**eAppendix**

**Regression discontinuity designs in epidemiology: causal inference without randomized trials**

**Jacob Bor, Ellen Moscoe, Portia Mutevedzi, Marie-Louise Newell, and Till Bärnighausen**

**Contents**

Summary 2

Generalizability of treatment effects in regression discontinuity designs 2

Fuzzy regression discontinuity as an instrumental variables approach 6

Applying fuzzy regression discontinuity to survival analysis 11

Robustness checks: alternate approaches to identifying complier causal effects 17

eAppendix references 20

# Summary

This eAppendix provides further details on two topics for which space limitations precluded in-depth discussion in the main text. First, regression discontinuity designs identify a local causal effect: the causal effect for the population close to the threshold; we discuss the generalizability of this estimand vis-à-vis the average causal effect identified in a randomized trial. Second, when a treatment is assigned probabilistically by a threshold rule (fuzzy regression discontinuity), regression discontinuity can be combined with an instrumental variables approach to identify complier causal effects, under some additional assumptions. These causal effects are “local” in two ways: they are local to the population close to the threshold; and they are local to compliers, i.e. those patients whose treatment status was determined by treatment assignment. We provide further details on the identification of complier causal effects in regression discontinuity designs with linear models, non-linear models, and survival analysis. Third, we describe the calculations used to obtain the complier causal effect estimates reported in the paper. Fourth, we assess the robustness of our fuzzy regression discontinuity results to alternate methods that do not utilize a survival analysis framework.

# Generalizability of treatment effects in regression discontinuity designs

When effects are heterogeneous, the causal effects identified in regression discontinuity designs are local to the population close to the threshold. That is, treatment effects are conditional on , and marginal with respect to all other observed and unobserved characteristics. If treatment effects are constant, as is often assumed in epidemiology (e.g., in non-saturated regression models), then the effect identified in a regression discontinuity design will be identical to the effect identified in an RCT and will be trivially generalizable to any other population. If treatment effects are constant across values of , but possibly heterogeneous across other observed or unobserved factors, then the local effect identified in a regression discontinuity design is equal to the average treatment effect identified in an RCT in the same population, but both may differ from treatment effects assessed in other populations. Indeed, an RCT can be thought of as a regression discontinuity design, where the assignment variable is a random number. Under randomization, the slopes of and will be zero, and the ratio or difference of these conditional expectations will be constant across .

In many real-life applications of regression discontinuity designs, generalizability to a wide range of may be plausible based on *a priori* knowledge. For instance, in a study looking at the fertility effects of a cash grant provided to girls born after a certain date, one would not expect fertility preferences to change rapidly with the calendar date of birth.1 Additionally, regardless of *a priori* knowledge, the presence of random error in measurements of will attenuate the slope of the conditional expectation functions, thus attenuating any effect heterogeneity. (As measurement error becomes large, the slopes of the conditional expectation functions will go towards zero.) Finally, the presence of effect heterogeneity in the area immediately around the threshold can be assessed in a regression context, by testing the interaction effect of a slope change in the conditional expectation function as it crosses the threshold. However, despite these considerations it remains true that although regression discontinuity estimates *may* be generalizable away from the threshold, causal identification occurs only at the threshold, and as such, inferences away from the threshold depend on untestable assumptions about the nature of treatment effect heterogeneity with respect to *Z*.

Given this caution, one might ask how regression discontinuity estimates can be interpreted without making untestable assumptions. One interpretation is the effect of a marginal, i.e., very small, change in the eligibility threshold for the study population or, equivalently, the effect of eligibility for persons close to the threshold. The extent of generalizability to other ranges of (and hence to larger threshold shifts) or to other study populations depends on the extent of treatment effect heterogeneity. If (as discussed above) treatment effects are approximately constant on the relevant scale, then the regression discontinuity estimate would also be the estimate of a larger change in the treatment threshold.

