Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____
Kevin Payton Delaney   3/31/2015

Approval Sheet
Describing and leveraging interaction effects for HIV prevention

By
Kevin Payton Delaney
Doctor of Philosophy
Epidemiology

_____ [Advisor's signature]
Patrick S. Sullivan, Phd, DVM
Advisor

_____ [Member's signature]
W. Dana Flanders, MD PhD
Committee Member

_____ [Member's signature]
Michael Kramer, PhD
Committee Member

_____ [Member's signature]
Lance Waller, PhD
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____ Date

Describing and leveraging interaction effects for HIV prevention

By

Kevin Payton Delaney
Master of Public Health, State University of New York at Albany, School of Public Health 2000
Bachelor of Science, The George Washington University, 1998

Advisor: Patrick S. Sullivan, PhD, DVM

An abstract of
A dissertation submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Epidemiology 2015

Abstract
Describing and leveraging interaction effects for HIV prevention
By Kevin Payton Delaney, MPH

Interaction can occur at many levels. People interact in conversations, sexual relations or virtually through the internet. In individuals, sometimes one factor can combine with another to cause disease that wouldn't have been observed otherwise; epidemiologists refer to this as causal interaction. In infectious disease epidemiology there is another level of interaction, in that the infected and uninfected populations must interact to transmit disease, and, because of this, at the population level public health interventions can also interact.

In my first study I collected data from 2,666 user profiles of men who use a social networking application, mostly to meet other men for sex. Overlapping circles defined by the geolocation data I extracted from the app covered the entire 132.4 square miles in the City of Atlanta and were analyzed with spatial statistics to highlight areas with higher densities of minority and young minority users. This simple method can describe the spatial density of users of a sexual networking app for future behavioral surveys and to identify areas of highest need for targeting prevention resources.

In the second study I used simulated data to compare 10 tests for statistical interaction and contrast these with 3 tests for causal interaction. I found that, at sample sizes typical for epidemiologic studies, the power to detect interaction is limited unless exposures have both strong individual effects and their combined effects are closer to multiplicative than additive. The power is even lower for tests specifically designed to detect causal interaction.

The aim of my third study was to describe population level interactions of interventions associated with HIV testing, in a model of the sexual networks of gay men. I found more frequent HIV testing will not result in reduced HIV incidence unless combined with improvements in effective HIV care. Only once care and viral suppression become the normative outcome of HIV diagnosis does additional focus on increasing HIV testing as the gateway to this outcome become warranted.

These three very different studies evaluate interaction on several levels. Together they emphasize the importance of studying and leveraging interaction effects for HIV prevention.

Describing and leveraging interaction effects for HIV prevention

By

Kevin Payton Delaney
Master of Public Health, State University of New York at Albany, School of Public Health, 2000
Bachelor of Science, The George Washington University, 1998

Advisor: Patrick S. Sullivan, PhD, DVM

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Epidemiology
2015

Contents

Page intentionally left blank

Table of Tables and Figures

Chapter 3

Chapter 4

Chapter 5

Additional background on infectious disease modelling

Manuscript submitted to Lancet HIV

Chapter 1: Overview, Objective and Specific Aims

Interaction

Merriam-Webster defines the term interaction as "mutual or reciprocal action or influence"

which is great because, there are four potential definitions in that conglomerate of terms. Interaction can occur at many levels, individuals can interact in personal conversations, sexual relations or only virtually through the internet. Different factors can interact in individuals, sometimes one factor can combine with another to lead to an effect that wouldn't have been observed from either factor alone[1,2]. This can happen at the population level as well, when public health interventions can interact to produce additional benefits[3-6]. In my dissertation I propose to evaluate interaction on several levels in three different studies. All of these use novel epidemiologic research methods to study different aspects of interactions, and apply information about interaction to HIV prevention.

Social networking websites and applications represent novel means for individual interaction. A variety of new social networking tools are now available for most "smart" phones. Many of these applications are specifically designed for men who have sex with men (MSM) to meet each other, often to engage in anonymous sex[7]. Combined applications designed for this purpose have more than 6 million users and 10,000 new users added daily.[7-11] Many of these applications build their services on the ability to use the geolocation features available on most phones and other communication devices

(iPods, iPads, and tablets) to provide location information for other application users, including their geographic proximity (in feet or miles) to the user's location.[8-11] Thus apps represent a technological advance in social interaction because they tell users how far away each other user and potential sex partner currently is, and update in real time. Social interactions on such apps occur in both space and time, in a way that was not possible even 10 years ago.

