BIOMETRICS 59, 1–?? DOI: 10.1111/j.1541-0420.2005.00454.x
December 2006

Web-based Supplementary Materials for Instrumental variables estimation of exposure effects on a time-to-event endpoint using structural cumulative survival models

by Torben Martinussen, Stijn Vansteelandt, Eric J. Tchetgen Tchetgen and David M. Zucker

This paper has been submitted for consideration for publication in *Biometrics*

1. Appendix: Large sample properties

Let $\mu(L;\theta) = E(G|L;\theta)$ be the conditional mean of the instrument given observed confounders L, which is function of an unknown finite-dimensional parameter θ. In the case of no observed confounders $\mu(\theta) = \theta = E(G)$ and $\hat{\theta} = \overline{G}$. We assume that $n^{1/2}(\hat{\theta} - \theta) = n^{-1/2}\sum_i \epsilon_i^\theta + o_p(1)$, where the ϵ_i^θ's are zero-mean iid variables. In the case of no observed confounders we have $\epsilon_i^\theta = G_i - \theta$. Let θ_0 denote the true value of θ.

We write $\|g\|_\infty = \sup_{t\in[0,\tau]} |g(t)|$ and use the notation $\mathcal{V}(g)$ to denote the total variation of g over the interval $[0,\tau]$. Let $B^\circ(t)$ denote the true value of $B(t)$, and let $M^\circ = \|B^\circ\|_\infty < \infty$.

Technical conditions:

(i) We assume that X and G are bounded, and denote the respective bounds by X_{max} and G_{max}.

(ii) Define $a(s,h) = E[R(s)XG^c e^{hX}]$. We assume that there exist $M > M^\circ$ and $\nu > 0$ such that $\inf_{s\in[0,\tau],h\in[-M,M]} a(s,h) \geqslant 1.01\nu$.

The quantities M° and M do not necessarily need to be known.

1.1 *Consistency*

Below we show that $\hat{B}_X(t,\theta_0)$ is uniformly consistent. In what follows we suppress θ_0 from the notation and write $B(t)$ instead of $B_X(t)$. The estimator is given by the recursion equation

$$\hat{B}_n(t) = \int_0^t \frac{\sum_i G_i^c e^{\hat{B}_n(s-)X_i} dN_i(s)}{\sum_i R_i(s)X_i G_i^c e^{\hat{B}_n(s-)X_i}} \tag{1}$$

It appears difficult to prove directly that $\hat{B}_n(t)$ is bounded. Instead we will take a different approach. We will modify the estimator in a way that will force it to be of bounded variation. We will then prove that the modified version of the estimator is consistent. If M is not known, the modified estimator is a theoretical construct that cannot actually be computed, but it will emerge that for large enough n the modified estimator is equal to the unmodified estimator.

We will use the Helly Selection Theorem in the following form.

Helly Selection Theorem: Let $\{f_n\}$ be a sequence of functions on $[0,\tau]$ such that $\|f_n\|_\infty \leqslant A_1$ and $\mathcal{V}(f) \leqslant A_2$, where A_1 and A_2 are finite constants. Then

a. There exists a subsequence $\{f_{n_j}\}$ of $\{f_n\}$ which converges pointwise to some function f.

b. If f is continuous, the convergence is uniform.

Then it follows that $\|\hat{B}_n - B^\circ\|_\infty \overset{\text{a.s.}}{\rightarrow} 0$.

Proof: For a function $H(t)$ on $[0,\tau]$, define

$$\Upsilon_n(H,t) = \int_0^t \frac{n^{-1}\sum_i G_i^c e^{H(s-)X_i} dN_i(s)}{A(s,H(s-))} \tag{2}$$

