Title
Measuring and Modifying Community Social Factors to Reduce Alcohol Use and HIV Risk

Author
Leslie, Hannah Hogan

Publication Date
2015

Peer reviewed|Thesis/dissertation

Measuring and Modifying Community Social Factors to Reduce Alcohol Use and HIV Risk

by

Hannah Hogan Leslie

A dissertation submitted in partial satisfaction
of the requirements for the degree of
Doctor of Philosophy
in
Epidemiology
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor Jennifer E. Ahern, Chair
Professor Maya Petersen
Professor Mark Wilson

Summer 2015

ABSTRACT

Measuring and Modifying Community Social Factors to Reduce Alcohol Use and HIV Risk

by

Hannah Hogan Leslie

Doctor of Philosophy in Epidemiology

University of California, Berkeley

Professor Jennifer E. Ahern, Chair

This dissertation addresses methodological and substantive questions around the estimation of contextual effects, with a focus on the influence of community factors on alcohol use in South Africa. Alcohol use contributes to a multitude of health conditions and is a particularly important risk factor for South African men due to heavy consumption and the high risk of alcohol-related harms such as violence and road accidents in this population. In addition, HIV transmission and progression to AIDS may be accelerated by alcohol use. Research into the causes of dangerous alcohol use to date has focused on individual-level factors. This work addresses contextual traits that may relate to alcohol use, particularly community social factors. Specifically, I evaluate analytic methods for the measurement and aggregation of individual perceptions of social factors such as collective efficacy (chapter 2), test the associations of alcohol outlet density and village collective efficacy with young men's drinking behaviors in rural South Africa (chapter 3), and assess the village-level effect of a randomized community mobilization trial intended to modify gender norms on men's alcohol use (chapter 4).

Measuring social factors that are not directly observed is a key challenge in the estimation of contextual effects. I consider methods for estimating the effect of latent group factors on health outcomes using observed item responses within individuals. I compare approaches where the group-level exposure is calculated as the overall mean of items within persons within groups (aggregation and regression) with latent variable methods, namely item response modeling within individuals and structural equation modeling within groups. In particular, I explore the creation of multiple plausible values for individual perception and the use of these measures in a multilevel structural equation model. Simulation studies across a range of conditions to assess the robustness of these methods suggest that latent variable models reduce bias in the estimation of contextual effects relative to consistent attenuation in approaches based on aggregation and regression. This bias correction incurs additional variability, however. The causal model linking the latent construct to item responses affects appropriate analytic choice, as one setting in which aggregation and regression approaches perform well is when the group latent trait is a composite of individual values (formative indicators) and the sampling fraction is high. Finally, consideration of the role of third variables in affecting measurement, exposure, and outcome suggests that aggregation and regression approaches can be highly prone to bias in these scenarios, with some bias correction possible when adjusting for the aggregate value of a variable that distorts measurement. Latent variable methods provide reasonable bias correction in these situations without control for the aggregate of the third variable; adjustment for the distortion factor in the measurement model did not contribute substantially to this correction. As a whole, this work suggests that latent variable methods are worthy of further

consideration in social epidemiologic analysis, with additional work warranted on reducing the variability of such methods and comparing them to a single-stage, fully latent model.

Chapter 3 presents a novel analysis of social and environmental factors related to alcohol outcomes in rural South Africa. I assess the relationship of collective efficacy and alcohol outlet density with heavy drinking and potential problem drinking using a population-representative survey of 581 young men from 22 villages in Agincourt, South Africa. In this cross-sectional analysis, informal social control and social cohesion show protective associations with men's heavy drinking but not problem drinking: a one standard deviation difference in each factor was associated with a -4% difference in expected prevalence of heavy drinking among young men. The number of formal and informal alcohol-serving establishments per square kilometer was associated with potential problem drinking but not heavy drinking. The expected difference in prevalence of potential problem drinking was 9% for a difference of one additional alcohol outlet per square kilometer. Although preliminary, these findings lay the groundwork for further investigation of contextual causes of alcohol use and suggest that such causes could be worthwhile sites for future intervention.

The final component of this work is an assessment of a randomized community mobilization intervention intended to modify gender norms in an effort to reduce HIV transmission in Agincourt. Although reducing alcohol consumption was not a primary aim, the intervention included alcohol-related content, as inequitable masculine gender norms have been linked to risky alcohol use, including drinking before sexual activity. I test the total effect of the intervention on village prevalence of heavy drinking, potential problem drinking, and alcohol use before sex based on a follow-up sample of 575 young men. I categorize intervention engagement into low, moderate, and high doses based on the proportion of young men participating in the intervention and test the association of each dose level with each outcome as an assessment of nonlinearity. A nonlinear relationship would be consistent with the mobilization element of the intervention, which was intended to generate spillover effects from engaged individuals to peers who did not participate in the intervention. The intervention was not significantly associated with village prevalence of any of the alcohol outcomes. However, high dose of village engagement was significantly positively associated with the prevalence of heavy drinking and of potential problem drinking (14% difference and 9% difference relative to moderate engagement respectively). In contrast, the village-level relationship between increasing intervention dose and pre-sex alcohol use was protective and nonlinear, with the lowest estimated prevalence at moderate engagement, but did not reach statistical significance. These results indicate a potential harmful ecological effect of high intervention engagement on heavy drinking and potential problem drinking and are compatible with no effect or a protective effect on alcohol use before sex. Subsequent analyses should assess individual-level pathways between intervention engagement and alcohol use and consider whether intervention activities such as tavern-based discussions and large soccer tournaments could have inadvertently increased alcohol consumption.

As a whole, this work advances the methodological tools available for the analysis of contextual effects and provides new evidence of the importance of community causes of alcohol use in the context of HIV risk, an area with limited prior research. Extensions of this research can help to solidify a causal relationship from community context to alcohol use and identify ways to optimize future interventions in this area.

To my mom, Mary Hogan Leslie, whose compassion and convictions first led me to this journey, and to my daughter Ada, whose tenacity inspired me to finish it.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The number of components comprising the sufficient cause of this dissertation is enough to make any proponent of multifactorial causality proud. The universally necessary cause is without question my advisor and dissertation chair Jennifer Ahern. Jen inspired me to pursue social epidemiology long before I became her advisee; since then, she has been in every way an exceptional mentor. Jen is incredibly accessible and responsive, providing incisive guidance within hours, if not minutes, of fielding my wordy emails. She provides trenchant advice to any challenge that I present, whether a knotty methodological concern or a potentially life-altering career decision. Jen's insights challenge me to deepen my understanding of the question at hand; her faith that I would finish never wavered. The only downside of graduating is no longer being Jen's advisee, but she will always serve as an exemplar of the academic I aspire to be: passionate about public health, dedicated to the highest quality of research, and unfailingly generous with her time and expertise to those lucky enough to be her mentees.

My hope of working with Maya Petersen was one of the main reasons I returned to Berkeley for a PhD; I am so grateful to have benefitted from her instruction and her insights on this dissertation research. Maya's enthusiasm for applying cutting edge statistical methods to consequential public health research is infectious; her willingness to translate those methods so others can do the same is boundless. I am especially obliged to her for finding time to review a very long dissertation at the last minute, and for providing critiques and suggestions to greatly improve the final version. Mark Wilson introduced me to the world of measurement and has helped to guide me through this unfamiliar territory. I am indebted to him for sharing his wealth of expertise and for ensuring that my enthusiasm for new methods did not outstrip my capacity to understand and properly apply them. I particularly appreciate his willingness to review chapters amidst travel and nominal vacation; I could not have attempted to extend from epidemiology to measurement and latent variable models without his patient guidance and constructive feedback.

Beyond my committee, Art Reingold and Ray Catalano provided advice as well as departmental support for my doctoral studies. My undergraduate advisor Anne Firth Murray helped me find my path in public health and has never stopped inspiring me to act for a better world. Starley Shade went out of her way to mentor me as a junior analyst at UCSF: she encouraged me to pursue graduate research, facilitated my return to Berkeley, and made sure I had any support I needed during my first two years of doctoral study.

I am deeply grateful to the Alcohol Research Group (ARG) for predoctoral support from the Graduate Research Training on Alcohol Problems program [T32AA007240]. I thank Lee Kaskutas and Sarah Zemore for welcoming me to ARG and ensuring I had access to the breadth of expertise on alcohol research housed there. Vicky Fagan fielded my logistical inquiries and enabled me to continue participating in the ARG seminar from afar. My fellow fellows, especially Kara MacLeod, Mina Subbaraman, Priscilla Martinez, Paul Gilbert, and Camillia Lui, provided insight on all things alcohol-related and advice on all aspects of surviving academia. I will miss our seminar discussions! My mentors at ARG were an invaluable resource: I thank Doug Polcin for his steady encouragement and Kate Karriker-Jaffe for her willingness to take me on despite my lack of experience in alcohol research, her generosity in sharing her expertise in survey and neighborhood research to strengthen my work, and her unfailing optimism that I was making progress even when I did not believe that to be the case. Many, many thanks.

A version of the third chapter of this dissertation was published in *Health & Place,* volume 34, under the title, "Collective efficacy, alcohol outlet density, and young men's alcohol use in rural South Africa." I am indebted to my co-authors for making this work possible and for

permitting its inclusion here. Jennifer Ahern oversaw the project with the rigor she brought to my entire dissertation. Audrey Pettifor generously shared not only the data resulting from her years of work in Agincourt, but also her insights to sharpen my conclusions. Rhian Twine, Kathleen Kahn, and Xavier Gómez-Olivé were kind enough to work with a virtual stranger and to provide a thorough grounding in the Agincourt Health and Demographic Surveillance System and associated research. I appreciate their expertise and their willingness to help me understand the work to which they have contributed so much. Although not co-authors, I thank Audrey's students Ann Gottert and Molly Rosenberg for their generosity in sharing their experiences and knowledge on alcohol use and community mobilization in Agincourt. I am fortunate to have been able to get to know this group of researchers and to build upon their many years of work in HIV prevention. I would never have had this opportunity were it not for Sheri Lippman. Sheri is the invisible member of my dissertation committee: she has welcomed me wholeheartedly to her research team the past few years without any direct benefit of her own. She has given me access to a wealth of data that made this dissertation possible as well as to an even greater trove of insights and inquiries. I have left each meeting with her invigorated by her incredible passion for this work and teeming with ideas and opportunities for moving forward. I am so grateful for her guidance throughout my doctoral studies and excited for ongoing collaborations in our future.

One of the pleasures of working on this dissertation has been the discovery of just the right person to resolve a problem that seemed insurmountable. Torsten Neilands shared his extensive knowledge of latent variable models and found time amidst a multitude of commitments to test my code against his own. I am incredibly obliged to Ellie Colson for enabling me to run many days of analysis on the server; Ellie went out of her way to get everything done as quickly as possible - checking on it in the wee hours of the morning - and to let me know as soon and as graciously as possible whenever my code hit an error. My friend Andrea Niles kindly put me in touch with Andy Lin at UCLA, who provided pro bono statistical consulting services on the intricacies of multilevel structural equation modeling in Stata.

The friendship and intellectual exchange among my fellow students in the School of Public Health has been one of the greatest benefits of completing this degree. I admire Laura Balzer's ability to understand causal inference methods and appreciate her willingness to share these insights without hesitation. I was lucky enough to work with Laura Forastiere and Wenjing Zheng on causal mediation (a project we will finish someday!). I will always be impressed by Laura's ability to code a new statistical estimator overnight and show up for an early morning meeting unruffled, and by Wen's capacity to explain concepts that may be simple to her but are far from straightforward to the rest of us. I am particularly fortunate to have joined a cohort of fellow Epi nerds. Thank you to my friends (and co-GSIs / office-mates / practice quals committee / wonderful people) Irene Headen, Divya Vohra, Rose Kagawa, Paul Wesson, and Jade Benjamin-Chung for enlivening these years with intellectual debates, late night grading sessions (campus-wide power outages notwithstanding), biostatistics help, and a camaraderie that has made it all worth it. Deb Karasek and I first spoke over a Skype connection spanning nearly 10,000 miles; we subsequently did our best to spend as much of grad school as possible within a few feet of each other. I could not be more grateful for her friendship, our incredibly productive 'work' days, and her frequently expressed conviction that I could do it, whatever it might be.

From farmers' marketing to hiking and biking, my friends in the Bay Area and beyond have reminded me of life beyond graduate school. I owe particular debts of gratitude to Onath Claridge for his willingness to explain math to me at the drop of a hat, to Ariel Sklar for providing much needed encouragement at the eleventh hour, and to Master Simon Pyle for

sharing the insanity of being a residency spouse with me and especially for sacrificing hundreds of hours of computer time to run simulations for me, taking whatever measures were needed to keep the analysis churning (see Appendix Figure A.1). Friends: thank you.

I have had the good fortune of being born into one incredible family and marrying into another. My grandfather James Leslie did not live to see this work, but never doubted I could be capable of it. The Hogan clan has loved and supported me my whole life, no matter how far I've wandered. Fiona and Ken Holloway provide weekly encouragement from across the pond; I can't wait until we can raise a glass together to celebrate finishing a degree made easier through their tireless support. Katherine Thomas is the most perfectly quirky sister-in-law I could hope for. I am truly grateful for her enthusiasm for my work and her delicious desserts; I am inspired by her cheerful but no less relentless determination. George Leslie-Waksman brings many talents to the role of brother-in-law; most notable for dissertation purposes are his gift of entertaining cranky infants until they succumb to exhaustion and his superb drink-making abilities. If a random sample of my relationship with my in-laws Susan and Carl happened to draw the week that they both took off of work to devote themselves entirely to grandparenting (and feeding and transporting me wherever I needed to go) so that I could frantically overhaul my dissertation, the resulting estimate would be completely unbiased. Beyond being role models in their dedication to their own careers, Susan and Carl are the most generous and loving in-laws I could imagine.

I can never repay my immediate family for the love and support they provide. My stepmother Mary Woodford is always interested in my work and eager to hear the latest updates; if she could have flown to Berkeley to convince the school to let me graduate, I have no doubt she would have done so in a heartbeat. My father Jim Leslie has encouraged intellectual exploration since before I could walk; his own pursuit of excellence sets the standard to which I aspire. His deep pride and quiet love sustain me. My sister Grace and I have been comrades in arms from the time we were kids ourselves to now that we have kids of our own; I could not imagine where I would be without her. I was awed by her dedication to completing her PhD (and determined as the little sister that I would not be left behind for too long!). She has provided unstinting encouragement and assistance at every step of this process, from proofreading graduate school essays to guiding me through the postdoctoral application process. She gamely soldiered through drafts of this work ('Hannah, I think you might be making up words again.') and left them more clear than she found them. She is and will always be my hero as we continue offwards, downwards, and backwards forever.

Molly Thomas is my better half. Little did I know when we first met in a whimsical old house in Oxford that this quirky, brilliant, and frighteningly determined woman would be by my side over a decade later. I am grateful for her equanimity in celebrating with me on good days and cheering me up on bad, her patience in tolerating my terrible puns and overlooking the causal diagrams scribbled on every piece of scrap paper in our apartment, her knowledge in explaining how human health actually works in, you know, real individuals, and her care in making sure I didn't forget to eat too often. I am inexpressibly lucky to have such an unflappable partner for all the good things in life, from bike rides to blizzards, dinner parties to diaper parties. And in perhaps the best thing in life: our daughter Ada. Thank you little bean for your aggressively cheerful giggle, your most welcome ability to sleep through the night from an early age, and your unmatched determination to sit up / take a step / rule the world. You remind me to take joy from my own accomplishments, to shake off the tough times with a hug and a kiss, and to always eat dessert first. From the very first moment, the expectation of the happiness you could bring has been increasing to infinity; for that, I will always be a most grateful Mama.

Chapter 1: Introduction

PROBLEM STATEMENT

In the fourth decade of the global epidemic of Human Immunodeficiency Virus (HIV), the necessary causal agent of Acquired Immunodeficiency Syndrome (AIDS), countries around the world are making tangible progress in reducing incidence and extending survival. Increasing access to highly active antiretroviral therapy has extended lifespans and reduced transmission (1). Despite these achievements, HIV and AIDS remain potent forces of morbidity and mortality worldwide: at the end of 2013, an estimated 35.0 million individuals were living with HIV and 2.1 million were newly infected, with 70% of global prevalence and incidence occurring in Sub-Saharan Africa (2). HIV risk remains concentrated among vulnerable and marginalized sub-populations, particularly sex workers, injection drug users, and men who have sex with men. In addition, in the generalized (population-wide) epidemic affecting much of Sub-Saharan Africa, young women are at extremely high risk (2,3). In South Africa, an estimated 13% of women are living with HIV by the time they are 24 years old (1,4). Combatting HIV globally and decreasing risk for individuals in vulnerable populations in particular requires a combination of prevention strategies, including not solely biomedical intervention but also efforts to modify the social, economic, political, and environmental factors that structure risk and vulnerability (5,6).

An increasingly recognized element of the HIV risk environment, as well as a detrimental health outcome on its own, is harmful alcohol use. Alcohol consumption contributes to over 200 health conditions, including injury and both communicable and non-communicable diseases (7). In South Africa, heavy alcohol consumption poses a serious risk to public health, accounting for nearly 40,000 deaths and thousands of disability-adjusted life years due to outcomes ranging from interpersonal violence to alcohol use disorder and cirrhosis (8–11). These estimates do not include morbidity and mortality due to HIV, despite increasing evidence that alcohol use may act as a catalyst for HIV transmission. Biologically, alcohol consumption disrupts immune function; animal models of HIV suggest that alcohol can increase viral replication and hasten the progression of disease (12), although evidence on accelerated disease progression in humans is limited to date (13). Alcohol consumption is associated with behavioral risks, including less effective condom use, multiple partners, and increased violence, sexual coercion and assault (14–21). Meta-analyses support the association of alcohol consumption, particularly heavy episodic drinking, with incident and prevalent HIV (19,20). A number of HIV-risk reduction interventions have taken place in Africa among alcohol-using populations; these studies showed mixed results, with short-term reductions in sexual and alcohol-related risk behaviors that were not consistent across groups or sustained over time (22). Additional randomized trials of individual-level or situational (bar-based) alcohol risk reduction are currently planned or underway (23,24).

Despite the growing attention to the contribution of alcohol use to HIV risk in Sub-Saharan Africa, and South Africa in particular, limited research addresses broader community factors and social processes that may influence alcohol consumption. Based on research largely outside of South Africa, three plausible community-level factors may shape alcohol use: alcohol availability, collective efficacy, and gender norms. The density of alcohol outlets is associated with alcohol consumption and related harms at the population level, although studies of individual behaviors have shown mixed results (25–27). It is not yet known whether alcohol availability is associated with drinking outcomes in South Africa. Epidemiologic research has identified collective efficacy, "the linkage of mutual trust and the willingness to intervene for the common good" (28) as a social factor associated with a broad array of health behaviors and

outcomes, including self-rated health, asthma, depression, and HIV risk behaviors such as condom use (29–37). A small number of studies have addressed at least one element of collective efficacy and alcohol use, with research in the United States identifying a protective association between social cohesion and binge drinking (38). Comprehensive assessment of collective efficacy as it relates to alcohol use is scarce in South Africa; brief (one- to two-item) measures of this construct have been shown to be associated with lower social disorder and reduced alcohol consumption in cross-sectional studies (39,40). A final community-level factor implicated in harmful alcohol use in the context of HIV risk is gender equity, with a considerable body of research suggesting that highly inequitable masculine gender norms are associated with heavy alcohol use (41–43). Individual-level interventions on masculine gender norms have demonstrated success in modifying men's beliefs and, to a lesser extent, behaviors (44–46). Evidence from a quasi-experimental trial suggests that gender-based interventions and alcohol-reduction programs can each reduce HIV risk; integrating them could result in synergistic prevention (47). A community-randomized trial of a gender-transformative intervention indicated that such approaches show promise in at least medium-term (one-year) reduction in problem drinking (48).

STUDY PURPOSE

The purpose of this research is to assess the role of community factors on alcohol use in the context of HIV risk in rural South Africa. Specifically, I evaluate analytic methods for the measurement and aggregation of individual perceptions of social factors such as collective efficacy (Aim 1); I test the associations of alcohol outlet density and village collective efficacy with young men's drinking behaviors in Agincourt, South Africa (Aim 2); and I assess the village-level effect on men's alcohol use of a randomized community mobilization trial intended to modify gender norms (Aim 3).

SPECIFIC AIMS

Aim 1: To compare analytic methods for the creation and analysis of latent group-level factors in situations of measurement error and confounding.

Determining an association between collective efficacy and alcohol use, for example, requires the measurement of an unobserved, group-level construct that may be best accessed through subjective individual perception. I conduct simulation studies to compare the accuracy and variability of commonly applied approaches to this question, including simple score aggregation, application of a measurement model for score creation, and use of multilevel latent variable analyses. Each approach offers distinct advantages in estimation depending on the setting of interest, but they have not been systematically compared in the epidemiologic literature. I simulate data following a range of plausible data-generating processes and explore the capacity of each approach to correct for random measurement error as well as systematic error introduced by a third variable.

Aim 2: To examine the relation of community social and physical environmental factors with heavy alcohol consumption and potential problem drinking in a population-based sample of young men in rural South Africa.

I conduct the first analysis of how both physical and social community factors relate to young men's alcohol use in South Africa. Drawing on a representative survey of 581 men aged 18 – 35 in the rural Mpumalanga province of South Africa, I assess alcohol outlet density and

collective efficacy at the village level as they relate to heavy drinking and potential problem drinking. I employ g computation to calculate an interpretable estimate of the potential outcome of an intervention on either exposure. I test the sensitivity of results to reverse causation between drinking and collective efficacy and, following the analytic comparison in Aim 1, to an alternative measurement model for individual perceptions of collective efficacy.

Aim 3: To assess the impact of a community-level HIV prevention intervention on young men's alcohol use, including drinking before sexual activity.

I analyze a second population-representative survey from Agincourt collected following implementation of a village-level intervention intended to mobilize villagers around gender equitable norms and HIV prevention, including reducing dangerous alcohol use. I estimate the total effect of the intervention on prevalence of three alcohol outcomes among young men: heavy drinking, potential problem drinking, and alcohol use before recent sexual activity. I quantify intervention engagement within each village. I then assess the shape of the relationship between thresholds of increasing engagement and alcohol use for nonlinearity, an indicator compatible with the presence of spillover effects from peer engagement.

SIGNIFICANCE

This work will contribute to epidemiologic research in several ways. First, despite the importance to social epidemiology in particular of latent constructs measured through proxy variables, no comprehensive assessment of analytic options for this challenging situation exists in the epidemiologic literature. Moreover, existing discussions outside of epidemiology do not extend to the complex polytomous scales common to epidemiologic research, nor do they address the robustness of assumptions underlying latent variable methods or their performance in the presence of systematic measurement error. The work in Aim 1 thus extends existing consideration of analytic approaches for unobserved group constructs in the context of epidemiologic research. Second, the recognition of alcohol use as a critical health risk in South Africa has not yet been matched with full consideration of its potentially modifiable community causes. By assessing both physical and social exposures related to alcohol use using a large and representative survey of young adults, I extend the research on structural causes of drinking to an understudied setting, specifically to rural South Africa. The results may help inform future interventions around alcohol use for young adults, including whether the increasing number of community-level HIV interventions can be modified to affect alcohol use as one pathway to HIV risk. Third, the work in Aim 3 represents a novel analysis of an innovative cluster-randomized HIV prevention trial. The trial provides the opportunity to consider whether a community-level gender-focused intervention may modify alcohol use and, if so, whether such an effect might multiply as increasing numbers of community members are involved. This work sheds light on the extent of engagement required to catalyze community-level impact of such an intervention while laying the groundwork for future analyses on particular mechanisms of effect.

In sum, the major contributions of this research are to deepen our knowledge of structural causes of alcohol use, including the susceptibility of risky drinking behaviors to community-level intervention, and to strengthen the methodological rigor of social epidemiologic analysis through improved understanding of available options for ecological metrics. These developments will provide the basis for similar analyses within ongoing trials in Agincourt and elsewhere to develop stronger evidence for causal effects of community social factors on alcohol use and HIV risk.

ORGANIZATION OF THE DISSERTATION

This dissertation is organized into five chapters. Chapter 1 introduced the study question, specific aims, and significance. Chapter 2 presents the results of the methodological exploration of analysis of unobserved group constructs through simulation studies. Chapter 3 details the associations between community-level factors and young men's alcohol use from a cross-sectional, population-representative survey in rural South Africa. Chapter 4 covers the analysis of intervention effects on prevalence of alcohol outcomes, including alcohol use prior to sexual activity. Chapter 5 concludes the dissertation by reviewing implications of the three studies conducted and priorities for future research.

Chapter 2: Analytic approaches to modeling contextual effects in situations of confounding and measurement error

ABSTRACT

Determining whether characteristics of a group, such as the social capital of a community, contribute to health outcomes is a major undertaking in social epidemiology. Analyses of such contextual effects are challenging due to the multilevel nature of the causal question and, for unobserved constructs such as social capital, the need to rely on error-prone measurements such as individual perceptions to determine exposure. The purpose of this study is to test the application of analytic methods used in educational and organizational research, namely item response and structural equation modeling, to a social epidemiologic context. In particular we assess a novel approach combining multiple item response estimators per person with structural equation modeling compared to existing methods. We address three questions: 1) how robust are such methods to violations of their underlying assumptions?, 2) How can both simpler and more complex approaches address systematic measurement error induced by a third variable?, and 3) How do such approaches perform under an alternative causal model in which individual perception of the latent traits creates the group value rather than the reverse?

We describe six combinations of analytic approaches. Three are within person summaries of the observed items: item mean score, expected *a posteriori* (EAP) estimator from an item response model, and plausible values drawn around the EAP. Two options are explored for between person summaries: mean value of the individual score employed in a multilevel regression model and each individual observation modeled in a latent variable structural equation model (SEM). We apply each combination to a range of simulation studies, all of which reflect a typical large social epidemiologic study: polytomous items assessing the group trait, moderate within group sample size (40), moderate number of groups (40), and high within group variance. First, we assess the robustness of each approach to changes in the underlying item model, person distribution, and group distribution. Second, we extend the causal model to incorporate a third variable that can function as a distortion of perceived cohesion, a component cause of the outcome, and/or, once aggregated, a component cause of the group characteristic. We test whether controlling for this variable within the measurement model and at the individual and group levels of the outcome model reduces bias in estimating of the contextual effect. Third, we explore an alternative causal model in which individual perceptions combine to create the group trait, a scenario that resembles the structure assumed by a mean-based approach more than that of a latent variable approach. We assess the performance of each method under this scenario in conditions of high and low sampling ratio. Following the simulation studies, we assess the relationship between village organizational capacity and heavy episodic drinking among young men in rural South Africa as an applied example to integrate the results of the simulations.

Mean and regression-based analyses provide estimates of the contextual effect that are biased towards the null, while combining plausible values with SEM corrects for measurement error when the underlying causal model shows group-to-individual effects. Using a summarized score, whether an item mean or an EAP, as a single metric of person location within a multilevel SEM does not reduce bias and can lead to highly variable results. The six analytic approaches prove fairly robust to violations of their underlying assumptions, with the plausible value and SEM method typically returning the least bias. Incorporation of a third variable that distorts measurement increases bias that is not corrected by including the covariate in the measurement

model. In some conditions, mean and regression approaches require adjustment for the covariate at the group level to reduce bias, while the plausible value-SEM models show minimal bias when adjusted at the individual level for covariates operating as confounders. Residual bias remains high and some results are contradictory, suggesting further exploration is required. Finally, applying these approaches to an individual-to-group causal model favors mean and regression approaches when the sampling fraction is high, but SEM approaches when the sampling fraction is low. The applied example illustrates how utilizing multiple methods can serve as a sensitivity analysis to bolster or weaken causal claims of any one analysis.

This work as a whole suggests that the novel combination of plausible values and SEM offers an effective means of bias reduction in estimating contextual effects of latent variables. This approach proved relatively robust to violations of underlying assumptions, followed most closely to causal graphs when incorporating third variables that acted as confounders, and performed well in a setting at odds with its structural premise, provided that sampling fraction is low. However, it was more variable than a mean and regression approach. Future social epidemiologic analyses should consider incorporating item response modeling and SEM into analyses of latent traits, particularly if the objective is to minimize bias in estimation. Subsequent methodological research is warranted on how to improve SEM analyses with a single individual metric, whether intermediate models mixing latent and observed traits offer the bias correction effects seen here with less effect on variability, and whether causal graphs should be modified to better suit multilevel analyses.

INTRODUCTION

Motivation

Among the major concerns of social epidemiology is the investigation of multilevel contextual effects: the impact of collective characteristics on the health of the individual beyond any effect of an individual analogue of that characteristic. Researchers theorize that macro social-structural conditions such as norms, poverty, and policy shape mezzo-level social network structure and characteristics (49). These in turn provide opportunities for micro-level psychosocial mechanisms like social support to develop and ultimately impact health through behavior, psychological factors, and physiologic response. A considerable body of research supports the plausibility of this framework as applied to a variety of health outcomes. In the field of HIV for example, early prevention interventions identified community mobilization as a potentially effective mechanism for shaping individual risk behavior (50,51). The structural factors that enable mobilization and the pathways linking it to individual change continue to be a key area of study in HIV prevention around the world (52–54). A larger literature examines collective efficacy, the capacity of a group to realize its shared goals, as a protective factor for neighborhood safety and health (28). To date, studies have documented associations between collective efficacy and self-rated health, asthma, depression, condom use, and alcohol consumption (29–32,38). Public health practice, including the development of interventions to modify such collective characteristics to improve health, depends upon accurate assessment of these complex causal effects of interest.

Challenges in identifying contextual effects

Three major challenges arise in research on the contextual effect of social factors: 1) the complexity of the causal processes across multiple levels of observation, 2) an inability to directly observe the underlying exposure, and 3) the necessity of measuring some exposures

through individual perception. By definition, contextual effect analysis is concerned with the effect of a group (level 2, L2) characteristic on an individual (level 1, L1) outcome, such as health status. Epidemiologists have described a number of possible causal links across these levels (55). These include a direct effect across levels; an indirect effect mediated by an individual-level exposure; and cross-level effect modification, where the relationship between an individual exposure and an individual outcome differs by a group-level covariate, or the reverse, when a group exposure interacts with an individual-level covariate. Causal relationships aside from the effect of interest are also important to consider in the multilevel context: even if a covariate does not confound the exposure-outcome relationship at the individual level, it can be a group-level confounder if both the confounder distribution and outcome incidence rate vary by group (56). Identifying the causal structure of exposures at the group level thus requires consideration of group characteristics that could shape the distribution of exposure, covariates, and disease states across groups.

When the exposure of interest in contextual effect analysis is a social process, additional complexity is introduced due to this factor not being directly observable. Social processes are latent variables, where latent is defined broadly as a "random variable whose realizations are hidden" (57). In lieu of the true value of a latent variable, researchers devise indicators intended to provide an observable metric for the underlying unobservable characteristic. Such indicators may exist at the same level as the variable of interest, such as percent of residents voting in an election as an indicator of community civic participation, or at the individual level, for example perceived cohesion. Here the multilevel nature of the effect of interest intersects with its unobservable state: although direct observation at L2 is preferable in measuring group characteristics (58), in many cases, social processes and other latent variables can be feasibly measured only through individual perception. The limited nature of what investigators can observe necessitates careful consideration of the hypothesized causal model underlying exposure, measures, and outcomes.

Tools to address these challenges

Considerable research in epidemiology, education, and psychometrics has addressed the conceptual and methodological challenges of contextual analyses. We briefly review causal models for multilevel questions from education research as well as two related analytic approaches to latent variables, item response modeling (IRM) and structural equation modeling (SEM), before assessing how these tools can be combined to social epidemiologic research questions.

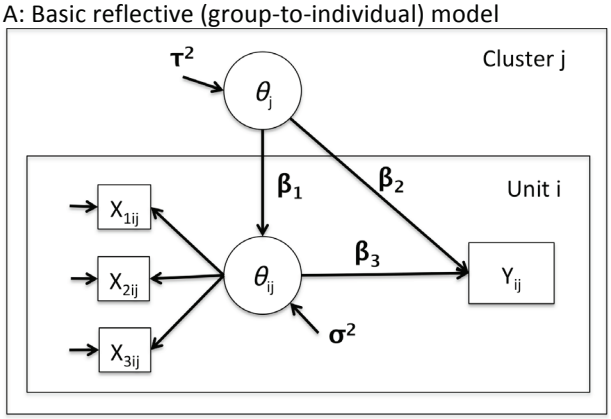
Causal models

Researchers in education, organizational research, and psychometrics have extended work on the measurement of latent individual traits through observed items to create a framework for measuring latent group traits through observed individual metrics (59–61). In conceptualizing the relationship between a group latent trait and individuals as 'indicators' or raters of that trait, researchers have proposed two broad classes of causal relationships: reflective indicators, which are downstream reflections of an underlying latent trait, and formative indicators, which are antecedent to the latent trait. Although 'measure' is typically conceptualized as for example a test or item assessing an individual-level characteristic, in the context of collective latent traits, individuals can serve as indicators or measures of the group characteristic, while their perception can be measured in turn (61). Structural models depicting a

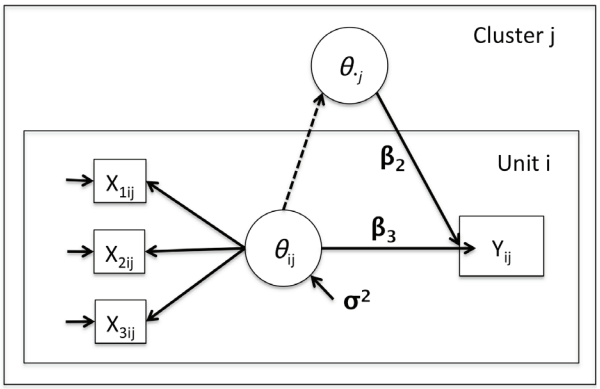
basic reflective model and a basic formative model are shown in Figure 2.1, panels A and B. (Throughout this document, latent variables are represented with Greek letters; observed variables with Roman. Schematic representations employ circles for latent variables and boxes for observed.) Figure 2.1A is a reflective model in that for individual i in cluster j, the latent group trait θ_j is a cause of individual perception θ_{ij}; the variability of θ_{ij} is represented as σ^2. Individual perspectives that serve as indicators of the group trait are measured through observed items X_{hij} (we depict three items for simplicity, though the actual number may be greater in practice). Both θ_j and θ_{ij} are component causes of the individual health outcome Y_{ij}. The contextual effect is represented by β_2. In contrast, panel B shows the causal model underlying formative indicators: the perceptions of individuals within a cluster form the components of the group trait θ_j (61). The dotted line represents the process by which θ_j is formed from θ_{ij}, such as summation. As in the reflective model, individual perception is measured by observed indicators X_{hij}; the contextual effect is represented by β_2.

Figure 2.1: Causal models of contextual effects of latent group traits measured through individuals

A: Basic reflective (group-to-individual) model



B: Basic formative (individual-to-group) model



We introduce here applied examples to facilitate understanding of these divergent causal structures. The reflective group-to-individual model depicts a situation in which for example, θ_j may represent the magnitude of informal organizational structures in a community (e.g., voluntary organizations, political advocacy groups), which speaks to the social capital of the cluster. Individuals perceive organizational density in their community as a function of the true group trait as well as individual variability. The actual social capital of the community as well as an individual's perception of it may both contribute to health behavior, such as alcohol consumption.

In the formative individual-to-group model, residents of an individual community possess opinions on the appropriate maintenance of collective order, perhaps as a function of historical processes that shaped the current community. The totality of those perspectives in turn shapes how social and anti-social behavior are monitored and potentially checked, which is θ_j, the group latent trait. Both individual perception and the functional summary of it can shape health outcomes such as drinking.

Distinguishing formative indicators from reflective indicators is a question of the process believed to underlie the group characteristic under study. Once hypothesized, this relationship has implications for measurement decisions. For reflective indicators, multiple assessments of a given indicator, such as ratings by different individuals within a cluster, are expected to be interchangeable since the indicator results from the common group construct (61). Discrepancies in ratings are ascribed to unreliability in the measurement. In considering how to aggregate

indicators from the observed level to an ecologic or group level, reflective indicators can be summarized as ratings subject to error, with a number of strategies available to account for error. In contrast, formative indicators are not assumed to be homogenous; deviation across individual responses is a substantive rather than nuisance finding. Formative indicators may be best summarized through metrics other than a simple mean, and can be considered reliable assessments of a latent group variable even when multiple ratings vary substantially from one another (61). While reflective indicators are assumed to be sampled from a population of infinite individuals, the underlying population individuals serving as formative indicators is assumed to be finite so that the individual values can combine to form the group latent trait; the proportion sampled from it is a key quantity in considering accuracy in estimating the latent trait (61).

Measurement models

Multiple analytic approaches exist for summarizing indicators both within individuals and within groups. Items can be averaged directly, assuming no measurement error, or can be modeled as indicators of a latent individual variable. Item response modeling (IRM) is a subset of SEM that provides a tool to fit nonlinear models to dichotomous or polytomous items by mapping individual location on the underlying latent trait and item difficulty onto a common metric (62). The probability that an individual at a certain location would endorse a specific item response is modeled as a function of individual location and fixed item traits (63). The true location of an individual can be predicted based on the assumed model. Further, the magnitude of error around a given individual location can be estimated; this enables generation of multiple plausible values for each individual estimate as a means of incorporating measurement error within estimation. In addition, individual covariates can be incorporated within the measurement model to improve predictions (64). IRM offers a number of strengths as a tool in developing and utilizing individual-level measures (65) and has been used in the measurement of group constructs as well, albeit less frequently than factor analysis (58). Extensions of IRM to multilevel settings include consideration of latent dependencies among subsets of items and the grouping of items within categories or other shared item stimuli (66–68). However, to our knowledge there has been no systematic inquiry on the potential contributions of IRM to measurement of latent group constructs, specifically the use of multiple plausible values for individual estimates and the ability to control for person characteristics in the measurement model.

Beyond the specific approach of IRM to estimating latent traits based on dichotomous observed items, structural equation modeling (SEM) more broadly is an estimation approach applicable to the use of effect indicators as metrics of a higher-level construct, whether the indicators are within person (items) or within groups (individuals). SEM treats group trait θ_j as an unobserved variable that can be decomposed into a total mean θ, group-specific deviations Υ_j, and individual deviations ϱ_{ij}. The contextual effect of interest is the link between true θ_j and observed Y_{ij} beyond any effect of individual-level θ_{ij} (the direct effect in epidemiologic terms). SEM enables estimation of θ_j by treating each individual rating X_{ij} (assuming here that X_{ij} is an observed variable at the individual level) as a measure of θ_j, akin to viewing individual items as a measure of an individual ability. The mean $\overline{X}_j$ of X_{ij} can then be corrected for unreliability based on the variation between X_{ij} (69). This approach can be extended when θ_j is not observed but is instead measured via items X_{hij} within individuals, in which case the correction for unreliability addresses both sampling and measurement error. While this correction can reduce

bias in the estimation of the effect of interest (61), parametric SEM requires placing distributional assumptions on θ, Υ_j, and ϱ_{ij}; simulation testing suggests the assumption of normality for the latent trait θ_j is the most fundamental to estimation (57,70,71). In addition, SEM estimation is a technically complex undertaking that may not be feasible for a given data set, particularly when attempting to model multiple levels of latent variables simultaneously, such as embedding a latent item response model within a latent group construct model.

A less complex alternative is to treat individual observations as error-free and to use the simple average of these observations as an unbiased estimate of the population mean in a standard regression model. This approach is implicitly based on a formative causal model, although this is infrequently recognized in practice. Applied epidemiologic research has employed both SEM and mean and regression methods (see for example references (31,72,73)).

To date, there has been no comprehensive assessment of the relative performance of these approaches in the context of epidemiologic research. We are in fact not aware of any epidemiologic research using plausible values for measurement. Systematic comparison of analytic options for aggregation of individuals as reflective indicators has been undertaken in education research, however. In two related studies, Lüdtke et al. investigated the use of SEM approaches at the item-to-individual and the individual-to-group levels (61,74). Relative to a simple aggregation approach, employing SEM reduced bias in the estimation of contextual effects in a finite sample. The cost of reduced bias was increased variability, particularly under conditions such as a small sample size, small number of groups, or low correlation of the exposure within groups (low intracluster correlation [ICC]). Employing SEM at just one level, such as modeling the group-level construct as a latent variable based on observed person scores, proved more reliable in these situations (61). These studies modeled items directly in a multilevel SEM without conducting an item response analysis. Other researchers have integrated IRM into multilevel SEM analyses. González et al. developed a double SEM by situating dichotomous item responses within models for latent traits of the individual respondents crossed with models of latent traits for the situation or context of particular item sets (68). In work more related to the question of individuals within groups, Cho et al. embedded IRM of dichotomous items within a two-level SEM when considering items within categories of items as the higher level grouping (67). This work included an initial simulation study suggesting the multilevel SEM model improved performance in item parameter estimation relative to a single-level fixed effects model; the authors recommended further testing of more complex situations, such as polytomous items and grouping by person rather than by item (67).