Statisticians have very specific mathematical formulas that they use to describe interaction, defined statistically as when the effects of two exposures are different when both are present than what would be expected based on adding or multiplying the effects of each exposure individually.[12,13] Epidemiologists generally hope to limit this definition to interactions of causal effects, pointing out that two factors can "interact" statistically, without acutally meeting the definition of causal interaction.[1,2,14-17] Although the concept of statistical interaction was described more than 3 decades ago,[12] and two models of statistical interaction were evaluated as far back as 1983,[13] there has a been a recent growth in the number of models and methods used to detect both statistical[18-23] and causal[14-17] interactions.

In infectious disease epidemiology there is another level of interaction, in that the infected and uninfected populations must "interact" to transmit disease. It was recently pointed out that this type of interaction is mathematically synonymous with statistical interaction[5], but occurs at the population level. In general infectious diseases are different from other diseases because exposures do not occur independent of one another.[2-5] For example, one person infected with the flu can cough on a train car or in

an elevator and expose many people all at once.   Conversely, my cholesterol does not in and of itself put you at risk for a heart attack, although this is of course relative as if you had lunch with me every day for 30 years our outcomes would likely be correlated. Conversely, if an intervention, such as vaccination, is applied in the community, every individual does not necessarily need to receive the intervention to receive protection. This is the concept of herd immunity.[3] Additionally the effects of two community level interventions may be greater than their individual effects. This concept is commonly referred to in the infectious disease modelling literature as "synergy."[6]

Novel methods for HIV Prevention

In a 2010 survey of pregnant women in South Africa[24] we collected data using a cell-phone based survey, an example of how this type of technology can be used to collect public health information. Even in Sub-Saharan Africa the majority of individuals have access to mobile phones and/or the internet[25], and methods for public health practitioners to use mobile and social networking technology to both collect and disseminate information about how populations at risk for HIV interact in these media is an area of growing research.[25-27]

For interactions of effects within a person we have identified 13 different proposed tests for interaction that have appeard in the literature in the last 5 years.[14-23] The theoretical framework and mathematical proofs provided by Vanderweele, Robins and Hernan have definitely sparked this advance.[14-17] The rapid expansion of proposed methods for estimating these quantities is likely also due to the availability of software packages that

can fit models necessary to quantify them.[14-23] However, this rapid advance comes with

many caveats and comments as to the appropriate study design and methods to identify

interaction effects.[28-30] Two recent HIV prevention studies I co-authored[24,31] found

significant interaction effects leading me to seek practical applications of the newly

proposed methods as guidance for how to evaluate interaction. Despite descriptions of

how to report causal interaction,[32-33] there has been surprisingly little in the way of

epidemiologic literature providing a practical guide to the performance, assumptions or

appropriate use of the abundance of tests proposed to detect interaction. Although

early work on interaction directly compared two of the many statistical tests[13], there

has not been a recent update that compares and contrasts the newer tests for detection

of additive interaction directly.

Despite the lack of guidance in this one area of growth resulting from advances in

computing power, the power of computers for complex data analysis is proving useful in

other areas of Infectious disease epidemiology. The complexity of describing even a

simple system of infectious disease transmission through epidemiologic methods grows

as the number of contacts and exposures change in a population over time.   The simple

systems of one or two household members experiencing a single-point exposure[3,5] do

not expand well to the study of airborne or sexually transmitted diseases particularly

among MSM. Although in theory it is possible to draw a causal diagram depicting the

transmission dynamics in a population would require a near infinite number of vertices

and edges to illustrate the changes in population probabilities of exposure over time[34].

Luckily the technologic advances of modern computing have provided a variety of

alternative models for infectious disease transmission, including compartmental or mean-field models that describe infectious disease transmission through a system of differential equations[35-36] as well as individual or agent-based models that use a series of rules to describe individual behavior and use computerized simulations to assess the effects of interventions at both individual and the population level.[36-38] These computationally intensive models that track invididuals over time have recently[39] been combined with models that describe network characteristics[40-43] to develop models that can be used to model sexual partnership formation and disease transmission through a network of sexual contacts over time. Advances in computing power that have occurred even over the time I have been working on my dissertation have cut the time to run these types of models from days to hours.