$$\Upsilon(H,t) = \int_0^t \frac{c(s,H(s))}{a(s,H(s))}\, ds \tag{3}$$

where

$$A(s,h) = \frac{1}{n}\sum_{i=1}^n R_i(s) X_i G_i^c e^{hX_i} \tag{4}$$

$$c(s,h) = E[R(s) G^c e^{hX} \lambda(s,L,G,X)] \tag{5}$$

with $\lambda(s,L,G,X) = (d/ds)E[N(s)|L,G,X]$, so that $E[R(s)G^c e^{hX} dN(s)] = c(s,h)ds$. The estimator $\hat{B}_n(t)$ is then the solution to $B(t) = \Upsilon_n(B,t)$. Let $\xi(y) = \operatorname{sgn}(y)\min(|y|,M)$. We then define the modified estimator $\tilde{B}_n$ to be the solution to the equation $B(t) = \Upsilon_n(\xi(B),t)$. Note that $\Upsilon(\xi(B^\circ),t) = \Upsilon(B^\circ,t) = B^\circ(t)$.

Define $q(s,h) = c(s,\xi(h))/a(s,\xi(h))$, so that

$$\Upsilon(\xi(H),t) = \int_0^t q(s,H(s))ds$$

The function $q(s,h)$ satisfies $\sup_{s\in[0,\tau],h\in\mathbb{R}} |q(s,h)| \leqslant 2G_{max} e^{MX_{max}} \lambda_{max} \nu^{-1}$, where λ_{max} is an upper bound on $\lambda(s,L,G,X)$ (which we assume exists). Moreover, $q(s,h)$ is Lipschitz with respect to h over $s \in [0,\tau]$ and $h \in \mathbb{R}$ with Lipschitz constant $\kappa = 2G_{max} e^{MX_{max}} \lambda_{max} \nu^{-1}(1 + X_{max} G_{max} e^{MX_{max}} \nu^{-1})$. Accordingly, by classical differential equations theory (Hartman, 1973, Thm. 1.1; Coddington, 1989, Sec. 5.8), B° is the *unique* solution to the equation $B(t) = \Upsilon(\xi(B),t)$ subject to $B(0) = 0$.

We note for later reference that for any two functions B_1 and B_2 we have

$$\|\Upsilon(\xi(B_1)) - \Upsilon(\xi(B_2))\|_\infty \leqslant \kappa\tau\|B_1 - B_2\|_\infty \tag{6}$$

Now, by the functional central limit theorem as given in Andersen and Gill (1982),

$$\sup_{s\in[0,\tau],h\in[-M,M]} |A(s,h) - a(s,h)| \stackrel{a.s.}{\to} 0 \tag{7}$$

Accordingly, from the the assumption that $\inf_{s\in[0,\tau],h\in[-M,M]} a(s,h) \geqslant 1.01\nu$, we get the result that $\inf_{s\in[0,\tau],h\in[-M,M]} A(s,h) \geqslant \nu$ for n sufficiently large. We thus find that the jumps in $\tilde{B}_n(t)$ are bounded by $n^{-1}D$ with $D = 2G_{max}e^{MX_{max}}/\nu$, implying that $\|\tilde{B}_n\|_\infty \leqslant D$ and $\mathcal{V}(\tilde{B}_n) \leqslant D$. Let $\mathbb{B}^*$ denote the class of functions $B(t)$ with these two properties. Further, let $\mathbb{H}$ denote the class of functions that are bounded by $\tilde{M} = \min(M, D)$ and have total variation less than D. Since $|\xi(y)| \leqslant |y|$ and ξ is Lipschitz(1), we find that $B \in \mathbb{B}^*$ implies that $\xi(B) \in \mathbb{H}$.

Next, define

$$\tilde{\Upsilon}_n(H,t) = \int_0^t \frac{n^{-1}\sum_i G_i^c e^{H(s-)X_i} dN_i(s)}{a(s, H(s-))} \tag{8}$$

From (7) it follows that

$$\sup_{s\in[0,\tau],H\in\mathbb{H}} |\Upsilon_n(H,s) - \tilde{\Upsilon}_n(H,s)| \stackrel{\text{a.s.}}{\to} 0 \tag{9}$$

For $U = (T, \delta, X, L, G)$, define

$$\psi_{H,t}(U) = \frac{\delta G^c e^{H(T-)X}}{a(T, H(T-))} \tag{10}$$

We then have $\tilde{\Upsilon}_n(H,t) = \mathbb{P}_n\psi_{H,t}$. We claim that the class of functions $\mathcal{F} = \{\psi_{H,t},\, H \in \mathbb{H}, t \in [0,\tau]\}$ is Donsker. This result is an immediate consequence of the following facts:

1. Sums and products of bounded Donsker classes are also Donsker.

2. For any finite K, the class of monotone functions mapping $[0,\tau]$ to $[-K,K]$ is Donsker (Kosorok, 2008, Thm. 9.24).

