kenyan study, for example, the interval $[-1, 1]$ for γ_s already allows the odds of response to be up to 2.72 times larger or smaller for HIV-positives than HIV-negatives. Combining this choice with Figure 1 yields an estimated HIV risk β_0 between 0.067 and 0.074. Logistic response models, like (6.2), are especially useful in this regard, because any restricted range of values for the sensitivity parameters indexing these models will (usually) produce bounded HEIRs for the target parameter, even when the outcomes are theoretically unbounded. When it is hard to pin down a single range, one may consider a growing set of ranges and observe how the ignorance region evolves accordingly. An indication of the deemed plausibility may be added by colouring the HEIRs correspondingly. This stops one step short of averaging in the Bayesian way, with the continuing goal of distinguishing data-based information from other sources.

In summary, we have proposed a formal, flexible and structured way of summarizing the results from a sensitivity analysis with incomplete outcomes. It makes the assumptions about the missing data explicit and shows how they affect inference. The clear separation of ignorance due to incompleteness and imprecision due to finite sampling may guide the trade-off between sample size and follow-up of nonrespondents at the design stage. The methods developed in this work have been applied beyond the missing data context, to investigate the sensitivity of causal inferences to untestable assumptions (Vansteelandt and Goetghebeur 2005). Because the information obtained from sensitivity analyses is often extremely detailed, we believe that well understood summaries like HEIRs and EUROs, can help augment the use and reporting of sensitivity analyses in practical investigations.

Acknowledgment

The authors are grateful to the editor and 2 referees for their constructive comments. The first author acknowledges support as a Postdoctoral Fellow of the Fund for Scientific Research - Flanders (Belgium) (F.W.O.). This work was partially supported by FWO-Vlaanderen Research Project G.0002.98: 'Sensitivity Analysis for Incomplete and Coarse Data'.

Appendix

A.1. Construction of uncertainty regions and proofs

For notational simplicity, we omit the subscript 0 and implicitly assume that probabilities, expectations and standard errors are taken with respect to the true observed data law.

Lemma 1 (Pointwise uncertainty intervals) *Let C_L and C_U be lower and upper $(1-\alpha^*)100\%$ confidence limits of β_l and β_u, respectively, based on CAN estimators for β_l and β_u and a critical value $c_{\alpha^*/2}$ that solves equation (4.3). Then the interval $[C_L, C_U]$ has asymptotic pointwise coverage probability $1-\alpha$.*

Proof: For univariate parameters, $(1-\alpha)100\%$ pointwise uncertainty limits for β_0 can be constructed by using that the pointwise coverage probability $1-\alpha$ equals $\inf_{\gamma\in\Gamma} \text{pr}(C_L \leq \beta(\gamma) \leq C_U) = \inf_{\gamma\in\Gamma}\{\text{pr}(C_L \leq \beta(\gamma)) - \text{pr}(C_U \leq \beta(\gamma))\}$ since $C_L \leq C_U$. Define $C_L = \hat{\beta}_l - c_{\alpha^*/2}\text{se}(\hat{\beta}_l)$ and $C_U = \hat{\beta}_u + c_{\alpha^*/2}\text{se}(\hat{\beta}_u)$ for some unknown $c_{\alpha^*/2} > 0$. For CAN estimators $\hat{\beta}_l$ and $\hat{\beta}_u$ and arbitrary $\beta \in [\beta_l, \beta_u]$,

we find that asymptotically

$$\begin{aligned}\mathrm{pr}(C_L \le \beta) - \mathrm{pr}(C_U \le \beta) &= \mathrm{pr}\left(\frac{\hat\beta_l - \beta_l}{\mathrm{se}(\hat\beta_\mathrm{l})} \le c_{\alpha^*/2} + \frac{\beta - \beta_l}{\mathrm{se}(\hat\beta_\mathrm{l})}\right) \\ &\quad -\mathrm{pr}\left(\frac{\hat\beta_u - \beta_u}{\mathrm{se}(\hat\beta_\mathrm{u})} \le -c_{\alpha^*/2} + \frac{\beta - \beta_u}{\mathrm{se}(\hat\beta_\mathrm{u})}\right) \\ &= \Phi\left\{c_{\alpha^*/2} + \frac{\beta - \beta_l}{\mathrm{se}(\hat\beta_\mathrm{l})}\right\} - \Phi\left\{-c_{\alpha^*/2} - \frac{\beta_u - \beta}{\mathrm{se}(\hat\beta_\mathrm{u})}\right\} \quad (3.1)\end{aligned}$$