Study questions

We extend existing investigations of methodological approaches to contextual effects and situate this comparison within a social epidemiologic context. Specifically, we consider three item-to-individual analytic approaches – mean item score, IRM-based estimator, and plausible values around the IRM estimator – combined with two individual-to-group approaches: mean and regression versus multilevel SEM. We consider these two-stage approaches potential alternatives to a one-stage model embedding IRM within a multilevel SEM (full SEM approach) that may offer practical benefits for complex data. We are particularly interested in the performance of plausible values-based approaches, as they have not been tested in prior contextual effect research. We consider conditions relevant to a social epidemiologic study, namely a polytomous scale within individuals, moderate number of groups (40) and observations within groups (40), and high within-group variation. We address three questions: 1) under a

reflective indicator model, how robust is each method to violations of the underlying assumptions?; 2) How can both simpler and more complex approaches address systematic measurement error induced by a third variable?; and 3) How do such approaches perform under an alternative formative model in which individual perception of the latent traits creates the group value rather than the reverse? The first and third questions represent translations and extensions of research conducted outside of epidemiology; the consideration of a third variable is a novel, exploratory analysis. We conclude with an applied example considering the association of village social capital, operationalized through individual perceptions of informal organizations and networks, with individual heavy episodic drinking in rural South Africa.

METHODS

The methods and results sections cover four elements: the three questions mentioned above are assessed through simulation studies (robustness checks, incorporation of a third variable, formative causal model instead of reflective) followed by an applied example integrating findings. We define below the structural causal model underlying all simulation studies to establish the setting and notation used subsequently; we then describe in greater detail the analytic approaches introduced above that are applied for within- and between-person aggregation. We subsequently present the purpose and details of the sets of simulations corresponding to the three questions of interest before introducing the study underlying the applied example on social capital and heavy episodic drinking.

Structural causal model: Notation and definitions

Groups are indexed by $j = \{1, \ldots, n_j\}$, individuals within a group are indexed by $i = \{1, \ldots, n_i\}$, and items within an individual are indexed by $h = \{1, \ldots, n_h\}$. Response options within items are indexed $k = \{1, \ldots, K+1\}$ such that there are k steps or transitions between response options. The variables of interest are defined as:

- θ_j: latent, continuous group-level exposure
- θ_{ij}: latent, continuous individual perception of θ_j
- X_{hij}: observed categorical response to n_h items assessing θ_{ij}
- W_{ij}: observed vector of covariates for individual *i* in group *j*
- W_j: observed vector of covariates for group *j*
- Y_{ij}: observed indicator that individual *i* in group *j* has disease outcome of interest
- δ_{hk}: fixed difficulty of step *k* on item *h* (for instance, the probability of choosing strongly agree instead of agree)
- α_h: fixed discrimination (slope) of item *h*

The full group data are $F^{group} = (F_j : j = \{1, \ldots, n_j\})$, with $F_j = (\theta_j, \boldsymbol{\theta_{ij}}, \mathbf{X_{hij}}, \mathbf{W_{ij}}, \mathbf{W_j}, \mathbf{Y_{ij}})$, where $\boldsymbol{\theta_{ij}}$, $\mathbf{X_{hij}}$, $\mathbf{W_{ij}}$, and $\mathbf{Y_{ij}}$ are vectors of one or more individual-level variables while $\mathbf{W_j}$ is a vector of cluster-level variables. The full person data are $F^{pers} = (F_i : i = \{1, \ldots, n_i\})$, with $F_i = (\theta_{ij}, \mathbf{X_{hij}}, \mathbf{W_{ij}}, Y_{ij})$. The full item data are $F^{item} = F_h : h = \{1, \ldots, n_h\})$, with $F_h = (\delta_{hk}, \alpha_h)$. Exogenous variables, defined as $U = (U_j : j = \{1, \ldots, n_j\}) \sim P_U$, with $U_j = \left(U_{\theta_j}, U_{\theta_{ij}}, U_{X_{hij}}, U_{W_{ij}}, U_{W_j}, U_{Y_{ij}}\right)$, are the vectors of unmeasured factors contributing to the group social processes and individual outcomes. (Although *U* encompasses latent variables, we follow conventions in epidemiology in using *U* rather than a Greek letter to distinguish this combination of unmeasured and unknown group- and individual-level processes from a single distinct latent construct, such as θ_j, that can

be defined and measured.) Item difficulty and discrimination are considered fixed. The structural equations comprising the general causal model are:

- For $j = \{1, \dots, n_j\}$:
  - $\mathrm{W_j} = f_{W_j}\left(U_{W_j}\right)$
  - $\theta_j = f_{\theta_j}\left(\mathrm{W_j}, U_{\theta_j}\right)$
- For $i = \{1, \dots, n_i\}$:
  - $\theta_{ij} = f_{\theta_{ij}}\left(\theta_j, W_j, U_{\theta_{ij}}\right)$
  - $\mathrm{Y_{ij}} = f_{Y_{ij}}\left(\theta_j, \theta_{ij}, \mathrm{W_{ij}}, U_{Y_{ij}}\right)$
- For $h = \{1, \dots, n_h\}$:
  - $\mathrm{X_{hij}} = f_{X_{hij}}\left(\theta_{ij}, \mathrm{W_{ij}}, \delta_{hk}, \alpha_h, U_{X_{hij}}\right)$

The observed data are $\mathrm{O} = (\mathrm{O_j} : j = \{1, \dots, \mathrm{n_j}\})$, where $\mathrm{O_j} = (\mathbf{X_{hij}}, \mathbf{W_{ij}}, \mathbf{W_j}, \mathbf{Y_{ij}})$.

Analytic notation and definitions

We estimate a number of quantities from the observed data. Applying an item response model to the items enables generation of $\hat{\theta}_{ij}$, the estimated person location of the latent trait θ_{ij}. We define below two options for this location, denoted $\hat{\theta}_{ij}^{EAP}$ and $\hat{\theta}_{ij}^{PV_r}$. The aggregate (mean) of any observed or calculated variable to a higher level is denoted through an overbar and a dot replacing the lower-level index *h* or *i*. We thus define:

$$\overline{\mathrm{X}}_{\cdot\mathrm{ij}} = \frac{1}{\mathrm{n_h}} \sum_{\mathrm{h}=1}^{\mathrm{n_h}} \mathrm{X_{hij}} \qquad 1$$

$$\overline{\mathrm{X}}_{\cdot\cdot\mathrm{j}} = \frac{1}{\mathrm{n_i}} \sum_{i=1}^{n_i} \overline{\mathrm{X}}_{\cdot\mathrm{ij}} \qquad 2$$

$$\overline{\hat{\theta}}_{\cdot j}^{EAP} = \frac{1}{n_i} \sum_{i=1}^{n_i} \hat{\theta}_{ij}^{EAP} \qquad 3$$

$$\overline{\hat{\theta}}_{\cdot j}^{PV_r} = \frac{1}{n_i} \sum_{i=1}^{n_i} \hat{\theta}_{ij}^{PV_r} \qquad 4$$

$$\overline{\mathrm{W}}_{\cdot\cdot\mathrm{j}} = \frac{1}{\mathrm{n_i}} \sum_{i=1}^{n_i} \mathrm{W_{ij}} \qquad 5$$

Variance of the group exposure is denoted $var\left(\theta_j\right) = \tau^2$ and variance of individual perception is $var\left(\theta_{ij}\right) = \sigma^2$; variance within each individual is denoted ε^2.

Target parameter

We wish to estimate the effect of manipulating θ_j on the vector of individual outcomes Y_{ij} while holding θ_{ij} constant. In other words, we target the controlled direct effect of θ_j not mediated by its effect via θ_{ij}. We further assume no interaction between θ_j and θ_{ij} such that this direct effect is constant across levels of θ_{ij} (75). Let $Y_{ij}(\vartheta_j, \vartheta_{ij})$ denote the counterfactual value that Y_{ij} would take under a manipulation to set $\theta_j = \vartheta_j$ and $\theta_{ij} = \vartheta_{ij}$. For comparability with prior research estimating a single coefficient in a logistic regression model, we target the direct effect summarized as a conditional odds ratio:

$$\frac{E\left(Y_{ij}(\vartheta_j, \vartheta_{ij})|W_{ij}\right)/\left(1 - E\left(Y_{ij}(\vartheta_j, \vartheta_{ij})|W_{ij}\right)\right)}{E\left(Y_{ij}(\vartheta_j - \tau, \vartheta_{ij})|W_{ij}\right)/\left(1 - E\left(Y_{ij}(\vartheta_j - \tau, \vartheta_{ij})|W_{ij}\right)\right)} \qquad 6$$

Under the causal model defined above and under the assumption of independent *U* (i.e., no unmeasured confounding), this causal target parameter is equivalent to the following associational parameter:

$$\frac{E\left(Y_{ij}\middle|\theta_j = \vartheta_j, \theta_{ij} = \vartheta_{ij}, W_{ij}\right)}{\left(1 - E\left(Y_{ij}\middle|\theta_j = \vartheta_j, \theta_{ij} = \vartheta_{ij}, W_{ij}\right)\right)} \Bigg/ \frac{E\left(Y_{ij}\middle|\theta_j = \vartheta_j - \tau, \theta_{ij} = \vartheta_{ij}, W_{ij}\right)}{\left(1 - E\left(Y_{ij}\middle|\theta_j = \vartheta_j - \tau, \theta_{ij} = \vartheta_{ij}, W_{ij}\right)\right)} \qquad 7$$

$$= \frac{E\left(Y_{ij}|\theta_j = \vartheta_j, \theta_{ij}, W_{ij}\right)}{\left(1 - E\left(Y_{ij}|\theta_j = \vartheta_j, \theta_{ij}, W_{ij}\right)\right)} \Bigg/ \frac{E\left(Y_{ij}|\theta_j = \vartheta_j - \tau, \theta_{ij}, W_{ij}\right)}{\left(1 - E\left(Y_{ij}|\theta_j = \vartheta_j - \tau, \theta_{ij}, W_{ij}\right)\right)} \qquad 8$$

The equality between equation 7 and equation 8 holds due to the assumption of no interaction. This contrast represents the conditional odds ratio for outcome Y_{ij} if θ_j is set to ϑ_j versus $\vartheta_j - \tau$, one standard deviation lower, and θ_{ij} is held constant. Other parameters of equal interest can be defined; this level of contrast was chosen to represent a conceivable degree of change in a latent trait, given its lack of natural scale. In an applied setting, this parameter could represent the odds of heavy episodic drinking (conditional on measured baseline covariates) if all young men lived in a community with social capital equal to level ϑ_j relative to the odds if the level of social capital were reduced by one standard deviation, with any effect of social capital on individual perception of social capital prevented.

Analytic approaches within person

This research explores two-stage estimation approaches to contextual effects: within-person summation of observed items and between-person aggregation to group estimates. For the first stage, we describe here the construction and underlying assumptions for three within-person

analytic approaches to summarizing observed items: mean score $\overline{X}_{ij}$, expected *a posteriori* (EAP) person location estimate $\hat{\theta}_{ij}^{EAP}$, and plausible values for person location estimates $\hat{\theta}_{ij}^{PV_r}$.

Each item response is assigned a numeric value from 0 to K indicating the count of completed item steps - in other words, the number of thresholds between the lowest item response and the selected response, ordered for example from strongest disagreement to strongest agreement. The most basic approach to estimating individual perception θ_{ij} is to average these item responses per person. Using a mean score to summarize the items can be justified statistically or through principles of item response theory, given certain assumptions. From a statistical viewpoint, the mean of a set of items would be an unbiased estimator of true person score (perception, θ_{ij}) as the number of items n_h increases to infinity only under the assumption that items are selected from a pool of independent and identically distributed (i.i.d.) items. Under item response theory, raw item score is a sufficient statistic for person location under a family of models known as Rasch models. These models propose a parametric form for the structural equation $X_{hij} = f_{X_{hij}}\left(\theta_{ij}, W_{ij}, \delta_{hk}, \alpha_h, U_{X_{hij}}\right)$. Models such as the one-parameter logistic model for dichotomous data or the partial credit model for polytomous data model the probability of a given item response as a function of person location θ_{ij} and item or step difficulty δ. Rasch models impose two critical assumptions: independence of items conditional on θ_{ij} and δ, and measurement invariance, whereby item difficulties $\alpha_h$do not vary across individual locations (63,76). The partial credit model, which extends the basic Rasch model to account for polytomous responses (77), expresses the probability of obtaining a score *x* on item *h* for individual *i* in cluster *j* as:

$$P\left(X_{hij} = x|\theta_{ij}\right) = \frac{exp\left(\sum_{k=0}^{x}\left(\theta_{ij} - \delta_{hk}\right)\right)}{\sum_{0}^{k} exp\left(\sum_{k=0}^{x}\left(\theta_{ij} - \delta_{hk}\right)\right)} \qquad 9$$

An alternative measurement model that does not assume invariance is:

$$P\left(X_{hij} = x|\theta_{ij}\right) = \frac{exp\left(\sum_{k=0}^{x}\alpha_h\left(\theta_{ij} - \delta_{hk}\right)\right)}{\sum_{0}^{k} exp\left(\sum_{k=0}^{x}\alpha_h\left(\theta_{ij} - \delta_{hk}\right)\right)} \qquad 10$$

In equation 10, the additional parameter α_h represents item discrimination; discrimination modifies the relationship between item difficulty and the probability of a response, such that the order of item thresholds may vary for individuals at different levels of the latent construct. Equation 10 is the generalized partial credit model, one of the two-parameter model family; setting α_h=1.0 for all items reduces the model to equation 9. The use of a simple mean based on raw item score, including the standardized average score, presumes that the underlying measurement model is of the Rasch form, where discrimination is held constant at 1.0 rather than allowed to vary per item.

Although raw score is a sufficient statistic for Rasch models, this does not mean that the raw score is an estimator for θ_{ij}. Item response methods provide a number of means of estimating individual location θ_{ij} that extend beyond raw score. Among the most common is the calculation of an EAP estimator also known as an empirical Bayes estimate. Estimation of the EAP requires specifying an assumed prior distribution of individual locations and discretizing it into quadrature points (78). This prior is combined with the likelihood function of observed response patterns to create a posterior distribution for each response pattern; the mean of this distribution is the estimated individual location. EAP estimates are computationally

straightforward to calculate and, if the prior is correct, minimize the mean squared error over the population. They can be biased due to the prior being incorrect or the prior shrinking estimates towards its mean when the number of items informing the likelihood is small (78). In addition, EAP estimates can be generated from a range of item response models, including models with varying item discrimination. Calculation of person location through the EAP also supports a useful extension: estimation of multiple plausible values per person. Under item response theory, a standard error of measurement can be calculated for each person location to assess the precision of estimated location. This error can be used to generate a distribution around each EAP estimate from which multiple estimates of location can be drawn. These plausible values (PV) help to capture the measurement error inherent in estimating latent person location: the more disparate the plausible values, the greater the measurement error and hence the greater uncertainty in any estimates based on person location. Plausible values are imputed data values and can be analyzed just as data imputed for missing values of observed covariates would be: analysis is repeated with each plausible value and the results combined to enable quantification of the impact of measurement error on any quantity of interest (79). A second extension of EAP calculations, which can also be applied in generation of plausible values, is the incorporation of other individual traits in estimation through latent regression. Controlling for external variables that are associated with individual perception can increase the precision of individual ability estimates (80).

The three methods described here for creating an individual score from a number of items – item mean, EAP estimator, and plausible values – are representative of the broad array of approaches that could be applied. The item mean is straightforward and asymptotically unbiased, although it may be biased given the small number of items on most measurement scales; the EAP estimate draws on item response theory to model potential item complexity but depends on appropriate choice of a prior distribution, and the plausible values approach extends from the EAP to incorporate measurement error. The performance of each approach in terms of optimizing accurate estimation of a contextual effect can be expected to vary based on the true underlying item model and individual traits.

Analytic approaches between persons

We now present two methods for the second stage of estimation, person to group: aggregation and regression and SEM. Combining the three within-person methods with these two between-person estimation approaches yields six analytic approaches for assessment. Individual scores can be averaged within groups to yield $\overline{\mathrm{X}}_{\cdot \mathrm{j}}$, $\widehat{\widehat{\theta}}_{\cdot j}^{EAP}$, and $\widehat{\widehat{\theta}}_{\cdot j}^{PV_r}$. The mean is an asymptotically unbiased estimate of the group latent trait θ_j, assuming the selected individuals are a random subset of the group population. It does not require distributional assumptions about θ_j. As an observed variable, $\overline{\mathrm{X}}_{\cdot \mathrm{j}}$, $\widehat{\widehat{\theta}}_{\cdot j}^{EAP}$, and $\widehat{\widehat{\theta}}_{\cdot j}^{PV_r}$ can be entered in traditional regression models, such as a hierarchical model regressing $\mathrm{Y_{ij}}$ on exposure and any covariates, with standard errors adjusted for the dependence of individuals within groups. This aggregation and regression approach follows the causal logic of the formative indicator model, in which individuals contribute their own latent trait value to the group composite, both of which may shape individual outcomes. In a finite population, however, the mean of the subset of observed individuals is an unreliable measure of the latent group construct, particularly for studies with small sample sizes and low ICC of the group trait, which may introduce bias into the estimated regression coefficients (61).

SEM offers an alternative to the mean and regression approach that treats each observation (individual in this case) as an error-prone measurement of the underlying latent truth. Following an effect indicator model, differences between individual responses are taken as indication of measurement error and used to correct effect estimates for attenuation due to measurement error. Reduced bias in estimation comes at the cost of increased variance (61) as well as potential sensitivity to violations of the underlying distributional assumptions.

Analytic approach summary

We have defined six distinct analytic approaches to the estimation of contextual effects from observed items within individuals. Table 2.1 summarizes the combinations of the three within-person and two between-person analytic methods, along with the components required for each method. Throughout the remainder of this research, we consider the primary contrast of interest to be between M and PV-based approaches, as the PV approaches, particularly PV-SEM are the most novel and the most distinct from prevailing social epidemiologic practice of M-MR. We assess EAP-based approaches in initial simulations to distinguish between implications of using an IRM-based score and using multiple metrics.

Table 2.1: Analytic approaches applied in simulation study: acronym, definition of exposure components (measures of θ_{ij}, θ_j) required

Within persons Between persons	Mean	EAP	PV (r=1, …, 5)
Mean & regression (MR)	M-MR $\overline{\mathrm{X}}_{\cdot \mathrm{ij}} = \frac{1}{\mathrm{n_h}} \sum_{\mathrm{h}=1}^{\mathrm{n_h}} \mathrm{X_{hij}}$ $\overline{\mathrm{X}}_{\cdot \cdot \mathrm{j}} = \frac{1}{\mathrm{n_i}} \sum_{i=1}^{n_i} \overline{\mathrm{X}}_{\cdot \mathrm{ij}}$	EAP-MR $\hat{\theta}_{ij}^{EAP}$ $\bar{\hat{\theta}}_{\cdot j}^{EAP} = \frac{1}{\mathrm{n_i}} \sum_{\mathrm{i}=1}^{\mathrm{n_i}} \hat{\theta}_{ij}^{EAP}$	PV-MR $\hat{\theta}_{ij}^{PV_r}$ $\bar{\hat{\theta}}_{\cdot j}^{PV_r} = \frac{1}{\mathrm{n_i}} \sum_{\mathrm{i}=1}^{\mathrm{n_i}} \hat{\theta}_{ij}^{PV_r}$
SEM	M-SEM $\overline{\mathrm{X}}_{\cdot \mathrm{ij}} = \frac{1}{\mathrm{n_h}} \sum_{\mathrm{h}=1}^{\mathrm{n_h}} \mathrm{X_{hij}}$	EAP-SEM $\hat{\theta}_{ij}^{EAP}$	PV-SEM $\hat{\theta}_{ij}^{PV_r}$ (r=1, …, 5)

Linking these approaches back to the target parameter, we note that, due to the latent nature of θ_j and θ_{ij}, equation 8 is not a parameter of the observed data distribution. We therefore consider estimation approaches based on targeting the following "proxy" target parameters:

$$\frac{E\left(Y_{ij}|\bar{X}_{\cdot\cdot j}=\bar{x}_{\cdot\cdot j},\bar{X}_{\cdot ij},W_{ij}\right)/\left(1-E\left(Y_{ij}|\bar{X}_{\cdot\cdot j}=\bar{x}_{\cdot\cdot j},\bar{X}_{\cdot ij},W_{ij}\right)\right)}{E\left(Y_{ij}|\bar{X}_{\cdot\cdot j}=\bar{x}_{\cdot\cdot j}-\hat{\tau},\bar{X}_{\cdot ij},W_{ij}\right)/\left(1-E\left(Y_{ij}|\bar{X}_{\cdot\cdot j}=\bar{x}_{\cdot\cdot j}-\hat{\tau},\bar{X}_{\cdot ij},W_{ij}\right)\right)} \quad 11$$

$$\frac{E\left(Y_{ij}|\bar{\hat{\theta}}_{\cdot j}^{EAP}=\bar{\vartheta}_{\cdot j}^{EAP},\hat{\theta}_{ij}^{EAP},W_{ij}\right)/\left(1-E\left(Y_{ij}|\bar{\hat{\theta}}_{\cdot j}^{EAP}=\bar{\vartheta}_{\cdot j}^{EAP},\hat{\theta}_{ij}^{EAP},W_{ij}\right)\right)}{E\left(Y_{ij}|\bar{\hat{\theta}}_{\cdot j}^{EAP}=\bar{\vartheta}_{\cdot j}^{EAP}-\hat{\tau}^{EAP},\hat{\theta}_{ij}^{EAP},W_{ij}\right)/\left(1-E\left(Y_{ij}|\bar{\hat{\theta}}_{\cdot j}^{EAP}=\bar{\vartheta}_{\cdot j}^{EAP}-\hat{\tau}^{EAP},\hat{\theta}_{ij}^{EAP},W_{ij}\right)\right)} \quad 12$$

$$\frac{E\left(Y_{ij}|\bar{\bar{\theta}}_{\cdot j}^{PV}=\bar{\bar{\vartheta}}_{\cdot j}^{PV},\hat{\theta}_{ij}^{PV},W_{ij}\right)/\left(1-E\left(Y_{ij}|\bar{\bar{\theta}}_{\cdot j}^{PV}=\bar{\bar{\vartheta}}_{\cdot j}^{PV},\hat{\theta}_{ij}^{PV},W_{ij}\right)\right)}{E\left(Y_{ij}|\bar{\bar{\theta}}_{\cdot j}^{PV}=\bar{\bar{\vartheta}}_{\cdot j}^{PV}-\hat{\tau}^{PV},\hat{\theta}_{ij}^{PV},W_{ij}\right)/\left(1-E\left(Y_{ij}|\bar{\bar{\theta}}_{\cdot j}^{PV}=\bar{\bar{\vartheta}}_{\cdot j}^{PV}-\hat{\tau}^{PV},\hat{\theta}_{ij}^{PV},W_{ij}\right)\right)} \quad 13$$

Finally, we consider the case that the true conditional distribution of Y_{ij} given $\overline{X}_{\cdot j}$, $\overline{X}_{ij}$, and W_{ij} is known *a priori* to be described by a main term logistical statistical model, such that

$$E\left(Y_{ij}|\overline{X}_{\cdot j},X_{\cdot j},W_{ij}\right)=\frac{1}{1+exp\left(-\left(\beta_0+\beta_2\overline{X}_{\cdot j}+\beta_3\overline{X}_{ij}+\beta_4W_{ij}\right)\right)} \quad 14$$

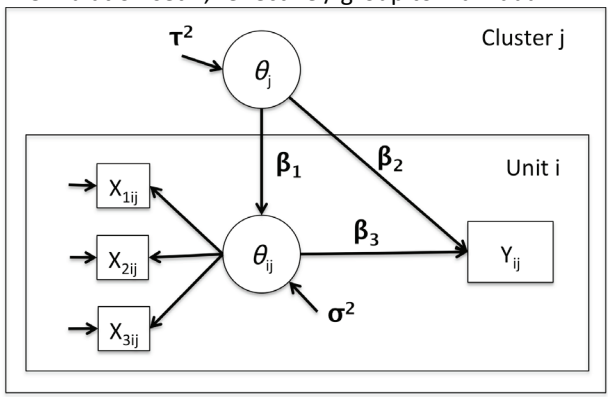
We thus consider estimators $\hat{\beta}_2$ for the contextual effect β_2 based on fitting a main term logistic regression according to model 14 across all analytic approaches (substituting $\hat{\theta}_{ij}$ for $\overline{X}_{ij}$ and $\bar{\bar{\theta}}_{\cdot j}$ for $\overline{X}_{\cdot j}$ for each analytic approach as noted in Table 2.1).

Monte Carlo simulations

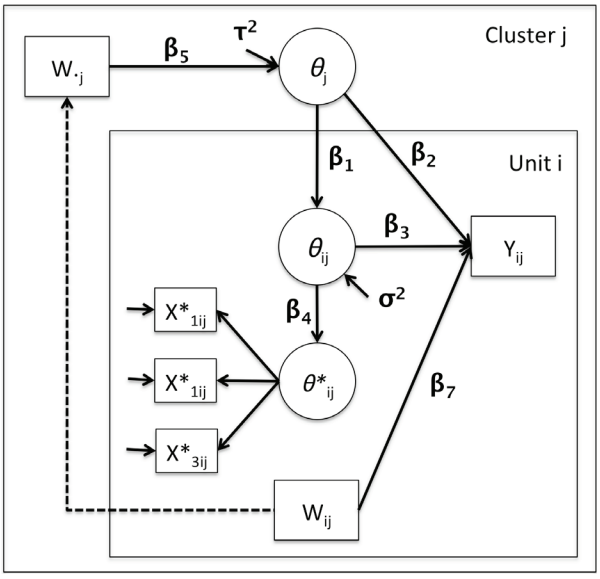
We undertake three sets of simulations to address the questions of interest using these six analytic approaches. Simulation set 1 assesses robustness to violation of underlying assumptions, set 2 explores analytic options for reducing bias due to systematic measurement error and confounding, and set 3 compares the performance of each analytic method under an alternative formative causal model. The general causal model for each set of simulations is shown in Figure 2.2, panels A through C respectively.

Figure 2.2: Causal models underlying simulation study

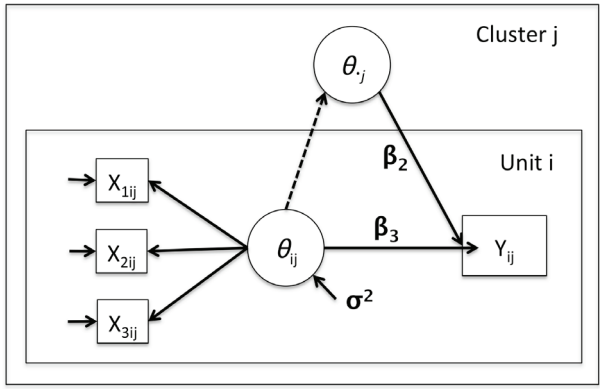
A: Simulation set 1, reflective / group to individual



B: Simulation set 2, reflective model with covariates



C: Simulation set 3, formative / individual to group



All simulations were intended to represent a feasible social epidemiologic study; we drew from a range of example studies to determine reasonable levels for the cross-simulation generating values (28,31,38,81,82). All simulations tested a balanced design with a large but attainable sample size: $n_j = 40$ groups, $n_i = 40$ individuals. To mimic a typical scale employed in large-scale surveys, $n_h = 7$ items were generated with K = 4 step parameters (5 response options). Difficulty levels were selected to create items with plausible but varying response distributions; the step difficulty matrix and sample distributions for each item are shown in Appendix Table B.1 and Figure B.1. All responses were complete (no missing data). Although the details of latent variable θ_j varied across simulations, the ICC was held at 0.10 unless specified otherwise, a value in keeping with those observed in social epidemiologic studies. The generating value of reliability for analyses using plausible values is 0.81.

For all simulations, we explore a simplistic setting (A) where $\beta_1 = 1$ (link from θ_j to θ_{ij} in reflective causal models), $\beta_2 = 0.3$ (contextual effect of θ_j on Y_{ij}), and $\beta_3 = 0$ (individual-level link from θ_{ij} to Y_{ij}) and a more realistic setting (B) where $\beta_1 = 1$, $\beta_2 = 0.3$, and $\beta_3 = 0.1$. In both cases the contextual effect of θ_j on Y_{ij} is of modest magnitude in keeping with results from past social epidemiologic studies. We analyze data for each setting for all sets of simulation parameters detailed below. Given the large number of simulations undertaken and analytic methods employed, we use setting A as a screening method to identify the analytic approaches worth full presentation and discussion. Setting A results are summarized below and presented in full in the Appendix; setting B results are presented and discussed in full.

Conditions for all simulations are summarized in Table 2.2; cross-simulation conditions are discussed below and conditions for each set detailed subsequently.

Table 2.2: Simulation characteristics

	Relationship of individual perception to group latent trait		Item-generating model		Distribution of θ_{ij}		Distribution of θ_j		W_{ij} role		
No.	Reflective (from group to individual)	Formative (from individual to group)	1PL	2PL	Normal	Non-normal	Normal	Non-normal	W_{ij} distorts θ_{ij}	W_{ij} affects Y_{ij}	$\bar{W}_{\cdot j}$ affects θ_j
Simulation set 1: Robustness checks – Figure 2.2A											
1.0	X		X		X		X				
1.1	X			X	X		X				
1.2	X		X			X	X				
1.3	X		X			X	X				
1.4	X		X		X			X			
1.5	X		X		X			X			
Simulation set 2: Role of third variable – Figure 2.2B											
2.0	X		X		X		X		X		
2.1	X		X		X		X		X	X	
2.2	X		X		X		X		X	X	X
2.3	X		X		X		X			X	X
Simulation set 3: Applicability of SEMs to formative indicator model – Figure 2.2C											
3.0†		X	X		X		X				
3.1†		X	X		X		X				

†The difference between these simulations is the sampling fraction: 0.50 in 3.0, 0.05 in 3.1.

Simulation set 1: Robustness checks under reflective causal model

The purpose of this group of simulations was to test the robustness of each analytic approach to violations of underlying assumptions in a simple setting without covariates. Specifically, after establishing a baseline comparison among methods, we vary the underlying item generating model, the distribution of θ_{ij}, and the distribution of θ_j. Following the reflective causal model in Figure 2.2A, the baseline data-generating process under each setting (A: $\beta_1 = 1$, $\beta_2 = 0.3$, and $\beta_3 = 0$, B: $\beta_1 = 1$, $\beta_2 = 0.3$, and $\beta_3 = 0.1$) is:

Simulation 1.0: Baseline

$$U_{\theta_j} \sim N(0, \tau^2), \tau^2 = 1$$
$$\theta_j = U_{\theta_j}$$
$$U_{\theta_{ij}} \sim N(0, \rho^2), \rho^2 = 8$$
$$\theta_{ij} = \beta_1\theta_j + U_{\theta_{ij}}$$
$$\mathrm{Y_{ij}} \sim Binom\left(1/\left[1 + \exp\left(-\left(-2 + \beta_2\theta_j + \beta_3\theta_{ij}\right)\right)\right]\right)$$ [1]
$$\alpha_h = (1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00)$$
$$\mathrm{X_{hij}} \sim Categ(p_x, x = (0, 1, 2, 3, 4))$$
$$p_x = P\left(X_{hij} = x|\theta_{ij}\right) = \frac{exp\left(\sum_{k=0}^{x}\left(\theta_{ij} - \delta_{hk}\right)\right)}{\sum_{0}^{k} exp\left(\sum_{k=0}^{x}\left(\theta_{ij} - \delta_{hk}\right)\right)}$$

We did not specify a total group size in this data-generating process, following the framework of the reflective causal model in assuming that indicators (individual observations) are a sample from an infinite population. Although this assumption is ubiquitous in such research and reasonable for most community-based studies, it bears further investigation in future work, particularly when considering smaller source populations such as schools, where a large fraction of observations can be sampled from a group of known and finite size (61,69).

This data-generating process could reflect the situation in which community social capital, itself an effect of historical processes, gave rise to individuals' beliefs regarding their community's capacity. Both the true social capital and perceptions of it shaped risk of anti-social behaviors such as heavy drinking. A subset of community residents were sampled and asked to rate the community capacity on a scale of items.

We predicted consistent performance in terms of bias for the three within-person analytic approaches since the underlying assumptions are met in this case; it is possible that each PV approach will show improved coverage of the 95% confidence interval due to incorporating measurement error in the variance calculation. Based on existing simulation studies of reflective causal models, the SEM approaches should reduce bias in estimation of the contextual effect

[1] An alternative parameterization would be $Y_{ij} \sim Binom\left(1/\left(1 + exp\left(-2 + \beta_2\theta_j + \beta_3\left(\theta_{ij} - \theta_j\right)\right)\right)\right)$, in which case β_2 represents the total between group effect, β_3 the within-group effect, and β_2-β_3 the contextual effect of interest. Although the difference is not critical for simulation purposes provided the correct estimate of the contextual effect is extracted from results, it is essential in real analyses to carefully consider calculation of the measures used for θ_{ij} in order to interpret the model coefficients correctly, especially in latent SEM models where the quantities used for θ_j are not directly calculated.

relative to MR approaches. The cost of this bias reduction is increased variability. We expect PV-SEM to outperform M-SEM and EAP-SEM due to increased capacity to control measurement error with multiple observations per person.

We permute this process in five ways for simulations 1.1 to 1.5. Each permutation shown below was specific to that simulation, such that all other elements of the data-generating process match simulation 1.0.

Simulation 1.1: Two parameter item model

$$\alpha_h = (-2.00, -1.25, -0.50, 0.25, 1.00, 1.75, 2.50)$$

$$\mathrm{X_{hij}} \sim Categ(p_x, x = (0, 1, 2, 3, 4))$$

$$p_x = P(X_{hij} = x|\theta_{ij}) = \frac{exp\left(\sum_{k=0}^{x} \alpha_h(\theta_{ij} - \delta_{hk})\right)}{\sum_{0}^{k} exp\left(\sum_{k=0}^{x} \alpha_h(\theta_{ij} - \delta_{hk})\right)}$$

Simulation 1.2: Non-normal θ_{ij}, Uniform

$$U_{\theta_{ij}} \sim Unif\left(-\frac{\sqrt{96}}{2}, \frac{\sqrt{96}}{2}\right)$$

$$\theta_{ij} = \beta_1 \theta_j + U_{\theta_{ij}}$$

Simulation 1.3: Asymmetric non-normal θ_{ij}, Chi squared

$$U_{\theta_{ij}} \sim \chi^2(1)$$

$$\theta_{ij} = \beta_1 \theta_j + \left(U_{\theta_{ij}} - 1\right)$$

Simulation 1.4: Non-normal θ_j, Uniform

$$U_{\theta_j} \sim Unif\left(-\sqrt{12}/_2, \sqrt{12}/_2\right)$$

$$\theta_j = U_{\theta_j}$$

Simulation 1.5: Asymmetric non-normal θ_j, Chi squared

$$U_{\theta_j} \sim \chi^2(2)$$

$$\theta_j = std\left(U_{X_j}\right)$$

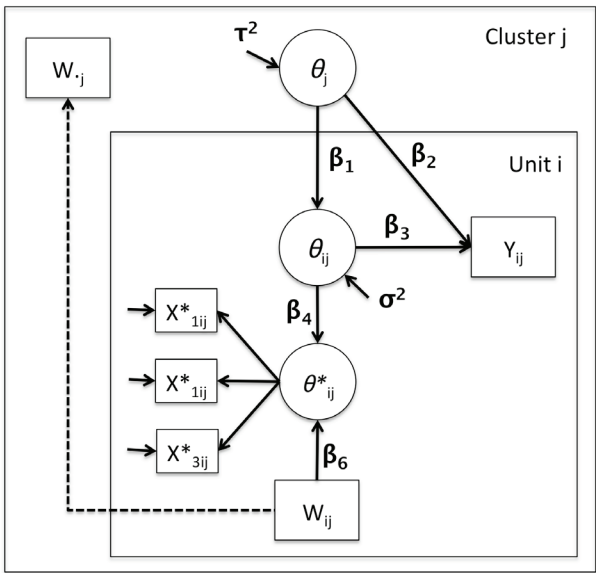
Simulation 1.1 deviates from baseline due to a more complex item model incorporating varying discrimination by item. This item model violates the assumption of i.i.d. items underlying the mean score. EAP and PV scores can appropriately reflect the more complex data-generating model if a two-parameter item response model is fit to the data. We hypothesize that EAP and PV will reduce bias relative to the mean approach in both MR and SEM analyses.

Simulations 1.2 and 1.3 include a non-normal θ_{ij} to test the sensitivity of the IRM-based approaches given the reliance of the EAP on a normal prior distribution. Simulation 1.2 is a uniform distribution and 1.3 a chi-squared distribution offset to ensure $E(\theta_{ij}) = \theta_j$. Simulations 1.4 and 1.5 return to a normally distributed θ_{ij} but vary the distribution of θ_j to a uniform and a chi squared distribution respectively to assess the sensitivity of SEM estimation when the key

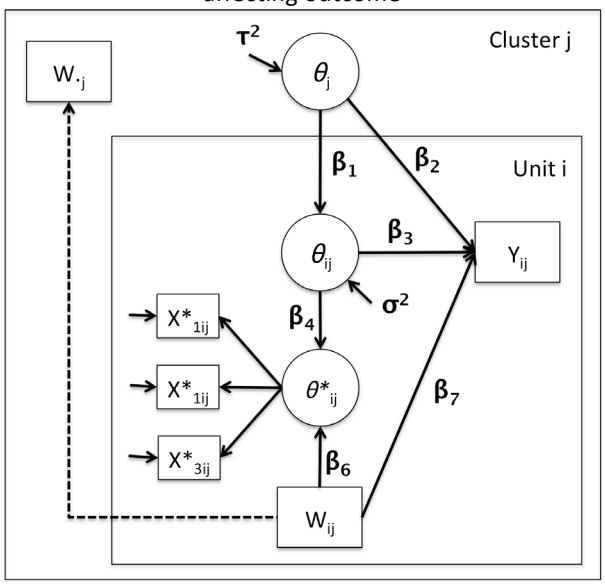
assumption of a normally distributed latent trait is violated.[2] These variants were chosen to represent a plausible non-normal distribution (uniform) and an implausible, asymmetric distribution that will most strongly test analytic robustness. We predict increased bias in the four IRM-based approaches (EAP-MR, PV-MR, EAP-SEM, PV-SEM) in simulations 1.2 and 1.3 relative to baseline. The M-MR and M-SEM analyses should be less sensitive. The magnitude of bias across analytic approaches should decline for simulation 1.3 given the higher ICC (see footnote). We anticipate performance of all SEM-based approaches to worsen relative to baseline in simulations 1.4 and 1.5, particularly the latter; MR approaches should be more robust.

Figure 2.3: Reflective causal models with covariates, null paths removed

A: Simulation 2.0, distorting measurement

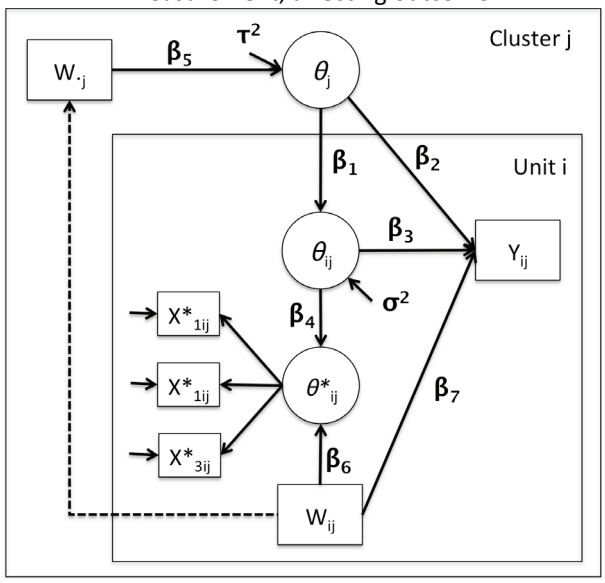


B: Simulation 2.1, distorting measurement and affecting outcome

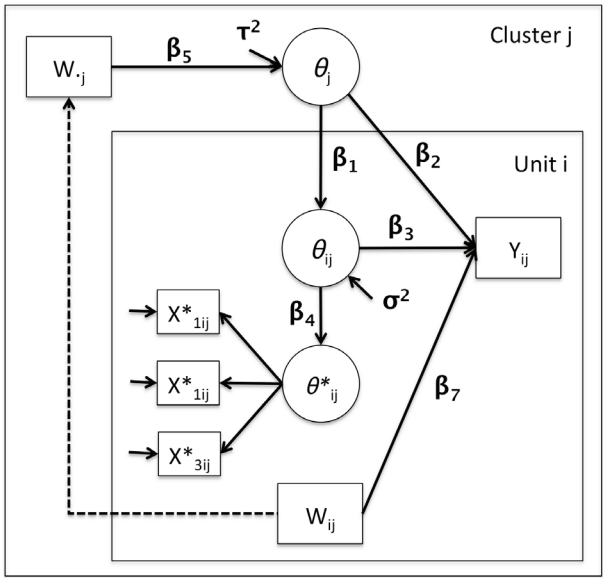


[2] Note that the distributions of $U_{\theta_{ij}}$ and U_{θ_j} were selected to ensure that $var\left(U_{\theta_{ij}}\right) = 8.0$ and $var\left(U_{\theta_j}\right) = 1.0$, maintaining $\sigma^2 = 9.0$, ICC=0.10. The exception is simulation 1.3. ICC in this instance is 0.25 ($\sigma^2 = var\left(U_{\theta_{ij}}\right) + \tau^2 = 2 + 1 = 3$), as a low variance of $U_{\theta_{ij}}$ is necessary to generate a strongly non-normal chi-squared distribution. θ_j in simulation 1.5 is scaled to have mean zero and variance 1.0 to maintain comparability across simulations (see Appendix Figure B.2 for the distribution of the resulting variable).