3. If H is bounded and has bounded variation, then H can be written as $H = H_1 - H_2$, where H_1 and H_2 are monotone increasing functions with $\|H_1\|_\infty \leqslant \|H\|_\infty + \mathcal{V}(H)$

and $\|H_2\|_\infty \leqslant \mathcal{V}(H)$ (Jordan decomposition). It follows that the class of functions H with $\|H\|_\infty \leqslant C_1$ and $\mathcal{V}(H) \leqslant C_2$ is Donsker.

4. If $H \in \mathbb{H}$, then the function $g(t) = a(t, H(t-)) = E[R(t)XG^c e^{H(t-)X}]$ is bounded and of bounded variation with $\|g\|_\infty \leqslant 2G_{max}e^{\tilde{M}X_{max}}$ and

$$\mathcal{V}(g) \leqslant 2X_{max}G_{max}e^{\tilde{M}X_{max}}(\mathcal{V}(r) + \mathcal{V}(H)),$$

where $r(s) = E[R(s)]$.

It follows that

$$\sup_{t\in[0,\tau],H\in\mathbb{H}} |\tilde{\Upsilon}_n(H,t) - \Upsilon(H,t)| \overset{\text{a.s.}}{\to} 0 \tag{11}$$

and therefore

$$\sup_{t\in[0,\tau],H\in\mathbb{H}} |\Upsilon_n(H,t) - \Upsilon(H,t)| \overset{\text{a.s.}}{\to} 0 \tag{12}$$

Now, by Helly's selection theorem, every subsequence of $\tilde{B}_n(t)$ has a further subsequence that converges to some limit. Since the jumps $\tilde{B}_n(t)$ are bounded by $n^{-1}D$ and the number of jumps in the interval $[t_1, t_2]$ divided by n converges uniformly to $E[N(t_2)] - E[N(t_1)] \leqslant C(t_2 - t_1)$ for some constant C, it follows that the limit of the sub-subsequence is continuous, and therefore (by the second part of Helly's theorem) the convergence of the sub-subsequence is uniform. Going further, the fact that $\tilde{B}_n = \Upsilon_n(\xi(\tilde{B}_n))$ in combination with (6) and (12) implies that the limit B of the sub-subsequence satisfies $B = \Upsilon(\xi(B))$. But we said before that B° is the unique continuous solution to this equation. We thus find that every subsequence of $\tilde{B}_n$ has a further subsequence that converges uniformly to B°. Consequently, $\tilde{B}_n$ itself converges uniformly to B°. Since $B^\circ \leqslant M^\circ$ and $\|\tilde{B}_n - B^\circ\|_\infty \overset{\text{a.s.}}{\to} 0$ (as just stated), for sufficiently large n we have $\|\tilde{B}_n\|_\infty \leqslant M^\circ + \frac{1}{2}(M - M^\circ)$ and therefore $\xi(\tilde{B}_n(t)) = \tilde{B}_n(t)$. So for n sufficiently large, $\tilde{B}_n$ solves $B = \Upsilon_n(B)$, or, in other words $\tilde{B}_n = \hat{B}_n$. We have thus shown that $\|\hat{B}_n - B^\circ\|_\infty \overset{\text{a.s.}}{\to} 0$, as desired.

The consistency of $\hat{B}_X(t, \hat{\theta})$ then follows immediately by a Taylor series expansion since $\hat{\theta}$ is consistent.

