Published in final edited form as:

Sex Transm Dis. 2010 August ; 37(8): 512–518. doi:10.1097/OLQ.0b013e3181d73feb.

Inverse probability weighting in STI/HIV prevention research: methods for evaluating social and community interventions

Sheri A. Lippman, PhD^{1,2}, Starley B. Shade, PhD², and Alan E. Hubbard, PhD³

¹Division of Epidemiology, University of California, Berkeley

²Center for Aids Prevention Studies, University of California, San Francisco

³Division of Biostatistics, University of California, Berkeley

Abstract

Background—Intervention effects estimated from non-randomized intervention studies are plagued by biases, yet social or structural intervention studies are rarely randomized. There are underutilized statistical methods available to mitigate biases due to self-selection, missing data, and confounding in longitudinal, observational data permitting estimation of causal effects. We demonstrate the use of Inverse Probability Weighting (IPW) to evaluate the effect of participating in a combined clinical and social STI/HIV prevention intervention on reduction of incident chlamydia and gonorrhea infections among sex workers in Brazil.

Methods—We demonstrate the step-by-step use of IPW, including presentation of the theoretical background, data set up, model selection for weighting, application of weights, estimation of effects using varied modeling procedures, and discussion of assumptions for use of IPW.

Results—420 sex workers contributed data on 840 incident chlamydia and gonorrhea infections. Participators were compared to non-participators following application of inverse probability weights to correct for differences in covariate patterns between exposed and unexposed participants and between those who remained in the intervention and those who were lost-to-follow-up. Estimators using four model selection procedures provided estimates of intervention effect between odds ratio (OR) .43 (95% CI:.22-.85) and .53 (95% CI:.26-1.1).

Conclusions—After correcting for selection bias, loss-to-follow-up, and confounding, our analysis suggests a protective effect of participating in the *Encontros* intervention. Evaluations of behavioral, social, and multi-level interventions to prevent STI can benefit by introduction of weighting methods such as IPW.

Keywords

Intervention Evaluation; Inverse Probability Weights; STI/HIV Prevention; Statistical Methods

Introduction

Over the last fifteen years STI/HIV prevention researchers have increasingly called for implementation of multi-level, multi-component, structural and social interventions aiming to modify environments that place populations at risk.¹⁻¹⁰ The enthusiasm for multi-level interventions is based on recognition that economic, political, and social context shapes sexual behavior. However, interventions that aim to change social structures and

environments using multiple prevention components are notoriously difficult to design, implement, and evaluate.^{6, 11} The gold standard for evaluation of community focused interventions to change social environments or structures is the randomized community trial. Given the cost of randomizing communities to a multi-level intervention, such efforts should only be undertaken based on findings from rigorously designed observational studies, in order to ensure that strategies tested in randomized community trials are those most likely be effective.

While observational research has been considered a less rigorous alternative to the randomized trial, observational designs have many strengths for evaluating prevention programs and examining the magnitude of their effects. Following a single community cohort over the course of an intervention prevention program offers a rigorous and cost-effective means to evaluate multi-level interventions, and, by using cutting edge statistical tools, it can provide estimates that are as close as possible to trial results without the cost. Randomizing individuals in a community is often difficult: one cannot enter a community and randomize friends and neighbors to participate or not in public forums, community events, or other mobilization efforts. Instead, evaluation in this context means comparing individuals in the community who participate to those who do not (or do so less). The challenge becomes comparability of the participants to the non-participants (internal study validity). Fortunately, there are weighting methods that can correct for the biases found in non-randomized studies and permit valid comparison.

This paper outlines the evaluation of a multi-component STI/HIV prevention project with the application of weighting methods to observational cohort data. The method, inverse probability weighting (IPW), is one estimating tool from a class of techniques called marginal structural models¹²⁻¹⁸ that facilitate causal inference with observational data.^{19, 20} IPW was first applied in clinical epidemiology as an analysis tool to address complex confounding structures in longitudinal AIDS treatment trials. The weighting methods have been well described in the AIDS treatment literature.^{13, 21-23} Whereas the complex theoretical and statistical underpinnings of these methods may have limited their accessibility initially, IPW is an invaluable tool in dealing with selection bias, loss-to-follow-up, and complex confounding in evaluation of prevention programs. We hope to facilitate wider use of IPW in STI/HIV intervention research by demonstrating, step by step, how IPW can be used in prevention studies.