C: Simulation 2.2, shaping exposure, distorting measurement, affecting outcome



D: Simulation 2.3, shaping exposure and affecting outcome



Simulation set 2: Incorporation of third variable

The purpose of this group of simulations was to extend the reflective causal model tested above to include a third variable acting to distort measurement and/or affect the exposure – outcome relationship. For each simulation and each analytic approach, we compare the bias-correcting strategies of controlling for covariate W_{ij} in the measurement model if there is one and controlling for W_{ij} and/or its group-level aggregate in the outcome model. The overall causal model is shown in Figure 2.2B; a different subset of the relationships between W_{ij} and exposure, measurement, and outcome are set to null in each of the four scenarios tested here, detailed in the data-generating processes below and shown in Figure 2.3.

As in simulation set 1, each simulation was run under setting A ($\beta_1 = 1, \beta_2 = 0.3, \beta_3 = 0$) and B ($\beta_1 = 1, \beta_2 = 0.3, \beta_3 = 0.1$).

Simulation 2.0: Covariate affects measurement

$$U_{W_{ij}} \sim Binom(0.25)$$
$$\mathrm{W}_{ij} = U_{W_{ij}}$$
$$\overline{\mathrm{W}}_{\cdot \mathrm{j}} = \frac{1}{\mathrm{n_i}} \sum_{\mathrm{i}=1}^{\mathrm{n_i}} \mathrm{W_{ij}}$$
$$U_{\theta_j} \sim N(0, \tau^2), \tau^2 = 1$$
$$\theta_j = \beta_5 \overline{W}_{\cdot j} + U_{\theta_j}, \beta_5 = 0$$
$$U_{\theta_{ij}} \sim N(0, \rho^2), \rho^2 = 8$$
$$\theta_{ij} = \beta_1 \theta_j + U_{\theta_{ij}}$$
$$\theta^*_{ij} = \beta_4 \theta_{ij} + \beta_6 W_{ij}, \beta_4 = 1, \beta_6 = -2\rho$$
$$\mathrm{Y_{ij}} \sim Binom\left(1/\left[1 + \exp\left(-\left(-2 + \beta_2 \theta_j + \beta_3 \theta_{ij} + \beta_7 W_{ij}\right)\right)\right]\right), \beta_7 = 0$$

$$\alpha_h = (1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0)$$
$$\mathrm{X_{hij}} \sim Categ(p_x, x = (0, 1, 2, 3, 4))$$
$$p_x = P\left(X_{hij} = x | \theta^*_{ij}\right) = \frac{exp\left(\sum_{k=0}^{x}\left(\theta^*_{ij} - \delta_{hk}\right)\right)}{\sum_0^k exp\left(\sum_{k=0}^{x}\left(\theta^*_{ij} - \delta_{hk}\right)\right)}$$

Simulation 2.1: Covariate affects measurement and outcome

$$\theta^*_{ij} = \beta_4\theta_{ij} + \beta_6 W_{ij}, \beta_4 = 1, \beta_6 = 2\rho$$
$$\mathrm{Y_{ij}} \sim Binom\left(1/\left[1 + \exp\left(-\left(-2 + \beta_2\theta_j + \beta_3\theta_{ij} + \beta_7 W_{ij}\right)\right)\right]\right),$$
$$\beta_7 = 0.45$$

In simulation 2.0, a third variable $\mathrm{W_{ij}}$ alters individual perception of θ_j such that the measurable quantity θ^*_{ij} is a function of both θ_{ij} and $\mathrm{W_{ij}}$, but $\mathrm{W_{ij}}$ is otherwise not influential. This situation is analogous to a survey where mode of administration, i.e. in person, by telephone, via a computer-assisted survey instrument, etc., may systematically influence individual responses on subjective questions such as neighborhood cohesion, but would not be expected to causally affect the outcome. Note that $\mathrm{W_{ij}}$ distorts perception upwards by nearly two standard deviations of θ_{ij} if $\mathrm{W_{ij}}$=1.[3] We select a sizable distortion to ensure that any differences in analytic approaches are evident beyond the bias expected due to measurement error. $\mathrm{W_{ij}}$ is not a confounder at the individual or group level as it is not a common cause of the exposure and outcome. Thus we would not expect controlling for $\mathrm{W_{ij}}$ or $\overline{\mathrm{W}}_{\cdot\mathrm{j}}$ to improve estimation in the regression models or SEM models where $\mathrm{W_{ij}}$ has not been incorporated into the individual measurement. However, where $\mathrm{W_{ij}}$ has been included in the measurement model, controlling for $\mathrm{W_{ij}}$ in the outcome model is expected to improve the accuracy of the overall estimation: including $\mathrm{W_{ij}}$ in the measurement model should improve the accuracy of estimating θ^*_{ij}; controlling for $\mathrm{W_{ij}}$ in the outcome model may help correct the estimated measure back to θ_{ij}.

We increase the role of covariates in the next three simulations; the data generating-processes noted include only those components that differ from simulation 2.0. In simulation 2.1 (Figure 2.3B), $\mathrm{W_{ij}}$ has a non-null effect on $\mathrm{Y_{ij}}$ in addition to distorting measurement. $\mathrm{W_{ij}}$ could represent being in an older age group that perceives community social capital differently and is at greater risk of disease. Although $\mathrm{W_{ij}}$ is not a confounder of the relationships among θ_j, θ_{ij}, and $\mathrm{Y_{ij}}$, in the observed data it does confound the contextual effect due to its association with θ^*_{ij}. $\mathrm{W_{ij}}$ can be viewed as a confounder of the mediator – outcome relationship, where θ^*_{ij} mediates the relationship between θ_j and $\mathrm{Y_{ij}}$; controlling for $\mathrm{W_{ij}}$ should thus be necessary to retrieve an unbiased estimate of the contextual effect (controlled direct effect of θ_j on $\mathrm{Y_{ij}}$). We further hypothesize that incorporating $\mathrm{W_{ij}}$ into the measurement model and adjusting for it in the outcome model will reduce bias due to improved estimation of θ_{ij}.

[3] Although the effect of $\mathrm{W_{ij}}$ on θ_{ij} increases within-individual variability in simulations 3.1 – 3.3, the true reliability changes only fractionally from 0.815 to 0.814.

Simulation 2.2: Covariates affect measurement, outcome and exposure

$$U_j \sim N(-1.85, 1.0)$$
$$U_{ij} \sim Unif(-3.0, 3.0)$$
$$U_{W_{ij}} = U_j + U_{ij}$$
$$\mathrm{W}_{ij} = Binom\left(1/\left[1 + \exp\left(-U_{W_{ij}}\right)\right]\right)$$
$$\overline{\mathrm{W}}_{\cdot \mathrm{j}} = \frac{1}{\mathrm{n_i}} \sum_{\mathrm{i}=1}^{\mathrm{n_i}} \mathrm{W_{ij}}$$
$$\theta_j = std\left(\beta_5 \overline{W}_{\cdot j} + U_{\theta_j}\right), \beta_5 = -4.0$$
$$\theta^*_{ij} = \beta_4 \theta_{ij} + \beta_6 W_{ij}, \beta_4 = 1, \beta_6 = -2\rho$$
$$\mathrm{Y_{ij}} \sim Binom\left(1/\left[1 + \exp\left(-\left(-2 + \beta_2 \theta_j + \beta_3 \theta_{ij} + \beta_7 W_{ij}\right)\right)\right]\right),$$
$$\beta_7 = 0.45$$

Simulation 2.3: Covariates affect outcome and exposure

$$U_{W_{ij}} = U_j + U_{ij}$$
$$\mathrm{W}_{ij} = Binom\left(1/\left[1 + \exp\left(-U_{W_{ij}}\right)\right]\right)$$
$$\overline{\mathrm{W}}_{\cdot \mathrm{j}} = \frac{1}{\mathrm{n_i}} \sum_{\mathrm{i}=1}^{\mathrm{n_i}} \mathrm{W_{ij}}$$
$$\theta_j = std\left(\beta_5 \overline{W}_{\cdot j} + U_{\theta_j}\right), \beta_5 = -4.0$$
$$\theta^*_{ij} = \beta_4 \theta_{ij} + \beta_6 W_{ij}, \beta_4 = 1, \beta_6 = 0$$
$$\mathrm{Y_{ij}} \sim Binom\left(1/\left[1 + \exp\left(-\left(-2 + \beta_2 \theta_j + \beta_3 \theta_{ij} + \beta_7 W_{ij}\right)\right)\right]\right),$$
$$\beta_7 = 0.45$$

Simulation 2.2 (Figure 2.3C) alters the causal model underlying simulation 2.0 such that $\mathrm{W_{ij}}$ affects exposure θ_j itself through aggregate $\overline{\mathrm{W}}_{\cdot \mathrm{j}}$ as well as the measurement of θ_{ij} and the outcome $\mathrm{Y_{ij}}$. An example of this scenario would be living in poverty, where the proportion of residents in poverty undermines social cohesion and, for a given individual, increases risk of disease. Note that across the full population, $\mathrm{P(W_{ij})}$[4]=0.25 in keeping with the prior two simulations, but in this case the probability of $\mathrm{W_{ij}}$ varies between groups. $\mathrm{W_{ij}}$ is a confounder of the exposure – outcome relationship and affects the measured exposure; $\overline{W}_{\cdot j}$ is an ecologic confounder as it varies between groups and is associated with $\overline{\mathrm{Y}}_{\cdot \mathrm{j}}$, which also varies by group. Because in this case the effect of $\overline{W}_{\cdot j}$ on $\mathrm{Y_{ij}}$ is mediated entirely by $\mathrm{W_{ij}}$, control for $\mathrm{W_{ij}}$ may be necessary and sufficient to retrieve an unbiased estimated of the effect of θ_j; we further

[4] The random variable $U_{W_{ij}}$ is the sum of a normal distribution $U_j \sim N(\mu, \sigma^2)$ and a uniform distribution $U_{ij} \sim Unif(a, b)$ as specified; its density is provided by the convolution $f(u) = \left.\Psi_0\left((b-u)/\sigma\right) - \Psi_0\left((a-u)/\sigma\right)\right/ b - a$, where Ψ_0 is the distribution function of the normal distribution (83). We derived the parameters to obtain P(0.25) empirically rather than evaluating the expectation of $1/\left[1 + \exp\left(-U_{W_{ij}}\right)\right]$; we provide the empirical distribution of this quantity and of $\mathrm{W_j}$ in Appendix Figure B.3.

hypothesize that incorporating W_{ij} into the measurement model will improve estimation of θ_{ij} and hence reduce bias.

Simulation 2.3 (Figure 2.3D) follows 2.2 except that W_{ij} no longer distorts measurement; it is designed to provide a check on appropriate control of confounding in a multilevel model. Controlling for W_{ij} in the outcome model alone should minimize bias in this case.

Simulation set 3: Formative (individual to group) causal model

For the final set of simulations, we alter the underlying causal model to a formative (individual to group) structure. This model, shown in Figure 2.2C, reflects the logic of the MR analytic approach: just as in MR a composite value is generated from equal contributions from each sampled observation, in the formative model a fixed population of individual values on the underlying social process combine to create the group value, which in turn affects disease status. Past research suggests that SEM approaches overcorrect in this type of setting, as individuals are not considered to be error-prone observations sampled from an infinite population, but instead are entities contributing equally to the composite group exposure (61). The degree of overcorrection and magnitude of bias in SEM analyses relative to the attenuation in mean-based approaches depended on the sampling fraction as well as the ICC and true group size. A small (<20%) proportion sampled from the fixed population introduced sufficient error to make SEM approaches reasonable options, while larger sampling fractions led to less bias in mean-based analyses. We test a sampling fraction of 0.50 in simulation 3.0, the upper bound of what one might expect in studies of small groups such as rural villages or schools, and of 0.05 in simulation 3.1, a reasonable to high estimate of that seen in social epidemiologic studies of neighborhoods and census tracts. To achieve these fractions, we generate data for $n_{i.full}$=80 individuals per group and $n_{i.full}$=800 individuals per group following the data-generating process below; we then sample n_i=40 observed individuals.

Simulation 3.0 and 3.1: Formative model

$$U1_j \sim N(0, 0.5)$$

$$U2_j \sim Unif\left(-\frac{\sqrt{6}}{2}, \frac{\sqrt{6}}{2}\right)$$

$$U_{\theta_{ij}} \sim N(0,\ \rho^2), \rho^2 = 8$$

$$\theta_{ij} = U1_j + U2_j + U_{\theta_{ij}}$$

$$\bar{\theta}_{.j} = \frac{1}{n_i}\left(\sum_{n=1}^{n_i} \theta_{ij}\right)$$

$$U_Y \sim Unif(-0.25, 0.25)$$

$$Y_{ij} \sim Binom\left(1/\left[1 + \exp\left(-\left(-2 + \beta_2\theta_j + \beta_3\theta_{ij} + U_Y\right)\right)\right]\right)$$

$$\alpha_h = (1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00)$$

$$X_{hij} \sim Categ(p_x, x = (0, 1, 2, 3, 4))$$

$$p_x = P\left(X_{hij} = x|\theta_{ij}\right) = \frac{exp\left(\sum_{k=0}^{x}\left(\theta_{ij} - \delta_{hk}\right)\right)}{\sum_0^k exp\left(\sum_{k=0}^{x}\left(\theta_{ij} - \delta_{hk}\right)\right)}$$

For these simulations, the variance of θ_j is 9.0 and variance of $\bar{\theta}_{.j}$ is approximately 1.0, keeping the ICC at 0.10 as in previous simulations. We predict the mean approach to show negative bias and the SEM approach positive bias in simulation 3.0; if the results from prior studies hold, with a low ICC and high sampling fraction, the mean-based models may outperform SEM in terms of bias and particularly MSE. With the smaller sampling fraction in simulation 3.1, we hypothesize reduced bias in the SEM models relative to mean-based.

Simulation analysis

Following data generation for each simulation, we fit a partial credit model with constant item discrimination to the observed items, with the exception of simulation 1.1, where we fit a two-parameter generalized partial credit model. We generate $\hat{\theta}_{ij}^{EAP}$ using a default normally distributed prior; we generate $\hat{\theta}_{ij}^{PV_r}$ for r={1,…, 5} around each EAP estimate. With the exception of simulation 1.1, we employ a non-parametric estimation approach for $\hat{\theta}_{ij}^{PV_r}$ that utilizes the empirical distribution of observations in lieu of a normal distribution.[5] This approach should provide greater robustness in the tests of a non-normal θ_{ij}. Data were simulated in R 3.1.3 (R Foundation for Statistical Computing); the Multidimensional Item Response package (mirt) was used to generate item responses from individual perception (84). Item fitting and person assessment were conducted with marginal maximum likelihood estimation using the Test Analysis Module (TAM) package version 1.5-2.0 (85) in R.

Individual item responses were aggregated into mean score $\overline{X}_{ij}$; $\overline{X}_{ij}$, $\hat{\theta}_{ij}^{EAP}$, and each set of $\hat{\theta}_{ij}^{PV_r}$ were averaged within group to create proxies for exposure $\overline{X}_{.j}$, $\hat{\bar{\theta}}_{.j}^{EAP}$, and $\hat{\bar{\theta}}_{.j}^{PV_r}$. For MR analyses, individual outcome Y_{ij} was regressed on each exposure estimate in turn (including the 5 sets of PV estimates) in generalized estimating equations (GEE) with a logit link and independent correlation structure, clustered by group. Simulations under setting B included the corresponding individual-level metric in the regression model. As a benchmark for each comparison, we also fit a GEE model on true θ_j (and true θ_{ij} for setting B) clustered by group. We fit each SEM analysis using $\overline{X}_{ij}$, $\hat{\theta}_{ij}^{EAP}$, and all of $\hat{\theta}_{ij}^{PV_{1-5}}$ respectively as individual-level ratings of an unmeasured group variable (see Appendix Figures B.4 and B.5 for schematics of the SEMs). Because $\overline{X}_{.j}$ or $\widehat{\theta_j}$ are not directly included in the SEM models, the coefficients returned are equal to the total between-group effect $(\beta_2 + \beta_3)$ and the pure within-group effect (β_3). We therefore calculate the contextual effect as the difference between these coefficients. (In setting A, the within group effect is null, rendering this calculation unnecessary.) Estimated variance of θ_j - $\widehat{\tau^2}$ - is constrained to equal 1, matching the generating value τ^2 so that the estimated contextual effect $\widehat{\beta_2}$ can be directly compared to true β_2. Due to the multilevel nature of the model, the integration method is mode and curvature adaptive Gauss-Hermite quadrature with 3 integration points, although in cases where this model will not converge after 20 iterations, the model is attempted with non-adaptive Gauss-Hermite quadrature instead.[6] The

[5] Fitting plausible values without the assumption of normality for this 2PL-based IRM should be possible in theory, but resulted in a technical error related to the iterative simulations in this case. Further attempts to resolve this issue for comparability will be undertaken in the future.

[6] A testing set of 50 simulations was run for all SEMs; for each of the analytic approaches, if fewer than 10% converged with mode and curvature Gauss-Hermite quadrature with 3 integration points within 20 iterations, non-adaptive Gauss-Hermite quadrature was employed for all simulations due to the time-

latter method is less reliable but can be more likely to converge in some cases. Initial values equal to the true simulation settings were provided to expedite model convergence. The model with plausible values constrains the relationship of each plausible value to the latent trait to be equal; the error variance estimates $\widehat{\varepsilon^2}$ for all plausible values are also constrained to be equal. In addition to the estimated coefficient and standard error, $\widehat{\sigma^2}$ and $\widehat{\varepsilon^2}$ are retained from each simulation to enable calculation of the estimated ICC and reliability, with n_{meas}=1 for models using $\overline{X}_{ij}$ and $\hat{\theta}_{ij}^{EAP}$ and 5 for the model using $\hat{\bar{\theta}}_{\cdot j}^{PV_{1-5}}$.

For simulation set 2, we explore approaches to controlling for covariates in the measurement model and at the individual and group levels of the outcome model to reduce systematic measurement error and confounding. We assess adjusting for W_{ij} in a latent regression within the IRM for EAP and PV approaches. We combine this adjustment with three adjustment strategies in the outcome model: no adjustment, adjustment for W_{ij}, and adjustment for aggregate $\overline{W}_{\cdot j}$ (see sample SEM schematics in Appendix Figures B.6 and B.7) Due to the large number of analytic permutations and the exploratory nature of this analysis, we confine our results to mean and plausible value-based approaches. In order to provide a reference for correct adjustment for the underlying causal relationships without the measurement and aggregation component, we regress Y_{ij} on the true exposure values in GEE models under each adjustment strategy. MR and SEM analyses were performed in Stata 13.0 and Stata 14.0 (StataCorp, Texas).

Each simulation was conducted 500 times, retaining the estimated coefficient of contextual effect $\widehat{\beta_2}$ and its standard error. We also retained the between-person variance estimate $\widehat{\sigma^2}$ and within-person variance estimate $\widehat{\varepsilon^2}$ from each SEM analysis. For plausible value analyses, $\widehat{\beta_2}$ was calculated by pooling the 5 estimates; $var(\widehat{\beta_2})$ incorporated variance between each of the 5 $\widehat{\beta_2}$ as well as the estimated variance within each model. Confidence intervals around the pooled plausible value estimate were calculated following a t statistic with γ degrees of freedom, where, with *r* as the number of plausible values, var_W the average of the estimated variance on the exposure coefficient within each regression, and var_B as empirical variance across the m estimates:

$$\gamma = (r-1)\left\{1 + \frac{var_W}{\left(1+\frac{1}{r}\right)var_B}\right\}^2 \quad (86)$$

The performance metrics reported for each analytic approach are defined in Table 2.3 below. We additionally report the proportion of trials in which the estimator could not be calculated and the coverage probability of the 95% confidence interval among non-missing results.

Table 2.3: Performance metrics for analytic approaches in simulation study

Metric	Formula	Applicable approaches
Bias	$\widehat{\beta_2} - \beta_2$	All
Relative bias	$\left(\widehat{\beta_2} - \beta_2\right) \Big/ \beta_2$	All
Empirical variance	$\frac{1}{n_{sim}} \sum \left(\widehat{\beta_2} - \overline{\beta_2}\right)^2$	All

intensive nature of mode and curvature estimation. Even with these settings favoring speed over reliability to a greater extent than an applied analysis, total computer processing time for all simulations combined exceeded 3,000 hours.

Mean squared error (MSE)	$\frac{1}{n_{sim}}\sum\left(\widehat{\beta_2}-\beta_2\right)^2$	All
ICC	$\tau^2 \Big/ (\tau^2+\widehat{\sigma^2})$	M-SEM, EAP-SEM, PV-SEM
Reliability	$\kappa = \frac{\tau^2}{\tau^2 + \widehat{\sigma^2}/n_i + \widehat{\varepsilon^2}/n_i * n_{meas}}$	M-SEM, EAP-SEM, PV-SEM

Motivating example

To demonstrate the application of these approaches to real data, we analyze the association of village organizations and networks, a metric of social capital, with heavy episodic drinking (HED) among young men in rural South Africa. A cross-sectional sample of adults aged 18 – 35 was drawn from 22 villages in Agincourt, South Africa; 600 women and 581 men provided data on personal characteristics and behavior as well as rating the contextual traits of their village (82). The role of organizations and networks was measured using an 8-item scale with 3 response options per question; an example item is 'How important are cultural groups for improving the lives of villages here' with response options 'Not important / no such groups,' 'A little important,' 'Very important'. HED is defined as consuming 6 standard drinks or more in a single sitting at least monthly in the past year. Data are also available on village-level covariates such as percent of the village living in female-headed households (a metric of poverty, as these households are typically poorer (87)) and other individual characteristics such as age and marital status. The causal model underlying this analysis is depicted in Appendix Figure B.8. We hypothesize that both true group organizations and networks as well as the difference between individual perception and group average may affect the outcome due to the possibility that greater knowledge of local resources may result in greater engagement and hence less risky behavior. We therefore mean-center individual perception and regress HED on group average and individual difference from average; we subtract the resulting coefficients to estimate the contextual effect. We do not standardize the group averages in order to keep them on the same scale as the individual differences; as a result the estimates from the regression-based models and the SEM analysis are scaled differently and can be compared based on direction and significance rather than magnitude.

Perceived importance of organizations and networks was calculated for each individual using a mean item score and by fitting a partial credit model to generate EAP estimators and PVs; we employ both a one-parameter model and two-parameter generalized partial credit model, comparing model fit using weighted mean square (infit statistic) for individual items and the log likelihood test for global fit. These individual scores served as the basis for MR and SEM analyses. We tested for measurement distortion through latent regression of the one-parameter and two-parameter versions of the organizations and networks scales on two variables: gender and language of survey administration. Because we limited the analysis to men only (HED is quite rare in women in this context), we based the analysis on the assumption that neither gender nor survey language affects the outcome. All analyses were run with sampling weights to ensure representativeness of the study sample to the target population. Weights accounted for non-response as well as number of eligible respondents per household, as only one individual could be selected in each household. Sampling weights were also included in measurement models for EAP and PV creation.

RESULTS

We present below the results addressing each question of interest in the corresponding simulation setting (robustness, incorporation of covariates, alternative causal model), followed by the applied example. Given the large number of results, in most cases we present full results from setting A (no individual-level effect: $\beta_1 = 1, \beta_2 = 0.3, \beta_3 = 0$) in the Appendix and focus on setting B ($\beta_1 = 1, \beta_2 = 0.3, \beta_3 = 0.1$), with a particular emphasis on the most divergent analytic approaches (M-MR and PV-SEM) to illustrate the greatest observed contrasts.

Simulation set 1: Robustness checks

Table 2.4 provides the results for the analytic approaches applied to the baseline condition (simulation 1.0) in both the simple setting A (no effect of θ_{ij} on Y_{ij}) and the more realistic setting B. In setting A, SEM analyses relying on a single individual-level measurement experienced difficulties in convergence using the mode-and-curvature adaptive quadrature method: fewer than 10% of iterations converged of the testing set of 50. Switching to non-adaptive Gauss-Hermite quadrature considerably improved convergence. Setting B similarly required the less reliable integration method, although approximately 10% of simulation runs failed to converge. All PV-SEM runs converged using mode-and-curvature adaptive quadrature with the exception of simulation 1.1, testing the two-parameter item-generating model. In setting B, 13.4% of runs required non-adaptive Gauss-Hermite quadrature to converge and 11.2% failed to converge under either integration method.

In setting A, MR results follow predictions from prior simulation studies: each estimate is attenuated by 10%, with approximately -4% undercoverage of the 95% CI due to this bias. Among the SEM approaches, M-SEM and EAP-SEM show even greater bias than the MR approaches (>-13.0%) as well as worse CI coverage, though the lower variance in these models results in equivalent MSE across all methods (0.009). The PV-SEM approach considerably reduced bias in estimation to -2.85% and provided accurate coverage of the 95% CI.

The same patterns pertain in setting B, with a glaring exception: M-SEM and EAP-SEM estimated extremely low variance within individuals, resulting in high and highly variable estimates of $\widehat{\beta_3}$, which under this setting is necessary to move from the estimated total effect to the contextual effect $\widehat{\beta_2}$. Alternative analytic specifications would be required to overcome this difficulty; given our goal of contrasting PV-SEM in particular with other approaches, we proceed without M-SEM and EAP-SEM results for setting B. Full results for these methods in setting A are presented in Appendix A.2 and referenced as relevant.

Table 2.4: Baseline performance of analytic approaches applied to reflective causal model

	Bias	Relative bias	Variance	MSE	Coverage probability	Convergence failure
Simulation 1.0: Baseline model, setting A						
True θ_j	-0.003	-0.94%	0.007	0.007	95.4%	NA
M-MR	-0.030	-10.06%	0.008	0.009	91.0%	NA
M-SEM	-0.042	-13.91%	0.007	0.009	89.4%	0.0%
EAP-MR	-0.030	-10.01%	0.008	0.009	90.6%	NA
EAP-SEM	-0.039	-13.00%	0.007	0.009	90.2%	2.4%
PV-MR	-0.032	-10.69%	0.008	0.009	91.6%	NA
PV-SEM	-0.009	-2.85%	0.009	0.009	95.0%	0.0%
Simulation 1.0: Baseline model, setting B						
True θ_j & θ_{ij}	0.003	0.87%	0.008	0.008	93.4%	NA
M-MR	-0.019	-6.48%	0.009	0.009	89.2%	NA

M-SEM	-2.516	-838.72%	5.502	11.820	96.6%	11.6%
EAP-MR	-0.022	-7.39%	0.009	0.009	89.6%	NA
EAP-SEM	-1.746	-581.97%	5.115	8.152	93.0%	9.2%
PV-MR	-0.019	-6.24%	0.009	0.009	90.6%	NA
PV-SEM	0.002	0.67%	0.010	0.010	94.4%	0.0%

Table 2.5 provides results for the robustness checks in setting B (results for setting A are in Appendix Table B.2; they are discussed only when contradictory to setting B results). All PV-SEM models in this set of simulations converged using mode-and-curvature adaptive quadrature.

Simulation 1.1 tests the robustness of each approach to an alternative item model, which we expected would affect M-MR more than EAP or PV-based analyses. Results differ between the two settings. In setting A, M-MR performs poorly, underestimating the true association by -20%. This is over twice the bias seen in EAP-MR and PV-MR, which were based on a two-parameter analytic model matching the underlying model. Contrary to prediction, PV-MR does not provide better coverage of the 95% confidence interval (89.8% for EAP-MR and PV-MR). In setting B, against expectations, M-MR outperformed the IRM approaches in terms of bias (-3.6% versus -8.1% for EAP-MR and -6.08 for PV-SEM). Higher variability in M-MR leads to lower MSE for EAP-MR and PV-MR, as well as better coverage of the 95% CI (90.0% for the plausible values approach). This simulation was also the only one of all those tested that posed considerable convergence problems for the PV-SEM method, with only 75% of runs converging under the preferred integration approach. These results imply that a mismatch between item structure and the analytic model could lead to considerable bias, but this was not the case as uniformly as expected.

The next two simulations test violations of the assumption of normal person distribution underlying EAP generation. In simulation 1.2 results, relative bias of the EAP-MR approach was -6.55%, 1% more than the other MR approaches and actually less than the corresponding bias in the baseline simulation. MSE was nearly equivalent across approaches. The PV-SEM approach minimizes relative bias (<1%) and provides the best coverage of the 95% CI. Simulation 1.3 shows a lower magnitude of bias across methods, as expected given the higher ICC used in this data-generating process. Again the EAP-MR and PV-SEM prove robust to the non-normal distribution of θ_{ij}; bias is highest in the PV-MR analysis (3.12%). Variance is higher in simulation 1.3 than 1.2; MSE is fractionally lower in the EAP-MR and PV-MR analyses (0.011) than M-MR and PV-SEM. Despite the strongly non-normal distribution, neither EAP nor PV-based approaches showed considerable bias. Violating the assumption of a normal distribution of the individual-level trait seemed to affect all approaches similarly, suggesting that sensitivity to non-normal distributions is not a reason to select the mean over methods that explicitly assume normality.

Table 2.5: Robustness results for analytic approaches applied to reflective causal model

	Bias	Relative bias	Variance	MSE	Coverage probability
Simulation 1.1: Two-parameter item model					
True θ_j & θ_{ij}	0.000	-0.06%	0.007	0.007	94.6%
M-MR	-0.011	-3.58%	0.010	0.010	87.4%
EAP-MR	-0.024	-8.14%	0.009	0.009	88.6%
PV-MR	-0.020	-6.80%	0.009	0.009	90.0%
PV-SEM	-0.018	-6.08%	0.013	0.014	88.5%
Simulation 1.2: Non-normal θ_{ij} distribution: uniform					
True θ_j & θ_{ij}	0.000	-0.15%	0.006	0.006	96.0%

M-MR	-0.017	-5.65%	0.008	0.009	91.6%
EAP-MR	-0.020	-6.55%	0.008	0.009	91.8%
PV-MR	-0.017	-5.55%	0.008	0.008	93.2%
PV-SEM	0.003	0.89%	0.009	0.009	96.8%
Simulation 1.3: Non-normal θ_{ij} distribution: chi squared, ICC = 0.25					
True θ_j & θ_{ij}	0.004	1.48%	0.008	0.008	93.0%
M-MR	-0.005	-1.77%	0.012	0.012	93.2%
EAP-MR	-0.004	-1.18%	0.011	0.011	92.2%
PV-MR	0.009	3.12%	0.011	0.011	94.4%
PV-SEM	-0.002	-0.68%	0.012	0.012	94.6%
Simulation 1.4: Non-normal θ_j distribution: uniform					
True θ_j & θ_{ij}	0.006	2.16%	0.008	0.008	94.6%
M-MR	-0.017	-5.50%	0.007	0.007	94.6%
EAP-MR	-0.019	-6.34%	0.007	0.007	94.4%
PV-MR	-0.015	-5.06%	0.007	0.007	94.4%
PV-SEM	0.005	1.50%	0.008	0.008	96.2%
Simulation 1.5: Non-normal θ_j distribution: chi squared					
True θ_j & θ_{ij}	-0.003	-0.86%	0.006	0.006	95.4%
M-MR	-0.004	-1.43%	0.013	0.013	82.0%
EAP-MR	-0.010	-3.26%	0.012	0.012	83.2%
PV-MR	-0.006	-1.90%	0.012	0.012	82.2%
PV-SEM	0.015	5.12%	0.013	0.013	90.4%

The final two robustness checks concern the distribution of θ_j, which is particularly important for SEM-based methods. None of the MR approaches are strongly affected by the non-normality in either 1.4 or 1.5; the magnitude of bias actually decreases relative to baseline, particularly for 1.5. Bias increases slightly for PV-SEM in simulation 1.4 relative to baseline, although it remains the least biased method, even including the true model results. This approach shows greater sensitivity to the strongly non-normal chi-squared distribution in simulation 1.5: bias increases to over 5%, making PV-SEM the most biased as well as the most variable of the methods in this setting. (This sensitivity was less pronounced in setting A, shown in Appendix Table B.2; otherwise results followed the pattern reported here.) Variance is higher across the methods in simulation 1.5 and coverage of the 95% CI quite low, from 82.0% for M-MR up to 90.4% for PV-SEM; within each simulation the MSEs are similar. The PV-SEM analysis provided accurate estimation of the ICC and reliability in simulation 1.4 at 0.101 (SD 0.020) and 0.811 (SD 0.036) respectively and slight underestimates in simulation 1.5: 0.094 (SD 0.036) and 0.782 (SD 0.076).

Simulation set 1 conclusions

The first set of simulations provided several insights into the performance of MR versus SEM approaches applied to a reflective causal model in conditions likely for a social epidemiologic study. At baseline, all MR-based approaches evidenced considerable attenuation, as predicted based on existing research. Of the SEM approaches, PV-SEM reduced the bias due to measurement error to near 0, although results were more variable under this method. The robustness of M-MR to a two-parameter item model was inconclusive, with one of two settings showing a high degree of bias; further testing is warranted. IRM-based results with the appropriate analytic model were robust to this change. Similarly, alterations to the distribution of θ_{ij} did not strongly perturb any analytic method, even those assuming a normal person

distribution. PV-SEM was robust to a uniform distribution of θ_j but did show moderate sensitivity to a strongly non-normal chi-squared distribution of θ_j in the more realistic setting. The results as a whole suggest that most methods are relatively robust to violations of their underlying assumptions, and that, with the exception of a fairly unrealistic distribution of θ_j, PV-SEM offers the best bias correction of the methods applied. The MSE provides a less clear demarcation between methods, as little difference emerged within any simulation when considering both bias and variance.

Simulation set 2: Role of third variable

The second set of simulations explores causal structures that include a third variable affecting measurement, exposure, and / or outcome. As in the prior set of simulations, the M-SEM approach experienced considerable convergence challenges, requiring use of the less reliable Gauss Hermite integration, and revealed considerable bias and variability in estimation. We present M-SEM results for setting A alone (Appendix tables B.3 – B.6) and restrict our focus here to the better performing methods applied to setting B.

Simulation 2.0: Covariate affects measurement

Simulation 2.0 tests adjustment approaches in the measurement and outcome models in the setting where W_{ij} affects perception θ_{ij} but is unassociated with Y_{ij}. The performance of each analytic and adjustment approach in this setting are summarized in Table 2.6. Using the true exposure θ_j, as expected there is no reduction in absolute bias by adjusting for W_{ij} or for W_{ij} and $\overline{W}_{\cdot j}$.

MR results contradict our first hypothesis that unadjusted outcome models would outperform adjusted models in that bias is reduced by controlling for both W_{ij} and $\overline{W}_{\cdot j}$ in the outcome model compared to neither or just W_{ij}. Even with this adjustment, residual bias remains non-negligible in M-MR results, over -4%. Unadjusted PV-SEM models also show considerable residual bias (nearly 8.6%), while adjusting for W_{ij} in the outcome model or both measurement and outcome models reduces bias to under -3%.

The second hypothesis, that controlling for W_{ij} in the outcome model would reduce bias if W_{ij} had been included in the measurement model, is supported so long as the outcome model in MR approaches also controls for $\overline{W}_{\cdot j}$. Adjusting for W_{ij} in the measurement model alone increases bias, at times considerably, in PV-SEM.

Table 2.6: Performance of adjustment strategies when W_{ij} distorts θ_{ij}, simulation 2.0

	Bias	Relative bias	Variance	MSE	Coverage probability
True θ_j & θ_{ij}					
Unadj.	-0.007	-2.31%	0.008	0.008	94.6%
Adj. for W_{ij}	-0.007	-2.23%	0.008	0.008	94.6%
Adj. for W_{ij} & $W_{\cdot j}$	-0.007	-2.31%	0.008	0.008	94.8%
M-MR					
NA / Unadj.	-0.017	-5.82%	0.009	0.009	87.2%
NA / Adj. W_{ij}	-0.040	-13.36%	0.009	0.010	86.0%
NA / Adj. W_{ij} & $W_{\cdot j}$	-0.012	-4.10%	0.010	0.010	89.6%
PV-MR					
Unadj. / Unadj.	-0.017	-5.70%	0.009	0.009	89.8%
Unadj. / Adj. W_{ij}	-0.038	-12.80%	0.008	0.010	88.2%
Unadj. / Adj. W_{ij} & $W_{\cdot j}$	-0.008	-2.64%	0.010	0.010	91.6%

Adj. / Unadj.	-0.020	-6.71%	0.009	0.009	87.6%
Adj. / Adj. W_{ij}	-0.046	-15.32%	0.008	0.010	86.6%
Adj. / Adj. W_{ij} & $W_{.j}$	-0.005	-1.54%	0.010	0.010	92.4%
PV-SEM					
Unadj. / Unadj.	0.026	8.59%	0.010	0.011	94.6%
Unadj. / Adj. W_{ij}	-0.005	-1.60%	0.010	0.010	94.6%
Unadj. / Adj. W_{ij} & $W_{.j}$	0.019	6.36%	0.011	0.011	95.4%
Adj. / Unadj.	0.206	68.74%	0.008	0.050	28.4%
Adj. / Adj. W_{ij}	-0.008	-2.69%	0.011	0.011	92.6%
Adj. / Adj. W_{ij} & $W_{.j}$	0.024	8.04%	0.011	0.012	94.4%

The MSE is similar across all methods (0.009 to 0.011) except when using M-SEM or adjusting in the measurement model only in PV-SEM, where it is considerably higher (0.51 in setting A for example). The approaches that minimize bias and MSE while optimizing CI coverage to within 1% of 95% are PV-SEM models that are adjusted for W_{ij} in outcome regression alone or adjusted for W_{ij} in outcome regression and measurement. Only this last approach adheres to expectations that bias will be minimized by not controlling for anything or by controlling for W_{ij} in both measurement and outcome models. The ICC and reliability estimates from the plausible values / SEM models are attenuated (ICC: 0.060, SD 0.018 and reliability: 0.701, SD 0.071). This attenuation is likely due to W_{ij} increasing the variance of θ^*_{ij} relative to θ_{ij}; no adjustment strategy corrected for this effect. These results suggest that systematic measurement error has multiple effects on analytic accuracy, and that adjustment in the measurement model alone results in severely biased estimation. Furthermore, existing graphical causal models may not provide sufficient guidance in selecting an adjustment approach.

Simulation 2.1: Covariate affects measurement and outcome

As shown in Table 2.7, models using the true θ_j and θ_{ij} show minimal bias (<-2.0%) regardless of adjustment approach, as expected given that W_{ij} is not a confounder of the relationships among θ_j, θ_{ij}, and Y_{ij}.

The observed data results follow the pattern of simulation 2.0: adjusting for W_{ij} in the outcome model does not reduce bias unless $\bar{W}_{.j}$ is also controlled for, with the exception of PV-SEM models. Adjusting for W_{ij} in the measurement model increases bias in most cases, particularly in SEM analysis with an unadjusted outcome model. However, the smallest magnitude of bias across all methods is in the PV-SEM approach adjusting for W_{ij} in both measurement and outcome (relative bias 0.5%), and the MR approaches adjusting for both W_{ij} and $\bar{W}_{.j}$ (relative bias -0.9% for M-MR, 0.8% for PV-MR).

Table 2.7: Performance of adjustment strategies when W_{ij} distorts θ_{ij} and affects Y_{ij}, simulation 2.1

	Bias	Relative bias	Variance	MSE	Coverage probability
True θ_j & θ_{ij}					
Unadj.	-0.005	-1.69%	0.007	0.007	93.4%
Adj. for W_{ij}	-0.006	-1.85%	0.007	0.007	93.6%
Adj. for W_{ij} & $W_{.j}$	-0.004	-1.28%	0.007	0.007	93.4%
M-MR					
NA / Unadj.	-0.030	-10.13%	0.009	0.009	88.4%
NA / Adj. W_{ij}	-0.028	-9.25%	0.009	0.010	89.4%
NA / Adj. W_{ij} & $W_{.j}$	-0.003	-0.93%	0.010	0.010	91.6%

PV-MR					
Unadj. / Unadj.	-0.029	-9.82%	0.008	0.009	90.6%
Unadj. / Adj. W_{ij}	-0.029	-9.51%	0.008	0.009	91.2%
Unadj. / Adj. W_{ij} & $W_{.j}$	0.002	0.80%	0.009	0.009	93.0%
Adj. / Unadj.	-0.031	-10.40%	0.008	0.009	89.0%
Adj. / Adj. W_{ij}	-0.036	-11.98%	0.008	0.010	88.8%
Adj. / Adj. W_{ij} & $W_{.j}$	0.008	2.65%	0.010	0.010	93.0%
PV-SEM					
Unadj. / Unadj.	0.007	2.41%	0.010	0.010	95.0%
Unadj. / Adj. W_{ij}	0.005	1.51%	0.010	0.010	95.2%
Unadj. / Adj. W_{ij} & $W_{.j}$	0.027	9.13%	0.011	0.012	94.6%
Adj. / Unadj.	0.183	60.96%	0.008	0.041	36.2%
Adj. / Adj. W_{ij}	0.001	0.45%	0.010	0.010	95.0%
Adj. / Adj. W_{ij} & $W_{.j}$	0.033	11.02%	0.011	0.012	94.8%

The MSE again clusters around 0.010 for most analytic options, with unadjusted MR approaches tending to minimize the MSE due to the larger variance of the SEM analyses. Coverage of the 95% CI is roughly 90% across all approaches except for the PV-SEM, where it is at least 94% except in the analyses adjusted in the measurement model alone. The estimated ICC from the PV-SEM analysis is inaccurate, at 0.059 (SD 0.019); reliability is underestimated, at 0.698 (SD 0.076). As a whole, these results suggest that considerable bias remains in the MR approaches unless both W_{ij} and $\overline{W}_{.j}$ are adjusted for in the outcome model. The PV-SEM approach results follow predictions more closely, with the analysis adjusted for W_{ij} in both measurement and outcome models minimizing bias and optimizing CI coverage, despite the similarity of the MSE between this model and the least biased of the MR results.