1.2 *Asymptotic normality*

Let $N(t) = \{N_1(t), \dots N_n(t)\}^T$ and $X = (X_1, \dots, X_n)$. For known θ we can write

$$\hat{B}_X(t,\theta) = \int_0^t H_\theta\{s, \hat{B}_X(s-,\theta)\} dN(s),$$

where the kth element of the n-vector $H_\theta\{t, \hat{B}_X(t-,\theta)\}$ is

$$\{G_k - \mu(L_k;\theta)\} e^{\hat{B}_X(t-,\theta)X_k} / \sum_{i=1}^n \{G_i - \mu(L_i;\theta)\} R_i(t) e^{\hat{B}_X(t-,\theta)X_i} X_i.$$

Let $V(t,\theta) = n^{1/2}\{\hat{B}_X(t,\theta) - B_X(t)\}$ and let $\dot{H}$ denote the derivative of H with respect to its second argument. It is then easy to see that

$$\begin{aligned} V(t,\theta) =& n^{1/2} \int_0^t H(s, B_X(s-)) \left[dN(s) - X dB_X(s) \right] \\ &+ \int_0^t V(s-,\theta)\{1 + o_p(1)\} \dot{H}(s, B_X(s-)) dN(s) \end{aligned}$$

which is a Volterra-equation, see Andersen et al. (1993), p. 91. The solution to this equation is given by

$$V(t,\theta) = \int_0^t \mathcal{F}(s,t) n^{1/2} H(s, B_X(s-)) \left[dN(s) - X dB_X(s) \right] + o_p(1),$$

where

$$\mathcal{F}(s,t) = \prod_{(s,t]} \left\{ 1 + \dot{H}(\cdot, B_X(\cdot)) dN(\cdot) \right\}$$

with the latter being a product integral that converges in probability to some limit. This leads to the iid-representation

$$V(t,\theta) = n^{-1/2} \sum_{i=1}^n \epsilon_i^B(t)$$

with the $\epsilon_i^B(t)$'s being zero-mean iid terms. Specifically

$$\epsilon_i^B(t) = \int_0^t \mathcal{F}(s,t) n^{1/2} \{H(s, B_X(s-))\}_i \left[dN(s) - X dB_X(s) \right]_i$$

with a_i being the ith element of the vector a. This together with

$$\begin{aligned} n^{1/2}\{\hat{B}_X(t,\hat{\theta}) - B_X(t)\} &= n^{1/2}\{\hat{B}_X(t,\theta) - B_X(t)\} + n^{1/2}\{\hat{B}_X(t,\hat{\theta}) - \hat{B}_X(t,\theta)\} \\ &= n^{1/2}\{\hat{B}_X(t,\theta) - B_X(t)\} + D_\theta(\hat{B}_X(t,\theta))_{|\hat{\theta}} n^{1/2}(\hat{\theta} - \theta) + o_p(1), \end{aligned}$$

where $D_\theta\{\hat{B}_X(t,\theta)\}$ is the first order derivative of $\hat{B}_X(t,\theta)$ w.r.t. θ gives an iid-decomposition of $n^{1/2}\{\hat{B}_X(t,\hat{\theta}) - B_X(t)\}$:

$$n^{1/2}\{\hat{B}_X(t,\hat{\theta}) - B_X(t)\} = n^{-1/2}\sum_{i=1}^{n}\epsilon_i^B(t,\theta) + o_p(1),$$

where

$$\epsilon_i^B(t,\theta) = \epsilon_i^B(t) + D_\theta(\hat{B}_X(t,\theta))_{|\theta}\epsilon_i^\theta. \tag{13}$$