Materials and Methods

Data source: *Encontros*

From 2003 to 2005, *Encontros*, or “coming together”, aimed to decrease incident STI and encourage adoption of consistent condom use among female, male, and transvestite sex workers in Corumbá, Brazil, by using joint clinical and social intervention strategies. The project was designed to engage the sex workers on an individual level through participation in STI/HIV counseling and testing, on an interpersonal level through peer-education, and on a community level through outreach and social activities. Community-based activities were designed to extend and strengthen collegial relationships through providing sex workers opportunities to engage in dialogue around sex work, discrimination, human rights, and prevention.²⁴

Participants included 420 men, women, and transvestites, 18 years of age or more, self-identifying as sex workers, who spoke Portuguese or Spanish and did not plan on leaving the study area permanently in the month following recruitment. Participation included an enrollment visit and four scheduled follow-up visits at 3, 6, 9 and 12 months following enrollment. Each visit included administration of a structured, interviewer-administered

questionnaire, counseling for STI/HIV, collection of biological samples for STI testing, treatment for STI if indicated, and testing for HIV if requested. Intervention activities were ongoing during rolling enrollment, follow-up visits, and through the final month of participation. Chlamydia and gonorrhea were diagnosed using COBAS AMPLICOR™ CT/NG PCR (Roche Molecular Diagnostics, Pleasanton, CA).

All enrolled participants were encouraged to participate in project sponsored events; as a result there is no “control” population. Instead we compare participants who were more and less exposed to the intervention, with the hypothesis that sex workers who actively participated in the project would present with fewer incident cases of chlamydia and gonorrhea. Project participation varied at each time point and included contact with peer educators and counselors and participation in community cultural or social events, workshops, educational activities, and organizations/associations. Participation was summarized into a dichotomous indicator of low or high participation at each visit.²⁴ Low exposure describes those who attended scheduled appointments but who participated very little in project activities. The high exposure group attended scheduled appointments and had more contact with educators or counselors and participated in project events, workshops, or organizations.

Biases and confounding

Participants in social or community interventions such as *Encontros* often can not be randomized. Instead, they choose how much to participate (or self-select) based on factors often associated with access to services, behaviors, and infection. For example, people who chose to participate more in *Encontros* activities tended to be older and already participating in community groups. Comparison of people participating more and less is likely to be confounded. Similarly, participants who remain in a prevention program or a research cohort are different from those who do not. Over half of the 420 participants in the *Encontros* cohort were lost-to-follow-up prior to completing a year of follow-up. Those who remained in the study were more likely to be married, were less mobile, and reported more STI symptoms. This loss-to-follow-up can be thought of as a missing data problem (absence of data on those who left) or as a selection bias problem (people who stay are different from those who do not). Estimates of effect based only on people who remain in the cohort may not reflect the true program effect on all participants.

Furthermore, in longitudinal studies one may also find time dependent confounding, whereby past outcomes, covariates and exposure to the intervention may impact subsequent exposure and subsequent infections.^{13, 14, 17} Attempting to eliminate the effects of time dependent confounding by treating prior participation as a confounder in a regression model can result in bias by controlling away the effect of interest. However, failure to account for prior participation can also bias the effect estimate. In this study, we explored confounding within intervals of data as well as time dependent confounding. The weighting approach presented herein can mitigate self-selection, loss-to-follow-up and confounding, including time dependent confounding, when relevant.

Weighting to balance populations

Our aim is to assess the effect of the intervention on incident STI among sex workers. We estimate the intervention effect by comparing the odds of presenting with an incident STI if all sex workers in our study population were a high participator, as compared to the odds of an STI if everyone were a low participator at each time point. This can be described as the marginal probability of the outcome under the different levels of participation (often generically called “treatments” in the causal inference literature). We assume that the characteristics that predict participation in the intervention are the same over all time points,

or intervals, so we simply pool data over time, analyzing all of the time points together in one model.