Using this result we now show that

$$\begin{aligned}\inf_{\gamma\in\Gamma}\{\mathrm{pr}(C_L \le \beta(\gamma)) - \mathrm{pr}(C_U \le \beta(\gamma))\} = \min\{&\mathrm{pr}(C_L \le \beta_l) - \mathrm{pr}(C_U \le \beta_l), \\ &\mathrm{pr}(C_L \le \beta_u) - \mathrm{pr}(C_U \le \beta_u)\} \quad (3.2)\end{aligned}$$

Note from (3.1) that

$$\mathrm{pr}(C_L \le \beta_l) - \mathrm{pr}(C_U \le \beta_l) = \Phi(c_{\alpha^*/2}) - \Phi\left\{-c_{\alpha^*/2} - \frac{\beta_u - \beta_l}{\mathrm{se}(\hat\beta_\mathrm{u})}\right\} \quad (3.3)$$

$$\mathrm{pr}(C_L \le \beta_u) - \mathrm{pr}(C_U \le \beta_u) = \Phi\left\{c_{\alpha^*/2} + \frac{\beta_u - \beta_l}{\mathrm{se}(\hat\beta_\mathrm{l})}\right\} - \Phi(-c_{\alpha^*/2}) \quad (3.4)$$

Asymptotically, it follows that for arbitrary $\beta \in [\beta_l, \beta_u]$ and a standard normal variate Z

$$\begin{aligned}\mathrm{pr}(C_L \le \beta) - \mathrm{pr}(C_U \le \beta) &= \mathrm{pr}(C_L \le \beta_l) - \mathrm{pr}(C_U \le \beta_l) \\ &\quad +\mathrm{pr}\left(c_{\alpha^*/2} \le Z \le c_{\alpha^*/2} + \frac{\beta - \beta_l}{\mathrm{se}(\hat\beta_\mathrm{l})}\right) \\ &\quad -\mathrm{pr}\left(-c_{\alpha^*/2} - \frac{\beta_u - \beta_l}{\mathrm{se}(\hat\beta_\mathrm{u})} \le Z \le -c_{\alpha^*/2} - \frac{\beta_u - \beta}{\mathrm{se}(\hat\beta_\mathrm{u})}\right) \quad (3.5) \\ &= \mathrm{pr}(C_L \le \beta_u) - \mathrm{pr}(C_U \le \beta_u) \\ &\quad -\mathrm{pr}\left(c_{\alpha^*/2} + \frac{\beta - \beta_l}{\mathrm{se}(\hat\beta_\mathrm{l})} \le Z \le c_{\alpha^*/2} + \frac{\beta_u - \beta_l}{\mathrm{se}(\hat\beta_\mathrm{l})}\right) \\ &\quad +\mathrm{pr}\left(-c_{\alpha^*/2} - \frac{\beta_u - \beta}{\mathrm{se}(\hat\beta_\mathrm{u})} \le Z \le -c_{\alpha^*/2}\right)\end{aligned}$$

Suppose first that $\mathrm{se}(\hat\beta_\mathrm{l}) \le \mathrm{se}(\hat\beta_\mathrm{u})$. Then $(\beta - \beta_l)/\mathrm{se}(\hat\beta_\mathrm{l}) \ge (\beta - \beta_l)/\mathrm{se}(\hat\beta_\mathrm{u})$. For a standard normal variate Z, it follows that

$$\mathrm{pr}\left(c_{\alpha^*/2} \le Z \le c_{\alpha^*/2} + \frac{\beta - \beta_l}{\mathrm{se}(\hat\beta_\mathrm{l})}\right) = \mathrm{pr}\left(-c_{\alpha^*/2} - \frac{\beta - \beta_l}{\mathrm{se}(\hat\beta_\mathrm{l})} \le Z \le -c_{\alpha^*/2}\right)$$

$$\geq \quad \text{pr}\left(-c_{\alpha^*/2} - \frac{\beta-\beta_l}{\text{se}(\hat{\beta}_\text{l})} - \frac{\beta_u-\beta}{\text{se}(\hat{\beta}_\text{u})} \leq Z \leq -c_{\alpha^*/2} - \frac{\beta_u-\beta}{\text{se}(\hat{\beta}_\text{u})}\right)$$
$$\geq \quad \text{pr}\left(-c_{\alpha^*/2} - \frac{\beta_u-\beta_l}{\text{se}(\hat{\beta}_\text{u})} \leq Z \leq -c_{\alpha^*/2} - \frac{\beta_u-\beta}{\text{se}(\hat{\beta}_\text{u})}\right)$$