We now argue that the process $V(t,\theta)$ converges in distribution as a process using arguments similar to what is done in Lin et al. (2000. p. 726). By taking the log to equation (4) in the main paper it is seen that $B_X(t)$ can be written as a difference of two monotone functions . Let $\tilde{H}_i(s)$ be the limit in probability of $\mathcal{F}(s,t)H_i(s,B_X(s-))$. Now, split $\tilde{H}_i(s)$ into its positive and negative parts, $\tilde{H}_i^+(s)$ and $\tilde{H}_i^-(s)$, and similarly with X_i, X_i^+ and X_i^-. Then $\int_0^t \tilde{H}_i(s)[dN_i(s) - X_i dB_X(s)]$ can be written as a difference of two monotone functions, and then we follow the arguments of Lin et al. (2000) (or use example 2.11.16 of van der Vaart and Wellner, 1996). Convergence in distribution for the process $V(t,\hat{\theta})$ also holds using the above Taylor expansion. It thus follows that

$$n^{1/2}\{\hat{B}_X(t,\hat{\theta}) - B_X(t)\}$$

converges to a zero-mean Gaussian process with a variance that is consistently estimated by

$$n^{-1}\sum_{i=1}^{n}\hat{\epsilon}_i^B(t,\hat{\theta})^2.$$

The derivative $D_\theta(\hat{B}_X(t,\theta))_{|\hat{\theta}}$ can be calculated recursively as $\hat{B}_X(t,\hat{\theta})$ is constant between the observed death times. Denote the jump times by $\tau_1,\ldots,\tau_m$. Hence

$$\hat{B}_X(\tau_j,\theta) = \hat{B}_X(\tau_{j-1},\theta) + d\hat{B}_X(\tau_j,\theta)$$

which then also holds for the derivative. Since $\hat{B}_X(0,\theta) = 0$ and the derivative of the increment in the first jump time, $d\hat{B}_X(\tau_1,\theta)$, is easily calculated we then have a recursive way of calculating the derivatives of $\hat{B}_X(\cdot,\theta)$.

2. Additional simulation results

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

3. Application to the HIP trial on effectiveness of screening on breast cancer mortality: additional results

[Table 4 about here.]

[Figure 1 about here.]

4. Alternative censoring condition

We assume in this section that the censoring time satisfies the following condition

$$\tilde{T} \perp\!\!\!\perp C|X, G, L \text{ and } P(C > t|X, G, L) = P(C > t|L) \qquad \text{(C)}$$

Recall that

$$\exp\{-B_X(t)x\} = \frac{P(\tilde{T} > t|X = x, G, L)}{P(\tilde{T}^0 > t|X = x, G, L)}. \qquad (14)$$

Proposition 1: Assume model (14) with the assumption that G is an instrumental variable, conditional on L, so that the causal diagram in Figure 1 holds, and further that the censoring time satisfies condition (C). Then

$$E\left[\{G - E(G|L)\} e^{B_X(t)X} R(t) \{dN(t) - dB_X(t)X\}\right] = 0, \qquad (15)$$

for each t.

Proof

First note that display (14) implies the following relationship between the conditional hazard functions $\lambda_{\tilde{T}^0}(t|X, G, L)$ and $\lambda_{\tilde{T}}(t|X, G, L)$:

$$\lambda_{\tilde{T}^0}(t|X, G, L)dt = \lambda_{\tilde{T}}(t|X, G, L)dt - dB_X(t)X.$$

It then follows by the independent censoring assumption (C) that

$$
\begin{aligned}
&E\left[\{G - E(G|L)\}\, e^{B_X(t)X} R(t)\, \{dN(t) - dB_X(t)X\}\right] \\
&= E\left[\{G - E(G|L)\}\, e^{B_X(t)X} I(C > t)\tilde{R}(t)\lambda_{\tilde{T}^0}(t|X, G, L)dt\right] \\
&= E\left[P(C > t|L)\, \{G - E(G|L)\}\, f_{\tilde{T}^0}(t|X, G, L)dt\right] \\
&= -E\left[P(C > t|L)\, \{G - E(G|L)\}\, \frac{d}{dt}P(\tilde{T}^0 > t|X, G, L)dt\right] \\
&= -\frac{d}{dt}E\left[P(C > t|L)\, \{G - E(G|L)\}\, P(\tilde{T}^0 > t|X, G, L)dt\right] \\
&= 0
\end{aligned}
$$

because, for any function $g_t(L)$, we have

$$
E\left[g_t(L)\, \{G - E(G|L)\}\, P(\tilde{T}^0 > t|X, G, L)dt\right] = 0
$$

since $G \perp\!\!\!\perp \tilde{T}^0|L$. This completes the proof.