Hypothetically speaking, to do away with the biases in this longitudinal data set, one would need to randomize people to low or high participation levels at every visit to ensure that past experiences and covariate distribution had no bearing on the exposure/outcome relationship at each interval. We simulate this repeat randomization using weights. Inverse probability weighting for exposure (or in this case for participation) assigns a weight for each subject equivalent to the inverse probability of being in his/her participation group, based on values of covariates, past outcomes, and exposures at each interval. This treatment probability is estimated using regression techniques. The exposure weights are then applied to the observed population, creating a new pseudo-population in which the joint distribution of confounders is balanced between the two exposure groups, essentially mimicking a randomization procedure at each time point.¹⁷

This weighting technique can also be used to deal with the issue of loss-to-follow-up or censoring. We assign each participant a weight equivalent to the inverse probability of remaining in the study at each interval, based on values of observed covariates and past outcomes and exposures. Probabilities are estimated using basic regression techniques. The censoring weights are applied to the observed population, creating a new pseudo-population in which censored subjects are “replaced” by up-weighting uncensored subjects with the same values of past exposures and covariates. Censoring weights and treatment weights are combined (multiplied) into final weights to allow for simultaneous adjustment.

Weighting Procedures

Data management—Cohort participants come for repeated visits; the data need to be in long or stacked format (Table 1) and outcomes must be linked with time points to create treatment weights and run analyses. In the *Encontros* study, not all subjects adhered to their scheduled visit dates, resulting in non-uniformity in data collection. Because visits were to take place every three months, we divided data into three month intervals, centered on the scheduled time points of 3, 6, 9, and 12 month visits, such that each participant's first visit was equivalent to time “0” and subsequent visits between 45 days and 135 days later (3 month intervals of 90 days centered on the next scheduled visit) were included in interval 1, and so on. We include five intervals of data in this analysis.

Table 1 displays a spreadsheet view of how the data should be set up for a weighted longitudinal analysis. A row has been created for each participant for each interval. Every row indicates whether the participant was present (used to create censoring weights) and, if so, his/her participation level and outcome. Note that there are no exposure data for the first (enrollment) visit; data for analyzing the exposure/outcome relationship begin in interval 1.

Variable and model selection—Once the data have been arranged, the next step is to model the association between confounders and the *exposure* (i.e. intervention participation.) When there is loss-to-follow-up or censoring, the association between confounders and presence at each visit must also be modeled. These models are the first step in calculating the weights and are generated using so-called black box techniques designed to do a good job predicting intervention exposure without concern about interpretability, at once attempting to trade-off variance and bias in the model predictions. Just as in a typical regression analysis, variables selected to control for confounding in these treatment models should be potential confounders related to the outcome. Including variables that are strongly related to treatment but not the outcome can severely hurt estimation. The resulting model has two elements: 1) the set of variables chosen, and 2) the functional form of these variables (e.g., main effects, polynomial terms, multiplicative interactions). These models

can be chosen by automated algorithms or by expert knowledge, although expert knowledge can rarely identify the proper functional form of the variables.

Most investigators confront missing data. We chose to carry forward values for those covariates deemed not to change drastically in the period of six months. For example, we considered that self-efficacy for condom use would not change at every visit and therefore need not be measured at every visit. To include the self-efficacy variable in the treatment and censoring models, values were carried forward from the last visit. Other methods of data imputation are available.²⁵ Variables listed in Table 2 were considered for inclusion in the weighting models.

Once a group of variables has been identified for weighting models, various approaches can be taken to model specification and selection.^{26, 27} For these data, we constructed four sets of weights for treatment and censoring using different criteria for each. (Table 3) First, we chose covariates based on expert knowledge of the subject matter. The second approach was based on an automated algorithm in the program R²⁸ called the deletion substitution algorithm (DSA). DSA is a data-adaptive estimation procedure based on cross-validation, which chooses among models trying to minimize the mean-squared error of prediction.^{29, 30} The third approach utilized another automated algorithm in R called polyclass, which builds a hierarchical set of models with main effects, linear spline terms and interaction terms and uses cross-validation to determine the optimal model size.³¹ We also constructed weights using a stepwise selection process, including all first order variables (main effects) and fractional polynomial terms (for continuous variables) with a p-value < 0.1.