The “local” or “contingent” nature of regression discontinuity effect estimates may at first appear to be a specific characteristic of this method. However, in the context of treatment effect heterogeneity, this property is not unique to discontinuity designs. In RCTs as well, effect estimates are “local” to the specific population under study, and may be dramatically different in other populations. An example may be instructive. Consider an RCT in which HIV patients with CD4 counts between 200 and 350 are assigned either to initiate ART immediately (treatment) or to delay ART until their CD4 count has fallen below 200 (control) (e.g., as in Severe, et al. 2010).2 Consider that the effect of ART on survival may vary depending on a patient’s baseline CD4 count. The average treatment effect identified in this RCT is analogous to a weighted average of CD4-count specific treatment effects across the range, 200 to 350 cells/μL (in a linear model, it is precisely so). Although the RCT provides information about the average treatment effect in this population, if effects are heterogeneous, we learn nothing about the effect of ART outside the 200-350 range. Moreover, we learn nothing about the specific effect of ART at different CD4 counts *within* the 200-350 cells/μL range (unless the study was sufficiently powered for subgroup analysis); by implication, the results of such an RCT may not be generalizable to populations with different CD4 count distributions across this range. In contrast, in the regression discontinuity design, at least we know to what (narrow) population the estimated effect applies.

This distinction may be important in practice. Consider the unlikely – but plausible – scenario in which the relationship between baseline CD4 count and the effect of immediate ART on mortality is non-monotonic: immediate ART is strongly protective at 200 cells/μL, but is possibly harmful at 350 cells/μL (e.g., because of lower long-run adherence). In this case, the protective average effect identified in an RCT might reflect the average of large protective effects for patients with lower CD4 counts and small harmful effects for patients with higher CD4 counts. A policy that raised the eligibility threshold from 200 to 350 cells/μL would be beneficial on average (in populations similar to the study population). But in fact, such a policy might unnecessarily harm some patients to help others. For the purpose of optimizing treatment eligibility thresholds, several regression discontinuity studies to obtain CD4-count specific effect estimates would provide more appropriate data than RCTs of large threshold changes.

# Fuzzy regression discontinuity as an instrumental variables approach

The scenario in which treatment is assigned probabilistically by a threshold rule is known as a “fuzzy regression discontinuity (FRD)” design. The “fuzziness” in FRD refers to the fact that some patients are affected by the threshold rule, while others are not. “Sharp regression discontinuity (SRD)” can be thought of as a special case of FRD, where the change in probability at the threshold is from zero to one. If there is no change in the probability of treatment at the threshold, then there is no discontinuity and a regression discontinuity design cannot be implemented. (To clarify a common misunderstanding, the “fuzziness” in FRD does not imply that the location of the treatment threshold is imprecise; in such a case there would be no discontinuity and the study design could not be implemented.)

As discussed in the main text, FRD can be thought of as an instrumental variables (IV) design, where the threshold rule is the instrument. IV approaches have recently become popular in epidemiology, with particular focus on Mendelian randomization as an instrument for phenotype.3 An instrument is a variable that is correlated with the treatment of interest (*valid first stage*), is as-good-as-randomly assigned (*quasi-randomization*), and is associated with the outcome only through its relationship with the treatment (*exclusion restriction*).4 The latter two conditions cannot be tested, but may be argued based on *a priori* substantive knowledge or study design. These conditions are most likely to hold when the instrument is a causal predictor of the treatment and it is exogenous to (comes from outside) the relationship between the treatment and outcome. For example, a natural candidate for an instrument is treatment assignment in a randomized experiment. Indeed, there is a large literature on using IV to adjust for non-compliance and contamination in RCTs and obtain causal effects of the treatment itself, as an alternative to intent-to-treat analysis.

In the FRD design described in this paper, the indicator , or “ART eligibility by CD4 count”, satisfies the requirements of an instrument. It is a very strong predictor of rapid ART initiation. Because of random noise in measured CD4 counts and the absence of systematic manipulation, ART eligibility is as good as randomly assigned for observations close to the threshold. Finally, the exclusion restriction is plausible: the decision to initiate a patient is made at the time when the patient gets their CD4 count results and nearly all of these patients would have initiated by three months; if it was decided that a patient would not initiate ART, then he or she would be instructed to come back in six months for another CD4 test.5 It is unlikely that ART eligibility would have any impact on mortality, except through rapid ART initiation.

IV techniques allow identification of a particular causal effect: the causal effect of the treatment on compliers, i.e., those induced to take up the treatment because of the instrument, and who would not otherwise have taken up the treatment. In linear models, this is known as – the complier average causal effect. ( is typically called “” for “local average treatment effect” in the economics literature.4 However, since regression discontinuity estimates are “local” in another way – i.e., local to the population at the threshold – we use the terminology, which is also more common in epidemiology.) If treatment effects are heterogeneous, then will generally differ from the effect of the treatment on the treated or the average treatment effect. IV approaches can also be used to identify complier causal relative risks, , as discussed below.