Simulation 2.2: Covariates affect exposure, measurement, and outcome

In simulation 2.2, W_{ij} acts as a true confounder, affecting θ_j through $\overline{W}_{.j}$ and Y_{ij} directly. As expected, controlling for W_{ij} reduces the bias due to the confounding when regressing Y on true θ_j and θ_{ij} (Table 2.8). Adjusting for $\overline{W}_{.j}$ in addition minimally reduces bias, but increases the variance such that the MSE is worse when controlling for both forms of W than neither (0.009 versus 0.007).

In the observed data analyses, traditional rules of confounder control pertain. Within each analytic strategy, controlling just for W_{ij} in the regression model minimizes bias; control for W_{ij} in the measurement model has minimal effect. MR approaches continue to be attenuated by nearly 10% even with the appropriate adjustment, however. As in most prior simulations, the PV-SEM model with adjustment as predicted by the causal graph (for W_{ij} in the regression model alone in this case) shows the least bias, although at nearly -6% the residual bias is non-negligible. Adjusting for W_{ij} in both models provides nearly identical results.

Table 2.8: Performance of analytic and adjustment strategies when W_{ij} distorts θ_{ij} and affects θ_j and Y_{ij}, simulation 2.2

	Bias	Relative bias	Variance	MSE	Coverage probability
True θ_j & θ_{ij}					
Unadj.	-0.036	-12.09%	0.007	0.009	90.4%
Adj. for W_{ij}	-0.004	-1.28%	0.007	0.007	93.4%
Adj. for W_{ij} & $W_{.j}$	-0.004	-1.22%	0.009	0.009	93.8%

M-MR					
NA / Unadj.	-0.042	-14.01%	0.009	0.010	86.4%
NA / Adj. W_{ij}	-0.030	-10.11%	0.009	0.009	89.8%
NA / Adj. W_{ij} & $W_{.j}$	0.056	18.82%	0.017	0.020	92.2%
PV-MR					
Unadj. / Unadj.	-0.041	-13.70%	0.009	0.010	87.6%
Unadj. / Adj. W_{ij}	-0.027	-9.07%	0.009	0.009	89.6%
Unadj. / Adj. W_{ij} & $W_{.j}$	0.068	22.50%	0.017	0.021	92.4%
Adj. / Unadj.	-0.041	-13.72%	0.009	0.010	87.6%
Adj. / Adj. W_{ij}	-0.027	-9.08%	0.009	0.009	89.4%
Adj. / Adj. W_{ij} & $W_{.j}$	0.068	22.73%	0.017	0.022	91.6%
PV-SEM					
Unadj. / Unadj.	-0.019	-6.17%	0.010	0.010	92.2%
Unadj. / Adj. W_{ij}	-0.016	-5.34%	0.010	0.010	92.6%
Unadj. / Adj. W_{ij} & $W_{.j}$	0.073	24.41%	0.018	0.023	95.0%
Adj. / Unadj.	0.415	138.48%	0.014	0.187	1.0%
Adj. / Adj. W_{ij}	-0.016	-5.39%	0.010	0.010	93.0%
Adj. / Adj. W_{ij} & $W_{.j}$	0.073	24.45%	0.018	0.024	95.0%

The MSE is minimized when MR approaches are adjusted for W_{ij} in the outcome model or both models, despite the high bias in these results. MSE is fractionally higher (0.010 versus 0.009) in the PV-SEM approaches that minimize bias. CI coverage is near 90% for MR models and approximately 92% for plausible values / SEM. The average estimated ICC and reliability are overestimated, at 0.122 (SD 0.030) and 0.838 (SD 0.037). Given the continued effect of Wij in increasing the variance of θ_{ij}^{*}, it is interesting to note improved estimation of the ICC and reliability relative to the prior two simulations, although still inaccurate. As a whole, the results confirm that adjustment for W_{ij} in the outcome model is necessary to reduce bias. Adjustment for W_{ij} in the measurement model shows minimal impact. Although the plausible values / SEM approach is the best option for minimizing bias, even the least biased models remain notably attenuated.

Simulation 2.3: Covariates affect exposure and outcome

The final simulation of this set tests how the analytic strategies applied when W_{ij} distorts measurement of θ_{ij} perform when W_{ij} does not affect measurement but does act as a confounder of θ_i and Y_{ij}. One run of the completely unadjusted PV-SEM approach failed to converge. As shown in Table 2.9, results of regressing Y on true θ_i and θ_{ij} are quite similar to simulation 2.2, reflecting the shared causal structure underlying W_{ij}, θ_i and θ_{ij} in these two simulations. Relative bias is minimal and nearly equivalent when controlling for W_{ij} or for both W_{ij} and $\overline{W}_{.j}$; MSE is minimized in the former analysis (0.007) due to higher variance with additional adjustment.

Within each method applied to the observed data, failure to adjust for W_{ij} results in considerable bias, over -17.0% in the MR approaches. As expected, adjusting for W_{ij} in the outcome model reduces bias; additional adjustment for $\overline{W}_{.j}$ is counterproductive. Across methods, the PV-SEM approach minimizes bias when adjusted for W_{ij} (relative bias < -2.0%), with a further fractional reduction in bias if the measurement model also includes W_{ij}.

Table 2.9: Performance of analytic and adjustment strategies when W_{ij} affects θ_j and Y_{ij}, simulation 2.3

	Bias	Relative bias	Variance	MSE	Coverage probability
True θ_j & θ_{ij}					
Unadj.	-0.035	-11.74%	0.007	0.008	93.0%
Adj. for W_{ij}	-0.004	-1.26%	0.007	0.007	96.0%
Adj. for W_{ij} & $W_{.j}$	-0.002	-0.63%	0.009	0.009	95.8%
M-MR					
NA / Unadj.	-0.054	-17.93%	0.008	0.010	85.4%
NA / Adj. W_{ij}	-0.027	-8.89%	0.008	0.009	92.0%
NA / Adj. W_{ij} & $W_{.j}$	-0.038	-12.56%	0.010	0.011	90.0%
PV-MR					
Unadj. / Unadj.	-0.052	-17.40%	0.007	0.010	87.6%
Unadj. / Adj. W_{ij}	-0.025	-8.42%	0.008	0.008	92.6%
Unadj. / Adj. W_{ij} & $W_{.j}$	-0.037	-12.35%	0.009	0.011	90.6%
Adj. / Unadj.	-0.052	-17.25%	0.007	0.010	87.8%
Adj. / Adj. W_{ij}	-0.025	-8.26%	0.008	0.008	92.4%
Adj. / Adj. W_{ij} & $W_{.j}$	-0.037	-12.17%	0.009	0.011	90.8%
PV-SEM					
Unadj. / Unadj.	-0.034	-11.35%	0.008	0.010	94.0%
Unadj. / Adj. W_{ij}	-0.005	-1.73%	0.009	0.009	95.8%
Unadj. / Adj. W_{ij} & $W_{.j}$	-0.021	-6.88%	0.011	0.011	95.4%
Adj. / Unadj.	0.288	95.87%	0.008	0.091	4.4%
Adj. / Adj. W_{ij}	-0.005	-1.59%	0.009	0.009	96.4%
Adj. / Adj. W_{ij} & $W_{.j}$	-0.020	-6.72%	0.010	0.011	95.4%

Higher variance in the SEM analyses leads to the MSE of 0.009 in the properly adjusted PV-SEM analyses fractionally exceeding that of the least biased MR approach, PV-MR adjusted for W_{ij} in at least the outcome model (MSE = 0.008). Coverage of the 95% CI reaches 92.6% in the best performing MR approach; SEM models show accurate or slightly conservative coverage except for the highly biased analysis when W_{ij} is included in the measurement model alone. The PV-SEM analysis accurately estimates the ICC and reliability, at 0.099 (SD 0.026) and 0.805 (SD 0.046) respectively. This simulation acts as a confirmation that when W_{ij} does not distort individual perception, traditional confounder rules apply. Moreover, the results of the first set of simulations pertain: properly adjusted MR approaches show considerable negative bias and modest undercoverage of the 95% CI. PV-SEM offers considerable bias reduction and improved CI coverage, at the cost of increased variability and hence nearly equivalent MSE.

Simulation set 2 conclusions

This exploratory assessment of analytic approaches to systematic measurement error and confounding yielded several insights and questions for further investigation. On the technical side, SEM models with a single individual metric encountered considerable convergence challenges; convergence problems for PV-SEM remained infrequent. In analytic models excepting PV-SEM, empirical results suggested that predictions for appropriate control of W_{ij} were inaccurate. Specifically, when W_{ij} affected measurement but was not a true confounder, controlling for both W_{ij} and $\overline{W}_{.j}$ reduced bias in several simulations, particularly when W_{ij} was also included in the measurement model. We explore this further in the discussion section. When W_{ij} was a true confounder, adjusting for it in the outcome model still improved estimation, but

substantial bias remained. In contrast, the PV-SEM approach required less adjustment to reduce bias, such that controlling for W_{ij} in the outcome model was most useful when W_{ij} was a true confounder. Controlling for W_{ij} in the measurement model improved estimation slightly when W_{ij} distorted measurement, although this required its inclusion in the outcome model as well. The reduction in bias resulting from including W_{ij} in the measurement model was considerably smaller than that achieved through the PV-SEM approach, despite the relatively large magnitude (nearly 2 SD of θ_{ij}) of the distortion from W_{ij}. Combining plausible values with SEM continued to provide the best option for minimizing bias and optimizing CI coverage, although the higher variance of these approaches led to similar MSE to the better performing of the MR approaches. ICC and reliability were accurately estimated in the plausible values / SEM model except when a third variable affected measurement, increasing person-level variance in a manner that was not corrected for in any adjustment. Further consideration of ways to address this phenomenon and of the implications of the reduction of bias when controlling for $\overline{W}_{.j}$ is warranted.

Simulations set 3: Alternative causal model

This set of simulations compared analytic performance under a formative (individual-to-group) causal model, which should favor the MR approaches, particularly when sampling fraction is high in simulation 3.0. The low (0.05) sampling fraction in simulation 3.1 may induce sufficient sampling error to allow the error correction of SEM to perform better than the MR approaches. Results from analysis of the true latent values show minimal bias in both settings (Table 2.10).

Table 2.10: Performance of MR versus SEM approaches under a formative causal model

	Bias	Relative bias	Variance	MSE	Coverage probability
Simulation 3.0: Large sampling fraction (0.50)					
True θ_j & θ_{ij}	0.003	1.13%	0.007	0.007	95.0%
M-MR	0.011	3.63%	0.008	0.008	93.0%
EAP-MR	0.008	2.78%	0.008	0.008	93.2%
PV-MR	0.012	4.09%	0.008	0.008	93.6%
PV-SEM	0.031	10.18%	0.009	0.010	95.8%
Simulation 3.1: Small sampling fraction (0.05)					
True θ_j & θ_{ij}	0.002	0.59%	0.008	0.008	93.6%
M-MR	-0.015	-5.04%	0.008	0.009	91.6%
EAP-MR	-0.018	-6.01%	0.008	0.009	91.2%
PV-MR	-0.014	-4.69%	0.008	0.008	91.8%
PV-SEM	0.004	1.39%	0.009	0.009	96.0%

The analytic results hew to predictions. In the setting where the aggregate characteristics of a fixed population affect the outcome and where 50% of individuals are sampled per group, use of the MR approach with any method within individuals reduces bias (to a minimum of 2.9% for the EAP-MR analysis). In contrast, the PV-SEM approach shows increased bias (10.2%) due to over-correction of θ_j. However, when the sampling fraction is low, the correction against attenuation in the PV-SEM approach reduces bias relative to the attenuated estimates from MR models: relative bias for PV-SEM is 1.4% compared to -4.7% for the next best approach, PV-MR.

The higher bias and greater variability of the PV-SEM analyses in simulation 3.0 result in the largest MSE (0.010). However, PV-SEM continues to provide excellent coverage of the 95% CI and to retrieve the ICC and reliability nearly exactly (0.102, SD 0.024 and 0.810, SD 0.042

respectively) in this simulation. In simulation 3.1, the slightly larger variance of PV-SEM leads an equivalent MSE to the more biased M-MR analyses (0.009); PV-MR minimizes MSE at 0.008 in this simulation. PV-SEM is conservative with respect to the 95% CI, with the MR approaches undercover the CI. The estimated ICC of 0.100 (SD 0.024) is accurate; estimated reliability is similarly close to the truth (0.808, SD 0.043).

Simulation set 3 conclusions

This set of simulations addressed the question of how MR and SEM approaches would perform when the underlying causal model matched the logic of MR analyses but sampling fraction varied. When sampling fraction is high, and thus the mean of sampled individuals is a reliable estimate of the true group composite, MR approaches outperform SEM, as expected. When sampling fraction is low, as is often the case in social epidemiologic research, the measurement-error correction of the SEM analyses proves useful in reducing bias. In keeping with prior results, higher variance of SEM analyses leads to fairly equivalent MSE across methods; PV-SEM does the best job of providing accurate coverage of the 95% CI.

Motivating example: Community organizations and heavy episodic drinking

We applied the insights from the simulation studies to the question of whether village social capital shapes individual drinking behavior among young men in rural South Africa. All 1181 survey respondents provided complete responses to the 8-item organizations and networks scale. They were most likely to consider church groups important in improving villagers' lives, with 846 (71.6%) considering church groups very important, and least likely to endorse men's groups as existing or important for the village (187 responding very important, 15.8%). Item location and fit are shown in full in Appendix Tables B.8 (1PL) and B.9 (2PL). Individual item fit statistics fall well within an acceptable range (0.75 to 1.33) in the one-parameter model, although the log likelihood test reveals improved global fit with a two-parameter model: deviance improves from 13861.3 to 12532.3.3 for a chi square statistic of 1329.0 with 15 degrees of freedom, $p<0.001$. For illustrative purposes, we proceed with both models; see the discussion for consideration of the content and statistical considerations around choice of measurement model. Both language and gender proved significantly associated with all versions of the organizations and networks scale (raw score, 1PL model, 2PL model) in separate linear regressions (results not shown). We therefore control for these variables in generating the final EAP and PV for analysis and include them at the individual-level and group-level in the outcome models. Villages ranged in size from 800 to several thousand people; the sampling fraction among young men is thus under 20% even in the smallest villages. Although we assume individual perception of organizations and networks function as reflective indicators, latent variable approaches would be justified in this setting for formative indicators as well, given the low sampling fraction per group.

Of the 581 young men in the survey, 118 (20.3%) reported HED in the past 12 months. Table 2.11 presents the association of village organizations and networks with HED among young men using the analytic strategies applied in the simulation studies. PV-SEM models converged using mode-and-curvature adaptive quadrature with 7 integration points when these models were fit drawing initial values from a non-adaptive Gauss-Hermite quadrature model; the mean / SEM model converged using non-adaptive Gauss-Hermite quadrature with 3 integration points. ICC and reliability estimates were obtained from models not adjusted for confounders, coefficient estimates and CIs from models with the confounders adjusted in the outcome model.

Table 2.11: Comparison of analytic approaches of adjusted association of village organizations and networks with heavy episodic drinking among young men (N=581)

	β (95% CI)	Estimated ICC	Estimated reliability
Mean / mean	-1.35 (-2.52, -0.19)	NA	NA
Mean / SEM	-8.64 (-48.38, 31.10)	0.98	1.00
1PL PV / mean	-0.85 (-1.63, -0.08)	NA	NA
2PL PV / mean	-1.13 (-2.04, -0.22)	NA	NA
1PL PV / SEM	0.32 (0.03, 0.62)	0.13	0.80
2PL PV / SEM	0.41 (0.11, 0.71)	0.13	0.80
1PL PV / SEM*	-0.32 (-0.62, -0.03)	0.13	0.80
2PL PV / SEM*	-0.41 (-0.71, -0.11)	0.13	0.80

*These models were run with initial values provided by models where the loading of each plausible value onto θ_j and θ_{ij} was constrained to 1.0.

All of MR approaches indicate a negative and significant association between organizations and networks and HED, regardless of underlying item model. The M-SEM analysis returns extremely high variance and hence an insignificant result. On initial fit, the two PV-SEM models differ from all other models in returning positive and significant estimates. However, these model results returned negative coefficients for PV loading onto the latent traits; providing positive starting values for the link from the measurement to latent traits[7] resulted in negative, significant estimates for the association of organizations and networks and HED that are the inverse of the initial model. Given the similarity of these estimates to the MR results and the plausibility of a positive measurement model, we consider the negative coefficients more likely to represent the direction of the true causal effect. As in the simulation studies, the M-SEM result is an outlier that is likely to be unreliable. The concordance of the results from the other models adds credibility to the interpretation of a modest negative association between village organizations and networks and young men's HED. In light of the simulation studies suggesting greater accuracy in the PV-SEM approach, we advocate for the use of the estimates from this approach as the best option for reporting. More broadly, the results demonstrate some of the challenges in applying SEM to complex data, despite its advantages in the simulation studies.

DISCUSSION

This work represents the first comprehensive assessment of analytic approaches for estimation of contextual effects of latent variables in the context of social epidemiologic research. We compared frequently employed mean and regression approaches to latent variable models incorporating IRM and SEM, with a specific exploration of the novel application of plausible values from IRM in multilevel SEM. Simulation studies mimicked social epidemiologic research in terms of group size and number as well as use of a scale of polytomous items. As predicted based on existing research, MR methods consistently underestimated the contextual effect of interest. The PV-SEM approach minimized bias and optimized accurate inference (coverage of the 95% CI) in most settings, including when underlying assumptions were violated, although it showed some sensitivity to strong non-normality of the latent trait. However, the PV-SEM approach was more variable than MR

[7] It is not uncommon to constrain the measurement model such that the coefficients relating the measures to the latent group variable are set to 1.0. An exploratory simulation of 125 runs suggests such constraints reduce the bias-correcting capacity of the plausible value / SEM approach and result in incorrect ICC estimation, although further testing is warranted.

analyses, such that these methods were often equivalent when balancing bias and variance. SEM approaches with a single individual score or estimator did not improve bias in a simple setting; efforts to refine these methods for more plausible causal structures may not be warranted. MR approaches are suitable for samples capturing a large portion of the target population if the underlying causal model is formative, implying individual latent traits precede the group latent trait. SEM analyses outperformed MR approaches under a formative model if the sampling fraction was small. Finally, testing of the role of a third variable led to several conclusions: adjusting for a variable in the measurement model must be accompanied by adjustment in the outcome model to avoid introducing significant bias; MR approaches are susceptible to bias that can be partially corrected through the inclusion of the aggregate value of the third variable in some settings, while PV-SEM did not require this adjustment to reduce bias; and the impact of adjusting for a distorting covariate in the measurement model was small relative to bias improvement across methods. As in the robustness checks, PV-SEM outperformed other approaches in terms of bias reduction and CI coverage but was more variable and hence generally equivalent in overall accuracy.

Taken as a whole, these results demonstrate the feasibility of a number of two-stage approaches to the estimation of contextual effects of social factors, particularly latent variable methods that directly incorporate the error inherent in the measurement of latent traits. The two-stage approaches may be particularly appealing as measurement models and outcome models grow more complex, rendering the application of a single full SEM technically challenging. The individual-level scores and estimators to summarize items applied here enabled use of linear models within individuals, a far simpler estimation problem for SEMs than categorical items (88). Moreover, modeling group constructs measured by a large number of polytomous items or by multiple sub-scales requires a considerable number of individuals and groups even before considering factors distorting measurement or covariates required for the outcome model. Nonetheless, some researchers caution against the potential bias and loss of information when scale scores or estimators replace items in estimation (60). Future research should consider the comparison of these methods to the full SEM approach in a range of applied contexts to test the bias-variance tradeoff achieved here against the potential of the full SEM analysis, as well as their suitability to data sets of varying size. An additional avenue worth exploring given the high variance of the PV-SEM analyses is a partially latent SEM in which for instance plausible values are modeled within individuals and the group average of those values is modeled within clusters; prior research suggests the reduced variance of this approach can compensate for increased bias relatively to the fully latent model in situations of small sample size or low ICCs (74).

Beyond choice of specific analytic approach, this work illuminates the need for greater consideration of the hypothesized causal model underlying the measurement of latent traits in epidemiologic research. As shown in prior research and confirmed here, which analytic method is most suitable varies based on the assumption of a reflective causal model or a formative causal model. This consideration is particularly critical for studies with large sampling fractions, although these are less common in epidemiologic research than education studies. Additional work on the implications of sampling fractions on analytic performance under a reflective causal model would address one of the remaining gaps in the literature (61). The broader question of how plausible these causal models are also warrants consideration in particular content applications. Imposing a reflective causal model to the measurement of a latent group trait such as collective efficacy of a geographic neighborhood implies that historical processes, including actions of past residents, created the group social capital at a given point in time, and that this

construct in term informed the perceptions of the current residents. An alternative view more akin to a formative model would be that a trait such as collective efficacy does not inherently exist outside of individual perceptions of it, so that a group's collective efficacy at any given time is the product of the individual actions and views of current residents. These two options represent the ends of a spectrum rather than mutually exclusive categories. Intermediate models, such as a reflective model without the assumption that individual perspectives are homogenous by definition but instead might truly differ by an observed covariate, are worth exploring. Social epidemiologists are just beginning to grapple with the stability of constructs such as collective efficacy over time, the potential contributions of structural versus more transient individual characteristics in creating such constructs, and the implications of neighborhood change over time for outcomes research (73,89). Linking these debates to the implications for measurement would enrich future research endeavors. Non-parametric SEM with more flexible definitions of group constructs, such as a vector of individual processes rather than a single summary like the mean, may represent a useful tool to apply in this endeavor.

Our results suggest that defining the causal model in terms of the role of covariates also has implications for analytic approach, with somewhat different implications than in analyses without measurement concerns. Introducing a covariate that distorted measurement led to considerable bias in MR approaches, some of which could be corrected by controlling for the aggregate of that confounder. We propose an explanation of the need to control for $\overline{W}_{\cdot j}$ in Appendix Figure B.9; this finding represents an important avenue of further research. Future studies should examine the magnitude of the bias when $\overline{W}_{\cdot j}$ is not controlled in settings with more modest levels of distortion, as this could represent an important and to-date unrecognized source of bias in epidemiologic analysis of contextual effects. Moreover, this result suggests that the single-level causal graphs employed in practice (90) may be an oversimplification leading to erroneous decisions for analytic adjustment when constructs are measured across multiple levels. The multilevel structural causal model approach employed in the appendix provides a more informative approach; methods of consulting both causal graphs and hierarchical causal models could present another option to explore. The fact that SEM approaches, which did not include an aggregated value of exposure in the model, did not require adjustment by $\overline{W}_{\cdot j}$ when W_{ij} distorted measurement provides another reason to consider such analyses if measurement distortion is a possibility in a given analysis. Finally, limited benefit was observed from including the covariate in the measurement model if it was also controlled in the regression model; the main implication of these results is that if a latent regression is necessary, for example as a strategy for dealing with differential item functioning (DIF), including the covariate from latent regression in the outcome model is necessary to correct the bias introduced.

On a more technical level, the simulation study and applied example raise the question of utilizing the appropriate measurement model in any IRM-based analyses, as the choice of measurement model can affect resulting parameter estimates. Under a construct measurement approach, models imposing constant discrimination provide conceptual benefits by creating a common metric of item difficulty across all measured individuals. In this theoretical framework, invariant discrimination is a prerequisite to a useful scale; responsible measurement entails revising items to meet this requirement rather than modifying the statistical model to accommodate variant items (63). Practically, however, this raises a question for researchers using established scales that have not been designed to ensure invariance or that display varying discrimination in a particular usage. Simulation study results indicate that the estimation of other relationships can be biased if the measurement model does not match the underlying data

generating process, for instance in the simple setting when a mean model was used for items with truly variant discrimination. It is thus important to identify in advance the key criteria – conceptual and statistical – for determining an optimal measurement model when the goal is accurate estimation of a contextual effect. For instance, in the applied example on social cohesion, there is no theoretical reason to select a two-parameter model over the constraints of a one-parameter model: the items were developed to cover a range of person locations with equal discrimination. The reasonable item fit statistics of the one-parameter analysis do not provide compelling evidence of statistical misfit, although the global fit does suggest that as a whole the two-parameter model better fits the data. Without strong evidence of individual item misfit, one could prioritize coherent measurement and use solely the results based on a one-parameter model. If specific items showed poor fit, they could be removed from the analysis to enforce uniform discrimination, though at the cost of precision of measurement. Alternatively, a conservative approach would be to use the two-parameter model as a sensitivity check and consider comparable results indicative of a robust association, with conflicting results inconclusive. This option at least provides a means of moving forward with imperfect measures while acknowledging the conceptual as well as statistical considerations involved in this decision.

This research was intended to ground latent variable analyses in the context of social epidemiologic research, to assess the suitability of latent variable analyses to social epidemiologic research compared to more commonly used methods, and to bring to bear epidemiological concern with confounding to the question of measurement. We conclude that the most novel approach attempted here, combining IRM-based plausible values with multilevel SEM, offers promise in reducing bias across a range of simulation settings, although it rarely outperforms mean and regression approaches in terms of both bias and variance. Although implementation of SEMs is technically challenging and can require more data than many social epidemiologic studies provide, these methods provide a means of addressing the attenuated effects observed in mean and regression approaches and hence preventing promulgation of false negatives in social epidemiologic research. Epidemiologists should extend their consideration of the causal model to incorporate measurement when latent variables are assessed, as the hypothesized model can help determine the most appropriate analytic method. There appear to be benefits to working through equations for structural causal models across multiple levels rather than single-level graphical simplifications in terms of identifying appropriate covariates for analytic control. Extensions to this work, including attempts to reduce the variability of the PV-SEM analyses and comparison to a full SEM approach, provide avenues for further exploration in translating such methods for applied epidemiologic work.

Chapter 3: Collective Efficacy, Alcohol Outlet Density, and Young Adults' Alcohol Use in Rural South Africa

ABSTRACT

Alcohol use contributes to morbidity and mortality in developing countries by increasing the risk of trauma and disease, including alcohol dependence. Limited research addresses determinants of alcohol use beyond the individual level in sub-Saharan Africa. We test the association of community collective efficacy and alcohol outlet density with young men's drinking in a cross-sectional, locally representative survey conducted in rural northeast South Africa. Informal social control and cohesion show protective associations with men's heavy drinking, while alcohol outlet density is associated with more potential problem drinking. These findings provide initial support for intervening at the community level to promote alcohol reduction.

INTRODUCTION

Alcohol use in South Africa

The harmful use of alcohol is a growing global public health priority. Alcohol consumption contributes to over 200 health conditions, including injury and both communicable and non-communicable diseases (7). Although the causal pathways are not fully elucidated, alcohol-related harms can be occasioned by the volume of alcohol consumed as well as through the particular pattern of drinking (91). The broad effects of alcohol on risk of injury as well as communicable and non-communicable diseases are of particular salience in developing countries where other component causes of such outcomes are prevalent. Although levels of drinking tend to be lower in developing countries, the associated harms of alcohol use are disproportionately high (92).

In South Africa, heavy alcohol consumption poses a serious risk to public health (93). Although over 40% of men in South Africa report abstinence from alcohol, consumption is high among drinkers; those who drink consume an average of over 30 liters of pure alcohol (ethanol) per year (7), which is equivalent to nearly 3.5 U.S. pints of 5% alcohol-by-volume beer every day. This concentrated use results in considerable morbidity and mortality, particularly among men. As of 2012, an estimated 39,000 deaths were attributable to alcohol in South Africa (6.4% of all deaths), the vast majority of them among men (10). The contribution of alcohol use to alcohol use disorder (AUD), road traffic accidents, and liver cirrhosis alone accounted for approximately 5% of disability-adjusted life years (DALYs) among South African males in 2012 (11). This represents only three of the health outcomes for which alcohol is a component cause and does not address morbidity and mortality from HIV, although increasing evidence of a role for alcohol in HIV transmission and progression to AIDS suggests that heavy alcohol consumption may be worsening South Africa's ongoing epidemic of HIV and AIDS (7,94–97). Preventing alcohol-related harms and dependence is therefore a critical means of improving population health in South Africa.

Determinants of alcohol use

Alcohol use is a product of factors ranging from national historical context to individual genetic predisposition. Globally, level of alcohol consumption is associated with greater economic development between countries and higher socioeconomic status within countries (7).

National and local policies on alcohol cost and availability as well as sanctions for alcohol-related offenses can shape individual consumption (98). Individual-level characteristics consistently associated with alcohol use include age and gender; in South Africa as well as globally, alcohol consumption tends to increase with age and is much more common in men than women (99). Between national policy interventions and individual characteristics lie a number of potentially modifiable community factors, such as social norms around alcohol consumption, that may shape individual drinking. Although there is a long history of community-based prevention strategies in developed countries (100), the relevance of this research to sub-Saharan Africa is only beginning to be assessed. Researchers recognize the need for prevention interventions that act on social and structural risk factors at the community level, particularly related to HIV (101–103). A more complete understanding of community causes of alcohol use in sub-Saharan Africa would facilitate effective population-level prevention of harmful alcohol use. We briefly review existing evidence, globally and in sub-Saharan Africa, of two potential community-level determinants of alcohol use: community collective efficacy and alcohol availability.

Collective efficacy and drinking

Motivated by theoretical work such as social disorganization theory, researchers have investigated links between community social context and drinking behavior (104). Social disorganization theory posits that neighborhood structural conditions such as poverty and residential instability shape health outcomes through social factors like collective efficacy (105). Two elements comprise collective efficacy: working trust among community members (social cohesion) and, based on that trust, a mutual expectation to take action for shared interests (informal social control) (28,106). Although social factors have been linked to adolescent drinking, limited research addresses collective efficacy and adult alcohol use (104). One study identified a protective association between informal social control and binge drinking among adults in Los Angeles, but found no association with cohesion (38). There is little research on this topic outside of the United States (104).

In South Africa, initial examinations of social disorganization theory have produced mixed results. A small number of studies on the context of adolescent alcohol use support the relevance of community factors such as neighborhood dereliction in drinking behavior (107,108), with one study documenting a potential protective association between community affirmation and adolescent consumption of home-brewed alcohol (109). Direct study of collective efficacy to date is scarce: a study in KwaZulu Natal employed a two-item measure of social cohesion that was correlated with lower social disorder (e.g., crime) and was associated with lower rates of adolescent sexual initiation, particularly for males (39). However, social cohesion was weakly positively correlated with neighborhood disadvantage in this study, contrary to theoretical predictions. Cain et al. measured perceived collective efficacy among men and women in Cape Town as an individual's belief in their community's capacity to prevent HIV and found this to be associated with reduced frequency and quantity of alcohol use (40). It remains to be determined if community collective efficacy shapes alcohol use in South Africa.

Alcohol availability and drinking

Structural conditions such as alcohol availability comprise a second major focus of investigation into community-level determinants of alcohol use. Alcohol outlet density increases physical access to alcohol, which may lower alcohol prices and shape social behavior around

drinking (110). Ecologic studies from developed countries have shown overall alcohol consumption and alcohol-related harms to be higher in areas with greater outlet density (25). Findings have been mixed when assessing individual alcohol consumption, with studies in New Zealand and the United States finding no association between density of off-premise alcohol outlets (i.e. liquor stores) and average individual consumption (26,27). A systematic review on availability of alcohol found the overall body of evidence to be inconclusive (111). Nonetheless, the U.S. Guide to Community Preventive Services deems regulation of alcohol outlets a useful public health tool (112).

It is not yet known whether alcohol outlet density affects alcohol use in South Africa, where a majority of alcohol is sold at informal taverns, or *shebeens*, as opposed to licensed on-premise (bar and restaurant) and off-premise alcohol outlets (113). A study from the Western Cape province found that socioeconomic deprivation is associated with a higher concentration of unlicensed outlets and fewer licensed outlets (114), suggesting that, as in existing studies of outlet density, surrounding poverty may act as a confounder (26,27,115). Few studies address both social and physical environmental predictors of alcohol use within communities, and none to our knowledge has addressed these questions in South Africa.

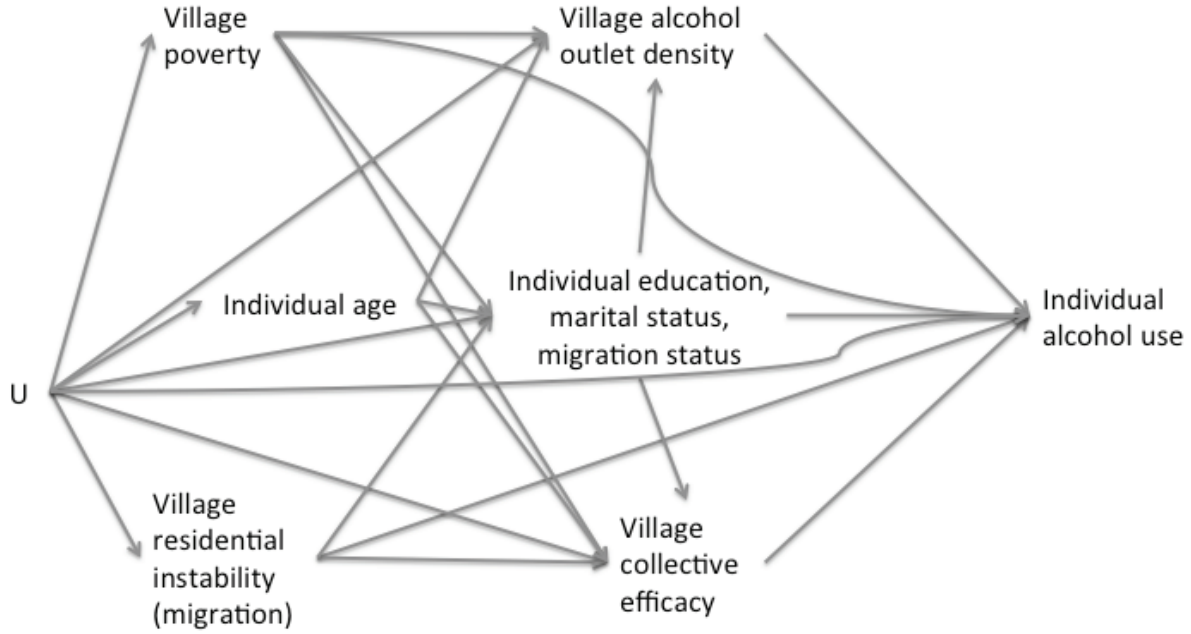
Study aims

We examine the relationship of community social and physical environmental factors with heavy alcohol consumption and potential problem drinking in a population-based sample of young men in rural South Africa. Heavy drinking is most consistently linked with alcohol-related morbidity and mortality, while the pattern and circumstances of drinking that comprise potential problem drinking are indicative of greater risk for future AUD. Figure 3.1 shows the proposed causal model underlying this study. As posited by social disorganization theory, village structural determinants of poverty and instability can undermine collective efficacy, while lower collective efficacy may increase individual alcohol use. Similarly, village poverty may affect the location of alcohol outlets; outlet density in each village plausibly increases individual drinking. The probability that an individual lives in a given village and hence is exposed to the local alcohol outlet density and collective efficacy is a function of individual characteristics such as age and education, which also affect alcohol consumption. Other individual characteristics, such as psychosocial factors, are excluded from the model due to the assumption they do not affect individual selection into a village and hence are independent of exposure.

We test whether 1) collective efficacy and 2) alcohol outlet density affect individual heavy drinking and potential problem drinking. We hypothesize a protective association between collective efficacy and drinking outcomes, particularly potential problem drinking due to its inclusion of elements beyond the individual (e.g., expressions of concern about one's drinking). We hypothesize a harmful association between alcohol outlet density and drinking outcomes, especially heavy drinking since outlet density facilitates access to alcohol. This research can inform structural interventions at the community level, like those being implemented for HIV reduction in this region. If social factors such as collective efficacy do impact drinking behavior, interventions at the community level provide an optimal platform for addressing alcohol use. If alcohol availability plays a critical role in consumption patterns, policy interventions or community action should be targeted at limiting or better regulating alcohol outlets. The efforts undertaken by the South African government to confront alcohol-related harms at the national level (113,116) could be complemented by community-level approaches if modifiable factors associated with alcohol use are identified.

Figure 3.1: Causal framework of contextual factors affecting alcohol use

Causal model of structural conditions, individual characteristics, and village factors shaping alcohol use. Individual characteristics are shown as antecedents of village factors following the convention that selection factors affecting individual probability of living in a particular village and being exposed to village characteristics can be depicted analogously to traditional confounders (90). U=Unmeasured causes.



MATERIALS AND METHODS

Study site

The study is situated in the rural Agincourt sub-district of the Mpumalanga province of South Africa, where the Medical Research Council and University of the Witswatersrand Rural Public Health and Health Transitions Research Unit (Agincourt) has been running a Health and Socio-demographic Surveillance System (HDSS) since 1992. At the time of this study, the area had approximately 90,000 people in 27 villages (117). Mozambican immigrants comprise a sizable minority of the population, many of them from the Shangaan ethnic group predominant among native-born residents (87). Unemployment is high, with only 29% of working age adults reporting employment in 2007 (118). HIV prevalence peaks at over 45% for 35-39 year old adults (119).

Study procedures

This study combines data from community and individual sources collected as part of a cluster randomized trial of an intervention called "One Man Can," which aims to reduce HIV risk among young men and women through community mobilization strategies (52). In 2010, a community asset mapping exercise took place as part of formative research prior to initiation of the trial. Key informants convened in each village and identified and mapped current physical infrastructure throughout the village, including schools, clinics, sports fields, and alcohol outlets. In 2012, a population-representative cross-sectional survey was conducted as the baseline for the trial. It consisted of a random sample of approximately 55 young adults (ages 18 – 35) per village from 22 of the sub-district villages, limited to one respondent per household. Individuals who had spent fewer than six months of the prior year as an area resident were ineligible. Visits

were made to 1,826 households of a total of 2,252 sampled for participation (81.1%); sample size was reached in some villages before the sample was exhausted. Sixty-nine percent (n=1,256) of households contacted included an eligible resident; 1181 of those eligible consented to enroll into the study (94.0%), 600 women and 581 men. Interviews were administered in English or Xitsonga (Shangaan) via computer-assisted personal interviews (CAPI) at the respondent's home. Sampling weights are applied to each response to account for probability of household selection and respondent selection within household.

The study was reviewed and approved by institutional review boards (IRBs) at the University of California, San Francisco; the University of North Carolina at Chapel Hill; and the University of the Witwatersrand, South Africa. The Mpumalanga Department of Health and Social Development Research Committee also approved the study. The analysis of de-identified data reported here was designated non-human subjects research by the IRB at the University of California, Berkeley.

Measures

Informal social control and cohesion were measured on the baseline survey using items based on Sampson's collective efficacy scales (28); items were added or adapted for local relevance, pilot tested, and revised to their final form (Table 3.1). Responses were coded from zero to two, with higher values representing increased likelihood on informal social control items and increased agreement with social cohesion items. Individual scores were calculated on each measure as the sum of standardized item responses; scores for respondents within each village were averaged to create continuous village-level metrics. We conducted sensitivity analyses using two alternative definitions of social exposures. First, to reduce the possibility of reverse causation due to drinkers perceiving village characteristics differently from others, village scores were recalculated excluding heavy drinkers and separately excluding potential problem drinkers. Second, to incorporate potential nonlinear response patterns, individual estimators for all respondents were created using the expected *a posteriori* (EAP) estimate from a generalized partial credit item response model for each measure with sampling weights included (77). This model extends the two-parameter logistic model to polytomous data; it allows both the step size - the probability of choosing a more highly coded response option (such as 'Agree' rather than 'Somewhat agree') – and discrimination - to vary within and across items. The flexibility of this model would accommodate nonlinear response patterns more easily than a simple mean score does; we chose the two-parameter model for sensitivity analysis because of this contrast with the mean score, despite the greater interpretability of a more restricted one-parameter model.

Two types of alcohol outlets were identified during the community mapping: taverns (both licensed and *shebeens*) and bottle stores. However, because bottle stores often function as informal gathering places for alcohol consumption in these villages, we calculate outlet density as the total of both types of outlets divided by village area in square kilometers (km^2).