There is no single best way to perform model selection. Some model selection routines require extensive programming and may only be accessible with substantial help. We recommend trying a variety of approaches and a variety of models; using different permutations of an expert knowledge model or stepwise approaches is a reasonable alternative to automated selection techniques. If results differ substantially by model selection procedure, further work is required to determine how and why the weights differ. Finally, one can derive inference using robust standard errors or techniques such as bootstrapping.³²

Weights should be assessed for their range – giving anyone a weight of 100 in a study with only 100 observations would mean that one person is contributing a large portion of the information towards the estimate of the parameter of interest. Extreme weights can be truncated; one can also truncate or transform variables that might be contributing to extreme or unstable weights, or one can restrict analyses to a subset of observations (see ETA assumption discussed below). One can also use so-called stabilized weights.¹⁷ Following sensitivity analyses for different truncation points, we chose to truncate our weights at a value of 20 to avoid over-representing any single observation or permitting extreme outliers to have very large weights.

Assumptions

The models we use require meeting a set of assumptions about the sampled data with respect to the larger population. In the case of IPW, the requisite assumptions have been laid out in detail.^{17, 26} First, one assumes there is no unmeasured confounding. There is no way to empirically test for no unmeasured confounding; collection of data on a complete set of covariates should be incorporated in the design phase. For the *Encontros* study, we collected data on a broad and inclusive set of measurable covariates.

Second, time-ordering is necessary for casual inference. Because there is no way to know exactly when someone acquires an infection and because we cannot collect samples every

day, we make the assumption in this study that variables reflecting activities, behaviors, and the social environment over the past three months predate infection at the end of the period (samples were collected at the end of each interval). Additionally, we have only included participants who were treated for previous infections, such that they are at risk in every interval included in the data.

Third, Experimental Treatment Assignment (ETA) or positivity assumes that groups defined by all possible combinations of covariates must have the potential to be in any (either) of the treatment groups. If there are covariate groups that will only be observed in one treatment state, then we cannot estimate the effect of the exposure within that group. There is also the practical ETA violation, which occurs when there are too few subjects who have one of the treatments in a covariate group to estimate the treatment effect in that group. Fundamentally ETA is not a problem of the estimator, but a problem pertaining to covariate groups for which there is no experimentation and is, therefore, of issue across estimation approaches. However, most typical regression analyses will not provide obvious alerts of this problem (it is invisible to these techniques), whereas the IPW estimator can “blow-up” as weights get very large. This is an advantage of the IPW estimator, though often disappointing to the researcher because it raises a red flag about a lack of information in the data to answer the question of interest. In our data, using 3 of 4 modeling methods almost all participants had $>.05$ chance of being in the high participation group. One can use rules of thumb to decide the importance of the bias (such as weights > 20), though there are also established computer intensive methods to examine bias due to a potential ETA violation.³³

Results

Overall, 420 cohort members contributed to 1,346 data collection visits, of which 1,306 occurred within the specified time frame for this analysis (up to 16 months following recruitment.) Valid results for chlamydia and gonorrhea testing were available for 1,262 visits, of which 848 visit occurring during the intervention (the other 414 results were results from the recruitment visit). Eight observations were excluded as infections were not treated at least 30 days before the subsequent visit (the participant was not at risk of infection). As such, there are 840 observations from 333 cohort members contributing to this analysis.

Modeling procedures, resulting weights, and estimates of effect are presented in Table 3. Because there are repeated measures, the final analysis was performed using generalized estimating equations (GEE) to account for the non-independence of repeated measures on individuals.³⁴ The crude estimate of effect was .51 (.26-.99): the odds that high participators would test positive for chlamydia or gonorrhea are half the odds of low participators. Estimates of the odds ratio generated using all of the modeling approaches are similar, between .43 and .53, and on the whole significant. Our results suggest a protective effect of participating in the Encontros intervention, reducing incident infections by an estimated 50%. Detailed results will be presented in a separate paper.