To interpret , we adopt the nomenclature of Angrist, Imbens, and Rubin (1996),6 who describe four possible latent patient “types” in terms of their potential treatment assignments (**eTable 1**). Always-takers are patients who always take up the treatment regardless of treatment assignment; never-takers never take up the treatment regardless of treatment assignment. Under the additional assumption that there are no treatment-defiers – i.e., people who take up the treatment when it is not assigned but refuse treatment when it is assigned – also known as the *monotonicity assumption*, all changes in treatment status occurring at the threshold in the FRD design will be due to changes in treatment status among compliers. Under the IV assumptions described above, all changes in outcomes at the threshold are due to changes in treatment status at the threshold; and all changes in treatment status at the threshold occur among compliers. Causal effects thus can be identified for compliers. Note that “patient type” is not a function of treatment assignment, but rather a latent characteristic which determines a subject’s response when treatment is assigned; the distribution of types is continuous at the threshold. The intent-to-treat effect is the weighted average of treatment effects induced by the instrument across these latent classes. However, since we have assumed no defiers and since there are no changes in treatment status for always-takers and never-takers, we have the following identity: , where is simply the difference in the probability of take up at the threshold.

**eTable 1. Latent patient types**

|  |  |  |  |
| --- | --- | --- | --- |
| **Patient Type** | **Treatment status if eligible by CD4 count, 1[Z<c]=1** | **Treatment status if ineligible by CD4 count, 1[Z<c]=0** | **Example: When to start ART** |
| Always-takers | T=1 | T=1 | Patients initiated based on stage-IV HIV disease. |
| Never-takers | T=0 | T=0 | Patients who refused ART due to religious reasons. |
| Compliers | T=1 | T=0 | Patients who initiated ART because they had CD4 < 200 but who would not have initiated if CD4 ≥ 200. |
| Defiers | T=0 | T=1 | Assumed not to exist. |

In our FRD set-up, *CACEFRD* is thus equal to:

*Equation 1*



can be estimated by obtaining the numerator and denominator in separate regression discontinuity models, and then dividing the two. Alternatively, and its standard error can be estimated directly using two-stage least squares (2SLS).7 A common critique of linear probability models such as 2SLS is that they make predictions outside the interval (0,1); but this will not be the case at the threshold, where the regression discontinuity model is saturated: as the neighborhood around *c* shrinks towards zero, the regression specification is equal to . At the cut-off, where the causal treatment effect is estimated, the model includes only an intercept and the indicator for the threshold rule.

With binary (count) outcomes, estimates risk (rate) differences. Epidemiologists are also often interested in relative risks. IV techniques have also been developed for multiplicative (log link) structural mean models, with independent derivations of a GMM estimator by Robins (1989)8 and Mullay (1997).9 IV techniques for binary outcomes have been reviewed elsewhere.3,10,11 Angrist (2001) shows that with a binary treatment, binary instrument, and no covariates, the multiplicative structural mean model identifies a proportional causal parameter for the complier population,11 i.e. the complier causal relative risk, or .

*Equation 2*



In the absence of covariates the components of and – i.e., the treated and control complier means – can be identified in data:12

*Equation 3*

where AT denotes “always-taker” and TC denotes “treated complier”. And, similarly: .

# Applying fuzzy regression discontinuity to survival analysis

IV techniques for survival analysis are currently being developed. Because survival times are often censored, the focus of regression analysis has traditionally been on the hazard, the instantaneous probability of death at time *t*, conditional on survival up to time *t*. The hazard is a conditional expectation, not an expectation, and this presents problems in settings where there is unobserved heterogeneity (in either the baseline hazard or proportional treatment effect). In general, if there is unobserved heterogeneity in individual specific hazards, then the composition of the surviving sample will change over time, leading to time-varying population hazards. Since patient “type” in FRD, i.e. complier vs. always-taker vs. never-taker, is unobserved, heterogeneity in hazards across types will lead to differential changes in sample composition across types. The result will be that the shares (proportions) of patients across different latent types at baseline will differ from the shares of *person-time* contributed by latent types to estimation of hazards; the two are equivalent only as . One could proceed with the assumption that hazards are identical across groups with the same treatment status (which is observed), however this would rarely be plausible: typically patients who would take up a therapy regardless of eligibility (always-takers) are at greater risk or more likely to benefit than treated patients who would not have taken up if ineligible (treated compliers).