Alcohol use was measured using the World Health Organization's Alcohol Use Disorders Identification Test (AUDIT), a well validated, 10-item screening tool for harmful and hazardous alcohol use that includes domains of alcohol consumption, symptoms of dependence, and related harms (120,121). Each question is scored from zero to four points. In accordance with past studies using the AUDIT in South Africa and elsewhere, we defined heavy drinking as a score of four or more on the subset of questions related to consumption alone (the AUDIT-C) and potential problem drinking as a score of eight or above on the full AUDIT (122–126). Other

alcohol variables included location respondents typically drank, such as at a tavern, restaurant, or home.

Additional covariates included age, education, marital status, being born outside of South Africa, and two metrics of individual poverty: earning no income in the past three months and experiencing food insecurity in the past 30 days. Individual risk behavior relevant to HIV was measured by querying respondents on up to three recent sexual relationships; concurrent sexual relationships were defined as at least one month of overlapping relationships in the past year. HDSS census data were used to determine percent of village residents who were temporary migrants, defined as having spent under six months in the area over the previous year; percent employed, and percent of households with a female head. The percent of residents who were temporary migrants was used as a proxy for residential instability in this population; percent employed and percent of female-headed household were used to capture village poverty level. Female-headed households are more likely to be poor (87).

Table 3.1: Items used to measure collective efficacy

Informal social control: *Would you say it is very likely, somewhat likely, or unlikely that your neighbors could be counted on to intervene in various ways if:*

- Children were skipping school and hanging out on a street corner?
- Children were breaking windows on a local building/destroying public property?
- Children were showing disrespect to an adult?
- A fight broke out at the pension point?
- The local school closed down the feeding scheme?
- A family didn't have enough food?
- The neighborhood water tank was broken?
- An elderly person was robbed?

Social cohesion: *For each of the following statements, please tell us if you agree a lot, somewhat agree, or do not agree at all with the statement:*

- People in this village are willing to help their neighbors.
- This is a close-knit village.
- People in this village can be trusted.
- People in this village generally get along well with each other.
- People in this village share the same values.
- People in this village look out for each other.

Analysis

We assessed correlation of the collective efficacy sub-scales to determine whether they reflected a single underlying construct and correlation of collective efficacy with outlet density to determine whether the physical environment and social context were interrelated. For the IRM approach, three models were tested: a unidimensional model treating all items as indicators of a single collective efficacy construct, a two-dimensional model assuming informal social control and cohesion are dimensions within a single construct, and a consecutive model treating each sub-scale as its own construct. Model fit was compared using the likelihood ratio test for difference in deviance within nested models and by calculating the Akaike Information Criterion (AIC), the deviance plus two times the number of parameters. The AIC provides a means of penalizing model complexity as well as of comparing non-nested models. The conceptual validity of the selected dimensionality was assessed using the correlation matrix of between-dimension relationships. IRM models were fit using marginal maximum likelihood estimation assuming a normal distribution of the latent variable(s) in the sampled population. Monte Carlo

integration was used unless otherwise specified; final estimation employed a convergence criterion of change in deviance <0.001.

Descriptive analysis included calculation of summary statistics of village characteristics and separate comparisons of characteristics of heavy-drinkers and potential problem drinkers respectively to all others using Chi square tests. To compare the relevance of each drinking outcome to HIV risk, we assessed relative importance of heavy drinking against current drinking and at-risk drinking in predicting concurrent sexual partnerships. Both exposure and outcome are measured at the individual level for this analysis; we accounted for clustering by village as a result of the sampling strategy to ensure appropriate variance estimates. As a measure of variable importance, we used the predicted change in percent of men reporting concurrent partnerships associated with each alcohol measure after adjusting for demographic covariates (age, marital status, education, recent income, food insecurity, immigrant status), which is a non-causal equivalent of the average treatment effect (127). Estimation of the predicted difference in concurrency associated with each drinking metric employed targeted maximum likelihood estimation (TMLE) to achieve the optimal bias-variance tradeoff for this parameter (128). TMLE is a double-robust estimation method that is asymptotically unbiased if either the outcome (concurrency) or exposure (alcohol measure) is modeled correctly. We employed machine learning to reduce dependence on correct model specification, using a library with a number of algorithms to provide a flexible approach to fitting models for both the exposure and the outcome. These included generalized linear models (logistic regression, step-wise logistic regression, logistic regression with interaction among covariates), generalized additive models, adaptive regression splines (earth), and random forest classification and regression. Cross-validation was used to identify a weighted combination of these algorithms that minimized the mean squared error in predicting the outcome being modeled (129).

We examined the association of each village-level exposure with heavy drinking and potential problem drinking as indicated in the causal framework. We adjusted the association between each collective efficacy sub-scale and drinking for village-level confounders suggested by social disorganization theory (poverty and residential instability); the associations of outlet density with alcohol outcomes controlled for village poverty. All analyses included individual characteristics likely to affect drinking and selection into a village: age, education, marital status, nationality, and poverty. We modeled age as years greater than 18 and as a quadratic term to best capture the nonlinear relationship between age and each drinking outcome. Analyses employed logistic regression with sampling weights and robust clustered standard errors as well as parametric g computation for marginal modeling. We did not use machine learning in this portion of the analysis because confidence intervals (CIs) generated using bootstrap resampling following g computation with machine learning rely on assumptions of normality for appropriate coverage. While TMLE does provide theory-based CIs using the influence curve of the estimator in question, there is not yet software to implement this analysis when the exposure of interest is continuous.

Marginal modeling enables estimation of an additive association that is interpretable in terms of population health (130). We estimated the expected difference in prevalence of each drinking outcome for one standard deviation (SD) difference in each collective efficacy sub-scale by setting the exposure measures for all villages to one-half SD above and below the grand mean and using the regression model to predict outcomes under each setting to capture this contrast. We used the same procedure for outlet density, manipulating density to capture a difference of one outlet per square kilometer. Bias-corrected CIs were generated from a clustered bootstrap

with 10,000 resamples (131). Regression and marginal modeling analyses were also run for sensitivity analyses incorporating alternative metrics for collective efficacy exposures.

We conducted one post-hoc analysis: we tested effect measure modification by computing the relative excess risk due to interaction (RERI) (132). The RERI translates statistical interaction to the additive scale; a significant RERI may be indicative of causal interaction (133). We consider $p<0.20$ statistically significant interaction. Analyses were conducted using the Test Analysis Module (TAM) package (85) in R 3.1.2 (R Foundation for Statistical Computing) for generating collective efficacy estimates using item response modeling, the TMLE and SuperLearner packages (134,135) in R 3.1.0 for variable importance analysis, and Stata 11.0 (StataCorp, College Station, TX, USA) for all other results.

RESULTS

Village characteristics

Village size ranged from 0.72 km^2 to 6.48 km^2, with populations between 800 and 9,000 at the time of the study. Taverns were more common than bottle stores, with up to six taverns or shebeens per village compared to no more than two bottle stores. Villages contained an average of 1.37 alcohol outlets (range 0 to 3.24) per km^2. Census data affirmed the high level of poverty in this region, with an average of 41.6% (SD 3.2%) of households headed by a female and only 19.7% of adults employed (SD 1.77%). An average of 17.7% of residents were temporary migrants (SD 2.6%).

For the household survey, contact was made with 1812 households of a total of 2159 sampled for participation and visited by the team (83.9%). Seventy-two percent (n=1313) of households included an eligible resident; 1181 of those eligible consented to enroll into the study (89.9%). All respondents provided complete responses to the collective efficacy sub-scales. Cronbach's alpha was 0.88 for informal social control and 0.81 for social cohesion. The 22 villages ranged from -0.28 to 0.36 on the standardized informal social control scale and from -0.39 to 0.38 on the standardized cohesion scale. Though the two sub-scales are theorized as dimensions of collective efficacy, they were not strongly related in this context: correlation was 0.34 at the village level. In the IRM analysis, the two-dimensional model fit significantly better than the unidimensional model (likelihood ratio test statistic 2239.3.8 with one degree of freedom, $p<0.001$) but showed no better fit based on AIC than separate models. Each sub-scale was thus considered an independent village characteristic and analyzed individually. EAP reliability was 0.86 for informal social control and 0.78 for social cohesion in separate models (63). Alcohol outlet density was not correlated with either collective efficacy subscale.

Individual characteristics

Nineteen of 600 women were classified as heavy drinkers (3.2%) and six as potential problem drinkers (1%); due to the rarity of these drinking outcomes among women, all multivariate analyses are limited to men. Of the 581 men in the baseline sample, 343 (59.0%) reported any lifetime alcohol use (Table 3.2). Two hundred and one men (34.6%) were heavy drinkers and 140 (24.1%) were potential problem drinkers. The youngest men were least likely to be heavy drinkers; heavy drinking increased with higher educational attainment and with recent income, but was not significantly related to marital status. Just over half (110 of 201, 54.7%) of heavy drinking men were potential problem drinkers. As shown in Table 3.2B, similar traits distinguished potential problem drinkers: men under 20 were less likely to be problem drinkers, as were those who had never been married and those earning no income in the past three months.

Both heavy drinkers and potential problem drinkers were more likely to report concurrent sexual partnerships than men without these alcohol outcomes, with similar estimates of variable importance (or standardized difference in probability of concurrency) for heavy and problem drinkers. As shown in Table 3.3, the standardized probability of concurrency was 24.2% higher (95% confidence interval [CI] 13.3, 35.1) among individuals who were heavy drinkers and 21.4% higher (95% CI 12.5, 30.3) among individuals who were potential problem drinkers compared to individuals who were neither heavy nor problem drinkers, after adjusting for individual confounders.

Table 3.2: Characteristics of young men by current drinking status (N=581)

A. Heavy drinkers compared to non-drinkers and light drinkers

	Non-drinker or light drinker (n=380) N (col %)	Heavy drinker (n=201) N (col %)	Chi square test p value
Age			
18-20	193 (50.8)	69 (34.3)	
21-25	109 (28.7)	74 (36.8)	
26-30	45 (11.8)	39 (19.4)	
31-35	33 (8.7)	19 (9.5)	0.001
Education			
Primary or less	45 (11.8)	18 (9.0)	
Some secondary	239 (62.9)	108 (53.7)	
Completed secondary or above	96 (25.3)	75 (37.3)	0.009
Marital status			
Never married	324 (85.3)	163 (81.1)	
Married (legal or traditional)	41 (10.8)	23 (11.4)	
Separated, divorced or widowed	15 (4.0)	15 (7.5)	0.176
Born outside South Africa	43 (11.3)	18 (9.0)	0.377
Earned no income within 3 months	278 (73.2)	128 (63.7)	0.018
Experienced food insecurity within 30 days	11 (2.9)	8 (4.0)	0.484
Concurrent sex partners in past year	108 (28.4)	108 (53.7)	<0.001
Ever used drugs	22 (5.8)	34 (16.9)	<0.001
Potential problem drinker	30 (7.9)	110 (54.7)	<0.001

B. Potential problem drinkers compared to non-drinkers and non-problem drinkers

	Non-drinker or non-problem drinker (n=441) N (col %)	Potential problem drinker (n=140) N (col %)	Chi square test p value
Age			
18-20	218 (49.4)	44 (31.4)	
21-25	132 (29.9)	51 (36.4)	
26-30	54 (12.2)	30 (21.4)	
31-35	37 (8.4)	15 (10.7)	0.001
Education			
Primary or less	45 (10.2)	18 (12.9)	
Some secondary	276 (62.6)	71 (50.7)	
Completed secondary or above	120 (27.2)	51 (36.4)	0.044

Marital status			
Never married	378 (85.7)	109 (77.9)	
Married (legal or traditional)	47 (10.7)	17 (12.1)	
Separated, divorced or widowed	16 (3.6)	14 (10.0)	0.009
Born outside South Africa	51 (11.6)	10 (7.1)	0.137
Earned no income within 3 months	326 (73.9)	80 (57.1)	<0.001
Experienced food insecurity within 30 days	13 (3.0)	6 (4.3)	0.438
Concurrent sex partners in past year	139 (31.5)	77 (55.0)	<0.001
Ever used drugs	29 (6.6)	27 (19.3)	<0.001

Table 3.3: Standardized probability of concurrent sex partners among men by drinking outcome (N=581)

Drinking outcome	Probability of concurrency [a] (95% CI)
Current drinker	19.9% (10.4, 29.4)
Heavy drinker	24.2% (13.3, 35.1)
Potential problem drinking	21.4% (12.5, 30.3)

[a] Variable importance quantified as average treatment effect controlled for age, education, marital status, birthplace, recent income, food insecurity

Collective efficacy

As shown in Table 3.4, village informal social control was significantly associated with lower odds of heavy drinking after adjusting for both village-level confounders and individual covariates (β=-1.18, 95% confidence interval [CI] -2.26, -0.09). Village social cohesion similarly showed a significant association with heavy drinking: β=-1.07 (95% CI -1.82, -0.31). However, neither informal social control nor cohesion was associated with potential problem drinking.

Table 3.4: Multivariate logistic models of the relationship between collective efficacy measures and alcohol use among men (N=581)

A. Heavy drinking

	Coeff. (95% CI)	Coeff. (95% CI)
Informal social control	-1.18 (-2.26, -0.09)	-----
Cohesion	-----	-1.07 (-1.82, -0.31)
Age (years over 18)	0.18 (0.01, 0.36)	0.18 (0.00, 0.36)
Age squared	-0.01 (-0.02, 0.00)	-0.01 (-0.02, 0.00)
Education		
Primary or less	REF	REF
Some secondary	-0.26 (-1.10, 0.59)	-0.25 (-1.07, 0.58)
Completed secondary or above	0.26 (-0.58, 1.10)	0.30 (-0.52, 1.12)
Marital status		
Never married	REF	REF
Married (legal or traditional)	0.24 (-0.55, 1.02)	0.19, (-0.61, 0.99)
Separated, divorced, widowed	-0.06 (-1.05, 0.93)	0.00 (-1.00, 1.00)
Born outside South Africa	-0.01 (-0.84, 0.82)	0.09 (-0.70, 0.89)
Earned no income within 3 months	-0.36 (-0.77, 0.05)	-0.32 (-0.73, 0.08)
Experienced food insecurity within 30 days	0.64 (-0.69, 1.98)	0.56 (-0.80, 1.92)
Village % female-headed households	0.03 (-0.06, 0.11)	-0.02 (-0.08, 0.05)
Village % employed	-0.05 (-0.25, 0.16)	-0.09 (-0.26, 0.08)
Village % migrant	0.12 (-0.04, 0.27)	0.17 (0.07, 0.27)

Intercept	-3.25 (-6.59, 0.09)	-1.48 (-3.85, 0.90)
B. Potential problem drinking		
	Coeff. (95% CI)	Coeff. (95% CI)
Informal social control	0.43 (-1.42, 2.28)	-----
Cohesion	-----	0.46 (-1.00, 1.92)
Age (years over 18)	0.16 (0.01, 0.31)	0.16 (0.01, 0.31)
Age squared	-0.01 (-0.02, 0.00)	-0.01 (-0.02, 0.00)
Education		
Primary or less	REF	REF
Some secondary	-1.02 (-2.24, 0.20)	-1.02 (-2.26, 0.21)
Completed secondary or above	-0.42 (-1.49, 0.65)	-0.44 (-1.50, 0.62)
Marital status		
Never married	REF	REF
Married (legal or traditional)	-0.24 (-1.26, 0.78)	-0.22 (-1.29, 0.85)
Separated, divorced, widowed	0.67 (-1.05, 2.40)	0.65 (-1.10, 2.41)
Born outside South Africa	-0.54 (-1.75, 0.68)	-0.58 (-1.78, 0.62)
Earned no income within 3 months	-0.19 (-0.76, 0.37)	-0.20 (-0.75, 0.35)
Experienced food insecurity within 30 days	0.16 (-0.93, 1.24)	0.18 (-0.92, 1.28)
Village % female-headed households	0.06 (-0.06, 0.18)	0.08 (-0.02, 0.18)
Village % employed	-0.02 (-0.29, 0.25)	0.00 (-0.21, 0.21)
Village % migrant	0.01 (-0.21, 0.23)	-0.01 (-0.16, 0.14)
Intercept	-3.22 (-7.72, 1.29)	-3.87 (-7.32, -0.41)

Table 3.5 displays the marginal modeling results for collective efficacy. A one SD higher level of community informal social control was associated with a -4.3% difference in prevalence of heavy drinking (95% CI -10.0, 0.7). One SD higher level of cohesion was associated with a difference in prevalence of -4.2% (-9.6, -0.4) in heavy drinking among men across all villages. Sensitivity analyses comparing different measurement strategies for the village-level variables supported these findings. The magnitude of the association of heavy drinking with informal social control remained robust and near significance after removing heavy drinkers from exposure assessment and employing an item response model (results not shown). The magnitude and significance of the association with cohesion were also unchanged.

Marginal modeling results suggest non-significant differences of 1.3% (95% CI -6.5, 8.1) and 1.5% (-4.2, 8.3) in prevalence of potential problem drinking associated with one SD higher levels of informal social control and cohesion respectively. These associations remained negligible and non-significant in both sensitivity analyses (not shown).

Table 3.5: Predicted population prevalence of alcohol use by level of collective efficacy

	Heavy drinking		Potential problem drinking	
Exposure:	Informal social control	Cohesion	Informal social control	Cohesion
High (0.5 SD above mean)	30.2%	29.8%	23.8%	24.1%
Mean	32.3%	31.9%	23.1%	23.3%
Low (0.5 SD below mean)	34.5%	34.0%	22.5%	22.6%
Difference (95% CI)	-4.3% (-10.0, 0.7)	-4.2% (-9.6, -0.4)	1.3% (-6.5, 8.1)	1.5% (-4.2, 8.3)

Alcohol outlet density

Results from multivariate regression of drinking outcomes on alcohol outlet density are shown in Table 3.6. Alcohol outlet density was not associated with heavy drinking in either multivariate regression analyses (Table 3.6) or in marginal modeling (Table 3.7). However,

higher outlet density was associated with increased risk of potential problem drinking. In multivariable regression analyses, we assessed whether this risk differed between men who primarily drank at alcohol establishments and all other men. We found that drinking location modified the association between outlet density and probability of problem drinking (RERI for interaction of outlet density and primarily drinking at taverns = 10.18, p=0.122). As a result, we present regression results stratified by drinking location in the right-hand panel of Table 3.6. Outlet density was positively associated with potential problem drinking only among the 281 men who drank in taverns (β=0.96, 95% CI 0.40, 1.52). Accounting for this interaction, the estimated prevalence of potential problem drinking was 27.6% under high alcohol outlet density and 18.4% under low density (Table 3.7). The marginal difference in potential problem drinking associated with a difference of one outlet per km^2 in all villages is therefore 9.2% (95% CI 2.2%, 16.7%).

Table 3.6: Multivariate logistic models of the relationship between alcohol outlet density and alcohol use (N=581)

	Heavy drinking	Potential problem drinking	
		Tavern drinkers (N=281)	Non-tavern drinkers (N=292)
	Coeff. (95% CI)	Coeff. (95% CI)	Coeff. (95% CI)
Alcohol outlet density	-0.13 (-0.64, 0.39)	0.96 (0.40, 1.52)	-0.36 (-1.19, 0.46)
Age (years over 18)	0.18 (0.02, 0.35)	0.04 (-0.13, 0.20)	0.25 (-0.10, 0.59)
Age squared	-0.01 (-0.02, 0.00)	0.00 (-0.01, 0.01)	-0.01 (-0.03, 0.02)
Education			
Primary or less	REF	REF	REF
Some secondary	-0.03 (-0.84, 0.78)	-1.09 (-2.42, 0.23)	-1.17 (-2.96, 0.61)
Completed secondary or above	0.46 (-0.35, 1.27)	-0.36 (-1.32, 0.61)	-1.01 (-2.91, 0.88)
Marital status			
Never married	REF	REF	REF
Married (legal or traditional)	0.10 (-0.72, 0.92)	0.14 (-1.45, 1.73)	-2.51 (-5.57, 0.55)
Separated, divorced, widowed	-0.16 (-1.15, 0.83)	0.86 (-0.91, 2.63)	-1.23 (-4.68, 2.23)
Born outside South Africa	0.07 (-0.76, 0.90)	-0.85 (-1.72, 0.02)	-1.25 (-3.87, 1.38)
Earned no income within 3 months	-0.23 (-0.63, 0.17)	-0.16 (-0.92, 0.59)	-0.10 (-1.71, 1.51)
Experienced food insecurity within 30 days	0.64 (-0.65, 1.93)	0.52 (-1.25, 2.30)	[a]
Village % female-headed households	0.03 (-0.08, 0.14)	0.19 (0.05, 0.34)	0.01 (-0.13, 0.15)
Village % employed	0.01 (-0.16, 0.19)	-0.12 (-0.31, 0.07)	0.03 (-0.18, 0.25)
Intercept	-2.53 (-6.76, 1.69)	-6.65 (-13.52, 0.21)	-3.09 (-8.49, 2.31)

[a] Variable omitted due to collinearity with outcome

Table 3.7: Predicted population prevalence of alcohol use by alcohol outlet density

	Heavy drinking	Potential problem drinking
High (0.5 outlets/km^2 above mean)	30.1%	27.6%
Mean	31.4%	22.8%
Low (0.5 outlets/km^2 below mean)	32.7%	18.4%
Difference (95% CI)	-2.6% (-12.6, 10.2)	9.2% (2.2, 16.7)

DISCUSSION

This population-based study provides evidence that community social and physical environmental factors shape alcohol consumption and potential problem drinking respectively in South Africa. As expected based on existing research on alcohol use and HIV, both heavy drinking and potential problem drinking were associated with concurrent sexual partnerships,

underscoring the importance of alcohol use as an element of the HIV risk environment (136). To our knowledge, no prior research has addressed the impact of both social and structural community characteristics on alcohol use within an adult population in South Africa. These results help to address this gap as well as the broader lack of research on community causes of alcohol use in low- and middle-income countries (104,111).

The association of collective efficacy measures with heavy drinking but not with potential problem drinking suggests that social disorganization theory may be relevant in explaining alcohol consumption in this context, albeit less relevant to alcohol dependence and harms. The association of each measure of collective efficacy with heavy drinking is unusual in research to date. Two studies in the United States and the Netherlands found some evidence of a protective association between moderate social cohesion and heavy drinking (81,137), while a study in among adults in Los Angeles identified a protective association of informal social control against binge drinking, but no association of social cohesion (38). In contrast, studies of collective efficacy among adolescents in the United States (138–140) and of social cohesion among adults in New Zealand (141) have found no direct association with alcohol use. It is difficult to determine whether differing results reflect an effect that varies by context or methodological variation in measurement and analysis across these studies. The lack of association between collective efficacy measures and potential problem drinking in our study indicates that social factors offer at best a partial explanation of drinking behavior. However, it is possible that the influence of the community social environment on young men is stronger in these rural villages, where employment opportunities are scarce compared to more urban, interconnected areas studied elsewhere. In addition, the community mapping exercise affirmed that village residents consider the village their community, ensuring that the units of analysis closely approximate individual perception of group identity and norms in the present study.

Further research on the application of social disorganization theory in this context is warranted. As part of the community mobilization intervention, study investigators have noted that the social cohesion measure fit their overall framework of mobilization, while informal social control was identified as a distinct construct (52). The ongoing community mobilization intervention provides an opportunity to test whether cohesion and social control change together or separately. Post-intervention assessment will enable testing of any relation of such changes to alcohol use. Moreover, some elements of the original social disorganization theoretical framework do not function identically in this setting. For example, regional patterns of migration and return are complex and have implications beyond residential turnover, such as the provision of remittances (87,142). The measure employed in this study may not capture the full range of influences of residential instability. As noted in other applications of social disorganization theory in South Africa, specific predictions based on theories developed in the United States may not hold true even if community factors do play a role in health behaviors (39). Refinement of the conceptual framework tested here would strengthen future research in understanding the role of community factors.

Alcohol outlet density was associated with potential problem drinking but not with heavy drinking in this study; men who drank primarily in taverns were responsible for the observed association. In other words, among men who primarily drink in taverns, those with greater access to alcohol outlets report more dangerous drinking behaviors. These findings indicate that while individual levels of consumption may not be associated with increased availability of alcohol within villages, symptoms of dependence and alcohol-related harms may be. Similar results were obtained in a nationally representative study in New Zealand, where outlet density was related to

alcohol-related harms without being associated with average consumption (26). One potential explanation for the lack of association between village outlet density and consumption is regional alcohol availability outside of taverns and liquor stores: one major site of drinking is weekly *muchongolo* (traditional) dance events, which rotate throughout the villages and at which residents of many villages congregate. Home-brewed alcohol as well as alcohol provided by informal vendors are available at the dances (143), providing a source of consumption independent of outlet density in one's home village.

The association of outlet density with potential problem drinking suggests that formal and informal taverns may shape social behavior around drinking in ways that result in greater perceived dependence and harms. A difference of just one outlet per square kilometer was associated with a meaningful difference in prevalence of potential problem drinking. This evidence bolsters the existing focus on shebeens in South Africa as critical sites of individual risk and of potential intervention (144,145). Additional metrics of drinking behavior and direct measurement of alcohol-related harms would provide greater insight on these relationships.

Several limitations should be considered in interpreting study findings. The measured confounders are unlikely to represent all shared antecedents of exposures and individual drinking. Results would be biased if outlet density or collective efficacy in one village affected drinking behavior in other villages in the study; such contamination is more plausible for outlet density than collective efficacy, as an individual could choose to travel to a village with greater availability of alcohol. Alcohol outlet density was measured prior to the individual survey; reverse causality could thus explain the association with alcohol use only if potential problem drinkers had relocated preferentially to villages with greater alcohol availability prior to the study. Collective efficacy was measured simultaneously with drinking. The observed association could therefore be due to reverse causation, with drinking behavior eroding collective efficacy. The sensitivity analysis excluding heavy drinkers from the calculation of the collective efficacy measure does account for how their perceptions of the village could impact results, but not for any effect heavy drinkers have on neighbors' perceptions or for their relocation to less cohesive villages.

Measurement error could bias the findings in a number of ways. Village collective efficacy is based on perceptions only of those aged 18 – 35. This age group was selected because the parent intervention study focuses on changing the social environment shaping sexual health for young women and their partners; however, it may result in incomplete measurement of village characteristics. The AUDIT is a screening tool and hence imperfectly sensitive and specific for heavy drinking (123). Self-report of alcohol use may be affected by social desirability bias and by uncertainty around standard drink size when consumption occurs in less formal settings. Site-specific research into drink size and patterns of consumption would strengthen future research (146). However, there is little reason to believe that responses to the AUDIT differ systematically by village factors, decreasing the chance of misclassification biasing the estimates unpredictably. Finally, the aggregation and regression-based approach employed here for the analysis of collective efficacy can lead to conservative estimates of the effect of latent social factors due to measurement error (74). Latent variable methods such as multilevel SEM offer promise in reducing this attenuation, at the cost of greater variability.

This study builds on several design and analysis strengths to provide new insight into the community context of drinking behavior. The results presented draw on a representative sample from a population-based sampling frame, rendering the findings more generalizable than data from studies using clinic-based populations or convenience samples. Alcohol outlet density was

assessed through community mapping in order to capture a full picture of drinking establishments, both licensed and unlicensed, in this region. Moreover, measures of collective efficacy employed were grounded in the theoretical work undertaken in the United States and adapted to this context to provide comprehensive, reliable metrics for use at the community level (52). The results proved robust to sensitivity analyses of potential reverse causality and exposure misclassification. Marginal modeling enabled calculation of population estimates of the difference in drinking corresponding to changes in exposure that could plausibly be effected through public health interventions, such as the one currently underway at the site. Estimates of potential change can help to guide choice of intervention components. This is particularly salient given the association of each drinking outcome with risky sexual behavior: both policy interventions and structural interventions seeking to modify social environments may offer promise as methods of reducing alcohol-related HIV risk.

The findings presented here provide the first evidence of associations between community social and physical environmental factors and young men's alcohol use in South Africa. They suggest that community social factors such as cohesion and perhaps informal social control are related to men's heavy drinking. Moreover, the results suggest that a modest difference in density of drinking establishments is associated with a substantial amount of potential problem drinking. Both drinking outcomes were correlated with high-risk sexual behavior. Identifying upstream factors that could mitigate direct harm from alcohol use as well as potentially affect HIV risk for young men and women opens new opportunities to improve population health in South Africa.

Chapter 4: Effects of a Randomized Community Mobilization Intervention on Alcohol Use in South Africa

ABSTRACT

Background: HIV prevention efforts in sub-Saharan Africa increasingly include modification of structural factors underlying the risk environment, such as inequitable gender norms. Alcohol use, particularly heavy use and drinking in the context of sexual activity, is a risk behavior related to hypermasculine gender norms that may fuel HIV transmission. We assess the impact of a randomized community mobilization intervention targeting gender norms and related behaviors on prevalence of alcohol outcomes in Agincourt, South Africa. We examine potential nonlinearity in the relation between intervention engagement levels and alcohol outcomes, which could be evidence of spillover effects on those who do not engage directly in the intervention.
Methods: Eleven of 22 villages in the study were randomized to receive a version of the One Man Can (OMC) campaign that focuses on mobilizing young men to modify gender norms and reduce risky behaviors. Volunteer mobilizers organized OMC activities such as street theater and soccer tournaments around specific risk domains, including alcohol use. A population-representative survey including 575 men aged 18 to 35 was conducted following the two-year intervention. Respondents were deemed to have engaged in OMC if they reported knowledge of the intervention and participation in at least two intervention activities. The survey measured heavy drinking and potential problem drinking with the Alcohol Use Disorder Identification Test (AUDIT) and queried alcohol use prior to most recent sex with each sexual partner. We assessed the effect of intervention assignment on village prevalence of heavy drinking, potential problem drinking, and drinking prior to sex among young men using targeted maximum likelihood estimation (TMLE). We tested a dose-response relationship between level of engagement among young men per village crossing specified thresholds (15%, 40%) and each outcome to assess nonlinearity in this association.
Results: The OMC intervention was not significantly related to any of the alcohol outcomes. High engagement with OMC was associated with higher prevalence of heavy drinking (estimated prevalence of 30.3%, 14.4% greater than estimated prevalence under moderate engagement, 95% confidence interval [CI] 4.7%, 24.2%) and of potential problem drinking (estimated prevalence 18.4%, 9.4% more than under moderate engagement, 95% CI 1.6%, 17.2%). These associations appeared nonlinear, as the differences between low and moderate engagement were negative and non-significant. Alcohol use before sex did not show a clear dose-response relationship with intervention engagement; moderate engagement was associated with lower prevalence of this outcome relative to both low and high engagement (11.9% versus 20.9% and 15.5% respectively). These differences were not statistically significant.
Conclusions: This study of a cluster-randomized community mobilization trial provided no evidence of a protective effect of the modified OMC intervention on men's alcohol use before sex at the community level, although the small study size of 22 villages led to imprecise estimates. Surprisingly, high dose of intervention engagement was associated with increased prevalence of heavy drinking and potential problem drinking among young men. No nonlinear dose effect was found for alcohol use before sex in ecological analyses. Individual-level assessment is warranted to substantiate this finding.

INTRODUCTION

South Africa has long been one of the countries most profoundly affected by the HIV epidemic, with the highest global burden of disease and an estimated adult prevalence of 17.9% (96). Prevention strategies increasingly include efforts to modify the social, economic, political, and environmental factors that structure risk and vulnerability (5,6,103,147). Understanding the structural elements that modify risk is thus an important element of successful prevention of HIV. Among social factors, gender inequality has been identified as a major driver of HIV transmission in sub-Saharan Africa, particularly the existence of a hegemonic masculinity that emphasizes dominance and aggression (44,148). While far from the only form of masculinity, enactment of this cultural ideal can entail violence, pursuit of multiple sexual partners and risky behaviors such as substance use, all of which contribute to HIV transmission (42,43,148,149). Qualitative and quantitative research from South Africa supports an association between individuals holding gender inequitable beliefs and engaging in sexual violence, partner violence, multiple concurrent partnerships, and heavy alcohol use (41–43).

Alcohol use is closely tied to the ideal of masculinity in South Africa in particular. Historically, alcohol use has been promoted as a masculine good and has been explicitly linked to male labor force participation through the *tot* system, where agricultural laborers were provided alcohol as a salary, and through the provision of taverns to miners as an implicit reward for labor (150). Moreover, the severely limited economic opportunities for Black South Africans under apartheid and the ongoing high unemployment rates for young men have rendered alcohol use as an attainable emblem of masculinity in a setting where other such indicators, like material possessions and the ability to support a family, are out of reach for many (42,144,151). Drinking patterns in South Africa are in keeping with the hypothesis that a subset of men pursue an ideal of masculinity that promotes excessive drinking. Although over 40% of men in South Africa report abstinence from alcohol, consumption is high among drinkers. Data from 2008 to 2010 indicate that men who drink consume an average of over 30 liters of pure alcohol (ethanol) per year (7), the equivalent of nearly 3.5 U.S. pints of 5% alcohol-by-volume beer every day and one of the highest levels of per-drinker consumption in the world (10).

Growing evidence indicates heavy alcohol use contributes to the HIV risk environment. A recent meta-analysis suggested an elevated HIV prevalence among those reporting using alcohol in sexual situations in Sub-Saharan Africa (97). This association could plausibly be due to both biological and behavioral mechanisms (94,95). Biologically, alcohol consumption disrupts immune function; animal models of HIV suggest that alcohol can increase viral replication and hasten the progression of disease (12). Evidence for a behavioral link between alcohol and HIV risk includes the association of drinking with multiple sexual partnerships, unprotected sex, coerced sex, and poor medication adherence among those on treatment (20,94,95,97,152). However, consideration of views of masculinity as a common cause of these outcomes is limited in this research. Alcohol use prior to sex, in particular, is associated with higher risk sexual encounters, including casual or transactional sex, and with diminished effectiveness of safe sex practices (e.g., increased chance of condom failure) (15). The recognition of alcohol's catalytic role in the HIV epidemic, especially in South Africa, has led to calls to increase intervention activities addressing the social factors giving rise to the dangerous use of alcohol, with gender inequity prominent among them (101).

Researchers have begun accruing evidence on the effectiveness of interventions designed to modify male gender norms to reduce the risks of inequity and hypermasculinity. A range of gender-transformative interventions have shown promise in modifying attitudes and behaviors,

although assessment of alcohol use as an outcome is uncommon (45,153). One study that did consider both gender norms and alcohol as elements of the HIV risk environment compared an individual-level intervention on gender norms to one on alcohol use (47). A short term impact on perpetration of gender-based violence and on alcohol use before sex was identified for each study arm respectively; the authors recommended future efforts to integrate alcohol and gender norms in HIV risk reduction counseling to optimize prevention impact. The Stepping Stones intervention is a community-based program comprising communication tools, relationship strategies, and life skills training with an emphasis on HIV prevention and reproductive well-being that has been implemented in sub-Saharan Africa and Asia (154). It includes a specific session on alcohol to emphasize the links from alcohol to violence and unsafe sex. A program evaluation identified reduced self-reported alcohol consumption in Ghana, Tanzania, and Uganda; bar owners in Uganda reported fewer alcohol-related harms, including serious drunkenness, violence, and casual sex in bars, following the intervention (154). When tested in a cluster-randomized trial, Stepping Stones resulted in a decrease in problem drinking among male participants one year post-baseline, although the difference was not sustained at two years (48).

The version of Stepping Stones implemented in the trial focused on the individual-level intervention for pragmatic reasons, although evidence from gender-based programming as a whole suggests that engaging the broader community may heighten program impact (45). A similar intervention program called One Man Can (OMC) was developed in South Africa in 2006 to transform men's gender norms in order to increase their involvement in children's lives – both their own and orphaned and vulnerable children in the community – and bolster their roles as advocates for gender equity and HIV prevention (44). OMC goals include drawing men's attention to the links between masculine ideals and harmful health behaviors such as heavy drinking, although content on alcohol is not a core component of the original intervention (153). OMC combines individual workshops with community-based activities in an attempt to effect change at multiple levels (46). Evaluations to date indicate that individual participants, particularly older men, alter their perceptions around masculine identities and relationships through the gender-transformative exercises (44). Participants also identified elements of hegemonic masculinity that are at odds with men's roles as family leaders, such as drinking to excess and shirking family responsibilities as a result (44). Whether fostering this existing recognition of an alternative masculinity has an effect on alcohol use is not yet known. To date, no study has rigorously assessed the impact of a community-based gender-transformative intervention on alcohol use.

We assess the effects of a randomized, community-based mobilization intervention employing OMC to modify gender norms in rural South Africa on alcohol use, including drinking prior to sex. Prior research using baseline trial data confirm an association between individual gender norms and HIV risk behaviors, including concurrent partnership, intimate partner violence, and alcohol abuse (42). The intervention was designed to modify the gender norms of individual men as well as equity between men and women in each village as a whole. Past research suggests a threshold of 15% of a target group engaging with an intervention as a critical mass for diffusion of behavior change messages (155). This intervention was intended to engage 40% of young men so that broader changes in norms would be effected, shaping health behaviors even among non-participants. We therefore test the total effect of the intervention on men's alcohol use at the village-level as well as the response of men's alcohol use to increasing levels of intervention engagement within the village; a nonlinear effect signals the potential for spillover effects on those who do not engage directly in the intervention.

MATERIALS AND METHODS

Study site

This intervention study took place in 22 villages in the Agincourt region of Mpumalanga province, approximately 300 miles northeast of Johannesburg, South Africa. The region is host to the Agincourt Health and Socio-Demographic Surveillance Site (HDSS) and was home to approximately 90,000 people at study onset (117). HIV is endemic in this region, reaching a prevalence of over 45% for adults aged 35 to 39 (119).

Study procedures

The intervention procedures are described in full elsewhere (82). Briefly, 11 of the 22 villages were randomized to a community mobilization intervention intended to increase awareness of the link from gender inequities to HIV and to engage community members, especially young men, in actions to reduce gender inequity and HIV risk. The intervention was modeled on OMC and developed in partnership with Sonke Gender Justice, the organization that designed the original OMC campaigns. Trained volunteer community mobilizers ran intervention activities, which included intensive workshops, community outreach, and the training of community action teams (CATs). Each activity was designed to bolster a particular element of community mobilization (e.g., cohesion, shared concern around HIV, collective action (52)) and to address at least one of seven content areas: 1) gender, power, and health; 2) gender and violence; 3) alcohol; 4) gender, HIV and AIDS; 5) healthy relationships; 6) human rights; and 7) actions for change. In addition to specific content on alcohol use, some intervention activities took place at *shebeens* (unlicensed drinking establishments) to maximize outreach to the target population. To extend the impact of the intervention beyond the individuals who directly engaged with it, mobilizers and community action teams organized community activities such as soccer tournaments, home visits, and film screenings. Due to the relative proximity of the intervention and control villages, mass media was not employed as a mobilization method.

The target population for the intervention was adults aged 18 to 35 (approximately 25,000 individuals), with a particular focus on men. The intervention goal was to reach 40% of men through at least one intervention activity within two years. However, all community members, including residents of other villages, were welcome to participate in intervention activities. The most recent Agincourt HDSS household census was used to create sampling frames for males and females aged 18-35 to enable population-representative sampling of approximately 1,200 young adults at baseline (March to June 2012) and again following intervention completion (July to September 2014). The target enrollment for each of the 22 communities was 55, with approximately 28 respondents of each gender. One individual was sampled per home. Individuals were eligible if they lived primarily at that home (spent a majority of nights each week there), were 18-35 at the time of the survey, were the gender for which the household was sampled, and had lived in the study area for the past year. Written informed consent was obtained from all participants in the survey. The survey was interviewer-administered in the local language Shangaan or English using a computer-assisted personal interview (CAPI) in which the interviewer directly entered responses into a tablet computer.

The study was reviewed and approved by institutional review boards (IRBs) at the University of California, San Francisco; the University of North Carolina at Chapel Hill; and the University of the Witwatersrand, South Africa. The Mpumalanga Department of Health and Social Development Research Committee also approved the study. The analysis of de-identified

data reported here was designated non-human subjects research by the IRB at the University of California, Berkeley.

Measures

Villages randomized to the intervention are defined as exposed for the total effect analysis; no mobilization activities were undertaken in control villages, although residents of these villages were able to engage with intervention activities in other villages of their own volition. Individual engagement with the intervention was assessed through a series of survey questions ranging from, "Have you heard of OMC or seen the OMC logo?" to "Are you a member of a community action team?". See Appendix Table C.1 for full scale and response options. We employed item response analysis to fit a one-parameter (partial credit) model to the engagement questions in order to assess item fit and scale reliability. The final scale showed good item fit (only one item indicating misfit) and reasonable reliability, with expected *a posteriori* [EAP] separation reliability of 0.812. Individuals with a total score of at least three out of a possible 20 were considered to have engaged with the intervention; this score would indicate having heard of OMC (a pre-condition for a positive response to subsequent items) and any two other forms of engagement, such as speaking with an OMC representative or attending a video presentation. A cut-point of exposure to at least one intervention activity had been proposed prior to study implementation; individuals who met this threshold were most likely to report seeing an OMC mural as their form of engagement. We chose a higher threshold of two forms of engagement to ensure that at least one would be an active undertaking with greater potential to influence health behaviors such as alcohol use that comprise indirect pathways to HIV risk.