5. R code

In the following we provide R code that simulate data and run the IV-estimation. The naive analysis is also carried out and both are plotted.

```
library(Rcpp)
library(survival)
library(timereg)
library(lava)

source("http://192.38.117.59/~tma/Rcode/sim_setup1.R")    # Simulate data and contains R-function
                                                          # running the IV-estimation
source("http://192.38.117.59/~tma/Rcode/test_IV_function.R")   # Contains R-function that also allows covariates L

#  X       # Exposure
#  G       # Instrument
#  L       # Covariates
#  time    # Time-variable
#  status  # Status-variable
```

```
length(time)
cbind(time,status,L,G,X)[1:10,] # First 10 lines of data

#Observational analysis (naive analysis):
fit.aalen=aalen(Surv(time,status==1)~X+G+L[,1:dim(L)[2]],max.time=3,n.sim=1000)
summary(fit.aalen)
# Naive estimate is plotted further below along with IV-estimate.


## IV-estimation:

fit.iv=iv_est1(time,status,G,X,L,max.time=3,max.time.bet=2.8,n.sim=1000,G_type=0)
# This runs the IV-analysis. 3 years of follow-up; the constant effect is estimated based on 2.8 years of follow-up
# 1000 resamplings are requested for testing goodness-of-fit of constant effects model.
# G_type gives the type of the instrument; 1 is continuous; 0 is binary
# The conditional mean E(G|L) is fitted using a linear model when G cont, and logistic regression when G binary

summary(fit.iv)

names(fit.iv) # res1 contains various results from the IV-analysis:
            # stime: is the ordered event times within (0,max.time)
            # B: is the IV-estimator \hat B_X(t)
            # se_B: the estimated standard errors of \hat B_X(t)
            # pval_0: p-value corresponding to supremum test of the null B_X(t)=0.
            # eps_B: is the iid-decomposition of \sqrt{n}(\hat B_X(t) - B_X(t))
            # beta: is the IV constant effects estimator \hat \beta_X
            # se_beta: the estimated standard error of \hat \beta_X
            # pval_beta: p-value corresponding to the null \beta_X=0.
            # pval_GOF_sup: p-value corresponding to supremum test of the null that const. effects model is ok.
            # pval_GOF_CvM: as pval_GOF_sup but now based on the Cramer Von Mises test statistic
            # GOF.resamp: a matrix with first row the ordered jump times in (0,max.time.bet),
            # second row the observed test process, and the remaining rows are 50 processes sampled under the null.


## Ploting estimates
par(mfrow=c(1,2))
plot(fit.iv)
lines(fit.aalen$cum[,1],fit.aalen$cum[,3],lty=2,type="s")
# Estimated constant effect  is the straight line (true value for this simulation is slope=0.1)
# Iv-estimate is given along with 95% pointwise conf.bands. (full lines)
```

```
# Naive estimate is given as dashed curve.

## GOF for const. eff model
plot(fit.iv,gof=T)
```

References

Andersen, P. K., and Gill, R. D. (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *Annals of Statistics*, **10**, 1100-1120.

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Berlin: Springer-Verlag.

Coddington, E. A. (1989). *An Introduction to Ordinary Differential Equations.* Mineola: Dover

Hartman, P. (1973). *Ordinary Differential Equations*, 2nd ed. (reprinted, 1982), Boston: Birkhauser.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Berlin: Springer-Verlag.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society - Series B*, **62**, 711-730.

test_proc_two_const.pdf

Figure 1. HIP-study. Observed goodness-of-fit test process $TST(t)$ (thick curve) along with 20 resampled processes under the null.

Table 1

Summary of simulations concerning the constant parameter estimator $\hat{\beta}_X$. Binary and continuous exposure case. Bias of $\hat{\beta}_X$, average estimated standard error, sd($\hat{\beta}_X$), empirical standard error, see($\hat{\beta}_X$), and coverage probability of 95% pointwise confidence intervals CP($\hat{\beta}_X$)) based on the instrumental variables estimator, in function of sample size n and at different strengths ρ (correlation) of the instrumental variable. Results for the 2SLS estimator $\check{\beta}_X$ of Tchetgen et al. (2015) are also given.