The weights generated using expert knowledge and the polyclass algorithm (Table 3: models 1 and 3) had a more restricted range. The DSA and stepwise approaches (models 2 and 4) yielded a broader range of weights, with some people having a very low probability of treatment. A smaller range of weights can imply fewer violations of ETA, but may imply more bias in the estimates of the treatment/censoring model. There is no theoretically best way to choose a modeling approach. Again, we recommend using multiple approaches.

We also ran a more traditional analysis using GEE to account for non-independence and controlling for confounding with pooled multivariable logistic regression. Controlling for covariables utilized in the “expert knowledge” model that remained significant at $p<.1$ in

multivariate analyses (including age at enrollment, self-efficacy, sex work outside of study city in the past 5 years, individual income, and number of unprotected sex acts in previous week), the estimate of effect was OR: 0.68 (95% CI:0.34-1.3). This estimate is not directly comparable to the estimates in table 3, which are marginal estimates of intervention effect over the distribution of covariates in the study population. Instead, this should be interpreted as the conditional effect of the intervention for a small subgroup of the study population: the youngest members of the cohort, with mean levels of self-efficacy, who did not work outside of the study area in the past 5 years, with the lowest tertile of income, and with no unprotected sex acts in the previous week.

Discussion

The hegemony of the randomized trial design has dampened enthusiasm for alternative designs that are sometimes more appropriate and definitely less expensive for evaluating STI/HIV prevention programs. As the statistical methods described in this manuscript become increasingly accessible, observational designs should experience a revived interest for evaluating STI/HIV prevention, especially as renewed calls for an evidence base in multi-faceted social and structural prevention initiatives gain momentum.^{11, 35, 36} The methods presented herein to re-weight the study population provide a means to utilize longitudinal, observational study designs for causal estimates of programmatic effects, correcting for the biases common in observational studies. Weighting has not been used widely in prevention research to date; we believe that evaluations of behavioral, social, and multi-level interventions can benefit from use of these methods.

In the *Encontros* study, despite the likelihood that self-selection and loss-to-follow-up would skew results, there was less bias than expected. There were some differences between participants who remained in the intervention and those who were lost and between those who participated and those who did not, but overall those discrepancies did not substantially alter the estimates of intervention effect. Because we corrected the potential biases through application of weights and used various model selection approaches, we can infer that our findings are not likely due to uncontrolled confounding and that the intervention reduced incident infections among actively participating sex workers. Even so, the possibility of residual confounding, particularly from unmeasured factors, cannot be dismissed, despite utilization of some of the best weighting tools available.

By using IPW we generated marginal estimates of effect: the average intervention effect across the entire study population, as if all participants were in the high exposure group compared to the counterfactual of having all participants in the low exposure group.[†] This estimate is more useful in terms of planning for a public health impact than conditional estimates, or those generated controlling for covariates. We ran a conditional multivariate analysis for purposes of comparison and found a very different odds ratio (OR: 0.68). It is not surprising that this fundamentally different effect estimate, which describes the intervention effect in only a subset of the study population, was not very close to the marginal estimates in table 3. Conditional estimates are often reported without acknowledgement that they are estimates of association or effect in a small subset of the study population. Had we run a traditional analysis (non-weighted) we may have interpreted our findings to indicate that the intervention was not successful. Care needs to be taken in reporting conditional estimates and interpreting them as global effect estimates. We also note that this traditional approach relies on the dubious assumption that an a priori specified

[†]Note that unlike in linear or log linear models, even in the absence of interaction, the conditional OR doesn't represent the marginal effect due to the non-collapsibility of the odds ratio. See reference 37: Jewell NP. *Statistics for Epidemiology*. Boca Raton: Chapman & Hall/CRC; 2004.

model is correct. Methods using machine learning to describe the treatment mechanism, such as the DSA described herein, make no such (extremely important) assumptions.