where we subtract $(\beta_u - \beta)/\text{se}(\hat{\beta}_\text{u})$ from both sides in the second step and use the inequality $(\beta - \beta_l)/\text{se}(\hat{\beta}_\text{l}) \geq (\beta - \beta_l)/\text{se}(\hat{\beta}_\text{u})$ in the third step. Using (3.5) it follows that $\text{pr}(C_L \leq \beta) - \text{pr}(C_U \leq \beta) \geq \text{pr}(C_L \leq \beta_l) - \text{pr}(C_U \leq \beta_l)$ for arbitrary $\beta \in [\beta_l, \beta_u]$. When instead $\text{se}(\hat{\beta}_\text{l}) > \text{se}(\hat{\beta}_\text{u})$, then $(\beta - \beta_l)/\text{se}(\hat{\beta}_\text{l}) > (\beta - \beta_l)/\text{se}(\hat{\beta}_\text{u})$. A similar argument as before then shows that

$$\text{pr}\left(-c_{\alpha^*/2} - \frac{\beta_u-\beta}{\text{se}(\hat{\beta}_\text{u})} \leq Z \leq -c_{\alpha^*/2}\right) \geq \text{pr}\left(c_{\alpha^*/2} + \frac{\beta-\beta_l}{\text{se}(\hat{\beta}_\text{l})} \leq Z \leq c_{\alpha^*/2} + \frac{\beta_u-\beta_l}{\text{se}(\hat{\beta}_\text{l})}\right)$$

so that $\text{pr}(C_L \leq \beta) - \text{pr}(C_U \leq \beta) \geq \text{pr}(C_L \leq \beta_u) - \text{pr}(C_U \leq \beta_u)$ for arbitrary $\beta \in [\beta_l, \beta_u]$. We conclude that (3.2) holds. The result (4.3) is now immediate from (3.2), (3.3) and (3.4).

Note that $\text{pr}(C_U \leq \beta_l) \approx 0$ and $\text{pr}(C_L \leq \beta_u) \approx 1$ when $\beta_u - \beta_l$ is large. Hence, it follows from (3.2) that $c_{\alpha^*/2}$ approximates the $(1-\alpha)100\%$ percentile of the standard normal distribution when there is much ignorance about the complete data parameter of interest and the intended sample size is large.

Lemma 2 (Strong uncertainty intervals) *Let C_L and C_U be lower and upper $(1-\alpha)100\%$ confidence limits of β_l and β_u, satisfying $pr(C_L > \beta_l) = pr(C_U < \beta_u) = \alpha/2$. Then the interval $[C_L, C_U]$ has a strong coverage probability between $1 - \alpha/2$ and $1 - \alpha$.*

Proof: For univariate parameters, $(1-\alpha)100\%$ strong uncertainty intervals for β can be constructed by using that the strong coverage probability $1-\alpha$ equals $1 - \text{pr}\{\exists \beta \in \text{ir}(\beta, \Gamma) : (C_L > \beta) \vee (C_U < \beta)\} = 1 - \text{pr}\{(C_L > \beta_l) \vee (C_U < \beta_u)\}$. Hence

$$\text{pr}(C_L > \beta_l) + \text{pr}(C_U < \beta_u) - \text{pr}\{(C_L > \beta_l) \wedge (C_U < \beta_u)\} \quad = \quad \alpha$$

Choose $C_L = \hat{\beta}_l - c_{\alpha_\text{l}^*/2}\text{se}(\hat{\beta}_l)$ and $C_U = \beta_u + c_{\alpha_\text{u}^*/2}\text{se}(\hat{\beta}_u)$, the lower and upper $(1-\alpha^*)100\%$ confidence limits for β respectively, calculated at the estimated ignorance limits $\hat{\beta}_l$ and $\hat{\beta}_u$. Here, we define the critical values $c_{\alpha_\text{l}^*/2}$ and $c_{\alpha_\text{u}^*/2}$

such that $\mathrm{pr}(C_L > \beta_l) = \mathrm{pr}(C_U < \beta_u) = \alpha^*/2$. Because $\mathrm{pr}\{(C_L > \beta_l) \wedge (C_U < \beta_u)\} \leq \mathrm{pr}(C_L > \beta_l) = \alpha^*/2$ we find $\alpha \in [\alpha^*/2, \alpha^*]$. When - as expected - $\mathrm{pr}(C_L > \beta_l | C_U < \beta_u) \leq \mathrm{pr}(C_L > \beta_l)$, then $\mathrm{pr}(C_L > \beta_l \wedge C_U < \beta_u) \leq \mathrm{pr}(C_L > \beta_l)\mathrm{pr}(C_U < \beta_u) = \alpha^{*2}/4$. It follows under this assumption that $\alpha \in [\alpha^* - \alpha^{*2}/4, \alpha^*]$ and that the choice $\alpha^* = \alpha$ yields slightly conservative $(1-\alpha)100\%$ strong uncertainty regions.