Continuous X	(n, ρ)			
	(1600,0.3)	(3200,0.3)	(800,0.5)	(1600,0.5)
Bias $\hat{\beta}_X$	-0.002	-0.004	-0.003	0.001
sd ($\hat{\beta}_X$)	0.107	0.074	0.082	0.057
see ($\hat{\beta}_X$)	0.113	0.073	0.084	0.057
95% CP($\hat{\beta}_X$)	97.2	95.5	96.1	95.5
Bias $\check{\beta}_X$	0.003	-0.001	-0.003	0.001
sd ($\check{\beta}_X$)	0.098	0.068	0.075	0.053
Binary X	(n, ρ)			
	(3200,0.3)	(6400,0.3)	(1600,0.5)	(3200,0.5)
Bias $\hat{\beta}_X$	-0.002	-0.004	-0.003	-0.000
sd ($\hat{\beta}_X$)	0.085	0.061	0.082	0.056
see ($\hat{\beta}_X$)	0.088	0.062	0.081	0.057
95% CP($\hat{\beta}_X$)	96.2	95.4	95.5	95.4
Bias $\check{\beta}_X$	0.001	-0.001	-0.001	-0.002
sd ($\check{\beta}_X$)	0.072	0.050	0.068	0.048

Table 2

Continuous exposure case. Time-dependent exposure effect. Bias of $\hat{B}_X(t)$, average estimated standard error, sd($\hat{B}_X(t)$), empirical standard error, see($\hat{B}_X(t)$)), and coverage probability of 95% pointwise confidence intervals CP($\hat{B}_X(t)$)) based on the instrumental variables estimator, in function of sample size n and at different strengths ρ (correlation) of the instrumental variable.

			$\rho = 0.3$				$\rho = 0.5$	
	n	$t = 1$	$t = 2$	$t = 3$	n	$t = 1$	$t = 2$	$t = 3$
Bias $\hat{B}_X(t)$	1600	0.005	0.008	0.001	800	-0.001	0.001	- 0.006
sd ($\hat{B}_X(t)$)		0.136	0.224	0.336		0.108	0.176	0.249
see ($\hat{B}_X(t)$)		0.138	0.228	0.363		0.107	0.176	0.264
95% CP($\hat{B}_X(t)$)		96.2	96.2	96.5		95.2	96.0	97.1
Bias $\hat{B}_X(t)$	3200	0.003	-0.001	-0.004	1600	0.001	0.005	0.003
sd ($\hat{B}_X(t)$)		0.097	0.156	0.224		0.076	0.122	0.175
see ($\hat{B}_X(t)$)		0.096	0.157	0.230		0.075	0.121	0.173
95% CP($\hat{B}_X(t)$)		95.1	95.4	96.6		94.8	95.0	95.5

Table 3

Summary of simulations concerning the constant parameter estimator $\hat{\beta}_X$ and two versions of the 2SLS estimator of Tchetgen et al. (2015). Binary exposure and continuous instrument. Bias of $\hat{\beta}_X$, average estimated standard error, sd($\hat{\beta}_X$), in function of sample size n. Results for two versions (see text for details) of 2SLS estimator $\check{\beta}_{1X}$ and $\check{\beta}_{2X}$ of Tchetgen et al. (2015) are also given.

n	bias $\hat{\beta}_X$	sd ($\hat{\beta}_X$)	bias $\check{\beta}_{1X}$	sd ($\check{\beta}_{1X}$)	bias $\check{\beta}_{2X}$	sd ($\check{\beta}_{2X}$)
1000	-0.002	0.117	0.069	0.117	0.039	0.100
2000	-0.002	0.079	0.067	0.079	0.038	0.068

Table 4
HIP-study.

	Control Group	Screening group		
		All	Compl.	Non-compl.
n	30565	30130	20146	9984