Additional estimation approaches are available to balance observational data, including the G-computation algorithm, which can be used for continuous exposures and is based on imputing the outcomes for individuals keeping all covariates fixed at their original values but modifying the treatment variable of interest.^{38, 39} Another iteration of the weighting approach is the use of propensity scores. With propensity scores the probability of treatment is modeled for each participant, much the same as it is with IPW; however, the resulting score for treatment probability is used to stratify populations or match participants to allow for estimation of association or effect within comparable strata or matched pairs.^{40,41} These estimation approaches can also be applied in other observational and in trial designs. For example, IPW can re-balance study populations at different time points in serial cross sectional designs in order to estimate intervention effects while adjusting for secular changes in population. In trials, estimation approaches such as IPW and G-computation can be applied to account for post-randomization differences in distribution of important covariates (i.e. empirical confounding).

Causal inference methods, including marginal structural models, have rarely been presented in forums specific to STI prevention. IPW is one approach that can help improve the quality of and maximize the use of data generated in observational studies and program evaluations of STI/HIV prevention. Given limited resources, rigorous evaluations of community-based interventions using these methods may provide more information regarding successful intervention approaches than funding a limited number of RCTs. Application of IPW is not beyond prevention researchers; however, users should be attentive about meeting the assumptions necessary to apply the methods correctly.

Acknowledgments

The authors gratefully acknowledge Juan Díaz and Magda Chinaglia, investigators on the *Encontros* study, Angela Donini, the study coordinator, and the study advisory committee and institutional partners (The Population Council, the Brazilian Ministry of Health, Pathfinder do Brasil, Rede Brasileira de Prostitutas, and the state and municipal Secretary of Health) for their dedication to the project that produced these data. We thank Adriana Pinho for data cleaning and Jennifer Ahern and Art Reingold for providing comments on this manuscript. Data collection for the *Encontros* project was supported by the Population Council and the Ministry of Health in Brazil. The first author received support from the Fogarty AIDS International Training and Research Program (AITRP) (grant 1 D43 TW00003) at the School of Public Health, University of California, Berkeley, while drafting this manuscript.

References

1. Kelly J. Community-level interventions are needed to prevent new HIV infections. *Am J Public Health*. Mar; 1999 89(3):299–301. [PubMed: 10076476]
2. Parker RG. Empowerment, Community Mobilization and Social Change in the face of HIV/AIDS. *AIDS*. 1996; 10(suppl 3):S27–S31. [PubMed: 8970709]
3. Parker RG, Easton D, Klein C. Structural Barriers and Facilitators in HIV Prevention: A Review of International Research. *AIDS*. 2000; 14(suppl 1):S22–S32. [PubMed: 10981471]
4. Waldo CR, Coates TJ. Multiple Levels of Analysis and Intervention in HIV Prevention Science: Exemplars and Directions for New Research. *AIDS*. 2000; 14(suppl 2):S18–S26. [PubMed: 11061638]
5. Beeker C, Guenther-Grey C, Raj A. Community empowerment paradigm drift and the primary prevention of HIV/AIDS. *Soc Sci Med*. Apr; 1998 46(7):831–842. [PubMed: 9541069]
6. Blankenship KM, Friedman SR, Dworkin S, Mantell JE. Structural interventions: concepts, challenges and opportunities for research. *J Urban Health*. Jan; 2006 83(1):59–72. [PubMed: 16736355]