Lemma 3 (Weak uncertainty intervals) *Let C_L and C_U be lower and upper $(1-\alpha^*)100\%$ confidence limits of β_l and β_u, respectively, based on CAN estimators for β_l and β_u and a critical value $c_{\alpha^*/2}$ that solves equation (5.2). Then the interval $[C_L, C_U]$ has an asymptotic weak coverage probability of $1-\alpha$.*

Proof: For univariate parameters, we construct $(1-\alpha)100\%$ weak uncertainty limits for β by using that

$$\begin{aligned}\frac{\mathrm{E}\|\mathrm{UR}_w(\beta,\Gamma) \cap \mathrm{ir}(\beta,\Gamma)\|}{\|\mathrm{ir}(\beta,\Gamma)\|} &= \frac{\mathrm{E}\{\min(\beta_u, C_U)\}}{\beta_u - \beta_l} - \frac{\mathrm{E}\{\max(\beta_l, C_L)\}}{\beta_u - \beta_l} \\ &+ \frac{\mathrm{E}(C_L - \beta_u | C_L > \beta_u)\mathrm{pr}(C_L > \beta_u)}{\beta_u - \beta_l} + \frac{\mathrm{E}(\beta_l - C_U | C_U < \beta_l)\mathrm{pr}(C_U < \beta_l)}{\beta_u - \beta_l}\end{aligned}$$

Defining $Z_u = (C_U - \beta_u)/\mathrm{se}(\hat{\beta}_\mathrm{u})$, we find that

$$\begin{aligned}\mathrm{E}\{\min(\beta_u, C_U)\} &= \int_{-\infty}^{\beta_\mathrm{u}} u f_{C_\mathrm{U}}(u)du + \int_{\beta_\mathrm{u}}^{+\infty} \beta_u f_{C_\mathrm{U}}(u)du \\ &= \int_{-\infty}^{0} \left\{\beta_u + z_u \mathrm{se}(\hat{\beta}_\mathrm{u})\right\} f_{Z_\mathrm{u}}(z_u)dz_u + \int_{0}^{+\infty} \beta_u f_{Z_\mathrm{u}}(z_u)dz_u \\ &= \beta_u + \mathrm{se}(\hat{\beta}_\mathrm{u}) \int_{-\infty}^{0} z_u f_{Z_\mathrm{u}}(z_u)dz_u\end{aligned}$$

and

$$\begin{aligned}\mathrm{E}(\beta_l - C_U | C_U < \beta_l)\mathrm{pr}(C_U < \beta_l) &= (\beta_l - \beta_u) \int_{-\infty}^{(\beta_\mathrm{l} - \beta_\mathrm{u})/\mathrm{se}(\hat{\beta}_\mathrm{u})} f_{Z_\mathrm{u}}(z_u)dz_u \\ &\quad -\mathrm{se}(\hat{\beta}_\mathrm{u}) \int_{-\infty}^{(\beta_\mathrm{l} - \beta_\mathrm{u})/\mathrm{se}(\hat{\beta}_\mathrm{u})} \mathrm{z_u f_{Z_u}(z_u) dz_u}\end{aligned}$$

Defining $Z_l = (C_L - \beta_l)/\mathrm{se}(\hat{\beta}_\mathrm{l})$, we find that

$$\mathrm{E}\{\max(\beta_l, C_L)\} = \int_{-\infty}^{\beta_\mathrm{l}} \beta_l f_{C_\mathrm{L}}(l)dl + \int_{\beta_\mathrm{l}}^{+\infty} l f_{C_\mathrm{L}}(l)dl$$

$$\begin{aligned} &= \int_{-\infty}^{0} \beta_l f_{Z_l}(z_l)dz_l + \int_{0}^{+\infty} \left\{\beta_l + z_l \mathrm{se}(\hat{\beta}_l)\right\} f_{Z_l}(z_l)dz_l \\ &= \beta_l + \mathrm{se}(\hat{\beta}_l) \int_{0}^{+\infty} z_l f_{Z_l}(z_l)dz_l \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}(C_L - \beta_u | C_L > \beta_u)\mathrm{pr}(C_L > \beta_u) &= (\beta_l - \beta_u) \int_{(\beta_u - \beta_l)/\mathrm{se}(\hat{\beta}_l)}^{+\infty} f_{Z_l}(z_l)dz_l \\ &\quad + \mathrm{se}(\hat{\beta}_l) \int_{(\beta_u - \beta_l)/\mathrm{se}(\hat{\beta}_l)}^{+\infty} \mathrm{z_l f_{Z_l}(z_l) dz_l} \end{aligned}$$