Objective and Specific Aims

Overall the objective of this dissertation is to use novel epidemiologic methods to study the concept of interaction at three different levels: social interaction, statistical and causal interactions in individuals, and the interaction or synergies that can occur in population level intervantions. All of these interactions are studied within the contect of the HIV epidemic, and we seek to describe and take advantage of interaction effects to inform HIV prevention efforts in the United States and abroad.

In order to evaluate different types of interaction with these novel technologies we have developed three specific areas of research in the hopes of advancing HIV prevention.

Specific Aim 1: Use the geolocation features of the a social networking application as a novel approach to calculating the population density of a population at high-risk for HIV infection.

Specific Aim 2: Compare proposed tests for statistical interaction of two exposures using simulated data in which two dichotomous variables interact to have effects that are greater than additive risk differences but less than multiplicative risk ratios, indicative of sufficient component cause interaction

Specific Aim 3: Use modern infectious disease modeling techniques to describe synergy between HIV testing interventions for men who have sex with men in the United States

The rest of this dissertation is presented in the following sections. In Chapter 2, we present an overview of the HIV epidemic, and HIV prevention generally. Then we

describe each of the three specific aims of this dissertation, providing an aim specific

background, study methods and findings in Chapters 3-5. Chapter 6 provides an overall

summary describing how the findings relate and next steps for these areas of research.

References

1) Rothman KJ. Causes. Am J Epidemiol. 1976;104:587–592.

2) Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott-Williams & Wilkins; 2012.

3) Hudgens MG, Halloran ME. Towards causal inference with interference. J Am Stat Assoc. 2008;103:832–842.

4) Halloran ME, Struchiner CJ. Causal inference for infectious diseases. Epidemiology. 1995;6:142–151.

5) VanderWeele TJ, Vandenbroucke JP, Tchetgen EJ, Robins JM. 26. A mapping between interactions and interference: implications for vaccine trials. Epidemiology. 2012 Mar;23(2):285-92.

6) Dodd PJ, White PJ, Garnett GP. Notions of synergy for combinations of interventions against infectious diseases in heterogeneously mixing populations. Math Biosci. 2010 Oct;227(2):94-104.

7) Rice E. Sex on Grindr: Young Men Who Have Sex with Men and Geo-Locating Social Networking Technology. ISIS Annual Conference on New Media, Youth and Sexual Health, SexTech 2012; 2012 March 31-April 2; San Francisco, CA.

8) www.grindr.com Archived at: http://www.webcitation.org/6LodfWVwq Accessed: 12/12/13.

9) http://www.jackd.mobi/ Archived at: http://www.webcitation.org/6LoduUIDX Accessed: 12/12/2013

10) www.adam4adam.com Archived at: http://www.webcitation.org/6Loe4v0I0 Accessed: 12/12/2013

11) www.boyahoy.com Archived at: http://www.webcitation.org/6LoePJ6e1

   Accessed: 12/12/2013

12) Miettinen OS. Causal and preventive interdependence: elementary principles.

   Scand J Work Environ Health 1982;8:159 –168.

13) Greenland S. Tests for interaction in epidemiologic studies: a review and a study

   of power. Stat Med. 1983 Apr-Jun;2(2):243-51.

14) VanderWeele TJ, Robins JM. The identification of synergism in the sufficient-

   component-cause framework. Epidemiology. 2007 May;18(3):329-39.

15) Hernán MA, Robins JM. Causal Inference. Chapman & Hall/CRC; 2015.

16) VanderWeele TJ. Sufficient cause interactions and statistical interactions.

   Epidemiology. 2009 Jan;20(1):6-13.

17) VanderWeele TJ. On the distinction between interaction and effect modification.

   Epidemiology. 2009 Nov;20(6):863-71.

18) Richardson DB, Kaufman JS. Estimation of the relative excess risk due to

   interaction and associated confidence bounds. Am J Epidemiol. 2009 Mar

   15;169(6):756-60.

19) Rose S, van der Laan M. A double robust approach to causal effects in case-

   control studies. Am J Epidemiol. 2014 Mar 15;179(6):663-9.

20) Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios

   and differences. Am J Epidemiol. 2005 Aug 1;162(3):199-200.

21) Brumback BA, Bouldin ED, Zheng HW, Cannell MB, Andresen EM. Testing and

   estimating model-adjusted effect-measure modification using marginal

structural models and complex survey data. Am J Epidemiol. 2010 Nov 1;172(9):1085-91.