In our application of FRD to the question of when to start ART for HIV, we rely on a “rare event / limited follow-up assumption”: when mortality rates are relatively low, heterogeneity is modest, and the duration of follow-up relatively short, there is little change in composition of the population over time (due to heterogeneity), and so the hazard estimated for a population is a close approximation to the average of the hazards across sub-populations. (For example, in a simulated population observed for up to five years, where 64% of patients had and 36% had , the pooled hazard was and the average hazard weighted by baseline shares was ; the bias was 1% of the difference in hazards.) We further assume that censoring times are not correlated with latent subject type.13

Under our rare event / limited follow-up assumption, the predicted hazard for treated patients to the left of the cut-off is approximately a weighted average of the hazards in the latent groups of always takers, treated compliers, and never-takers:

*Equation 4*

and symmetrically for . The complier causal difference in hazards can thus be approximated as . Estimating the corresponding ratio measure – the complier causal hazard ratio, – requires individual estimates of treated and control complier hazards. To obtain these, we use an additional decomposition: the predicted hazard for *treated* patients to the left of the cut-off is approximately the average of the hazards for always-takers and treated compliers:

*Equation 5*

This is simply the “conditional-on-treated” version of Equation 4 and can be estimated in a regression discontinuity model limiting the sample to the treated, as explained below.

In our FRD analysis, we proceeded as follows. First, we estimated predicted hazard rates, and , for patients presenting just above vs. below the 200-cells/μL cut-off in an “intent-to-treat” exponential regression discontinuity hazard model (Table, main text). Second, we estimated probabilities of ART initiation approaching the threshold from above vs. below in a “first stage” linear probability regression discontinuity model, with ART initiation in three months as the outcome. Based on this model, we predicted:

By our rare event / limited follow-up assumption, the “intent-to-treat” difference in hazards divided by the “first stage” difference in probabilities at the threshold approximates the complier causal difference in hazards, i.e .

Third, we estimated another exponential regression discontinuity hazard model, limiting the sample to the treated, , and obtained predictions for and . The latter prediction is the hazard among always-takers at the threshold; this is identical on either side of c: . Plugging estimates of and into Equation 5 along with estimates of and from the first stage regression, we solved for , the treated complier hazard at the threshold. We calculated the control complier hazard by subtracting off the complier causal difference in hazards, . Dividing the two, we obtained an estimate of the complier causal hazard ratio, . (Nearly identical results were obtained through the symmetrical procedure of estimating an exponential regression discontinuity hazard model for , computing the control complier hazard and adding to obtain the treated complier hazard.) The results are presented in **eTable 2** and described in the main text.

Under the rare event / limited follow-up assumption we have invoked, the bias in our approach is likely to be small; however, it is still an approximation and this method may perform badly in other settings with higher failure rates, larger effect heterogeneity, or longer follow-up. Another approach would be to focus directly on the population survival curve, which – unlike the hazard – is an unconditional probability and is not biased by unobserved heterogeneity. Abbring & van den Berg (2005) show that in the absence of covariates, the Kaplan-Meier estimator can be plugged into the numerator of the Wald IV estimator to obtain a *CCRR*.14 The Kaplan-Meier estimator is non-parametric and will estimate unbiased population survival curves even when there is unobserved heterogeneity across units or heterogeneous treatment effects and still estimate unbiased survival curves. However, the Kaplan-Meier estimator cannot accommodate continuous covariates; in regression discontinuity applications, the relationship between the continuous assignment variable and outcome (e.g., survival) must be modeled. Further theoretical development of IV methods for survival analysis may lead to alternative and improved estimators for FRD designs with time-to-event data in the presence of censoring.