We thus obtain

$$\alpha = \frac{\mathrm{se}(\hat{\beta}_l)}{\beta_u - \beta_l} \int_{0}^{+\infty} z_l f_{Z_l}(z_l)dz_l - \frac{\mathrm{se}(\hat{\beta}_u)}{\beta_u - \beta_l} \int_{-\infty}^{0} z_u f_{Z_u}(z_u)dz_u + \epsilon$$

where

$$\begin{aligned} \epsilon &= \frac{\mathrm{E}(C_L - \beta_u | C_L > \beta_u)\mathrm{pr}(C_L > \beta_u)}{\beta_u - \beta_l} + \frac{\mathrm{E}(\beta_l - C_U | C_U < \beta_l)\mathrm{pr}(C_U < \beta_l)}{\beta_u - \beta_l} \\ &= \int_{-\infty}^{(\beta_l - \beta_u)/\mathrm{se}(\hat{\beta}_u)} f_{Z_u}(z_u)dz_u + \int_{(\beta_u - \beta_l)/\mathrm{se}(\hat{\beta}_l)}^{+\infty} f_{Z_l}(z_l)dz_l \\ &\quad + \frac{\mathrm{se}(\hat{\beta}_u)}{\beta_u - \beta_l} \int_{-\infty}^{(\beta_l - \beta_u)/\mathrm{se}(\hat{\beta}_u)} z_u f_{Z_u}(z_u)dz_u - \frac{\mathrm{se}(\hat{\beta}_l)}{\beta_u - \beta_l} \int_{(\beta_u - \beta_l)/\mathrm{se}(\hat{\beta}_l)}^{+\infty} z_l f_{Z_l}(z_l)dz_l \end{aligned}$$

This is a very small correction term because the events $C_L > \beta_u$ and $C_U < \beta_l$ are extremely rare (unless there is little ignorance about the target parameter and/or the sample size is very small). Define $C_L = \hat{\beta}_l - c_{\alpha^*/2}\mathrm{se}(\hat{\beta}_l)$ and $C_U = \hat{\beta}_u + c_{\alpha^*/2}\mathrm{se}(\hat{\beta}_u)$ for some unknown $c_{\alpha^*/2}$. Then, we show how $c_{\alpha^*/2}$ can be found. For CAN estimators $\hat{\beta}_l$ and $\hat{\beta}_u$, Z_l and Z_u follow asymptotically normal distributions with mean $-c_{\alpha^*/2}$ and $c_{\alpha^*/2}$, respectively, as the sample size approaches infinity. Indeed,

$$Z_l = \frac{\hat{\beta}_l - \beta_l - c_{\alpha^*/2}\mathrm{se}(\hat{\beta}_l)}{\mathrm{se}(\hat{\beta}_l)} = \frac{\hat{\beta}_l - \beta_l}{\mathrm{se}(\hat{\beta}_l)} - c_{\alpha^*/2}$$

where $(\hat{\beta}_l - \beta_l)/\mathrm{se}(\hat{\beta}_l)$ converges to a standard normal distribution in law. A $(1-\alpha)100\%$ weak uncertainty interval can now be estimated by solving $c_{\alpha^*/2}$ from equation (5.2). In this way, we obtain a weak uncertainty interval with asymptotic coverage probability $1-\alpha$.