We assess three alcohol-related outcomes that could plausibly be affected by the OMC intervention. Overall alcohol use was measured on the baseline and follow-up survey using the Alcohol Use Disorders Identification Test (AUDIT), a validated 10-item screening tool for harmful and hazardous alcohol use (120,121). The AUDIT enables identification of heavy drinkers (individuals with scores of four or above out of a possible twelve on questions related to frequency and intensity of alcohol consumption) (122–124,126) as well as individuals at risk of alcohol use disorder (potential problem drinkers), defined as those with a score of eight or above out of the 40 possible points on the full AUDIT (125,156). AUDIT questions on alcohol consumption were asked of individuals' current typical usage; questions on dependence and harms used a timeframe of the past year. Alcohol use in the context of sex was measured as part of a series of questions on each individual's three most recent sexual partners. For each partner, individuals were asked whether they or their partner had used any amount of alcohol just before their most recent sexual encounter (157). For this outcome, we restricted analysis to men with at least one reported sexual encounter in the three months prior to the interview to reduce recall errors and to maximize the time period between intervention start and outcome measurement. Each of these outcomes represents specific high-risk alcohol-related behaviors that could be modified by the OMC intervention.

In addition to the AUDIT scale, survey responses pre- and post-intervention also included individual demographic data that we aggregated by village, namely age, educational attainment, marital status, recent income, and experiencing food insecurity. Village-level covariates were obtained from two additional data sources. Agincourt HDSS census data were used to determine percent of village residents who were temporary migrants, defined as having spent under six months in the area over the previous year, percent of residents who were employed, and percent of households with a female head. We employed percent employed and percent of female-headed

household to capture village poverty level, as female-headed households are more likely to be poor (87). Secondly, a pre-randomization community mapping exercise led by key informants in each village provided information on infrastructure in each village, including alcohol outlets (bottle stores and shebeens). We calculated alcohol outlet density per kilometer squared (km^2) as a metric of alcohol availability by village.

Causal model

The questions of interest are whether the intervention affected alcohol use outcomes per protocol and whether there is evidence of spillover, whereby increasing engagement of individuals in a village achieved a catalytic (nonlinear) effect in reducing alcohol use. This section reviews causal models that enable translation of these questions into quantities that can be estimated from the observed data.

Notation

Variables measured at the village level are indexed *j* for j=1, 2, … n_j=22 villages; individual-level variables are indexed *ij* for the n_i individuals in village *j*. We defined the following variables:

- W_j: baseline covariates for village *j*, such as percent of female-headed households and density of alcohol outlets
- W_{ij}: covariates measured at follow-up for individual *i* in village *j*, such as age
- A_j: indicator that village *j* was assigned to intervention
- Z_{ij}: indicator that an individual male respondent engaged in the intervention
- $Y1_{ij}$: indicator that individual *i* in village *j* reported heavy drinking
- $Y2_{ij}$: indicator that individual *i* in village *j* screened positive for potential problem drinking
- $Y3_{ij}$: indicator that individual *i* in village *j* reported alcohol use before sex with at least one sexual partner in the three months prior to the survey

Vectors of individual variables are denoted through bold text. The mean of individual observations within a group *j* is indicated with an overbar, such that the proportion of men with the outcome in question (such as a sexually active man reporting pre-sex alcohol use, $Y3_{ij} = 1$) is $\overline{Y3}_{.j}$. We defined exogenous variables $U = \left(U_j : j = \{1, \ldots 22\}\right) \sim P_U$, with $U_j = \left(U_A, U_{W_j}, U_{W_{ij}}, U_Z, U_Y\right)$ encompassing the unmeasured factors that contributed to each measured covariate. U_A was assumed independent from the other unmeasured variables due to randomization of treatment; we consider implications of dependence among other *U* in the discussion. For each outcome, we assumed the following causal relationships among the variables defined:

For j ={1, …, 22}, $W_j = f_{W_j}\left(U_{W_j}\right)$; $A_j = f_A(U_A)$

For i ={1, …, n_i}, $W_{ij} = f_{W_{ij}}\left(W_j, U_{W_{ij}}\right)$; $Z_{ij} = f_X\left(A_j, W_j, W_{ij}, U_Z\right)$;

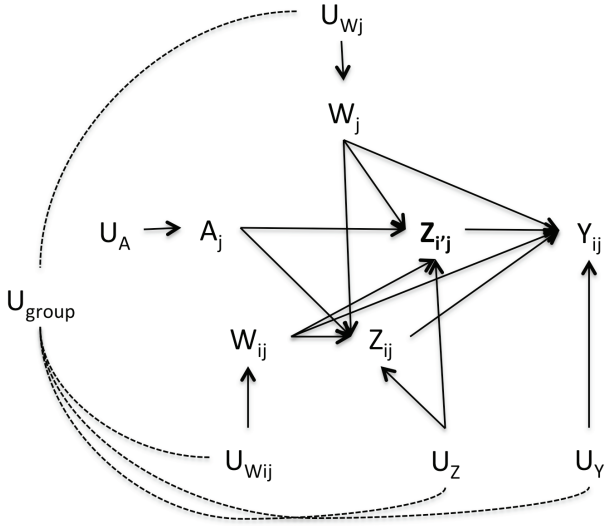
$Y_{ij} = f_Y\left(A_j, W_j, W_{ij}, Z_{ij}, U_Y\right)$

Figure 4.1 presents two related causal models of the relationships among these variables; arrows indicate causal relationships while dashed curved connectors represent correlation potentially explained by unmeasured common causes. Panel A is a multilevel model including

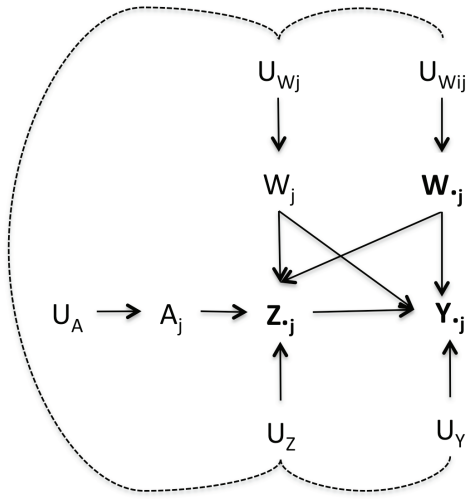
nodes at the village level, such as A_j, as well as individual-level nodes such as Y_{ij}. The individual-level nodes are dependent within groups (shown here through common cause U_{group}), meaning this model cannot be manipulated like a single-level directed acyclic graph to inform analytic choices. However, this figure is helpful in directly illustrating two hypothesized paths from intervention to outcome: via individual engagement Z_{ij} and via engagement by other village residents $\mathbf{Z_{i'j}}$, representing the intended spillover effect of the intervention. We limited the causal relationship between intervention and alcohol outcomes to these two paths. However, the spillover effect violates the assumption that a given individual's outcome is not affected by another individual's exposure status, known as the stable unit treatment value assumption (SUTVA). This assumption is required to cleanly identify individual-level effects, whether that be the total effect or the proportion of effect mediated through individual engagement (158).

Figure 4.1: Causal models of intervention effects on engagement and alcohol outcomes

A: Multilevel level



B: Village level



An alternative means of addressing the questions of interest is to summarize individual variables to the village level, as shown in panel B. Defining the village as the unit of interest results in plausibly independent and identically distributed (i.i.d.) observations, albeit a small number of them. Note that in this depiction, the entirety of the intervention effect on village-level outcomes is mediated by village-level engagement $\mathbf{Z_{\cdot j}}$ (i.e., as in panel A, we excluded the possibility of a direct effect of A_j on Y). Summarizing engagement to the village level combines the effects mediated by direct individual engagement with effects due to peer engagement; it permits assessment of the nonlinearity of this relationship as a test of potential spillover effects. We followed this village-level analytic strategy as an initial investigation. Let $\overline{Y}_{\cdot j}(a_j)$ denote the counterfactual value that $\overline{Y}_{\cdot j}$ would take setting A_j to a_j. We assessed the total effect of the intervention on each alcohol outcome by pursuing the target parameter $E\left[\overline{Y}_{\cdot j}(1) - \overline{Y}_{\cdot j}(0)\right]$, the expected change in village prevalence of the alcohol outcome among men (limited to sexually active men for pre-sex alcohol use) when setting exposure to receiving the intervention for all villages versus not receiving it for all villages. Under the causal model in Figure 4.1B, this causal parameter can be estimated through the statistical parameter $E\left[\left(\overline{Y}_{\cdot j} | A_j = 1\right) - \left(\overline{Y}_{\cdot j} | A_j = 0\right)\right]$.

We assessed the presence of a nonlinear synergistic effect of increasing engagement with the OMC campaign by summarizing $\mathbf{Z_{.j}}$ into proportion of individuals engaged and using two thresholds of interest - the level suggested for critical mass of a community intervention (15%) and the target identified for this specific intervention (40%) – to categorize this proportion into low ($\overline{Z}_{.j} \leq 0.15$), moderate ($0.15 < \bar{Z}_{.j} \leq 0.40$), and high ($\overline{Z}_{.j} > 0.40$) engagement. We created indicators for each category of intervention dose, denoted X_j^0, X_j^{15}, and X_j^{40}.

We estimated the expected prevalence of each outcome $\overline{Y}_{.j}$ if all villages were set to each category of intervention dose and plotted these estimates to depict the shape of the relationship. Lastly, we considered the relative location of each estimate as an indicator of dose-response between increasing levels of OMC engagement and alcohol outcomes among young men. We defined the target parameters $E\left[E\left(\bar{Y}_{.j}\left(X_j^{15}=1\right)\right)-E\left(\bar{Y}_{.j}\left(X_j^{0}=1\right)\right)\right]$ and $E\left[E\left(\bar{Y}_{.j}\left(X_j^{40}=1\right)\right)-E\left(\bar{Y}_{.j}\left(X_j^{15}=1\right)\right)\right]$ to quantify the difference in outcome prevalence setting all villages to moderate versus low engagement and high versus moderate engagement respectively.

Analysis

We conducted descriptive analyses of village characteristics pre-randomization and of the demographic characteristics and alcohol outcomes of the follow-up survey respondents. To estimate the target parameters of interest, we employed targeted maximum likelihood estimation (TMLE), a semi-parametric estimation approach that employs models for both exposure and outcome to reduce bias (128). TMLE is implemented by estimating the outcome conditional on exposure and covariates (Q_Y) and then drawing on the treatment mechanism (G_A) to update[8] the fit of the initial Q_Y to minimize the bias of the specific parameter of interest (159). TMLE offers several advantages in this context: it is robust to mis-specification of the model for exposure or outcome, it provides theoretically justified variance estimation even after model-fitting procedures by relying on the asymptotic theory of the efficient influence curve, and it is highly efficient, a particular benefit given the small sample size. In this analysis, TMLE is guaranteed to be consistent in the total effect estimation because G_A is known: each village had a fixed 50% probability of being randomized to intervention. As a substitution estimator, TMLE respects the bounds of the original parameter. The use of TMLE enabled us to perform model selection without over-fitting or understating the resulting variance.

Given the small sample size and hence limited capacity to adjust for confounders, we undertook data-adaptive selection of covariates for adjustment from a list of candidates considered to be confounders based on the causal model. For total effect estimation, we tested whether inclusion of a covariate associated with the outcome could render estimation more efficient by performing leave-one-out cross-validation (160). Specifically, we removed one village at a time from the dataset and used the remaining n_j-1 (21) villages to fit linear regression models of each outcome on intervention initially unadjusted and then combined with each candidate covariate in turn (alcohol outlet density per village; the proportion of houses in the village headed by women; the proportions of village residents who are migrants, employed, or born outside South Africa; the proportions of baseline respondents who are high school graduates, married, or earning income; and the proportion of male baseline respondents classified

[8] In this case, the update will be 0 when estimation of total effects is not adjusted for covariates. G_A is known for all total effect estimates, and in the absence of covariates, Q_Y is a linear regression of Y on the exposure.

as heavy drinkers and potential problem drinkers). We predicted the outcomes on the village held out from this initial fit and calculated the mean squared error (MSE) comparing the true outcome to the prediction in this validation sample. Final TMLE estimation employed the adjustment approach with the lowest average MSE across 22 validation sets. We defined the treatment mechanism as P(A=0.5) due to randomization and did not estimate G_A.

For the dose-response analysis, we repeated the leave-one-out cross validation using a logistic regression model with each dose indicator (X_j^0, X_j^{15}, X_j^{40}) as the outcome in turn to select the best covariate for control in each treatment model. We assessed each dose indicator alone and with each candidate covariate to identify the most efficient (optimal bias-variance tradeoff) outcome models for heavy drinking, potential problem drinking, and pre-sex alcohol use. The identified covariates were included in the treatment mechanism and outcome model in TMLE. (Note that due to our hypothesized causal structure, intervention status is an instrumental variable rather than a confounder of the engagement – alcohol use relationship, and hence is not included in this portion of the analysis.) We predicted the estimated prevalence of each outcome setting all villages to low, moderate, and high dose in turn and plotted the three estimates together to assess linearity of the relationship between increasing intervention engagement and prevalence of each alcohol outcome. We calculated the ATE for moderate versus low engagement and high versus moderate engagement by subtracting the estimated prevalence at the lower threshold from the next higher. For all estimates, we calculated 95% confidence intervals (CI) using the influence curve-based standard error and a *t* distribution with 21 degrees of freedom.

As a means of assessing sensitivity to measurement error, we repeated total effect and dose-response analyses of alcohol use prior to sex excluding men who reported pre-sex drinking but no lifetime alcohol use. All analyses were conducted in R 3.1.3 (R Foundation for Statistical Computing) using the longitudinal TMLE (ltmle) package 0.9-5 (161).

RESULTS

At baseline, 2,252 households were sampled for participation; 1,826 (81.1%) were visited before the sample size per village was attained. Sixty-nine percent (n=1,256) of households contacted included an eligible resident; 1181 of those eligible consented to enroll into the study (94.0%), 600 women and 581 men. Baseline characteristics of these young adults and of the villages as a whole indicate balanced covariates in the control and intervention arms, although alcohol consumption, particularly heavy drinking, was more common among men in control villages than intervention villages (Table 4.1).

Table 4.1: Characteristics of villages at baseline, by randomization group

	Control (N=11)	Intervention (N=11)
	Mean ± SD	Mean ± SD
Percent female-headed households^	41.2 ± 3.7	42.0 ± 2.7
Percent migrants^	18.3 ± 2.9	17.2 ± 2.3
Alcohol outlet density (per km^2)^	1.59 ± 0.85	1.14 ± 0.73
Respondent age	24.4 ± 5.4	24.1 ± 5.3
Percent respondents completing secondary school	31.8 ± 8.5	34.7 ± 10.9
Percent respondents currently married	23.9 ± 8.4	21.1 ± 6.7
Percent respondents reporting recent income	32.9 ± 9.1	36.5 ± 6.5
Percent respondents reporting ever consuming alcohol	36.5 ± 7.1	32.9 ± 7.1
Percent male respondents classified as heavy drinkers	39.3 ± 12.9	30.3 ± 11.1
Percent male respondents classified as potential problem drinkers	26.8 ± 11.6	21.8 ± 9.9

[^] Measures obtained from Agincourt Health and Socio-demographic Surveillance System (DHSS) census prior to study initiation

[v] Calculated based on pre-randomization community mapping exercise to identify all alcohol outlets, including taverns, unlicensed taverns and bottle stores, which often function as gathering places.

The community survey following intervention was based on an initial sampling frame of 2,384 individuals, of whom 1,167 were ineligible (due to eligibility criteria such as age or the enrollment of a co-resident in the survey), 38 declined to participate, and 4 were deceased. A total of 1,175 young adults (96.5% of the eligible sample) enrolled in the survey: 575 men and 600 women. Respondents indicated high levels of poverty, with the majority earning no income within three months and nearly 10% of women reporting food insecurity in the past month (Table 4.2). Engagement in the OMC intervention was high, with 36.2% of men and 22.2% of women reporting knowledge of the intervention and exposure to at least two other intervention elements. As intended, proportion of men engaging in the intervention was significantly higher in intervention villages, ranging from 36.0% to 81.0%, compared to 0.0% to 20.0% in control villages (Figure 4.2). Nine villages were categorized as low dose, three as moderate, and 10 as high.

Table 4.2: Characteristics of respondents in follow-up survey

	Men (N=575) N (Col. %)	Women (N=600) N (Col. %)
Age		
18-20	213 (37.0)	114 (19.0)
21-25	178 (31.0)	173 (28.8)
26-30	113 (20.0)	153 (25.5)
31-35	71 (12.4)	160 (26.7)
Education		
Primary or less	74 (12.9)	110 (18.3)
Some secondary	314 (54.6)	285 (47.5)
Completed secondary or above	187 (32.5)	205 (34.2)
Marital status		
Never married	507 (88.2)	377 (62.8)
Married (legal or traditional)	58 (10.1)	194 (32.3)
Separated, divorced or widowed	10 (1.7)	29 (4.8)
Born outside South Africa	16 (2.8)	23 (3.8)
Earned no income within three months	429 (74.6)	397 (66.2)
Experienced food insecurity within 30 days	34 (5.9)	58 (9.7)
Engaged in OMC intervention	208 (36.2)	133 (22.2)
Reported ever consuming alcohol	288 (50.1)	63 (10.5)
Heavy drinker	141 (24.5)	17 (2.8)
Potential problem drinker	92 (16.0)	13 (2.2)
Reported any sexual partner in past 3 months	388 (67.5)	430 (71.7)
Reported consuming alcohol prior to sex in past 3 months (N=388 / N=430)	66 (17.0)	62 (14.4)

Figure 4.2: Men's engagement in One Man Can by randomized intervention status (N=22 villages)

As expected, men were far more likely than women to report ever drinking alcohol (50% versus 10%) as well as heavy drinking and potential problem drinking (24.5% versus 2.8% and 16.0% versus 2.2% respectively). Of the 818 individuals reporting at least one sexual relationship in the three months before the follow-up survey, 128 (15.6%) reported consuming alcohol prior to most recent sex. In contrast to other alcohol outcomes, men and women were equally likely to report pre-sex alcohol use. However, a majority of women (48, 77.4%) reporting that they drank prior to most recent sex provided a conflicting response of no lifetime alcohol use on the AUDIT questionnaire. Ten of 66 men (15.2%) similarly provided discrepant responses. The sizable proportion of women providing conflicting responses on lifetime drinking versus pre-sex alcohol use and the low prevalence of other drinking outcomes in women reinforce our decision to focus on men's outcomes in the outcome analyses.

The optimal strategies for estimation of the total effect of the intervention on heavy drinking, potential problem drinking, and alcohol use before sex (main and sensitivity analyses) selected through the cross-validation analysis were: unadjusted, adjusted for baseline prevalence of potential problem drinking, and adjusted for alcohol outlet density respectively. Post-intervention outcome prevalence and the results of total effect estimation are shown in Table 4.3. Heavy drinking and potential problem drinking were more common in intervention villages, though also more variable. The prevalence of alcohol use before sex was equivalent across intervention and control villages, with a high degree of variability. Limiting this outcome to men reporting lifetime alcohol use decreased the prevalence by two to three percent.

Both AUDIT-based metrics were positively but not statistically significantly associated with the intervention, with an average of 9.2% greater prevalence of heavy drinking (95% CI -0.5%, 18.9%) and 3.3% greater prevalence of potential problem drinking (95% CI -3.6%, 10.3%) among men if all villages had been assigned to the intervention versus all villages set to the control condition. Alcohol use before sex showed the more expected negative association with intervention assignment (-5.5%), although this association also did not reach statistical significance. Results did not differ in the sensitivity analysis. As a whole, the total effect analyses did not reveal any statistically significant impacts of the intervention on alcohol outcomes at the village level.

Table 4.3: Difference in prevalence of alcohol outcomes associated with OMC intervention

	Outcome prevalence, unadjusted		Average treatment effect
	Intervention villages (N=11) Mean ± SD	Control villages (N=11) Mean ± SD	(95% CI)
Heavy drinking	0.289 ± 0.140	0.197 ± 0.075	0.092 (-0.005, 0.189)
Potential problem drinking	0.184 ± 0.107	0.135 ± 0.065	0.033 (-0.036, 0.103)
Alcohol use before sex [a]	0.168 ± 0.127	0.187 ± 0.154	-0.055 (-0.162, 0.052)
Alcohol use before sex, sensitivity analysis [b]	0.149 ± 0.114	0.158 ± 0.157	-0.051 (-0.150, 0.049)

[a] Among those sexually active in past three months

[b] Among those sexually active in past three months reporting lifetime alcohol use

The baseline covariates that minimized MSE in estimating exposure models for each level of intervention dose were prevalence of potential problem drinking for low dose, proportion of employed respondents for moderate dose, and proportion of respondents earning income for high dose. Baseline adjustment variables for the outcome models matched those in the total effects analysis with several exceptions: heavy drinking with low dose (prevalence of potential problem drinking), heavy drinking and alcohol use before sex with moderate dose (proportion female-headed households) and potential problem drinking with moderate dose (unadjusted).

The relationships between intervention dose and estimated prevalence of each outcome are shown in Figures 4.3 and 4.4. Increasing engagement in the OMC intervention showed a slightly nonlinear but harmful association with both AUDIT-based measures, contrary to expectation (Figure 4.3). The estimated prevalence of heavy drinking was 19.4% at low engagement, 15.9% at moderate engagement, and 30.3% at high engagement. As shown in Table 4.4, the contrast between high and moderate engagement was positive and statistically significant at 14.4% (95% CI 4.7%, 24.2%). Potential problem drinking followed the same pattern, with a decrease from an estimated prevalence of 13.6% at the lowest dose to 9.0% at the moderate dose but a peak of 18.4% at high engagement. This difference of 9.4% across the 40% threshold was statistically significant (95% CI 1.6%, 17.2%). The nonlinear relationship between intervention engagement and each of these outcomes could indicate a threshold effect or a spillover effect such that high engagement is associated with considerably greater prevalence of harmful alcohol use than low and moderate engagement.

Figure 4.3: Dose-response relationship of intervention engagement with heavy drinking and potential problem drinking

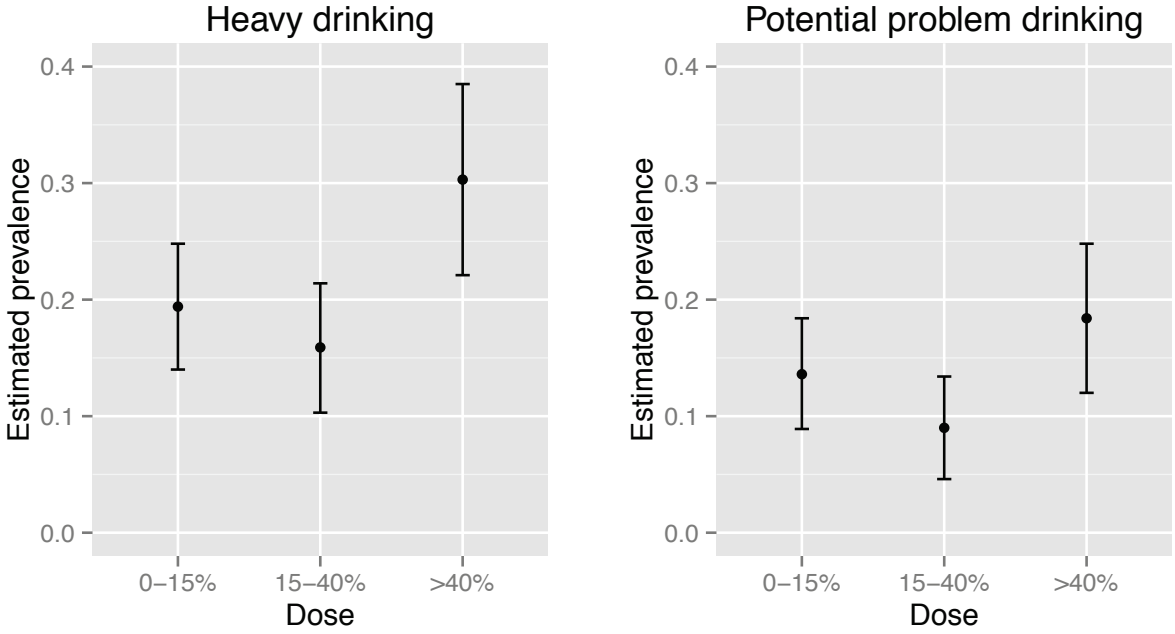
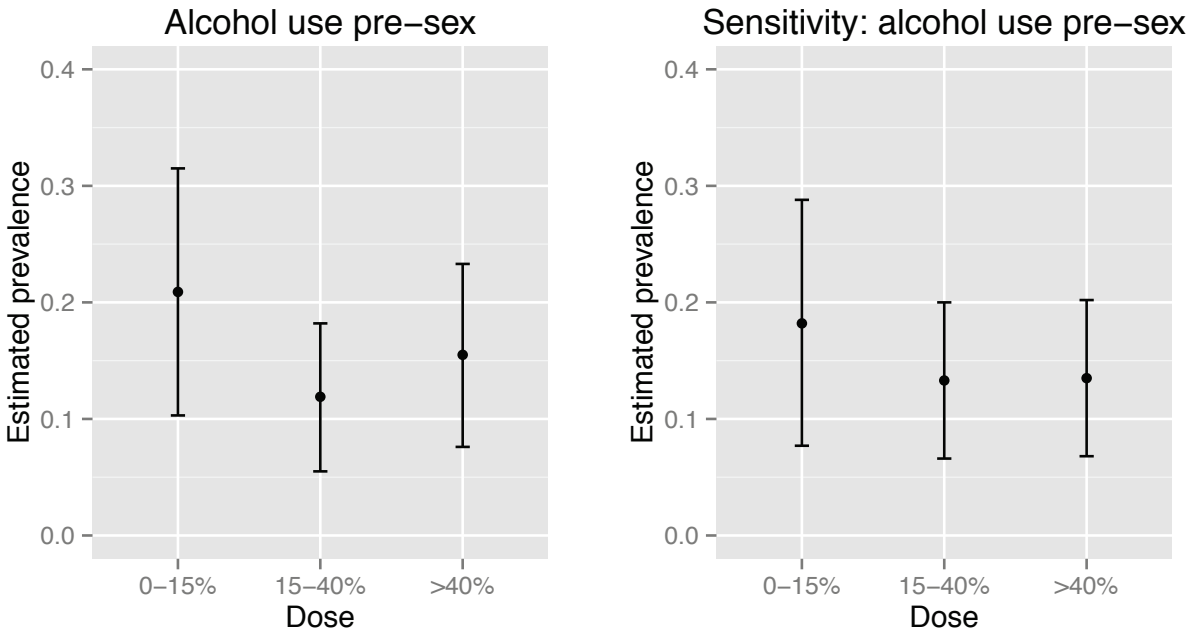


Table 4.4: Difference in prevalence of alcohol outcomes across thresholds of intervention engagement

	Moderate versus low engagement ATE (95% CI)	High versus moderate engagement ATE (95% CI)
Heavy drinking	-0.035 (-0.112, 0.042)	0.144 (0.047, 0.242)
Potential problem drinking	-0.046 (-0.111, 0.019)	0.094 (0.016, 0.172)
Alcohol use before sex	-0.090 (-0.209, 0.028)	0.036 (-0.060, 0.132)
Alcohol use before sex, sensitivity analysis	-0.049 (-0.164, 0.065)	0.003 (-0.085, 0.090)

In keeping with the total effect results, degree of engagement did not show a statistically significant association with alcohol use before sex. The ATE for moderate versus low engagement was -9.0% (estimated prevalence of 11.9% versus 20.9%), although this difference was not significant (Table 4.4, row 3). This protective effect did not obtain when contrasting high engagement (15.5% estimated prevalence of alcohol use before sex) against moderate (ATE 3.6%, 95% CI -6.0%, 13.2%). The sensitivity analysis followed the same pattern. These results are suggestive of a nonlinear relationship in that a protective association diminished at higher levels of engagement, but the confidence intervals are too wide to draw firm conclusions on the existence or shape of this relationship.

Figure 4.4: Dose-response relationship of intervention engagement with alcohol use before sex



DISCUSSION

This population-based study of a randomized community mobilization intervention identified no protective effects of the intervention as well as evidence of a potentially harmful association between intervention uptake and alcohol consumption at the village level. In this analysis, the OMC intervention itself and having over 15% of the target audience of young men engaged with the intervention were negatively but not statistically significantly associated with the proportion of men reporting alcohol use prior to recent sex. Moderate (15% - 40%) engagement was associated with the lowest prevalence of all outcomes, although the small number of villages observed in this category limit inference from this result. However, high collective engagement in the intervention was associated with higher prevalence of men's heavy drinking and potential problem drinking. There are a number of possible explanations for this finding. It could represent a truly harmful effect of the intervention, potentially mediated by OMC activities occurring in alcohol-serving establishments as well as including large gatherings such as soccer tournaments that provide a milieu for greater alcohol consumption. Reverse causation could also contribute to the observed association, such that individuals who consume more alcohol were drawn to the social activities within the OMC campaign, increasing intervention engagement in villages with more drinkers. A third possibility is a change less in behavior than in perception, that reporting of alcohol consumption might differ among those exposed to the intervention and more aware of harmful repercussions of alcohol use. This hypothesis is perhaps more plausible for potential problem drinking, which includes symptoms of alcohol dependence such as hearing from a friend or relative that one's drinking is a cause for concern. Increased peer awareness of dangerous alcohol use could in particular help explain the sharp increase in potential problem drinking at the highest threshold of intervention engagement. Finally, the results may be due to random or systematic error. Avenues of further research to disentangle these competing explanations are discussed subsequently.

Inferring causality from these results require several conditions to hold: temporality, ignorability, SUTVA, and positivity (158,162–165). We consider the plausibility of these assumptions separately for the total effect and dose-response analyses. Temporality requires exposure to precede the outcome; we limited reports of alcohol use before sex to the three months prior to survey date to ensure temporal ordering of the OMC campaign and this outcome. However, the AUDIT-based measures are less clearly indexed in time: the heavy drinking metric assesses typical drinking while potential problem drinking includes both typical consumption and alcohol-related consequences in the past 12 months, making analyses of these outcomes more vulnerable to bias due to reverse causation. Ignorability, or the absence of unmeasured confounders, is reasonably likely to be met in the total effects analysis given that exposure is randomized. Randomization is the gold standard method of minimizing confounding, although the small sample size precludes complete confidence in the balance of unmeasured covariates between study arms. Although we limit our analyses to the village level to negate individual-level interference and maintain SUTVA, spillover between villages did occur: residents from control villages were able to and did participate in OMC activities in small numbers. This contamination could bias total effect estimates towards the null by rendering control and intervention villages more similar. Finally, positivity is the condition that each level of treatment is possible for any combination of observed covariates; non-positivity requires truncating data or extrapolating over areas that are not supported in the observed data (166). The randomized intervention fulfills the requirement of positivity because the probability of each village being randomized to either condition is equal (50%) regardless of baseline covariates.

Our analysis of the dose-response relationship between intervention engagement and alcohol outcomes requires stronger assumptions since the exposure is observed rather than randomized. The considerations noted above for temporality pertain in this analysis as well. Ignorability is a serious concern, as we were forced to limit adjustment for measured confounders due to our small sample size in order to avoid overfitting; it is possible that important confounders affecting both the level of engagement and the prevalence of alcohol outcomes remain unmeasured and/or uncontrolled. The major threat to positivity in the dose-response analysis would be intervention assignment itself, given the strong relationship between intervention and proportion of engagement. However, because we assert no direct effect of intervention assignment on alcohol outcomes at the village level unmediated by engagement, intervention is not a confounder and does not require analytic control. This assertion can be called into question, particularly as our implementation of the causal model relies on the proportion of respondents engaged as a single summary of the complex vector of individual engagement. It is possible that intervention effects are mediated by engagement in ways not captured by the proportion, such as through engagement of specific highly influential individuals. To the extent this is the case, intervention assignment may operate as a confounder. All other measured covariates show relatively even distributions across each dose threshold, as might be expected given the baseline balance following randomization. Other summary measures could be considered in the future to relax the assumption that proportion engaged captures the entirety of the mediated effect of intervention on alcohol outcomes.

This consideration of the required assumptions illuminates many of the strengths and limitations of this research. These findings represent a novel test of the hypothesis of community-wide effects resulting from a cluster-randomized intervention intended to transform gender norms. Community-randomized trials offer a rare opportunity to overcome many of the analytic challenges of social epidemiologic research (163); this trial attempts to address critical

structural risk factors fueling the ongoing HIV epidemic. Beyond study design, uptake of the intervention was strong in practice. We drew on population-representative surveys pre- and post-intervention to assess impact and employed reliable measures of alcohol consumption. The analytic approaches helped to maximize efficiency and robust inference despite the small sample size.

Beyond the challenges noted in meeting the required assumptions, particularly for the dose-response analysis, the major limitations of this work are the small sample size and the use of a repeated panel survey rather than a longitudinal cohort. However, the high levels of migration in this setting rendered a cohort both impractical and a less reliable representation of the individuals present for the period of intervention activities. The results are also subject to measurement error, as the AUDIT is a screening tool rather than a diagnostic instrument or an intensive alcohol measurement tool such as a diary of recent use (167). Indeed, the discrepancies between lifetime alcohol use and alcohol use at last sex signal problematic measurement in this case, potentially due to respondents' misunderstanding or misreporting one or both quantities. The sensitivity analysis suggested in this case that results were robust to removal of men with conflicting responses. Although not the focus of this analysis, the high proportion of women reporting consuming alcohol before recent sex appears anomalous in the context of this research and other data on alcohol use in this setting. Further investigation on the wording and interpretation of this question is warranted. As noted above, there is a possibility of systematic measurement error in these results if the OMC campaign raised awareness of harmful alcohol use in intervention villages and led to these respondents either overstating or more accurately reporting alcohol use in contrast to underreporting in control villages.

Several extensions to this research would help to illuminate and extend the findings. The potential for a harmful effect of intervention activities on alcohol consumption can be investigated using monitoring data of the intervention to assess whether a higher frequency of activities in alcohol-serving establishments is associated with drinking outcomes. Monitoring data also provide a useful means of triangulating the measurement of reported dose in intervention villages with delivered dose as measured by the intervention staff. A second and more substantive extension is to undertake an individual-level analysis with explicit consideration of effects mediated through individual engagement with OMC compared to those transmitted through engaged peers. Existing research details the assumptions required to undertake such an analysis and outlines one possible analytic strategy to pursue (158,168). The study described here should provide a relatively sound foundation for this complex analysis, with its randomized design and careful dose measurement. In addition, the use of semi-parametric estimation methods for mediation, including multiple mediators, is an area of active research (159,169–171). These tools have not yet been applied to spillover-mediated effect, but they could extend the analytic options in this nascent area. Drawing on individual-level data would strengthen this analysis by improving the power to detect effects, potentially rendering the association with alcohol use before sex significant should the result hold on the individual level. It would also enable greater control of confounding and examination of a more finely grained assessment of the relationship between intervention engagement and increased alcohol consumption observed at the village level, such as whether specific types of participation in OMC may explain this link.

The findings presented here represent an important contribution to the growing literature on gender norms, alcohol use, and HIV risk in South Africa. Based on this village-level analysis, the OMC intervention did not show a significant protective effect on alcohol use before sex,

although the limited power of the analysis should be considered when inferring from this result. High engagement with the intervention evidenced a harmful relationship with prevalence of heavy drinking and potential problem drinking, a concerning finding that bears further assessment and, if born out at the individual level, suggests modifications to the intervention may be needed in future implementation.

Chapter 5: Conclusion

Identifying the effects of community context on health outcomes is a critical undertaking in social epidemiology, one that to date has largely been confined to research in developed countries. This dissertation focuses on social and structural causes of alcohol use in South Africa, where dangerous levels of drinking among men represent a threat to population health directly as well as indirectly through a possible increase in risk of HIV transmission. In this concluding section, I summarize the major implications of the results presented, identify avenues for subsequent research, and consider broader future directions for this area of study.

Statistical methods in multilevel modeling and latent variable modeling have advanced rapidly over the past two decades, with ever greater synthesis between these sub-fields (172,173). As many others have noted, it is critical that enthusiasm over advances in technical capacity to implement complex multilevel models does not subsume the careful development of causal frameworks for the questions of interest. The importance of clear modeling has been a theme in epidemiologic journals coincident with the increasing availability of advanced analytic tools, with particular emphasis on the refinement of causal graphs and counterfactual reasoning to facilitate the expression of research questions (174,175). To date, however, consideration of causal models in epidemiology has not extended to a thorough examination of measurement of constructs that are not directly observable. The research presented here underscores the importance of defining the causal relationships between the latent construct and observed measures and of considering any covariates that could affect this relationship. Proposing such a model is a necessary first step to identifying an appropriate analytic tool.

Our results suggest that, except if the researcher advances a formative causal model with high sampling fraction, latent variable methods such as SEM should be considered in addition to the more familiar aggregation and regression approaches. The combination of plausible values from an item response model with a multilevel SEM offered considerable bias correction across a range of simulation settings testing robustness. Although less variable than the latent variable approaches, mean and regression analyses consistently underestimated contextual effects. Further research, such as comparison of a full SEM and a partially latent plausible value model, might help to identify a more clear choice in optimizing the bias-variance tradeoff. Extensions to the exploration of analytic adjustment for third variables presented here are also needed: the results suggest analyses employing mean and regression approaches could be considerably biased by measurement distortion, particularly if the aggregate value of the distorting factor is not included in the outcome model. A range of realistic levels of distortion should be tested to investigate the relevance of this finding to applied research.

The applied work presented in chapters 3 and 4 breaks new ground in investigating contextual factors associated with alcohol use in rural South Africa, including factors that can be modified by interventions such as the community mobilization trial described in chapter 4. The baseline analysis in chapter 3 provides evidence of an association between contextual factors and alcohol use. Although preliminary, the results linking social cohesion to heavy drinking and alcohol outlet density to potential problem drinking suggest scope for interventions to reduce dangerous alcohol use through modifying community norms as well as by strengthening regulatory policies on alcohol availability. These results could be strengthened by application of the latent variable methods tested in chapter 2. Village-level analysis of the follow-up data collected after the implementation of the community mobilization intervention provided somewhat troubling results, as the intervention did not demonstrate a protective effect on alcohol

use before sex (although power to detect such an effect was low). The intervention was related to higher prevalence of heavy and potential problem drinking as proportion of men engaging with the intervention increased. Understanding the results from the two studies in concert requires combining the strengths of each study: the individual-level analysis of the baseline study with the longitudinal follow-up and randomized intervention of the second. Future research will consider whether social cohesion increased as a result of the mobilization intervention and whether any changes were associated with drinking outcomes, a rare opportunity to study modification to community social factors over time. A second priority area of investigation is the mechanism linking intervention to alcohol outcomes, including individual-level mediation analysis of intervention engagement and alcohol outcomes, as well as consideration of changes in individual norms and perceptions as they relate to drinking behaviors. The research completed to date justifies continued investigation into contextual effects of social and structural factors on alcohol use in this context. Further work will be critical to substantiating these findings to inform any future iterations of the mobilization intervention as well as novel interventions at the community or regional and national policy levels.

Alcohol use in sub-Saharan Africa, and South Africa in particular, is emerging as a potentially critical catalyst for the HIV epidemic. Research in this setting on alcohol use in particular and health outcomes in general has generally focused on individual-level determinants. As researchers and policy makers seek to identify effective tools to sustain progress made in curbing HIV over the decades to come, more broadly focused interventions are gaining in importance, including community-based efforts. Taken together, these factors indicate that it is a critically important time to determine the contributions of social and structural factors to alcohol consumption. This dissertation provides the methodological tools and conceptual foundation to serve as a basis for advancing this vital work.

References

1. UNAIDS: Joint UN Program on HIV/AIDS. UNAIDS Report on the Global AIDS Epidemic 2012 [Internet]. Geneva, Switzerland: UNAIDS; 2012. Available from: http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_UNAIDS_Global_Report_2012_with_annexes_en.pdf

2. UNAIDS: Joint UN Program on HIV/AIDS. The GAP Report [Internet]. Geneva, Switzerland: UNAIDS; 2014 Jul. Available from: http://www.unaids.org/sites/default/files/en/media/unaids/contentassets/documents/unaidspublication/2014/UNAIDS_Gap_report_en.pdf

3. Pettifor AE, Hudgens MG, Levandowski BA, Rees HV, Cohen MS. Highly efficient HIV transmission to young women in South Africa. AIDS. 2007 Apr;21(7):861–5.

4. UNAIDS: Joint UN Program on HIV/AIDS. AIDSinfo [Internet]. AIDS Information. [cited 2015 Jul 13]. Available from: http://www.unaids.org/en/dataanalysis/datatools/aidsinfo/

5. Coates TJ, Richter L, Caceres C. Behavioural strategies to reduce HIV transmission: how to make them work better. The Lancet. 2008 Aug 23;372(9639):669–84.

6. Gupta GR, Parkhurst JO, Ogden JA, Aggleton P, Mahal A. Structural approaches to HIV prevention. The Lancet. 2008 Aug 30;372(9640):764–75.

7. World Health Organization. Global status report on alcohol and health 2014 [Internet]. Geneva, Switzerland: World Health Organization; 2014 May [cited 2014 Jun 16]. Available from: http://www.who.int/substance_abuse/publications/global_alcohol_report/en/

8. Parry CDH. South Africa: alcohol today. Addiction. 2005;100(4):426–9.

9. Schneider M, Norman R, Parry C, Bradshaw D, Pluddemann A, South African Comparative Risk Assessment Collaborating Group. Estimating the burden of disease attributable to alcohol use in South Africa in 2000. S Afr Med J. 2007 Sep 19;97(8):664–72.