**eTable 2. Estimates from fuzzy regression discontinuity hazard analysisa**

|  |  |  |  |
| --- | --- | --- | --- |
| **Row** | **Parameter** | **Interpretation** | **Estimate** |
| 1 |  | Hazard among eligibles | 0.027 |
| 2 |  | Hazard among non-eligibles | 0.038 |
| 3 |  | Hazard in always-takers and treated compliers | 0.020 |
| 4 |  | Hazard in always-takers | 0.036 |
| 5 |  | Pr(ART initiation) among eligibles | 0.561 |
| 6 |  | Pr(ART initiation) among non-eligibles | 0.201 |
| 7 |  | Pr(complier); difference in ART at the threshold | 0.360 |
| 8 |  | Difference in hazards at the threshold | -0.010 |
| 9 |  | Complier causal difference in hazards | -0.029 |
| 10 |  | Treated complier hazard | 0.011 |
| 11 |  | Control complier hazard | 0.040 |
| 12 |  | Complier causal hazard ratio | 0.283 |
| Note: Rows 1 and 2 estimated in intent-to-treat regression discontinuity model, using exponential hazard regression model. Rows 3 and 4 estimated in regression discontinuity exponential hazard model for sample of treated patients only. Rows 5 and 6 estimated in regression discontinuity linear probability model with rapid (3 month) ART initiation as the outcome. All regression models were estimated for patients presenting with CD4 counts between 50 and 350, and included separate linear terms in first CD4 count on either side of the threshold. In contrast to the intent-to-treat results presented in Table 1 of the main paper, all models excluded patients who died or were lost to follow-up in the first three months, for a regression sample of n = 3508 for rows 1 and 2 and n = 1386 for rows 3 and 4. The remaining estimates were calculated as follows: row 7 = row 5 – row 6. Row 8 = row 1 – row 2. Row 9 = row 8 / row 7. Row 10 estimated by Equation 5, using rows 3,4,5, and 6. Row 11 = row 10 – row 9. Row 12 = row 10 / row 11. | | | |

# Robustness checks: alternate approaches to identifying complier causal effects

As a robustness check on our FRD estimates, we limited the sample to subjects observed for at least three years and estimated FRD models for the probability of death by three-years. With a binary outcome for three-year mortality, and no censored observations, we are able to use established IV methods for binary outcomes.10,11 This approach has been used widely in epidemiologic studies of survival times with non-compliance, even though it discards censored observations and changes the estimation sample.15

As a yardstick against which to assess the results from our exponential hazard models, we first estimated intent-to-treat regression discontinuity models for the relative risk of death in three-years using robust Poisson regression models, limiting the sample to persons with at least three years of potential follow-up. We estimated similar models for the absolute risk of death in three years using linear probability (OLS) regression models. In the Poisson models, we included separate linear terms in first CD4 count on either side of the threshold; in the OLS models, we modeled first CD4 count on the log scale, which was a better fit to the data. Similar to our “preferred” exponential hazard model – Table 1, Model 2a – we estimated these regressions limiting the sample to patients presenting with a first CD4 count between 50 and 350 cells/μL. **eTable 3** presents the results. ART eligibility, as determined by a CD4 count less than 200 cells/μL, was protective against three-year mortality (), implying that patients who were eligible at baseline were 36% (5 percentage points) less likely to have died in three years than persons who were not eligible at baseline. These results were nearly identical to our estimated intent-to-treat hazard ratio of 0.65, although the confidence intervals were somewhat wider as a result of discarding observations with less than three years of potential follow-up.

To obtain , we scaled the difference in three-year mortality at the threshold by the difference in the probability of ART initiation in 2SLS models. We used models with separate linear terms in log-first CD4 count on either side of the threshold, estimated on the range 50, 350 cells/μL. This approach yielded a causal risk difference of lower probability of three-year mortality for patients who initiated because they were CD4-count eligible, compared to those who were precluded from initiating because they were ineligible (**eTable 3**). We estimated the complier causal risk ratio using Stata’s *ivpois* command, which implements a generalized method of moments estimator for binary and limited dependent variable IV models;9,16 this has the same moment conditions as the multiplicative structural mean model.3,8 The was (**eTable 3**), very close to the of 0.28 estimated under our rare event assumption. Rapid ART initiation causally reduced three-year mortality by 76% among patients who initiated ART because their CD4 count was below 200 cells/μL, relative to those who were barred from initiating because they were ineligible.