22) Kuss O, Schmidt-Pokrzywniak A, Stang A. Confidence intervals for the interaction contrast ratio. Epidemiology. 2010 Mar;21(2):273-4.

23) Cheung YB. A Modified Least-Squares Regression Approach to the Estimation of Risk Difference. Am. J. Epidemiol. 2007; 166(11): 1337-1344. Greenland S. Tests for interaction in epidemiologic studies: a review and a study of power. Stat Med. 1983 Apr-Jun;2(2):243-51.

24) Goga AE, Dinh TH, Jackson DJ, Lombard C, Delaney KP, et al for the South Africa PMTCT Evaluation (SAPMCTE) Team. First population-level effectiveness evaluation of a national programme to prevent HIV transmission from mother to child, South Africa. J Epidemiol Community Health. 2014 Nov 4. pii: jech-2014-204535.

25) Zurovac D, Otieno G, Kigen S, Mbithi AM, Muturi A, Snow RW, Nyandigisi A. Ownership and use of mobile phones among health workers, caregivers of sick children and adult patients in Kenya: cross-sectional national survey. Global Health. 2013 May 14;9:20.

26) Nsanzimana S, Ruton H, Lowrance DW, Cishahayo S, Nyemazi JP, Muhayimpundu R, Karema C, Raghunathan PL, Binagwaho A, Riedel DJ. Cell phone-based and internet-based monitoring and evaluation of the National Antiretroviral Treatment Program during rapid scale-up in Rwanda: TRACnet, 2004-2010. J Acquir Immune Defic Syndr. 2012 Feb 1;59(2):e17-23.

27) van Heerden A, Norris S, Tollman S, Richter L, Rotheram-Borus MJ. Collecting maternal health information from HIV-positive pregnant women using mobile phone-assisted face-to-face interviews in Southern Africa. J Med Internet Res. 2013 Jun 10;15(6):e116.

28) Skrondal A. Commentary: much ado about interactions. Epidemiology. 2014 Sep;25(5):723-6.

29) Greenland S. Interactions in epidemiology: relevance, identification, and estimation. Epidemiology. 2009 Jan;20(1):14-7.

30) Kaufman JS. Interaction reaction. Epidemiology. 2009 Mar;20(2):159-60.

31) Delaney KP, Rurangirwa J, Facente S, Dowling T, Janson M, et al. Using a multi-test algorithm to improve the positive predictive value of rapid HIV testing and linkage to HIV care in non-clinical HIV test sites. JAIDS Submitted.

32) Knol MJ, Vanderweele TJ, Groenwold RH et al. Estimating measures of interaction on an additive scale for preventive exposures. Eur J Epidemiol 2011;(26(6) 433-438.

33) Knol MJ, Vanderweele TJ. Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol 2012; 41(2): 514-520.

34) Lauritzen SL, Richardson TS. Chain graph models and their causal interpretations. J R Stat Soc Ser B 2002; 64: 321-348.

35) Anderson RM, May RM. Infectious Diseases in Humans: Dynamics and Control. New York, New York: Oxford University Press; 1991.

36) Keeling MJ, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton, NJ: Princeton University Press; 2008.

37) Railsback SF, Grimm V. Agent-based and Individual-based Modeling: A Practical Introduction. Princeton, NJ: Princeton University Press; 2012.

38) Van Dyke Parunak H, Savit R, Rialo RL. Agent-based modeling vs. equation-based modeling: A case study and users' guide. Proceedings of the First International Workshop on Multi-Agent Systems and Agent-Based Simulation. London, UK: Springer-Verlag; 1998.

39) Krivitsky PN, Handcock MS. A separable model for dynamic networks. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2014; 76.1: 29-46.

40) Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. statnet: Software tools for the Statistical Modeling of Network Data. Seattle, WA; 2003. Version 2.0. Available from: http://statnetproject.org.

41) Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M (2008). statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. Journal of Statistical Software, 24(1). Available from: http://www.jstatsoft.org/v24/i01/paper Accessed: 11/20/14

42) Goodreau SM, Handcock MS, Hunter DR, Butts CT, and Morris M. A statnet Tutorial. Journal of Statistical Software. 24(9). Available from: http://www.jstatsoft.org/v24/i09/paper Accessed 11/20/2014

43) Jenness SM, Goodreau SM, Morris M. EpiModel – Mathematical Modeling of

Infectious Disease Version 1.1.1, A Tutorial. Available from: www.epimodel.org

Accessed 11/20/14

Chapter 2: The HIV Epidemic

After more than three decades since the identification of the Human Immunodeficiency virus (HIV) as the causative agent of the Acquired Immune Deficiency syndrome (AIDS), transmission of HIV infection is still a major health problem.[1-4] This despite the fact that modes of virus transmission are well-defined.[1-5] One interesting aspect of the worldwide HIV epidemic is that its characteristics vary widely throughout the world.[1] Heterosexual sex is the most common exposure pathway in sub-Saharan Africa and most of Southeast Asia. However, in parts of Southeast Asia, Eastern Europe and the former Soviet Republics exposure through injection drug use is thought to be the source of sustained and increasing HIV transmission. In the United States, men who report having sex with another man (MSM) as their main risk for HIV infection have been, and continue to be, the most heavily impacted risk group.[3-4]

According to the US Centers for Disease Control and Prevention the MSM risk group represented the highest percentage of both new diagnoses and prevalent (previously diagnosed) infection, and the only group in the United States for whom HIV incidence is estimated to be increasing.[3-4] The CDC reported that 63% of all incident HIV infections estimated to have occurred in 2010 in the US occurred in MSM,[3] despite the fact that this group is estimated to account for less than 2% of the US population.[6] White MSM accounted for 11,200 new HIV infections while black/African American MSM accounted for 10,600 in 2010,[3] even though there are more than 5 times as many white men as black men in the US.[7] Whereas new HIV infections were relatively stable among MSM overall

from 2007–2010, they increased 22% among young MSM; among young MSM black men accounted for 55% of all new infections in 2010.[3]

Among heterosexuals, the risk of new HIV infection is highest among African American (AA) women.[3] Much like the disparities among African American MSM, in 2010 the rate of new HIV infections among black women was 20 times that of white women, and over 4 times the rate among Hispanic/Latina women.[3] Some women become infected because they may be unaware of a male partner's risk factors for HIV infection. AA MSM are more likely to report having had sex with a woman in the past year than are MSM of other races, although MSM who also have sex with women (MSMW) have been reported to have lower risk for and prevalence of HIV infection than MSM only (MSMO).[8] Relationship dynamics may also play a role in the increased risk for AA women. For example, some women may not insist on condom use because they fear that their partner will physically abuse or leave them.[9] Among all races combined, it is estimated that more than 220,000 women in the United States are infected with HIV.[4] Nearly one out of four of these women don't know they have HIV. This puts them at high risk of passing the virus to their babies.[10]

Women can pass HIV to their babies during pregnancy, while the baby is being delivered, or through breast-feeding.[1] Mother-to-child transmission is the most common way children become infected with HIV.[1] Nearly all AIDS cases in U.S. children are because of mother-to-child transmission.[4] Because black women are disproportionately affected by HIV, the rate of HIV diagnosis per 100,000 live births is also markedly higher in the black population than among whites.[4]

HIV Epidemic Worldwide

Overall, the US HIV epidemic, with an estimated 1.2 million people infected (<0.02% of the population) pales in comparison to the global HIV epidemic. In many sub-Saharan African countries more than 20% of the total population is infected with HIV.[1] This is part of the reason that it appears the US HIV epidemic, concentrated in MSM, is so different from that in the rest of the world. It is estimated that, globally 2-4% of the male population has had sex with another man.[11] Worldwide, prevalence rates of HIV are consistently an order of magnitude higher for MSM than other populations.[11] It is believed that this disparity is driven by the increased probability of HIV transmission for receptive anal sex relative to insertive anal sex or vaginal sex.[10-12] However, in countries with so-called "generalized" HIV epidemics,[1] much more of the overall population is infected, e.g. if 10% of a 98% heterosexual population has HIV and 30% of a 2% MSM population has HIV, then overall there would be 9.8% of the total population represented by heterosexual infections, compared to 0.67% of the population being infected MSM. In the US where only 0.02% of the heterosexual population is infected, it is clear why MSM represent the majority of the epidemic.