10. World Health Organization. Global Health Observatory Data Repository [Internet]. World Health Organization. 2014 [cited 2015 May 19]. Available from: http://apps.who.int/gho/data/view.main.52200

11. World Health Organization. Health statistics and information systems: Estimates for 2000-2012 [Internet]. World Health Organization. 2015 [cited 2015 May 21]. Available from: http://www.who.int/healthinfo/global_burden_disease/estimates/en/

12. Pandrea I, Happel KI, Amedee AM, Bagby GJ, Nelson S. Alcohol's Role in HIV Transmission and Disease Progression. Alcohol Res Health. 2010;33(3):203–18.

13. Hahn JA. Unhealthy alcohol use and HIV disease progression in two ART naive Ugandan cohorts [Internet]. Research Society on Alcoholism 38th Annual Scientific Meeting; 2015

Jun 22; San Antonio, Texas. Available from: http://www.rsoa.org/2015meet-2-RSA-DAILY-SCHEDULE.pdf

14. Shuper PA, Joharchi N, Irving H, Rehm J. Alcohol as a Correlate of Unprotected Sexual Behavior Among People Living with HIV/AIDS: Review and Meta-Analysis. AIDS Behav. 2009 Dec 1;13(6):1021–36.

15. Fisher JC, Cook PA, Kapiga SH. Alcohol Use Before Sex and HIV Risk: Situational Characteristics of Protected and Unprotected Encounters Among High-Risk African Women. Sex Transm Dis. 2010 Sep;37(9):571–8.

16. Townsend L, Rosenthal SR, Parry CDH, Zembe Y, Mathews C, Flisher AJ. Associations between alcohol misuse and risks for HIV infection among men who have multiple female sexual partners in Cape Town, South Africa. AIDS Care. 2010 Dec 1;22(12):1544–54.

17. Medley A, Seth P, Pathak S, Howard AA, DeLuca N, Matiko E, et al. Alcohol use and its association with HIV risk behaviors among a cohort of patients attending HIV clinical care in Tanzania, Kenya, and Namibia. AIDS Care. 2014 Oct 3;26(10):1288–97.

18. Pithey A, Parry C. Descriptive systematic review of sub-Saharan African studies on the association between alcohol use and HIV infection. J Soc Asp HIVAIDS Res Alliance. 2009;6(4).

19. Fisher JC, Bang H, Kapiga SH. The Association Between HIV Infection and Alcohol Use: A Systematic Review and Meta-Analysis of African Studies. Sex Transm Dis. 2007 Nov;34(11):856–63.

20. Baliunas D, Rehm J, Irving H, Shuper P. Alcohol consumption and risk of incident human immunodeficiency virus infection: a meta-analysis. Int J Public Health. 2010 Jun 1;55(3):159–66.

21. Zablotska IB, Gray RH, Koenig MA, Serwadda D, Nalugoda F, Kigozi G, et al. Alcohol Use, Intimate Partner Violence, Sexual Coercion and HIV among Women Aged 15–24 in Rakai, Uganda. AIDS Behav. 2009 Apr 1;13(2):225–33.

22. Morojele NK, Ranchod C. Review of Interventions to Reduce Alcohol Use-Related Sexual Risk Behavior in Africa. Afr J Drug Alcohol Stud. 2011;10(2):95–106.

23. Parry CD, Morojele NK, Myers BJ, Kekwaletswe CT, Manda SO, Sorsdahl K, et al. Efficacy of an alcohol-focused intervention for improving adherence to antiretroviral therapy (ART) and HIV treatment outcomes – a randomised controlled trial protocol. BMC Infect Dis. 2014 Sep 12;14(1):500.

24. Morojele NK, Kitleli N, Ngako K, Kekwaletswe CT, Nkosi S, Fritz K, et al. Feasibility and acceptability of a bar-based sexual risk reduction intervention for bar patrons in Tshwane, South Africa. J Soc Asp HIVAIDS Res Alliance. 2014;1–9.

25. Popova S, Giesbrecht N, Bekmuradov D, Patra J. Hours and Days of Sale and Density of Alcohol Outlets: Impacts on Alcohol Consumption and Damage: A Systematic Review. Alcohol Alcohol. 2009 Sep 1;44(5):500–16.

26. Connor JL, Kypri K, Bell ML, Cousins K. Alcohol outlet density, levels of drinking and alcohol-related harm in New Zealand: a national study. J Epidemiol Community Health. 2011 Oct 1;65(10):841–6.

27. Pollack CE, Cubbin C, Ahn D, Winkleby M. Neighbourhood deprivation and alcohol consumption: does the availability of alcohol play a role? Int J Epidemiol. 2005 Aug 1;34(4):772–80.

28. Sampson RJ, Raudenbush SW, Earls F. Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy. Science. 1997 Aug 15;277(5328):918–24.

29. Browning CR, Cagney KA. Neighborhood Structural Disadvantage, Collective Efficacy, and Self-Rated Physical Health in an Urban Setting. J Health Soc Behav. 2002 Dec 1;43(4):383–99.

30. Browning CR, Olinger-Wilbon M. Neighborhood Structure, Social Organization, and Number of Short-Term Sexual Partnerships. J Marriage Fam. 2003;65(3):730–45.

31. Ahern J, Galea S. Collective Efficacy and Major Depression in Urban Neighborhoods. Am J Epidemiol. 2011 Jun 15;173(12):1453–62.

32. Cagney KA, Glass TA, Skarupski KA, Barnes LL, Schwartz BS, Mendes de Leon CF. Neighborhood-Level Cohesion and Disorder: Measurement and Validation in Two Older Adult Urban Populations. J Gerontol B Psychol Sci Soc Sci. 2009 May;64B(3):415–24.

33. Kerrigan D, Witt S, Glass B, Chung S, Ellen J. Perceived Neighborhood Social Cohesion and Condom Use Among Adolescents Vulnerable to HIV/STI. AIDS Behav. 2006 Nov 1;10(6):723–9.

34. Kerrigan D, Telles P, Torres H, Overs C, Castle C. Community Development and HIV/STI-Related Vulnerability Among Female Sex Workers in Rio De Janeiro, Brazil. Health Educ Res. 2008 Feb 1;23(1):137–45.

35. Lippman SA, Chinaglia M, Donini A, Diaz J, Reingold A, Kerrigan D. Findings from Encontros: A Multilevel STI/HIV Intervention to Increase Condom Use, Reduce STI, and Change the Social Environment Among Sex Workers in Brazil. Sex Transm Dis. 2012 Mar;39(3).

36. Pronyk PM, Harpham T, Busza J, Phetla G, Morison LA, Hargreaves JR, et al. Can social capital be intentionally generated? A randomized trial from rural South Africa. Soc Sci Med. 2008 Nov;67(10):1559–70.

37. Pronyk PM, Harpham T, Morison LA, Hargreaves JR, Kim JC, Phetla G, et al. Is social capital associated with HIV risk in rural South Africa? Soc Sci Med. 2008 May;66(9):1999–2010.

38. Carpiano RM. Neighborhood social capital and adult health: An empirical test of a Bourdieu-based model. Health Place. 2007 Sep;13(3):639–55.

39. Burgard SA, Lee-Rife SM. Community Characteristics, Sexual Initiation, and Condom Use among Young Black South Africans. J Health Soc Behav. 2009 Sep 1;50(3):293–309.

40. Cain D, Pitpitan EV, Eaton L, Carey KB, Carey MP, Mehlomakulu V, et al. Collective Efficacy and HIV Prevention in South African Townships. J Community Health. 2013 Oct 1;38(5):885–93.

41. Dworkin SL, Colvin C, Hatcher A, Peacock D. Men's Perceptions of Women's Rights and Changing Gender Relations in South Africa Lessons for Working With Men and Boys in HIV and Antiviolence Programs. Gend Soc. 2012 Feb 1;26(1):97–120.

42. Gottert AL. Gender norms, masculine gender-role strain, and HIV risk behaviors among men in rural South Africa [Internet]. The University of North Carolina at Chapel Hill; 2014 [cited 2015 May 18]. Available from: http://gradworks.umi.com/36/68/3668474.html

43. Townsend L, Ragnarsson A, Mathews C, Johnston LG, Ekström AM, Thorson A, et al. "Taking Care of Business": Alcohol as Currency in Transactional Sexual Relationships Among Players in Cape Town, South Africa. Qual Health Res. 2011 Jan 1;21(1):41–50.

44. van den Berg W, Hendricks L, Hatcher A, Peacock D, Godana P, Dworkin S. "One Man Can": shifts in fatherhood beliefs and parenting practices following a gender-transformative programme in Eastern Cape, South Africa. Gend Dev. 2013 Mar 1;21(1):111–25.

45. Barker G, Ricardo C, Nascimento M, Olukoya A, Santos C. Questioning gender norms with men to improve health outcomes: Evidence of impact. Glob Public Health. 2010 Sep 1;5(5):539–53.

46. Dworkin SL, Hatcher AM, Colvin C, Peacock D. Impact of a Gender-Transformative HIV and Antiviolence Program on Gender Ideologies and Masculinities in Two Rural, South African Communities. Men Masculinities. 2013 Jun 1;16(2):181–202.

47. Kalichman SC, Simbayi LC, Cloete A, Clayford M, Arnolds W, Mxoli M, et al. Integrated Gender-Based Violence and HIV Risk Reduction Intervention for South African Men: Results of a Quasi-Experimental Field Trial. Prev Sci. 2009 Apr 8;10(3):260–9.

48. Jewkes R, Nduna M, Levin J, Jama N, Dunkle K, Puren A, et al. Impact of Stepping Stones on incidence of HIV and HSV-2 and sexual behaviour in rural South Africa: cluster randomised controlled trial. BMJ. 2008 Aug 7;337:a506.

49. Berkman LF, Glass T. Social integration, social networks, and health. Social epidemiology. New York: Oxford University Press; 2000. p. 137–73.

50. Kegeles SM, Hays RB, Pollack LM, Coates TJ. Mobilizing young gay and bisexual men for HIV prevention: a two-community study. AIDS. 1999 Sep 10;13(13):1753–62.

51. Hays RB, Rebchook GM, Kegeles SM. The Mpowerment Project: Community-Building with Young Gay and Bisexual Men to Prevent HIV. Am J Community Psychol. 2003 Jun 1;31(3-4):301–12.

52. Lippman SA, Maman S, MacPhail C, Twine R, Peacock D, Kahn K, et al. Conceptualizing Community Mobilization for HIV Prevention: Implications for HIV Prevention Programming in the African Context. PLoS ONE. 2013 Oct 11;8(10):e78208.

53. Swendeman D, Basu I, Das S, Jana S, Rotheram-Borus MJ. Empowering sex workers in India to reduce vulnerability to HIV and sexually transmitted diseases. Soc Sci Med. 2009 Oct;69(8):1157–66.

54. Kerrigan D, Moreno L, Rosario S, Gomez B, Jerez H, Barrington C, et al. Environmental–structural interventions to reduce HIV/STI risk among female sex workers in the Dominican Republic. Am J Public Health. 2006;96(1):120–5.

55. Blakely TA, Woodward AJ. Ecological effects in multi-level studies. J Epidemiol Community Health. 2000 May 1;54(5):367–74.

56. Morgenstern H. Ecologic Studies in Epidemiology: Concepts, Principles, and Methods. Annu Rev Public Health. 1995;16(1):61–81.

57. Skrondal A, Rabe-Hesketh S. Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models. Boca Raton, FL: Chapman & Hall / CRC; 2004. 523 p.

58. Raudenbush SW, Sampson RJ. Ecometrics: Toward a Science of Assessing Ecological Settings, With Application to the Systematic Social Observation of Neighborhoods. Sociol Methodol. 1999;29(1):1–41.

59. Bollen K, Lennox R. Conventional wisdom on measurement: A structural equation perspective. Psychol Bull. 1991 Sep;110(2):305–14.

60. Bollen KA, Bauldry S. Three Cs in measurement models: Causal indicators, composite indicators, and covariates. Psychol Methods. 2011 Sep;16(3):265–84.

61. Lüdtke O, Marsh HW, Robitzsch A, Trautwein U, Asparouhov T, Muthén B. The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. Psychol Methods. 2008;13(3):203–29.

62. Rijmen F, Tuerlinckx F, De Boeck P, Kuppens P. A nonlinear mixed model framework for item response theory. Psychol Methods. 2003;8(2):185–205.

63. Wilson M. Constructing Measures: an Item Response Modeling Approach. New York, NY: Taylor & Francis Group; 2005.

64. Wilson M, De Boeck P. Descriptive and explanatory item response models. Explanatory item response models: A generalized linear and nonlinear approach. New York, NY: Springer-Verlag; 2004.

65. Embretson SE. The new rules of measurement. Psychol Assess. 1996;8(4):341–8.

66. Ip EH, Wang YJ, Boeck P de, Meulders M. Locally dependent latent trait model for polytomous responses with application to inventory of hostility. Psychometrika. 2004 Jun;69(2):191–216.

67. Cho S-J, Boeck PD, Embretson S, Rabe-Hesketh S. Additive Multilevel Item Structure Models with Random Residuals: Item Modeling for Explanation and Item Generation. Psychometrika. 2013 Dec 12;79(1):84–104.

68. González J, De Boeck P, Tuerlinckx F. A double-structure structural equation model for three-mode data. Psychol Methods. 2008;13(4):337–53.

69. Marsh HW, Lüdtke O, Robitzsch A, Trautwein U, Asparouhov T, Muthén B, et al. Doubly-Latent Models of School Contextual Effects: Integrating Multilevel and Structural Equation Approaches to Control Measurement and Sampling Error. Multivar Behav Res. 2009 Nov 30;44(6):764–802.

70. Raudenbush S, Bryk AS. Hierarchical Linear Models. Second edition. California: Sage Publications, Inc.; 2002.

71. StataCorp. Stata Structural Equation Modeling Reference Manual. 13th ed. College Station, Texas: Stata Press; 2013.

72. Mujahid MS, Roux AVD, Morenoff JD, Raghunathan T. Assessing the Measurement Properties of Neighborhood Scales: From Psychometrics to Ecometrics. Am J Epidemiol. 2007 Apr 15;165(8):858–67.

73. Schmidt NM, Tchetgen Tchetgen EJ, Ehntholt A, Almeida J, Nguyen QC, Molnar BE, et al. Does Neighborhood Collective Efficacy for Families Change Over Time? The Boston Neighborhood Survey. J Community Psychol. 2014 Jan 1;42(1):61–79.

74. Lüdtke O, Marsh HW, Robitzsch A, Trautwein U. A 2 × 2 taxonomy of multilevel latent contextual models: Accuracy–bias trade-offs in full and partial error correction models. Psychol Methods. 2011;16(4):444–67.

75. Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. Epidemiology. 2006;17(3):276–84.

76. Wilson M, Adams RJ. Rasch models for item bundles. Psychometrika. 1995 Jun 1;60(2):181–98.

77. Masters GN, Wright BD. The Partial Credit Model. In: van der Linden WJ, Hambleton RK, editors. Handbook of Modern Item Response Theory [Internet]. Springer New York; 1997

[cited 2014 Feb 28]. p. 101–21. Available from: http://link.springer.com/chapter/10.1007/978-1-4757-2691-6_6

78. Embretson SE, Reise SP. Measuring Persons: Scoring Examinees with IRT Models. Item Response Theory for Psychologists. 1st ed. New Jersey: L. Erlbaum Associates; 2000. p. 159–86.

79. OECD. PISA Data Analysis Manual. OECD, Switzerland; 2009.

80. Adams RJ, Wilson M, Wu M. Multilevel Item Response Models: An Approach to Errors in Variables Regression. J Educ Behav Stat. 1997 Mar 20;22(1):47–76.

81. Echeverría S, Diez-Roux AV, Shea S, Borrell LN, Jackson S. Associations of neighborhood problems and neighborhood social cohesion with mental health and health behaviors: The Multi-Ethnic Study of Atherosclerosis. Health Place. 2008 Dec;14(4):853–65.

82. Pettifor AE, Lippman SA, Selin A, Peacock D, Gottert AL, Maman S, et al. A cluster randomized-controlled trial of a community mobilization intervention to change gender norms and reduce HIV risk in rural South Africa: Design, implementation and baseline findings. 2015.

83. Brandt S. Data Analysis: Statistical and Computational Methods for Scientists and Engineers. Springer Science & Business Media; 2014. 532 p.

84. Chalmers RP. mirt: A multidimensional Item Response Theory Package for the R Environment. J Stat Softw. 2012 May 24;48(6):1–29.

85. Kiefer T, Robitzsch A, Wu M. Test Analysis Modules [Internet]. 2015. Available from: cran.r-project.org/web/packages/TAM/TAM.pdf

86. Carlin JB, Li N, Greenwood P, Coffey C. Tools for analyzing multiple imputed datasets. Stata J. 2003;3(3):226–44.

87. Collinson MA. Striving against adversity: the dynamics of migration, health and poverty in rural South Africa. Glob Health Action [Internet]. 2010 Jun 3 [cited 2014 Jul 1];3. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2882287/

88. Braeken J, Kuppens P, Boeck PD, Tuerlinckx F. Contextualized Personality Questionnaires: A Case for Copulas in Structural Equation Models for Categorical Data. Multivar Behav Res. 2013 Nov 1;48(6):845–70.

89. Margerison-Zilko C, Cubbin C, Jun J, Marchi K, Fingar K, Braveman P. Beyond the Cross-Sectional: Neighborhood Poverty Histories and Preterm Birth. Am J Public Health. 2015 Apr 16;105(6):1174–80.

90. Fleischer NL, Diez Roux AV. Using directed acyclic graphs to guide analyses of neighbourhood health effects: an introduction. J Epidemiol Community Health. 2008 Sep 1;62(9):842–6.

91. Rehm J, Baliunas D, Borges GLG, Graham K, Irving H, Kehoe T, et al. The relation between different dimensions of alcohol consumption and burden of disease: an overview. Addiction. 2010 May 1;105(5):817–43.

92. Room R, Jernigan D, Carlini Marlatt B, Gureje O, Mäkelä K, Marshall M, et al. Alcohol and the developing world : a public health perspective [Internet]. Helsinki: Finnish Foundation for Alcohol Studies in collaboration with World Health Organization; 2002 [cited 2015 Mar 17]. Available from: http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-57181

93. Ferreira-Borges C, Dias S, Babor T, Esser MB, Parry CDH. Alcohol and public health in Africa: can we prevent alcohol-related harm from increasing? Addiction [Internet]. 2015 May 1 [cited 2015 May 7]; Available from: http://onlinelibrary.wiley.com/doi/10.1111/add.12916/abstract

94. Hahn JA, Woolf-King SE, Muyindike W. Adding Fuel to the Fire: Alcohol's Effect on the HIV Epidemic in Sub-Saharan Africa. Curr HIV/AIDS Rep. 2011 Sep 1;8(3):172–80.

95. Shuper PA, Neuman M, Kanteres F, Baliunas D, Joharchi N, Rehm J. Causal Considerations on Alcohol and HIV/AIDS — A Systematic Review. Alcohol Alcohol. 2010 Mar 1;45(2):159–66.

96. UNAIDS: Joint UN Program on HIV/AIDS. UNAIDS Report on the global AIDS epidemic 2013 [Internet]. Geneva, Switzerland: UNAIDS; 2013 [cited 2014 Jan 13] p. 198. Available from: http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS_Global_Report_2013_en.pdf

97. Woolf-King SE, Steinmaus CM, Reingold AL, Hahn JA. An update on alcohol use and risk of HIV infection in sub-Saharan Africa: Meta-analysis and future research directions. Int J Alcohol Drug Res. 2013 Apr 3;2(1):99–110.

98. Anderson P, Chisholm D, Fuhr DC. Effectiveness and cost-effectiveness of policies and programmes to reduce the harm caused by alcohol. The Lancet. 2009 Jun 27;373(9682):2234–46.

99. Parry CD, Plüddemann A, Steyn K, Bradshaw D, Norman R, Laubscher R. Alcohol Use in South Africa: Findings from the First Demographic and Health Survey (1998). J Stud Alcohol Drugs. 2005 Jan 1;66(1):91.

100. Aguirre-Molina M, Gorman DM. Community-Based Approaches for the Prevention of Alcohol, Tobacco, and Other Drug Use. Annu Rev Public Health. 1996;17(1):337–58.

101. Fritz K, Morojele N, Kalichman S. Alcohol: The Forgotten Drug in HIV/AIDS. Lancet. 2010 Aug 7;376(9739):398–400.

102. Kalichman SC. Social and Structural HIV Prevention in Alcohol-Serving Establishments. Alcohol Res Health. 2010;33(3):184–94.

103. Khumalo-Sakutukwa G, Morin SF, Fritz K, Charlebois ED, van Rooyen H, Chingono A, et al. Project Accept (HPTN 043): A Community-Based Intervention to Reduce HIV Incidence in Populations at Risk for HIV in Sub-Saharan Africa and Thailand. J Acquir Immune Defic Syndr. 2008 Dec;49(4):422–31.

104. Bryden A, Roberts B, Petticrew M, McKee M. A systematic review of the influence of community level social factors on alcohol use. Health Place. 2013 May;21:70–85.

105. Fulkerson JA, Pasch KE, Perry CL, Komro K. Relationships Between Alcohol-related Informal Social Control, Parental Monitoring and Adolescent Problem Behaviors Among Racially Diverse Urban Youth. J Community Health. 2008 Dec 1;33(6):425–33.

106. Sampson RJ. The neighborhood context of well-being. Perspect Biol Med. 2003;46(3 Suppl):S53–64.

107. Brook DW, Rubenstone E, Zhang C, Morojele NK, Brook JS. Environmental stressors, low well-being, smoking, and alcohol use among South African adolescents. Soc Sci Med. 2011 May;72(9):1447–53.

108. Parry CD, Morojele NK, Saban A, Flisher AJ. Brief report: Social and neighbourhood correlates of adolescent drunkenness: a pilot study in Cape Town, South Africa. J Adolesc. 2004 Jun;27(3):369–74.

109. Onya H, Tessera A, Myers B, Flisher A. Community influences on adolescents' use of home-brewed alcohol in rural South Africa. BMC Public Health. 2012 Aug 11;12(1):642.

110. Campbell CA, Hahn RA, Elder R, Brewer R, Chattopadhyay S, Fielding J, et al. The Effectiveness of Limiting Alcohol Outlet Density As a Means of Reducing Excessive Alcohol Consumption and Alcohol-Related Harms. Am J Prev Med. 2009 Dec;37(6):556–69.

111. Bryden A, Roberts B, McKee M, Petticrew M. A systematic review of the influence on alcohol use of community level availability and marketing of alcohol. Health Place. 2012 Mar;18(2):349–57.

112. Task Force on Community Preventive Services. Recommendations for Reducing Excessive Alcohol Consumption and Alcohol-Related Harms by Limiting Alcohol Outlet Density. Am J Prev Med. 2009 Dec;37(6):570–1.

113. Parry CDH. Alcohol policy in South Africa: a review of policy development processes between 1994 and 2009. Addiction. 2010 Aug 1;105(8):1340–5.

114. Bowers Y, Rendall-Mkosi K, Davids A, Nel E, Jacobs N, London L. Liquor outlet density, deprivation and implications for foetal alcohol syndrome prevention in the Bergriver municipality in the Western Cape, South Africa. South Afr Geogr J. 2014 Apr 14;96(0):1–13.

115. Ahern J, Margerison-Zilko C, Hubbard A, Galea S. Alcohol Outlets and Binge Drinking in Urban Neighborhoods: The Implications of Nonlinearity for Intervention and Policy. Am J Public Health. 2013 Apr;103(4):e81–7.

116. Parry C, London L, Myers B. Delays in South Africa's plans to ban alcohol advertising. The Lancet. 2014 Jun;383(9933):1972.

117. Kahn K, Collinson MA, Gómez-Olivé FX, Mokoena O, Twine R, Mee P, et al. Profile: Agincourt Health and Socio-demographic Surveillance System. Int J Epidemiol. 2012 Aug 1;41(4):988–1001.

118. Hunter LM, Nawrotzki R, Leyk S, Maclaurin GJ, Twine W, Collinson M, et al. Rural Outmigration, Natural Capital, and Livelihoods in South Africa. Popul Space Place. 2014 Jul 1;20(5):402–20.

119. Gómez-Olivé FX, Angotti N, Houle B, Klipstein-Grobusch K, Kabudula C, Menken J, et al. Prevalence of HIV among those 15 and older in rural South Africa. AIDS Care Psychol Socio-Med Asp AIDSHIV. 2013;25(9):1122–8.

120. Saunders JB, Aasland OG, Babor TF, De La Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. Addiction. 1993;88(6):791–804.

121. Babor TF, Biddle-Higgins JC, Saunders JB, Monteiro M. AUDIT: The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Health Care. Geneva, Switzerland: World Health Organization; 2001. Report No.: Second edition.

122. Bradley KA, DeBenedetti AF, Volk RJ, Williams EC, Frank D, Kivlahan DR. AUDIT-C as a Brief Screen for Alcohol Misuse in Primary Care. Alcohol Clin Exp Res. 2007;31(7):1208–17.

123. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA, ACQUIP. The audit alcohol consumption questions (AUDIT-C): An effective brief screening test for problem drinking. Arch Intern Med. 1998 Sep 14;158(16):1789–95.

124. Desmond K, Milburn N, Richter L, Tomlinson M, Greco E, van Heerden A, et al. Alcohol consumption among HIV-positive pregnant women in KwaZulu-Natal, South Africa: Prevalence and correlates. Drug Alcohol Depend. 2012 Jan 1;120(1–3):113–8.

125. Myer L, Smit J, Roux LL, Parker S, Stein DJ, Seedat S. Common Mental Disorders among HIV-Infected Individuals in South Africa: Prevalence, Predictors, and Validation of Brief Psychiatric Rating Scales. AIDS Patient Care STDs. 2008 Feb;22(2):147–58.

126. Peltzer K, Preez NF, Ramlagan S, Anderson J. Antiretroviral treatment adherence among HIV patients in KwaZulu-Natal, South Africa. BMC Public Health. 2010 Mar 5;10(1):111.

127. Hubbard A, Munoz ID, Decker A, Holcomb JB, Schreiber MA, Bulger EM, et al. Time-dependent prediction and evaluation of variable importance using superlearning in high-dimensional clinical data. J Trauma Acute Care Surg July 2013 [Internet]. 2013 [cited 2014 Feb 28]; Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=ovfto&AN=01586154-201307001-00009

128. van der Laan MJ, Rose S. Targeted Learning: Causal Inference for Observational and Experimental Data. Springer; 2011. 678 p.

129. van der Laan MJ, Polley EC, Hubbard AE. Super Learner. Stat Appl Genet Mol Biol [Internet]. 2007 Sep 16 [cited 2013 Jul 24];6(1). Available from: http://www.degruyter.com/view/j/sagmb.2007.6.1/sagmb.2007.6.1.1309/sagmb.2007.6.1.1309.xml

130. 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. 2009 May 1;169(9):1140–7.

131. DiCiccio TJ, Efron B. Bootstrap Confidence Intervals. Stat Sci. 1996 Aug 1;11(3):189–212.

132. VanderWeele TJ, Knol MJ. A Tutorial on Interaction. Epidemiol Methods. 2014 May;3(1):1–40.

133. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia: Wolters Kluwer; Lippincott Williams & Wilkins; 2008.

134. Gruber S. tmle: Targeted Maximum Likelihood Estimation [Internet]. 2012 [cited 2013 Jul 24]. Available from: http://cran.r-project.org/web/packages/tmle/index.html

135. Polley E, van der Laan MJ. SuperLearner: Super Learner Prediction [Internet]. 2012 [cited 2013 Jul 24]. Available from: http://cran.r-project.org/web/packages/SuperLearner/index.html

136. Scribner R, Theall KP, Simonsen N, Robinson W. HIV Risk and the Alcohol Environment. Alcohol Res Health. 2010;33(3):179–83.

137. Kuipers MAG, van Poppel MNM, van den Brink W, Wingen M, Kunst AE. The association between neighborhood disorder, social cohesion and hazardous alcohol use: a national multilevel study. Drug Alcohol Depend. 2012 Nov 1;126(1-2):27–34.

138. De Haan L, Boljevac T, Schaefer K. Rural Community Characteristics, Economic Hardship, and Peer and Parental Influences in Early Adolescent Alcohol Use. J Early Adolesc. 2010 Oct 1;30(5):629–50.

139. Ennett ST, Foshee VA, Bauman KE, Hussong A, Cai L, Reyes HLM, et al. The Social Ecology of Adolescent Alcohol Misuse. Child Dev. 2008 Nov 1;79(6):1777–91.

140. Maimon D, Browning CR. Underage drinking, alcohol sales and collective efficacy: Informal control and opportunity in the study of alcohol use. Soc Sci Res. 2012 Jul;41(4):977–90.

141. Lin E-Y, Witten K, Casswell S, You RQ. Neighbourhood matters: Perceptions of neighbourhood cohesiveness and associations with alcohol, cannabis and tobacco use. Drug Alcohol Rev. 2012;31(4):402–12.

142. Collinson M, Tollman SM, Kahn K, Clark S, Garenne M. Highly prevalent circular migration: households, mobility and economic status in rural South Africa. Afr Move Afr Migr Urban Comp Perspect. 2006;194–216.

143. Niehaus I, Stadler J. Muchongolo Dance Contests: Deep Play in the South African Lowveld. Ethnology. 2004 Oct 1;43(4):363–80.

144. Morojele NK, Kachieng'a MA, Mokoko E, Nkoko MA, Parry CDH, Nkowane AM, et al. Alcohol use and sexual behaviour among risky drinkers and bar and shebeen patrons in Gauteng province, South Africa. Soc Sci Med. 2006 Jan;62(1):217–27.

145. Scott-Sheldon LAJ, Carey KB, Carey MP, Cain D, Simbayi LC, Kalichman SC. Alcohol use disorder, contexts of alcohol use, and the risk of HIV transmission among South African male patrons of shebeens. Drug Alcohol Depend. 2014 Jul 1;140:198–204.

146. Nayak MB, Kerr W, Greenfield TK, Pillai A. Not All Drinks Are Created Equal: Implications for Alcohol Assessment in India. Alcohol Alcohol. 2008;43(6):713–8.

147. Harrison A, Colvin CJ, Kuo C, Swartz A, Lurie M. Sustained High HIV Incidence in Young Women in Southern Africa: Social, Behavioral, and Structural Factors and Emerging Intervention Approaches. Curr HIV/AIDS Rep. 2015 Apr 9;12(2):207–15.

148. Jewkes R, Morrell R. Gender and sexuality: emerging perspectives from the heterosexual epidemic in South Africa and implications for HIV risk and prevention. J Int AIDS Soc. 2010 Feb 9;13(1):6.

149. Brown J, Sorrell J, Raffaelli M. An exploratory study of constructions of masculinity, sexuality and HIV/AIDS in Namibia, Southern Africa. Cult Health Sex. 2005 Nov 1;7(6):585–98.

150. Mager A. "One Beer, One Goal, One Nation, One Soul": South African Breweries, Heritage, Masculinity and Nationalism 1960–1999. Past Present. 2005 Aug 1;188(1):163–94.

151. Mager A. "White liquor hits black livers": meanings of excessive liquor consumption in South Africa in the second half of the twentieth century. Soc Sci Med. 2004 Aug;59(4):735–51.

152. Woolf-King SE, Maisto SA. Alcohol Use and High-Risk Sexual Behavior in Sub-Saharan Africa: A Narrative Review. Arch Sex Behav. 2011 Feb 1;40(1):17–42.

153. Dworkin SL, Treves-Kagan S, Lippman SA. Gender-Transformative Interventions to Reduce HIV Risks and Violence with Heterosexually-Active Men: A Review of the Global Evidence. AIDS Behav. 2013 Aug 10;17(9):2845–63.

154. Interagency Gender Working Group. Three case studies: involving men to address gender inequities [Internet]. Washington, DC: Population Reference Bureau; 2003 Jul [cited 2015 May 17] p. 72. Report No.: 186075. Available from: http://www.igwg.org/pdf/InvolvMenToAddressGendr.pdf

155. NIMH Collaborative HIV/STD Prevention Trial Group. The community popular opinion leader HIV prevention programme: conceptual basis and intervention procedures. AIDS Lond Engl. 2007 Apr;21 Suppl 2:S59–68.

156. Peltzer K. Prevalence of Alcohol Use by Rural Primary Care Outpatients in South Africa. Psychol Rep. 2006 Aug 1;99(1):176–8.

157. Brown JL, Vanable PA. Alcohol use, partner type, and risky sexual behavior among college students: Findings from an event-level study. Addict Behav. 2007 Dec;32(12):2940–52.

158. VanderWeele TJ. Direct and Indirect Effects for Neighborhood-Based Clustered and Longitudinal Data. Sociol Methods Res. 2010 May 1;38(4):515–44.

159. Zheng W, van der Laan MJ. Targeted Maximum Likelihood Estimation of Natural Direct Effects. Int J Biostat [Internet]. 2012 [cited 2013 May 30];8(1). Available from: http://works.bepress.com/mark_van_der_laan/308

160. van der Laan M, Dudoit S. Unified Cross-Validation Methodology For Selection Among Estimators and a General Cross-Validated Adaptive Epsilon-Net Estimator: Finite Sample Oracle Inequalities and Examples. UC Berkeley Div Biostat Work Pap Ser [Internet]. 2003 Nov 26; Available from: http://biostats.bepress.com/ucbbiostat/paper130

161. Schwab J, Lendle S, Petersen M, van der Laan M, Gruber S. ltmle: Longitudinal Targeted Maximum Likelihood Estimation [Internet]. 2015 [cited 2015 Jul 27]. Available from: https://cran.r-project.org/web/packages/ltmle/index.html

162. van der Laan MJ, Robins JM. Unified Methods for Censored Longitudinal Data and Causality. New York: Springer Verlag; 2003. 424 p.

163. Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. Soc Sci Med. 2004 May;58(10):1929–52.

164. Petersen ML. Compound treatments, transportability, and the structural causal model: the power and simplicity of causal graphs. Epidemiol Camb Mass. 2011 May;22(3):378–81.

165. Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. Stat Methods Med Res. 2012 Feb 1;21(1):31–54.

166. Ahern J, Cerdá M, Lippman SA, Tardiff KJ, Vlahov D, Galea S. Navigating non-positivity in neighbourhood studies: an analysis of collective efficacy and violence. J Epidemiol Community Health. 2013 Feb 1;67(2):159–65.

167. Greenfield TK, Kerr WC. Alcohol measurement methodology in epidemiology: recent advances and opportunities. Addiction. 2008;103(7):1082–99.

168. Vanderweele TJ, Hong G, Jones SM, Brown JL. Mediation and Spillover Effects in Group-Randomized Trials: A Case Study of the 4Rs Educational Intervention. J Am Stat Assoc. 2013 Jun 1;108(502):469–82.

169. Imai K, Keele L, Yamamoto T. Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. Stat Sci. 2010 Feb;25(1):51–71.

170. Tchetgen Tchetgen EJ. Inverse odds ratio-weighted estimation for causal mediation analysis. Stat Med [Internet]. 2013 [cited 2013 Jul 31]; Available from: http://onlinelibrary.wiley.com/doi/10.1002/sim.5864/abstract

171. Nguyen QC, Osypuk TL, Schmidt NM, Glymour MM, Tchetgen EJT. Practical Guidance for Conducting Mediation Analysis With Multiple Mediators Using Inverse Odds Ratio Weighting. Am J Epidemiol. 2015 Mar 1;181(5):349–56.

172. Rabe-Hesketh S. Multilevel Latent Variable Modeling [Internet]. Statistical Modelling and Inference Conference; 2010 Feb [cited 2014 Nov 19]; Brisbane, Australia. Available from: http://eprints.ncrm.ac.uk/1677/1/Multilevel_Latent_Variable_Modeling.pdf

173. Rabe-Hesketh S, Skrondal A, Pickles A. Generalized multilevel structural equation modeling. Psychometrika. 2004 Jun 1;69(2):167–90.

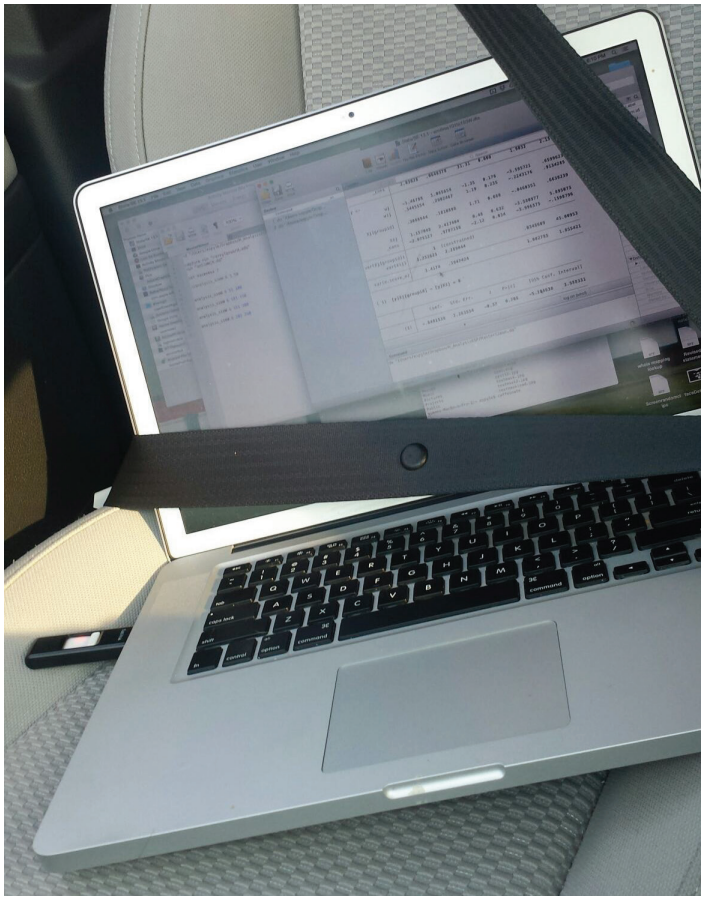
174. Pearl J. Causality: Models, Reasoning, and Inference. 2nd ed. Cambridge, UK: Cambridge University Press; 2009. 484 p.

175. Greenland S, Brumback B. An overview of relations among causal modelling methods. Int J Epidemiol. 2002 Oct 1;31(5):1030–7.