**eTable 3. Regression Results for Three-Year Mortalitya**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Outcome** | **Parameter** | **Estimate** | **(95% CI)** | **N** |
| *Full sample* | | | | | |
| Exponential (Table 1, 2a) | Time to death | Hazard ratio | 0.65 | (0.45, 0.94) | 3710 |
| *Sample: patients with >3y potential follow-up* | | | | | |
| Poisson | 3y mortality | Risk ratio | 0.64 | (0.41, 1.01) | 2276 |
| MSMM/IVPOIS | 3y mortality | Risk ratio | 0.24 | (0.04, 1.49) | 2209 |
| OLS | 3y mortality | Risk difference | -0.05 | (-0.10, 0.01) | 2209 |
| OLS (first stage) | Rapid ART | Risk difference | 0.38 | (0.31, 0.45) | 2209 |
| 2SLS | 3y mortality | Risk difference | -0.12 | (-0.26, 0.02) | 2209 |
| aEach row displays results from a separate regression model. All models were estimated for the CD4 count range 50-350 cells/μL. All models control for CD4 count as a continuous predictor, with separate linear trends on either side of the 200-cells/μL threshold. The OLS and 2SLS regressions modeled CD4 count in log-units. The exponential and Poisson intent-to-treat models included patients with less than three months of follow-up. The final four rows of the table exclude these patients since treatment status (rapid ART) is not observed for them. Rapid ART = ART within 3 months of a first CD4 count; 2SLS = two stage least squares; MSMM/IVPOIS = multiplicative structural mean model, estimated using Stata’s *ivpois* command.16 The F-statistic on the first stage regression was 123.2, indicating that there is no weak instrument problem.17 Standard errors are robust to heteroskedasticity in all models except MSMM/IVPOIS. 3y = three-year. | | | | | |

# eAppendix references

1. Bor J. Cash transfers and teen pregnancy in an HIV endemic setting: a regression discontinuity study, in: *Essays on the Economics of HIV/AIDS in Rural South Africa*. Dissertation. Boston, MA: Harvard University, School of Public Health, May 2013.

2. Severe P, Juste MA, Ambroise A, et al. Early versus standard antiretroviral therapy for HIV-infected adults in Haiti. *N Engl J Med*. 2010;363(3):257-265.

3. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist’s dream? *Epidemiology.* 2006;17:360-372.

4. Imbens GW, Angrist JD. Identification and estimation of Local Average Treatment Effects. *Econometrica*. 1994;62(2):467-75.

5. Houlihan CF, Bland RM, Mutevedzi P, et al. Cohort profile: Hlabisa HIV Treatment and Care Programme*. Int J Epidemiol.* 2011;40:318.

6. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association.* 1996;91(434):444-455.

7. Imbens GW, Lemieux T. Regression discontinuity designs: a guide to practice. *J Econom*. 2008;142(2):615-35.

8. Robins JM. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest L, Freeman H, Mulley A, eds. Health Services Research Methodology: A Focus on AIDS. NCHRS, U.S. Public Health Service; 1989:113–59.

9. Mullahy J. Instrumental-variable estimation of count data models: applications to models of cigarette smoking behavior. *Rev Econ Stat*. 1997;79(4):586-593.

10. Clarke PS, Windmeijer F. Instrumental variable estimators for binary outcomes. *J Am Stat Assoc.* 2012;107(500):1638-52.

11. Angrist JD. Estimation of limited dependent variable models with dummy endogenous regressors. *J Bus Econ Stat.* 2001;19(1):2-28.

12. Imbens GW, Rubin DB. Estimating outcome distributions for compliers in instrumental variable models. *Rev Econ Stud,*1997;64:555-574.

13. McIntosh M. Instrumental variables when evaluating screening trials: estimating the benefit of detecting cancer by screening. *Stat Med*. 1999; 18(20): 2775–2794.

14. Abbring JH, van den Berg GJ. Social experiments and instrumental variables with duration outcomes. *Tinbergen Institute Discussion Paper 2005-047/3*, 2005*.*

15. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol.* 2000;29(4):722-729.

16. Nichols A. ivpois: Stata module for IV/GMM Poisson regression. 2007. [http://ideas.repec.org/c/boc/bocode/s456890.html\*/](http://ideas.repec.org/c/boc/bocode/s456890.html*/)

17. Bound J, Jaeger DA, Baker R. 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *J Am Stat Assoc.* 1995;90(430), 443-450.