A.2. Proof of Theorem 1

Given that all individual point estimators in $\hat{\text{ir}}_N(\boldsymbol{\beta}, \Gamma)$ are weakly consistent uniformly over $\boldsymbol{\gamma} \in \Gamma$

$$(\forall \epsilon, \delta > 0)(\exists N_0(\epsilon, \delta))(N > N_0(\epsilon, \delta) \Rightarrow \text{pr}\{\sup_{\gamma \in \Gamma} \|\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma}) - \boldsymbol{\beta}(\boldsymbol{\gamma})\| < \delta\} > 1 - \epsilon)$$

Since

$$\begin{aligned} \sup_{\hat{\beta}_N(\gamma) \in \hat{\text{ir}}_N(\beta, \Gamma)} \inf_{\beta(\gamma) \in \text{ir}(\beta, \Gamma)} \|\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma}) - \boldsymbol{\beta}(\boldsymbol{\gamma})\| &\leq \sup_{\gamma \in \Gamma} \|\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma}) - \boldsymbol{\beta}(\boldsymbol{\gamma})\| \\ \sup_{\beta(\gamma) \in \text{ir}(\beta, \Gamma)} \inf_{\hat{\beta}_N(\gamma) \in \hat{\text{ir}}_N(\beta, \Gamma)} \|\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma}) - \boldsymbol{\beta}(\boldsymbol{\gamma})\| &\leq \sup_{\gamma \in \Gamma} \|\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma}) - \boldsymbol{\beta}(\boldsymbol{\gamma})\| \end{aligned}$$

This implies that

$$\begin{aligned} & (\forall \epsilon, \delta > 0)(\exists N_0(\epsilon, \delta)) \left\{ N > N_0(\epsilon, \delta) \Rightarrow \text{pr} \left(\sup_{\hat{\beta}_N(\gamma) \in \hat{\text{ir}}_N(\beta, \Gamma)} \inf_{\beta(\gamma) \in \text{ir}(\beta, \Gamma)} \|\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma}) - \boldsymbol{\beta}(\boldsymbol{\gamma})\| < \delta \right.\right. \\ & \left.\left. \wedge \sup_{\beta(\gamma) \in \text{ir}(\beta, \Gamma)} \inf_{\hat{\beta}_N(\gamma) \in \hat{\text{ir}}_N(\beta, \Gamma)} \|\hat{\boldsymbol{\beta}}_N(\boldsymbol{\gamma}) - \boldsymbol{\beta}(\boldsymbol{\gamma})\| < \delta \right) > 1 - \epsilon \right\} \\ \Rightarrow \quad & (\forall \epsilon, \delta > 0)(\exists N_0(\epsilon, \delta)) \left(N > N_0(\epsilon, \delta) \Rightarrow \text{pr}(\|\hat{\text{ir}}_N(\boldsymbol{\beta}, \Gamma) - \text{ir}(\boldsymbol{\beta}, \Gamma)\| < \delta) > 1 - \epsilon \right). \end{aligned}$$

References

Cochran, W. (1977). *Sampling Techniques (3rd ed.)*. New York: Wiley.

Copas, J.B. and Li, H.G. (1997). Inference for non-random samples (with discussion). *J. Royal Statist. Soc. B* **59**, 55-77.

Diggle, P. and Kenward, M.G. (1994). Informative drop-out in longitudinal data-analysis (with discussion). *Appl. Statist.* **43**, 49-93.

Goetghebeur, E., Molenberghs, G. and Kenward, M.G. (1999). Sense and sensitivity when intended data are missing. *Kwantitatieve technieken* **62**, 79-94.

Horowitz, J.L. and Manski, C.F. (2000). Nonparametric Analysis of Randomized Experiments With Missing Covariate and Outcome Data. *J. Am. Statist. Assoc.* **95**, 77-88.

Horowitz, J.L., Manski, C.F., Ponomareva, M. and Stoye, J. (2003). Computation of Bounds on Population Parameters when the Data are Incomplete. *Reliable Computing* **9**, 419-440.

Imbens, G.W. and Manski, C.F. (2004). Confidence Intervals for Partially Identified Parameters. *Econometrica* **72**, 1845-1857.

Joffe, M.M. (2001). Using information on realized effects to determine prospective causal effects. *J. Royal Statist. Soc. B* **63**, 759-774.

Kenward, M.G., Molenberghs, G. and Goetghebeur, E. (2001). Sensitivity analysis for incomplete categorical data. *Statistical Modelling* **1**, 31-48.

Kooreman, P. (1993). Bounds on the Regression-Coefficients when a Covariate is Categorized. *Comm. Statist. - Theory and Methods* **22**, 2373-2380.

Little, R.J. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Molenberghs, G., Kenward, M.G. and Goetghebeur, E. (2001). Sensitivity analysis for incomplete contingency tables. *Appl. Statist.* **50**, 15-29.

Newey, W.K. and McFadden, D. (1994). *Large Sample Estimation and Hypothesis Testing.* Handbook of Econometrics Vol. 4. Elsevier: Amsterdam.