7. Sweat MD, Denison JA. Reducing HIV incidence in developing countries with structural and environmental interventions. *AIDS*. 1995; 9 A:S251–257. [PubMed: 8819593]
8. Tawil O, Verster A, O'Reilly KR. Enabling approaches for HIV/AIDS prevention: can we modify the environment and minimize the risk? *AIDS*. Dec; 1995 9(12):1299–1306. [PubMed: 8605048]
9. Sumartojo E, Doll L, Holtgrave D, et al. Enriching the mix: incorporating structural factors into HIV prevention. *AIDS*. Jun.2000 14 1:S1–2. [PubMed: 10981468]
10. Merson M, Padian N, Coates TJ, et al. Combination HIV prevention. *Lancet*. Nov 22; 2008 372(9652):1805–1806. [PubMed: 19027478]
11. Gupta GR, Parkhurst JO, Ogden JA, et al. Structural approaches to HIV prevention. *Lancet*. Aug 30; 2008 372(9640):764–775. [PubMed: 18687460]
12. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*. Jul; 2006 17(4):360–372. [PubMed: 16755261]
13. Hernan MA, Brumback BA, Robins JM. Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Stat Med*. Jun 30; 2002 21(12):1689–1709. [PubMed: 12111906]
14. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. Sep; 2004 15(5):615–625. [PubMed: 15308962]
15. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. Jul; 2006 60(7):578–586. [PubMed: 16790829]
16. 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., editors. *Health Service Research Methodology: A Focus on AIDS*. Washington, D.C.: U.S. Public Health Service, National Center for Health Services Research; 1989. p. 113-159.
17. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. Sep; 2000 11(5):550–560. [PubMed: 10955408]
18. van der Laan, MJ.; Robins, JM. *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer; 2003.
19. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health*. 2000; 21:121–145. [PubMed: 10884949]
20. Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol*. Apr; 2002 31(2):422–429. [PubMed: 11980807]
21. Brumback BA, Hernan MA, Haneuse SJ, Robins JM. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Stat Med*. Mar 15; 2004 23(5):749–767. [PubMed: 14981673]
22. Cole SR, Hernan MA, Anastos K, et al. Determining the effect of highly active antiretroviral therapy on changes in human immunodeficiency virus type 1 RNA viral load using a marginal structural left-censored mean model. *Am J Epidemiol*. Jul 15; 2007 166(2):219–227. [PubMed: 17478436]
23. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*. Sep; 2000 11(5):561–570. [PubMed: 10955409]
24. Lippman SA, Donini A, Diaz J, et al. Social-Environmental Factors and Protective Sexual Behavior Among Sex Workers: The Encontros Intervention in Brazil. *Am J Public Health*. Oct 15.2009 epub.
25. Little, RJA.; Rubin, DB. *Statistical Analysis with Missing Data*. 2nd. New York: John Wiley & Sons; 2002.
26. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. Sep 15; 2008 168(6):656–664. [PubMed: 18682488]
27. Mortimer KM, Neugebauer R, van der Laan M, Tager IB. An application of model-fitting procedures for marginal structural models. *Am J Epidemiol*. Aug 15; 2005 162(4):382–388. [PubMed: 16014771]
28. Vienna, Austria: R Foundation for Statistical Computing; 2009. R: A Language and Environment for Statistical Computing [computer program]. <http://www.R-project.org>

29. Sinisi SE, van der Laan MJ. Deletion/substitution/addition algorithm in learning with applications in genomics. *Stat Appl Genet Mol Biol*. 2004; 3 Article18.
30. DSA: Data-Adaptive Estimation with Cross-Validation and the D/S/A Algorithm. R package version 3.1.3. [computer program]. <http://www.stat.berkeley.edu/~laan/Software/>
31. polyspline: Polynomial spline routines. R package version 1.1.1. [computer program]. 2008.
32. Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap*. Chapman & Hall/CRC; 1994.
33. Wang, Y.; Petersen, ML.; Bangsberg, D.; van der Laan, MJ. Diagnosing Bias in the Inverse Probability of Treatment Weighted Estimator Resulting from Violation of Experimental Treatment Assignment. UC Berkeley Division of Biostatistics Working Paper Series. 2006. <http://www.bepress.com/ucbbiostat/paper211>
34. Fitzmaurice, GM.; Laird, NM.; Ware, JH. *Applied longitudinal analysis*. Hoboken, N.J.: Wiley-Interscience; 2004.
35. Coates TJ, Richter L, Caceres C. Behavioural strategies to reduce HIV transmission: how to make them work better. *Lancet*. Aug 23; 2008 372(9639):669–684. [PubMed: 18687459]
36. Merson MH, O'Malley J, Serwadda D, Apisuk C. The history and challenge of HIV prevention. *Lancet*. Aug 9; 2008 372(9637):475–488. [PubMed: 18687461]
37. Jewell, NP. *Statistics for Epidemiology*. Boca Raton: Chapman & Hall/CRC; 2004.
38. Ahern J, Hubbard A, Galea S. Estimating the effects of potential public health interventions on population disease burden: a step-by-step illustration of causal inference methods. *Am J Epidemiol*. May 1; 2009 169(9):1140–1147. [PubMed: 19270051]
39. Petersen ML, Wang Y, van der Laan MJ, Bangsberg DR. Assessing the effectiveness of antiretroviral adherence interventions. Using marginal structural models to replicate the findings of randomized controlled trials. *J Acquir Immune Defic Syndr*. Dec 1.2006 43 1:S96–S103. [PubMed: 17133209]
40. Kurth T, Walker AM, Glynn RJ, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol*. Feb 1; 2006 163(3):262–270. [PubMed: 16371515]
41. Yanovitzky I, Zanutto E, Hornik R. Estimating causal effects of public health education campaigns using propensity score methodology. *Evaluation and Program Planning*. 2005; 28:209–220.