Appendices

A. APPENDIX TO ACKNOWLEDGEMENTS

Figure A.1



B. APPENDIX TO CHAPTER 2

Table B.1: Item step difficulty matrix

$\delta_{1k=0\text{-}4}$ 0.00, 0.90, 1.95, 3.00, 4.40

$\delta_{2k=0\text{-}4}$ 0.00, 3.25, 4.90, 7.70, 6.90

$\delta_{3k=0\text{-}4}$ 0.00, 2.40, 5.20, 2.40, 0.10

$\delta_{4k=0\text{-}4}$ 0.00, 2.50, 3.40, 2.60, 0.10

$\delta_{5k=0\text{-}4}$ 0.00, 1.70, 1.70, 0.85, -1.60

$\delta_{6k=0\text{-}4}$ 0.00, 1.75, -0.10, -1.75, -5.00

$\delta_{7k=0\text{-}4}$ 0.00, 0.80, 0.30, 0.00, -3.00

Figure B.1: Item response distributions sampled from baseline simulation 1.0 (N=1600)

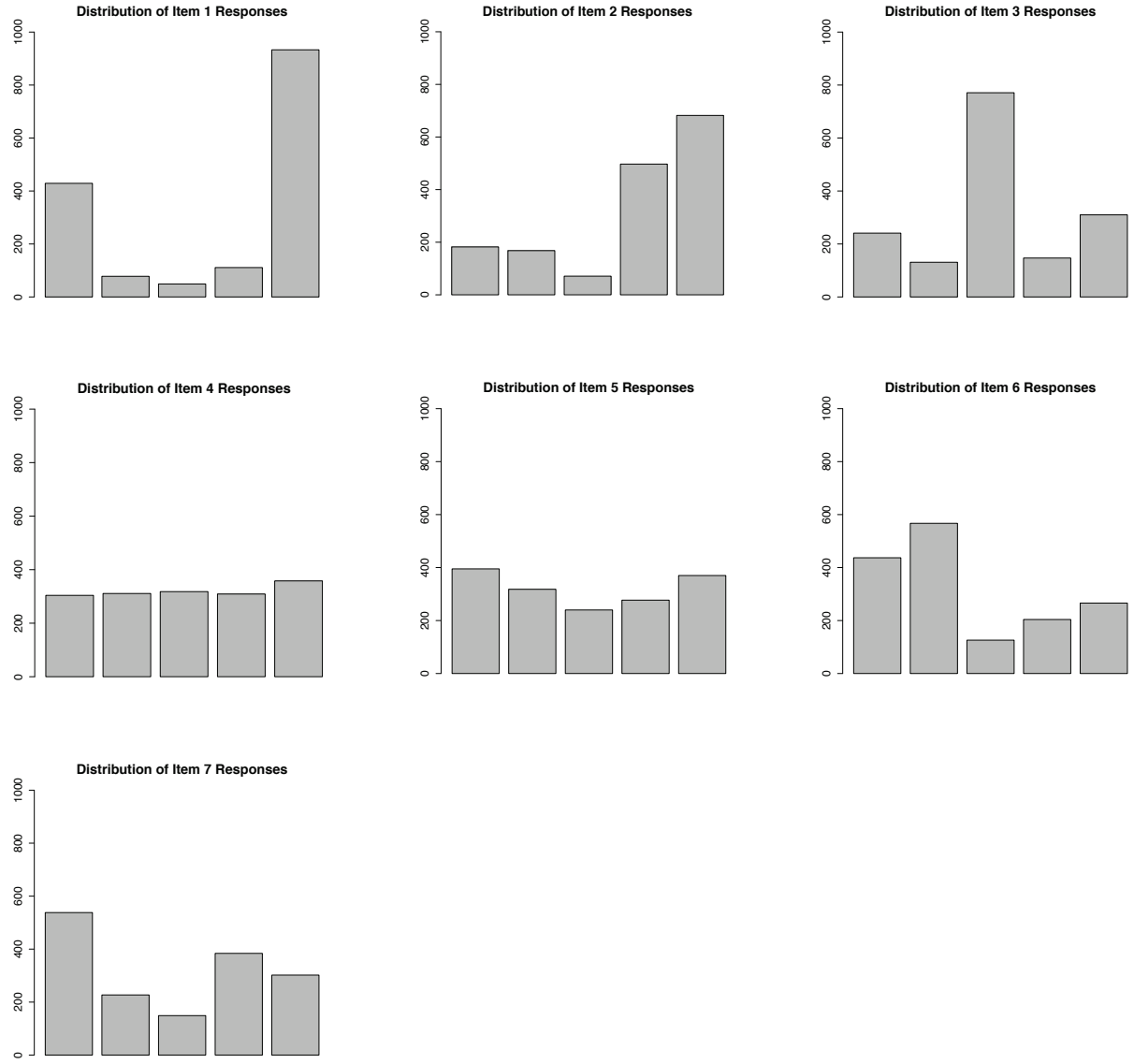


Figure B.2: Distribution of θ_j in simulation 1.5

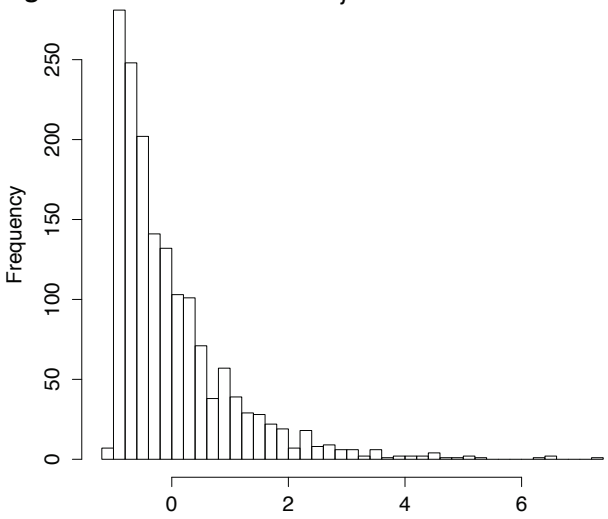
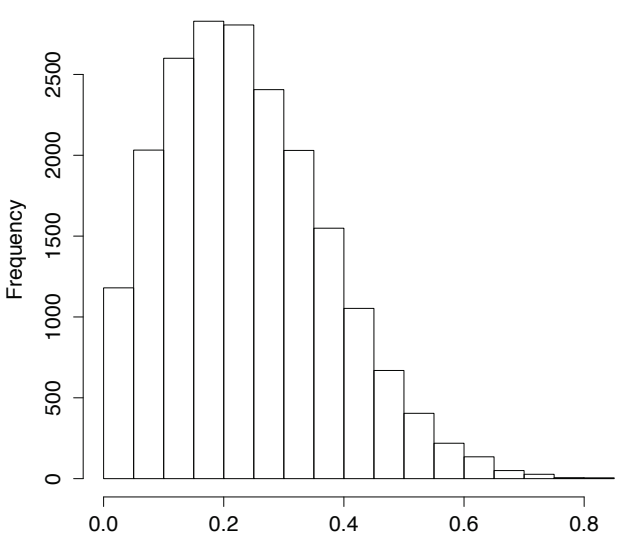
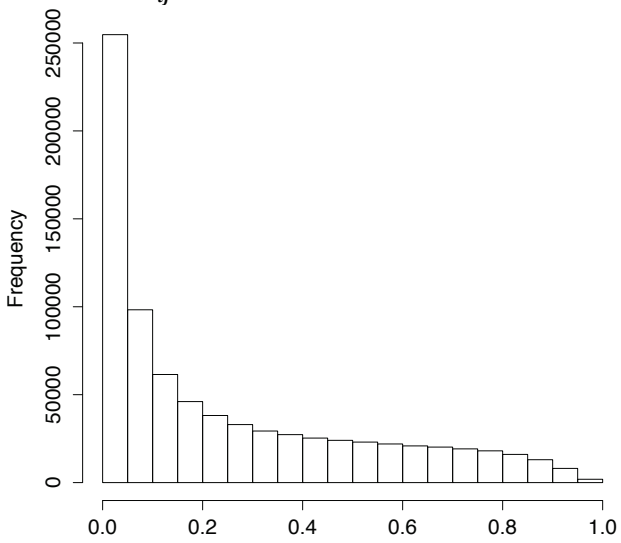


Figure B.3: Distributions of W_j and of inverse logit function of $U_{W_{ij}}$, simulations 2.2 and 2.3



Min: 0.000, Max: 0.875, Median: 0.225, Mean: 0.247

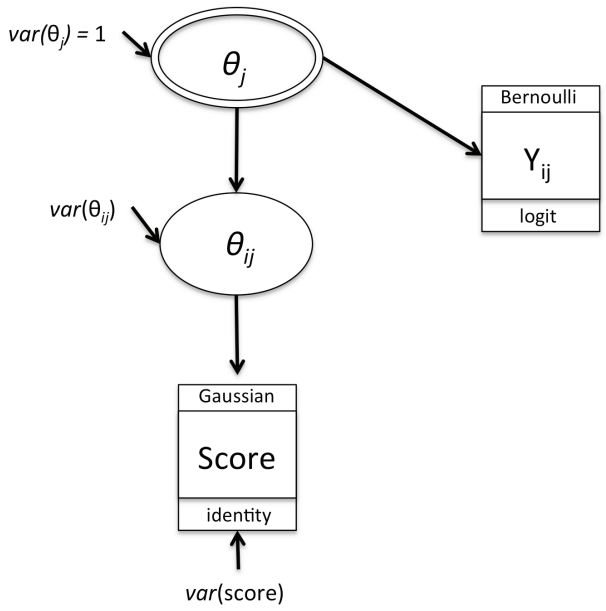


Min: 0.000, Max: 0.994, Median: 0.133, Mean: 0.247

Figure B.4: SEM schematics, single measure per person (mean score or EAP)

Variables enclosed in circles are latent; double rings represent higher-level traits. Paths with the same label are constrained to be equal; variances with an assigned value are constrained to that value.

Setting A



Setting B

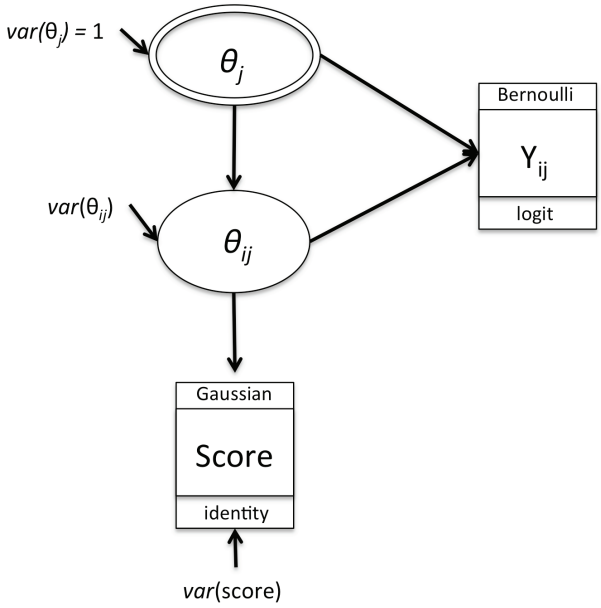
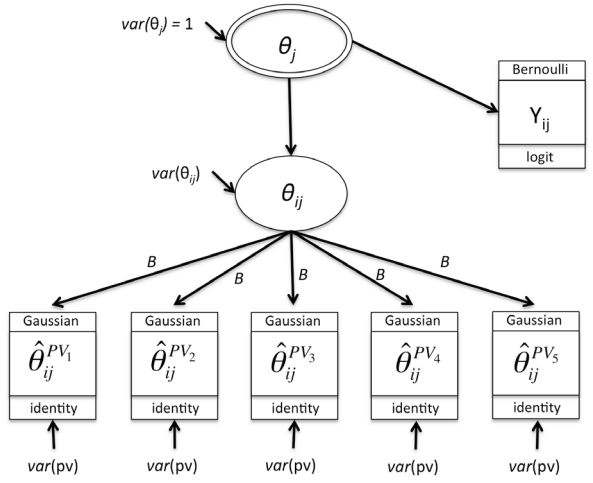


Figure B.5: SEM schematics, multiple measures per person (PVs)

Setting A



Setting B

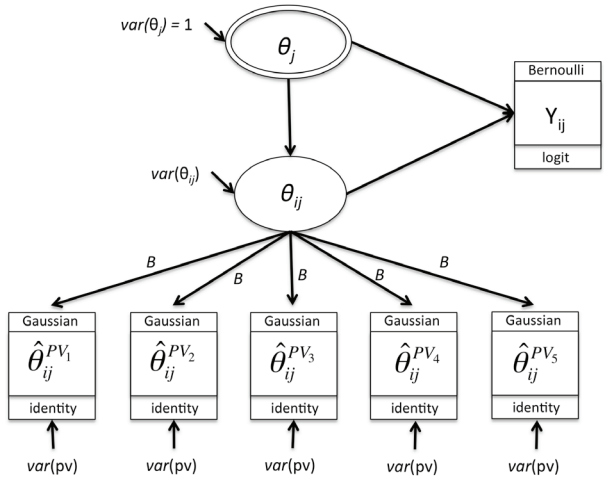
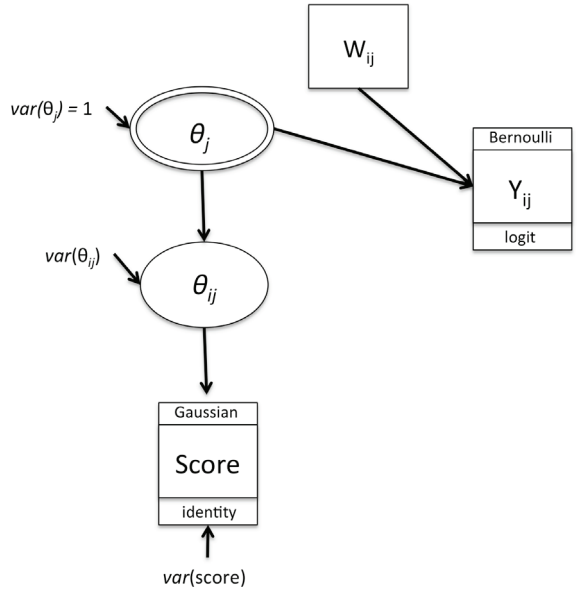


Figure B.6: SEM schematics adjusting for W_{ij} with a single measure per person

Setting A



Setting B

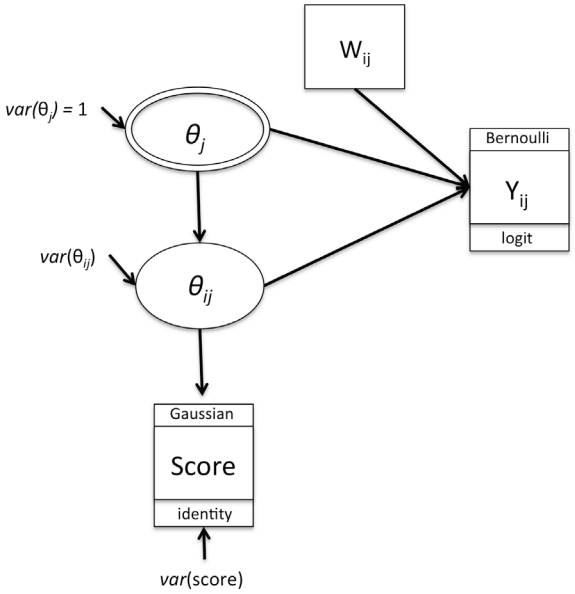
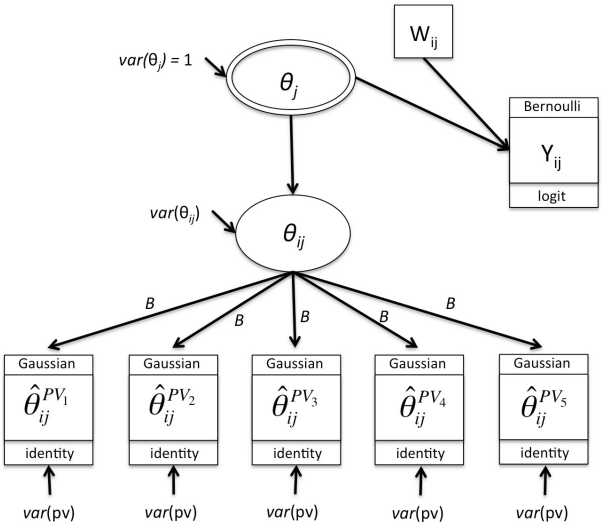


Figure B.7: SEM schematics adjusting for W_{ij} with multiple measures per person

Setting A



Setting B

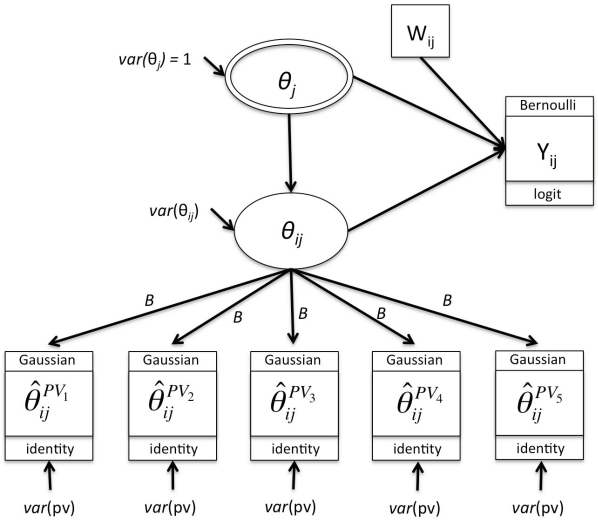
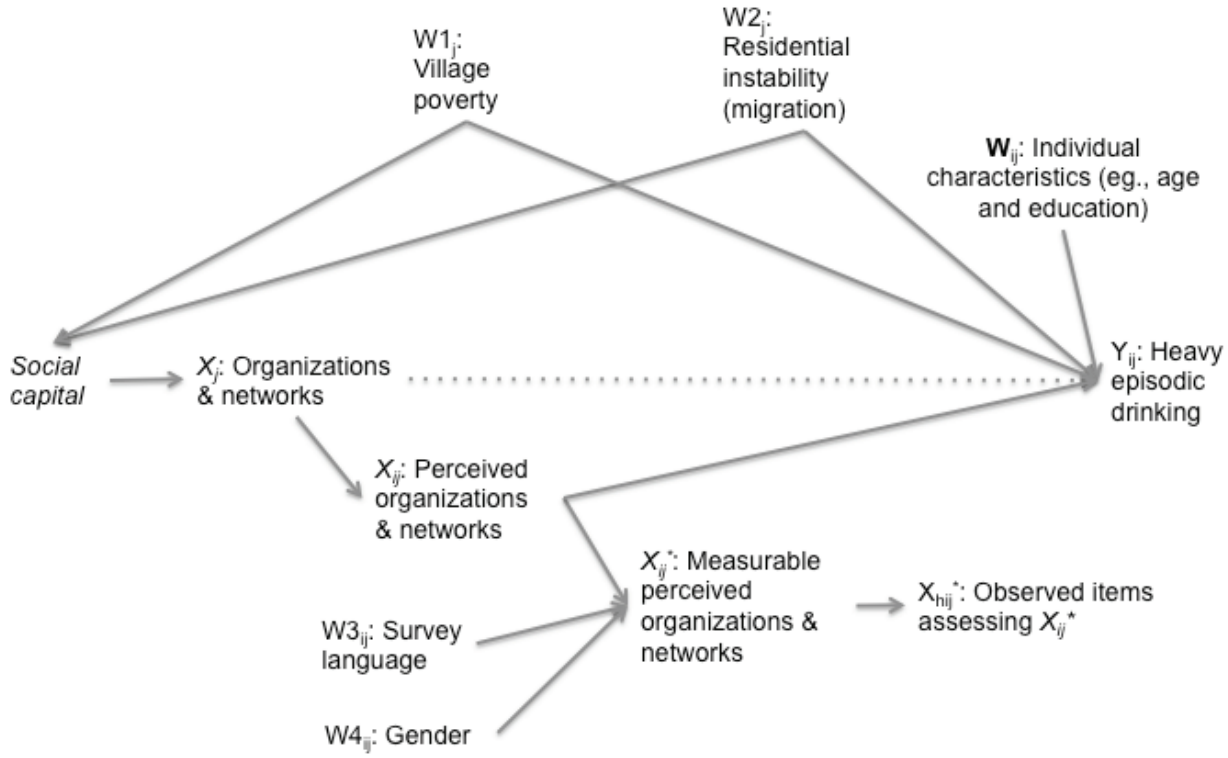


Figure B.8: Conceptual model for motivating example – village social capital and heavy episodic drinking among young men



The dotted line represents the contextual effect of interest. The link between gender and alcohol use is excluded because, although all participants are included in the calculation of the social measure, the outcome is restricted to men.

Table B.2: Robustness results for analytic approaches applied to reflective causal model, setting A

	Bias	Relative bias	Variance	MSE	Coverage probability	Convergence failure
Simulation 1.1: Two-parameter item model						
True θ_j	0.003	0.86%	0.007	0.007	94.4%	NA
M-MR	-0.058	-19.19%	0.008	0.011	84.0%	NA
M-SEM	-0.086	-28.72%	0.359	0.365	83.4%	7.4%*
EAP-MR	-0.027	-8.91%	0.008	0.009	89.8%	NA
EAP-SEM	-0.035	-11.51%	0.009	0.010	89.1%	1.0%*
PV-MR	-0.029	-9.54%	0.008	0.009	89.8%	NA
PV-SEM	-0.023	-7.57%	0.011	0.011	90.7%	9.2%**
Simulation 1.2: Non-normal θ_{ij} distribution: uniform						
True θ_j & θ_{ij}	0.000	-0.15%	0.006	0.006	95.6%	NA
M-MR	-0.041	-13.68%	0.007	0.009	90.2%	NA
M-SEM	-0.037	-12.17%	0.007	0.009	92.4%	0.0%*
EAP-MR	-0.036	-11.91%	0.007	0.008	91.2%	NA
EAP-SEM	-0.042	-13.90%	0.007	0.009	90.8%	0.0%*
PV-MR	-0.036	-12.15%	0.007	0.008	92.0%	NA
PV-SEM	-0.011	-3.50%	0.008	0.008	95.6%	0.0%
Simulation 1.3: Non-normal θ_{ij} distribution: chi squared, ICC = 0.25						
True θ_j & θ_{ij}	0.004	1.29%	0.007	0.007	94.2%	NA
M-MR	-0.003	-0.87%	0.008	0.008	92.4%	NA

M-SEM	-0.176	-58.50%	0.041	0.072	54.2%	9.8%*
EAP-MR	-0.004	-1.26%	0.008	0.008	92.2%	NA
EAP-SEM	-0.071	-23.52%	0.006	0.011	76.4%	0.0%*
PV-MR	-0.005	-1.72%	0.008	0.008	92.6%	NA
PV-SEM	-0.002	-0.55%	0.008	0.008	94.4%	0.0%
Simulation 1.4: Non-normal θ_j distribution: uniform						
True θ_j & θ_{ij}	0.007	2.35%	0.007	0.007	95.8%	NA
M-MR	-0.022	-7.22%	0.007	0.008	93.4%	NA
M-SEM	-0.051	-17.05%	0.006	0.009	87.0%	0.0%*
EAP-MR	-0.022	-7.20%	0.007	0.007	93.4%	NA
EAP-SEM	-0.048	-16.05%	0.006	0.008	87.4%	0.2%*
PV-MR	-0.023	-7.78%	0.007	0.007	93.0%	NA
PV-SEM	0.000	0.13%	0.008	0.008	96.6%	0.0%
Simulation 1.5: Non-normal θ_j distribution: chi squared						
True θ_j & θ_{ij}	0.001	0.26%	0006	0.006	95.2%	NA
M-MR	-0.019	-6.46%	0.010	0.010	83.6%	NA
M-SEM	0.028	9.28%	0.024	0.024	82.5%	0.4%*
EAP-MR	-0.019	-6.40%	0.010	0.010	83.8%	NA
EAP-SEM	0.015	4.92%	0.017	0.018	87.6%	2.0%*
PV-MR	-0.021	-6.91%	0.010	0.010	84.8%	NA
PV-SEM	0.005	1.76%	0.010	0.010	91.8%	0.0%

*Using non-adaptive Gauss-Hermite quadrature based on failure of testing set of 50 simulations to show reasonable convergence under mode-and-curvature adaptive quadrature.

**Using non-adaptive Gauss-Hermite quadrature for 17.2% of total runs

Table B.3: Performance of adjustment strategies when W_{ij} distorts θ_{ij}, simulation 2.0A

	Bias	Relative bias	Variance	MSE	Coverage probability	Convergence failure
True θ_j						
Unadj.	0.000	-0.08%	0.006	0.006	95.4%	NA
Adj. for W_{ij}	0.000	-0.03%	0.006	0.006	95.4%	NA
Adj. for W_{ij} & $W_{.j}$	0.000	-0.06%	0.007	0.007	95.4%	NA
M-MR						
NA / Unadj.	-0.045	-14.92%	0.009	0.011	86.2%	NA
NA / Adj. W_{ij}	-0.044	-14.73%	0.009	0.010	86.2%	NA
NA / Adj. W_{ij} & $W_{.j}$	-0.021	-7.11%	0.009	0.010	89.8%	NA
M-SEM						
NA / Unadj.	-0.026	-8.52%	0.011	0.012	94.5%	9.8%*
NA / Adj. W_{ij}	-0.033	-10.96%	0.013	0.014	92.0%	0.2%*
NA / Adj. W_{ij} & $W_{.j}$	-0.018	-6.00%	0.014	0.014	93.7%	1.2%*
PV-MR						
Unadj. / Unadj.	-0.050	-16.65%	0.008	0.011	87.0%	NA
Unadj. / Adj. W_{ij}	-0.049	-16.42%	0.008	0.011	85.8%	NA
Unadj. / Adj. W_{ij} & $W_{.j}$	-0.021	-7.14%	0.009	0.010	89.8%	NA
Adj. / Unadj.	-0.055	-18.41%	0.008	0.011	84.8%	NA
Adj. / Adj. W_{ij}	-0.054	-18.12%	0.008	0.011	85.4%	NA
Adj. / Adj. W_{ij} & $W_{.j}$	-0.015	-4.98%	0.009	0.010	89.6%	NA
PV-SEM						
Unadj. / Unadj.	-0.007	-2.37%	0.010	0.010	94.2%	0.0%
Unadj. / Adj. W_{ij}	-0.007	-2.17%	0.010	0.010	94.2%	0.0%
Unadj. / Adj. W_{ij} & $W_{.j}$	0.016	5.48%	0.010	0.010	96.0%	0.0%
Adj. / Unadj.	0.205	68.43%	0.009	0.051	31.7%	0.2%**

Adj. / Adj. W_{ij}	-0.007	-2.27%	0.010	0.010	94.4%	0.0%
Adj. / Adj. W_{ij} & $W_{.j}$	0.025	8.44%	0.011	0.011	95.6%	0.0%

*Using non-adaptive Gauss-Hermite quadrature based on failure of testing set of 50 simulations.
**0.2% also converged using non-adaptive Gauss-Hermite quadrature instead of mode-and-curvature.

Table B.4: Performance of adjustment strategies when W_{ij} distorts θ_{ij} and affects Y_{ij}, simulation 2.1A

	Bias	Relative bias	Variance	MSE	Coverage probability	Convergence failure
True θ_j						
Unadj.	0.001	0.46%	0.006	0.006	94.8%	NA
Adj. for W_{ij}	0.003	0.95%	0.006	0.006	94.4%	NA
Adj. for W_{ij} & $W_{.j}$	0.004	1.34%	0.006	0.006	93.8%	NA
M-MR						
NA / Unadj.	-0.034	-11.25%	0.008	0.009	89.0%	NA
NA / Adj. W_{ij}	-0.041	-13.80%	0.008	0.009	87.0%	NA
NA / Adj. W_{ij} & $W_{.j}$	-0.018	-5.95%	0.009	0.009	88.4%	NA
M-SEM						
NA / Unadj.	0.019	6.21%	0.332	0.332	92.3%	12%*
NA / Adj. W_{ij}	-0.012	-3.95%	0.159	0.159	90.5%	0.6%*
NA / Adj. W_{ij} & $W_{.j}$	0.003	0.97%	0.166	0.165	93.0%	0.6%*
PV-MR						
Unadj. / Unadj.	-0.038	-12.75%	0.007	0.009	88.4%	NA
Unadj. / Adj. W_{ij}	-0.047	-15.59%	0.008	0.010	85.8%	NA
Unadj. / Adj. W_{ij} & $W_{.j}$	-0.018	-6.07%	0.009	0.009	89.0%	NA
Adj. / Unadj.	-0.042	-13.89%	0.007	0.009	88.0%	NA
Adj. / Adj. W_{ij}	-0.052	-17.29%	0.007	0.010	85.0%	NA
Adj. / Adj. W_{ij} & $W_{.j}$	-0.012	-3.83%	0.009	0.009	88.8%	NA
PV-SEM						
Unadj. / Unadj.	0.003	0.84%	0.009	0.009	94.7%	2.4%**
Unadj. / Adj. W_{ij}	-0.003	-1.13%	0.009	0.009	94.5%	2.2%
Unadj. / Adj. W_{ij} & $W_{.j}$	0.020	6.56%	0.010	0.010	94.1%	2.2%
Adj. / Unadj.	0.204	67.90%	0.007	0.049	29.9%	2.2%
Adj. / Adj. W_{ij}	-0.003	-1.15%	0.009	0.009	94.9%	2.0%
Adj. / Adj. W_{ij} & $W_{.j}$	0.029	9.55%	0.010	0.011	93.3%	2.0%

*Using non-adaptive Gauss-Hermite quadrature based on failure of testing set of 50 simulations.
**0.2% also converged using non-adaptive Gauss-Hermite quadrature instead of mode-and-curvature.

Table B.5: Performance of analytic and adjustment strategies when W_{ij} distorts θ_{ij} and affects θ_j and Y_{ij}, simulation 2.2A

	Bias	Relative bias	Variance	MSE	Coverage probability	Convergence failure
True θ_j						
Unadj.	-0.029	-9.57%	0.006	0.007	94.0%	NA
Adj. for W_{ij}	0.003	1.08%	0.006	0.006	93.8%	NA
Adj. for W_{ij} & $W_{.j}$	0.006	2.09%	0.008	0.009	93.4%	NA
M-MR						
NA / Unadj.	-0.094	-31.49%	0.007	0.016	72.2%	NA
NA / Adj. W_{ij}	-0.042	-13.88%	0.007	0.009	87.2%	NA
NA / Adj. W_{ij} & $W_{.j}$	0.044	14.76%	0.017	0.019	91.4%	NA
M-SEM						
NA / Unadj.	-0.101	-33.62%	0.007	0.017	68.1%	4.8%*
NA / Adj. W_{ij}	-0.055	-18.29%	0.007	0.010	84.6%	0.0%*
NA / Adj. W_{ij} & $W_{.j}$	-0.007	-2.46%	0.013	0.013	93.5%	1.4%*

PV-MR						
Unadj. / Unadj.	-0.097	-32.32%	0.007	0.016	71.0%	NA
Unadj. / Adj. W_{ij}	-0.043	-14.41%	0.007	0.009	86.6%	NA
Unadj. / Adj. W_{ij} & $W_{.j}$	0.048	16.14%	0.017	0.019	89.8%	NA
Adj. / Unadj.	-0.097	-32.18%	0.007	0.016	73.6%	NA
Adj. / Adj. W_{ij}	-0.043	-14.25%	0.007	0.009	87.2%	NA
Adj. / Adj. W_{ij} & $W_{.j}$	0.049	16.50%	0.017	0.019	89.2%	NA
PV-SEM						
Unadj. / Unadj.	-0.077	-25.65%	0.008	0.014	82.4%	0.0%**
Unadj. / Adj. W_{ij}	-0.021	-7.05%	0.008	0.009	93.0%	0.0%
Unadj. / Adj. W_{ij} & $W_{.j}$	0.069	23.11%	0.018	0.022	93.4%	0.0%
Adj. / Unadj.	0.379	126.39%	0.014	0.158	1.0%	0.0%**
Adj. / Adj. W_{ij}	-0.020	-6.78%	0.008	0.009	93.0%	0.0%
Adj. / Adj. W_{ij} & $W_{.j}$	0.071	23.52%	0.017	0.022	93.4%	0.0%

*Using non-adaptive Gauss-Hermite quadrature based on failure of testing set of 50 simulations.

** 0.2% and 0.4% converged using non-adaptive Gauss-Hermite quadrature.

Table B.6: Performance of analytic and adjustment strategies when W_{ij} affects θ_j and Y_{ij}, simulation 2.3A

	Bias	Relative bias	Variance	MSE	Coverage probability	Convergence failure
True θ_j						
Unadj.	-0.034	-11.46%	0.006	0.007	93.0%	NA
Adj. for W_{ij}	-0.002	-0.82%	0.006	0.006	95.4%	NA
Adj. for W_{ij} & $W_{.j}$	-0.004	-1.36%	0.008	0.008	95.2%	NA
M-MR						
NA / Unadj.	-0.063	-20.95%	0.007	0.011	82.4%	NA
NA / Adj. W_{ij}	-0.035	-11.72%	0.008	0.009	88.0%	NA
NA / Adj. W_{ij} & $W_{.j}$	-0.049	-16.40%	0.009	0.012	86.8%	NA
M-SEM						
NA / Unadj.	-0.073	-24.35%	0.007	0.013	79.4%	4.8%*
NA / Adj. W_{ij}	-0.046	-15.25%	0.007	0.009	89.0%	0.2%*
NA / Adj. W_{ij} & $W_{.j}$	-0.064	-21.46%	0.009	0.013	85.0%	1.4%*
PV-MR						
Unadj. / Unadj.	-0.064	-21.26%	0.007	0.011	83.2%	NA
Unadj. / Adj. W_{ij}	-0.036	-12.11%	0.008	0.009	87.0%	NA
Unadj. / Adj. W_{ij} & $W_{.j}$	-0.051	-16.90%	0.009	0.012	85.6%	NA
Adj. / Unadj.	-0.064	-21.48%	0.007	0.011	81.2%	NA
Adj. / Adj. W_{ij}	-0.037	-12.32%	0.008	0.009	87.2%	NA
Adj. / Adj. W_{ij} & $W_{.j}$	-0.051	-17.16%	0.009	0.012	85.2%	NA
PV-SEM						
Unadj. / Unadj.	-0.043	-14.20%	0.008	0.010	92.0%	0.4%**
Unadj. / Adj. W_{ij}	-0.013	-4.30%	0.008	0.008	95.6%	0.4%
Unadj. / Adj. W_{ij} & $W_{.j}$	-0.030	-10.13%	0.010	0.011	93.8%	0.4%
Adj. / Unadj.	0.308	102.61%	0.010	0.105	4.8%	0.6%***
Adj. / Adj. W_{ij}	-0.013	-4.37%	0.008	0.008	95.8%	0.6%
Adj. / Adj. W_{ij} & $W_{.j}$	-0.031	-10.23%	0.010	0.011	94.2%	0.6%

*Using non-adaptive Gauss-Hermite quadrature based on failure of testing set of 50 simulations.

**0.2% and ***0.04% also converged using non-adaptive Gauss-Hermite quadrature instead of mode-and-curvature.

Table B.7: Performance of MR versus SEM approaches under a formative causal model, setting A

	Bias	Relative bias	Variance	MSE	Coverage probability	Convergence failure
Simulation 3.0: Large sampling fraction (0.50)						
True θ_j & θ_{ij}	0.002	0.63%	0.007	0.007	95.0%	NA
M-MR	-0.004	-1.47%	0.008	0.008	91.0%	NA
M-SEM	-0.022	-7.31%	0.007	0.008	91.9%	0.8%*
EAP-MR	-0.004	-1.42%	0.008	0.008	91.6%	NA
EAP-SEM	-0.020	-6.78%	0.008	0.008	92.1%	1.4%*
PV-MR	-0.006	-2.07%	0.008	0.008	91.8%	NA
PV-SEM	0.018	6.08%	0.009	0.009	95.2%	0.0%
Simulation 3.1: Small sampling fraction (0.05)						
True θ_j & θ_{ij}	0.000	0.00%	0.006	0.006	96.4%	NA
M-MR	-0.026	-8.79%	0.007	0.008	92.4%	NA
M-SEM	-0.041	-13.59%	0.007	0.008	92.2%	0.0%*
EAP-MR	-0.026	-8.71%	0.007	0.007	92.4%	NA
EAP-SEM	-0.039	-12.87%	0.007	0.008	91.4%	2.4%*
PV-MR	-0.028	-9.49%	0.007	0.008	92.0%	NA
PV-SEM	-0.002	-0.79%	0.008	0.008	96.4%	0.0%

Table B.8: Organizations & networks scale item location and fit, unadjusted one-parameter model

Item stem	Threshold 1	Threshold 2	Item fit (p value)	Item step fit (p value)
School groups	-0.92	-0.52	0.98 (0.157)	0.87 (<0.001)
Political groups	-0.06	0.15	1.29 (<0.001)	1.14 (<0.001)
Youth groups	0.29	0.49	1.10 (0.001)	0.97 (0.096)
Sports groups or leagues	-0.95	-0.67	0.90 (0.005)	0.87 (<0.001)
Women's groups / groups that aim to support women	-0.11	0.11	0.97 (0.106)	0.88 (<0.001)
Men's groups or groups that aim to support men	1.19	1.37	1.16 (0.001)	1.06 (0.069)
Church groups	-1.13	-0.86	0.88 (0.002)	0.84 (<0.001)
Cultural groups	-0.48	-0.22	1.02 (0.144)	0.98 (0.155)

Thresholds are the locations at which an individual has a probability of 50% of selecting that response category or a higher one. Threshold 1 is thus the location at which a respondent has a 50% probability of selecting A little important or Very important versus Not important / no such groups; Threshold 2 is the location at which a respondent has a 50% probability of selecting Very important. Higher thresholds indicate 'harder' items or steps – response options that are less likely to be selected relative to lower levels of endorsement on that item.

Table B.9: Organizations & networks scale item discrimination, location, and fit, unadjusted two-parameter model

Item stem	Item discrimination (step discrimination)	Threshold 1	Threshold 2	Item fit (p value)	Item step fit (p value)
School groups	-1.83 (2.24)	-12.00	-0.41	0.94 (0.016)	0.99 (0.198)
Political groups	-1.57 (1.03)	0.14	0.20	1.00 (0.216)	1.01 (0.199)
Youth groups	-1.80 (1.73)	0.51	0.53	0.97 (0.068)	1.00 (0.237)
Sports groups or leagues	-2.17 (3.59)	-12.00	-0.47	0.95 (0.045)	0.99 (0.213)
Women's groups / groups that aim to support women	-2.13 (2.55)	0.11	0.12	0.96 (0.061)	0.99 (0.202)
Men's groups / groups that aim to support men	-1.45 (1.29)	1.73	1.73	0.99 (0.210)	1.01 (0.208)

Church groups	-3.28 (3.35)	-12.00	-0.63	0.95 (0.061)	1.00 (0.248)
Cultural groups	-1.46 (2.28)	-12.00	-0.16	0.94 (0.010)	0.99 (0.213)

Figure B.9: Explanation for reduction in bias controlling for $\overline{W}_{.j}$ when W_{ij} distorts measurement

For the purpose of this illustration, we remove the item model and assume that the latent trait θ_{ij} is observed as X^*_{ij} without the need to calculate from items back to person scores. The structural models are:

$$U_{W_{ij}} \sim Binom(0.25)$$
$$W_{ij} = U_{W_{ij}}$$
$$\overline{W}_{.j} = \frac{1}{n}\sum_{i=1}^{n_i} W_{ij}$$
$$U_{\theta_j} \sim N(0, \tau^2)$$
$$\theta_j = U_{\theta_j}$$
$$U_{\theta_{ij}} \sim N(0,\ \rho^2)$$
$$\theta_{ij} = \beta_1\theta_j + U_{\theta_{ij}}$$
$$X^*_{ij} = \beta_4\theta_{ij} + \beta_5 W_{ij}$$

We decompose $\bar{X}^*_{.j}$ and express it in terms of the quantity it is intended to capture, θ_j:

$$\bar{X}^*_{.j} = \frac{1}{n}\sum_{i=1}^{n} X^*_{ij}$$
$$\bar{X}^*_{.j} = \frac{1}{n}\sum_{i=1}^{n}\left(\beta_4\theta_{ij} + \beta_5 W_{ij}\right)$$
$$\bar{X}^*_{.j} = \frac{1}{n}\sum_{i=1}^{n}\left(\beta_4\theta_{ij}\right) + \frac{1}{n}\sum_{i=1}^{n}\left(\beta_5 W_{ij}\right)$$
$$\bar{X}^*_{.j} = \beta_4\frac{1}{n}\sum_{i=1}^{n}\left(\theta_{ij}\right) + \beta_5\overline{W}_{.j}$$
$$\bar{X}^*_{.j} = \beta_4\frac{1}{n}\sum_{i=1}^{n}\left(\beta_1\theta_j + N(0,\rho^2)\right) + \beta_5\overline{W}_{.j}$$
$$\bar{X}^*_{.j} = \beta_4\left(\beta_1\theta_j + N\left(0, {}^{\rho^2}/_{n}\right)\right) + \beta_5\overline{W}_{.j}$$
$$\bar{X}^*_{.j} = \beta_4\beta_1\theta_j + \beta_4\left(N\left(0, {}^{\rho^2}/_{n}\right)\right) + \beta_5\overline{W}_{.j}$$

$$\bar{X}^*_{.j} - \left(\beta_4\left(N\left(0, {}^{\rho^2}/_{n}\right)\right) + \beta_5\overline{W}_{.j}\right) = \beta_4\beta_1\theta_j$$

$$\frac{\bar{X}_{\cdot j}^{*} - \left(\beta_4\left(N\left(0, {}^{\rho^2}/_{n}\right)\right) + \beta_5\overline{W}_{\cdot j}\right)}{\beta_4\beta_1} = \theta_j$$

In a general case, the difference between the proxy $\bar{X}_{\cdot j}^{*}$ and θ_j is a function of:

- β_1 - the path coefficient between θ_j and θ_{ij}
- β_4 - the path coefficient between θ_{ij} and X_{ij}^{*}
- Sample variance of θ_{ij}
- β_5 - the path coefficient between W_{ij} and X_{ij}^{*}
- Aggregate $\overline{W}_{\cdot j}$.

In the specific simulations in this research, $\beta_1 = \beta_4 = 0$, so this difference simplifies to:

$$N\left(0, {}^{\rho^2}/_{n}\right) + \beta_5\overline{W}_{\cdot j}$$

C. APPENDIX TO CHAPTER 4

Table C.1: Intervention engagement questionnaire

Item	Response options
335 Heard / seen OMC	Dichotomous (Yes/no)
336 Spoken with OMC rep	Dichotomous (Yes/no)
337 OMC rep came to my house	Dichotomous (Yes/no)
338 Received OMC literature	Dichotomous (Yes/no)
339 Received condoms from OMC	Dichotomous (Yes/no)
340 / 347 Attended or seen an OMC talk or discussion group	Dichotomous (Yes/no)
342 Number of times participated in OMC workshops	Categorical: 0, 1, 2-6, 7+
343 Know of an OMC CAT in this village	Dichotomous (Yes/no)
344 Member of OMC CAT	Dichotomous (Yes/no)
345 Seen OMC performance	Dichotomous (Yes/no)
346 Been to an OMC video presentation	Dichotomous (Yes/no)
348 Seen any OMC murals in village	Dichotomous (Yes/no)
349 Participated in any way to create mural	Dichotomous (Yes/no)
350 Participated in Red Card Campaign with OMC	Dichotomous (Yes/no)
351 Participated in OMC soccer	Dichotomous (Yes/no)
352 Been to OMC protest / rally	Dichotomous (Yes/no)
357 Leaders spoken to you about OMC	Dichotomous (Yes/no)
355 Feel part of OMC	Dichotomous (Yes/no)

OMC: One Man Can; CAT: community action team