Nordheim, E.V. (1984). Inference from nonrandomly missing categorical data: an example from a genetic study on Turner's syndrome. *J. Am. Statist. Assoc.* **79**, 772-780.

Robins, J.M. (1989). *The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies*, in Health Service Research Methodology: A focus on AIDS, pp. 113-159. NCHSR, U.S. Public Health Service.

Robins, J.M. (1997). Non-response models for the analysis of non-monotone non-ignorable missing data. *Statist. Med.* **16**, 21-37.

Rosenbaum, P.R. (1995). Quantiles in nonrandom samples and observational studies. *J. Am. Statist. Assoc.* **90**, 1424-1431.

Rubin, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581-592.

Scharfstein, D.O., Robins, J.M. and Rotnitzky, A. (1999). Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models (with discussion). *J. Am. Statist. Assoc.* **94**, 1096-1146.

Scharfstein, D.O., Daniels, M.J. and Robins, J.M. (2003). Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. *Biostatistics* **4**, 495-512.

Van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge: Cambridge University Press.

Vansteelandt, S. and Goetghebeur, E. (2001). Generalized Linear Models with Incomplete Outcomes: the IDE Algorithm for Estimating Ignorance and Uncertainty. *J. Graph. Comput. Statist.* **10**, 656-672.

Vansteelandt, S. and Goetghebeur, E. (2005). Sense and sensitivity when correcting for observed exposures in randomized clinical trials. *Statist. Med.* **24**, 191-210.

Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E. and Kenward, M.G. (2001). Sensitivity analysis for nonrandom dropout: A local influence approach. *Biometrics* **57**, 7-14.

Verstraeten, T., Farah, B., Duchateau, L. and Matu, R. (1998). Pooling sera to reduce the cost of HIV surveillance: a feasibility study in a rural Kenyan district. *Trop. Med. Int. Health* **3**, 747-750.

Department of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281, S9, 9000 Ghent, Belgium

E-mail: Stijn.Vansteelandt@ugent.be

Department of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281, S9, 9000 Ghent, Belgium; and Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston MA 02115, U.S.A.

E-mail: Els.Goetghebeur@ugent.be

Medical Statistics Unit, London School of Hygiene and Tropical Medicine, 129 Keppel Street, London WC1E 7HT, U.K.

E-mail: Mike.Kenward@lshtm.ac.uk

Center for Statistics, Limburgs Universitair Centrum, Universitaire Campus - building D, 3590 Diepenbeek, Belgium

E-mail: Geert.Molenberghs@luc.ac.be

Table 1. Simulation results: Estimated coverage probabilities (two-sided P-values of the null hypothesis of no bias in the estimates), average length ($\overline{\hat{C}_{\mathrm{U}} - \hat{C}_{\mathrm{L}}}$) of the uncertainty interval and average adjusted critical values ($\overline{c}_{\mathbb{R}^{*=2}}$).

Type	Coverage	Average Length	$\overline{c}_{\mathbb{R}^{*=2}}$
Strong	0.949 (0.293)	0.332	1.960
Weak	0.955 (0)	0.244	0.803
Pointwise	0.948 (0.221)	0.308	1.645