Table 1

Data set up

ID	Visit date	Interval	Present in interval	Exposure (participation) level	Time dependent covariates: self-efficacy *	Time independent covariates: age	Outcome (STI)	Treatment probability p(tx w)	Censoring probability p(cn w)	Inverse probability weight **
1	21-7-03	0	1		-1.48	47	0			
1	5-11-03	1	1	low	-1.48 *	47	0	.476	.782	2.69
1	9-2-04	2	1	high	-1.87	47	1	.524	.693	2.76
1		3	0	-	-1.87 *	47			.587	
1	27-7-04	4	1	high	-0.72	47	0	.524	.473	4.04
1		5	0	-	-0.72 *	47			.361	
2	21-4-03	0	1		0.72	33	0			
2	4-8-03	1	1	low	0.72 *	33	0	.711	.697	2.02
2	9-11-03	2	1	low	0.91	33		.711	.592	2.37
2		3	0	-	0.91 *	33	0		.478	
2	24-5-04	4	1	low	1.01	33	0	.733	.366	3.73
2	24-8-04	5	1	high	1.02	33	0	.289	.217	15.96

* carried forward from previous interval when data missing

** $IPW = \frac{1}{p(\alpha x | w)} \times \frac{1}{p(cn | w)}$

Table 2
Variable selection for weighting models

	Non-time-varying variables: measured at recruitment (time 0)	Time-varying variables: measured at multiple intervals
Continuous	Age	Number of sex acts in previous week
	Age began sex work	Number of unprotected sex acts in previous week
		Self-efficacy score
		Social cohesion score
		STI stigma score
Categorical	Gender	Possession of official documents
	Performed sex work outside of study city in the past 5 years	Report of any unprotected sex act with a new client in past 3 months
	Previously tested for HIV	Report any STI symptoms in past 3 months
	Knows someone w/ HIV or someone who died of AIDS	Changed residence since last visit
	Individual income (in tertiles)	
	Ever used cocaine or crack	
	Prevalent CT or NG infection	
	Frequency of drinking	

Comparison of modeling approaches for constructing inverse probability weights and estimating the odds of chlamydia or gonorrhea infection given high vs. low participation in the intervention.

TABLE 3

Model selection	Description	Median (IQR) weights*	Range of weights	Estimated OR	95% CI (robust)
Crude estimate				.51	(.26-.99)
1: Expert knowledge	Based on bivariate data and knowledge of STI determinants	7.7 (5.9-11.2)	2.8, 36.7	.43	(.22-.85)
2: Deletion Substitution Algorithm in R	Automated algorithm – includes power terms and interaction terms	7.5 (5.9-11.2)	2.0, 237	.49	(.25-.98)
3: Polyclass Algorithm in R	Automated algorithm – polychotomous regression including linear splines	3.4 (3.0-4.8)	1.9, 20.6	.50	(.25-1.0)
4: Stepwise selection procedure	Stepwise selection for confounders when $p < .1$, includes fractional polynomials	7.7 (5.1-13.2)	1.9, >1000	.53	(.26-1.1)

* IQR= inter quartile range; all weights truncated at 20 (equivalent to probability .05)