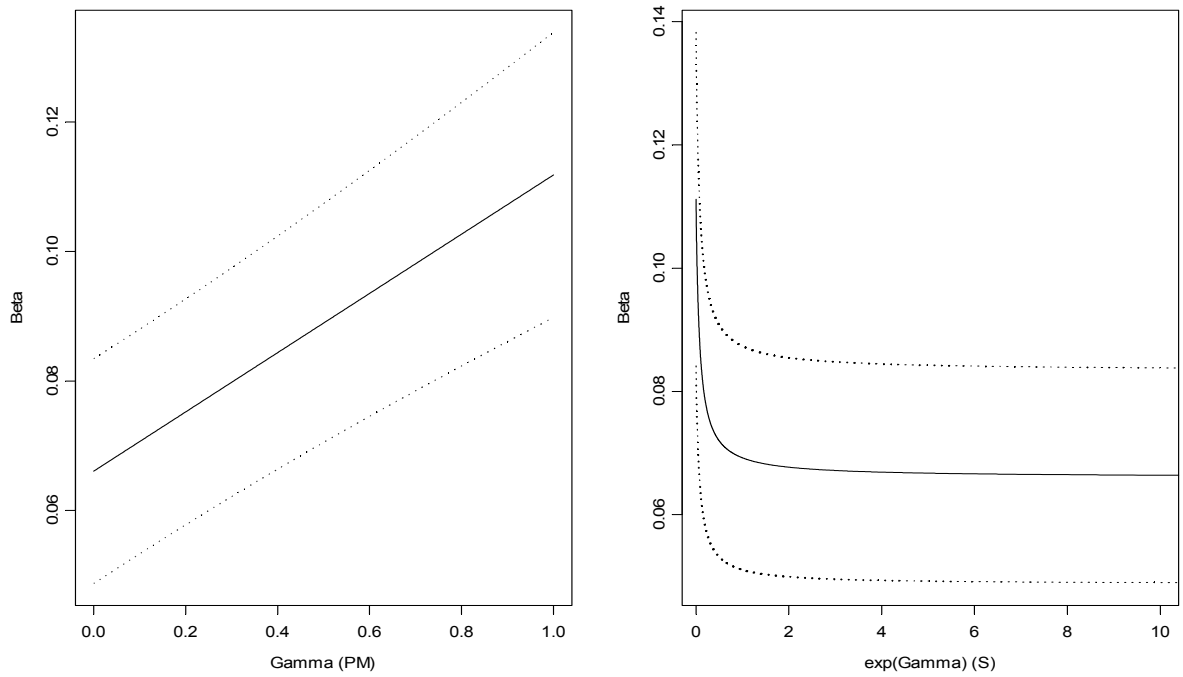

Figure 1. Estimates (solid line) and 95% confidence intervals (dotted lines) for $\beta_0 = \mathrm{pr}(Y = 1)$ in function of $\gamma_{\mathrm{pm}} = \mathrm{pr}(Y = 1|R = 0)$ (left) and the odds ratio $\exp(\gamma_{\mathrm{s}})$ of response for HIV positives to HIV negatives (right).

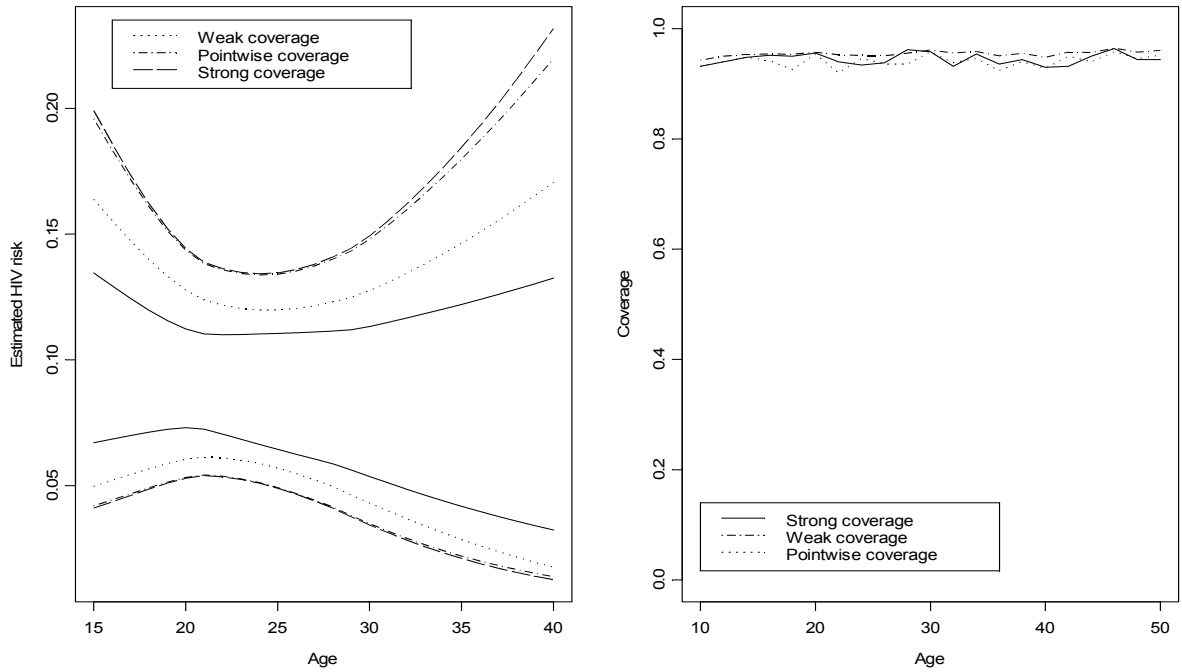

Figure 2. Left: HEIRs (solid lines) and 95% EUROs (dotted lines) for age-specific HIV risk; Right: Bootstrap-based estimated coverage of 95% EUROs for age-specific HIV risk.