In much of the rest of the world however, the HIV epidemic is centered in the heterosexual population. WHO estimated that, in 2013, 35 million people were infected with HIV worldwide, with 5700 new infections occurring each day. Most of these were in low or middle income countries, and nearly half (48%) among women. Nearly 71% of all HIV infected persons live in sub-Saharan Africa. The disparity is even more striking when you consider infections among children. One in seven new infections in 2013

were estimated to occur in children <15 years of age, the vast majority of these

occurring due to mother to child transmission during pregnancy or the first year of life.

Of an estimated 3.2 million children <15 years infected 91% live in sub-saharan Africa.[1]

This occurs primarily because of the nearly 1.5 million HIV-infected women who

become pregnant annually worldwide, all of whom need antiretroviral medications for

prevention of mother to child transmission, less than half receive these drugs that have

been shown in a variety of clinical trials to reduce the risk of transmission from mother

to infant significantly.[1]

HIV Prevention Strategies/Interventions

This failure of the public health and healthcare infrastructure to provide even the

simplest of HIV therapies to people who need it argues against the effectiveness of what

has been touted as the next great innovation in HIV prevention. Granich[13] recently used

a mathematical model to suggest that significant reductions in transmissions and

therefore incident HIV cases would be gained from identifying all HIV-infected

individuals, and starting them on anti-retroviral therapy to reduce the concentration of

circulating HIV virus to levels would make transmission to HIV-negative sex partners

unlikely if not impossible. Cohen[14] then implemented this strategy in randomized

controlled trial within a cohort of heterosexual couples in which one of the two was HIV-

infected and the other was HIV-uninfected. This study found that, of the 28

transmissions that could be virologically linked to the enrolled partner, only 1 occurred

in the group receiving HIV drug therapy. This lead to rapid expansion from the

mathematical model, to proof of concept to global recommendations.[13-16] In 2013 WHO

revised recommendations for HIV therapy such that they now recommend that ART be

initiated for all patients with CD4 ≤500 cells/ mm3, and initiated immediately regardless

of CD4 for children up to five years old, people with active TB or coinfected with

hepatitis B virus with severe chronic liver disease and people living with HIV in

serodiscordant partnerships.[16] Many countries in Africa have begun to offer therapy to

their population, but as we have seen for HIV-infected mothers, coverage at even the

lower threshold of clinically indicated therapy has been challenging.[1] Even in the United

States it is estimated that only 35% of HIV-infected persons currently have viral

suppression to the point of an undetectable HIV viral load.[17-18]

An alternative to early-initiation of HIV therapy is the use of antiretroviral medication by

uninfected persons to prevent infection, called pre-exposure prophylaxis.[19] This

strategy has recently been shown to be effective in reducing HIV incidence among gay

men,[20] and discordant heterosexual couples,[21-22] but was not effective in a clinical trial

which enrolled high –risk women.[23] Despite these conflicting results, the CDC recently

released guidelines for clinician with recommendations for how PreP should be used in

the US.[24] Because this intervention targets uninfected persons at high-risk of acquiring

HIV, an even greater number of individuals could potentially require both ongoing HIV

therapy, and frequent medical visits for monitoring while they are taking antiretrovirals

for PreP.

For pregnant women, WHO has recently revised their guidelines for treatment, so that

HIV-infected pregnant women should be started on a combination of antiretrovirals as

early as 14 weeks of pregnancy.[16] To implement such an intervention requires that all

pregnant women who are unaware of their HIV status be offered an HIV test, ideally

early in pregnancy. However, WHO estimates that globally on 54% of HIV-infected

pregnant women received an HIV test in 2013.[1] Recently there have been multiple

reports of pregnant women testing negative for HIV antibodies early in pregnancy, only

to be identified as HIV-infected after their baby is born through infant testing.[25-27] This

suggests that these women were infected later in their pregnancy than when they were

offered their routine HIV test during pre-natal care. As a result WHO recommends

rescreening pregnant women at or after 32 weeks, in order to identify the maximum

number of women infected in time to start antiretroviral therapy before they deliver.[28]

Still, many women do not receive this test, and as a result their infants are needlessly

exposed to HIV.[25]

For all of these biomedical strategies HIV diagnosis and linkage to HIV medical care are

necessary but not sufficient first steps. In the United States HIV testing is recommended

during pregnancy, at least once in their lifetime for the entire US population, and

annually for MSM.[29,30] Recent data suggests that the uptake of testing is increasing,

particularly among black MSM, but most persons newly diagnosed with HIV infection

have been infected for over 5 years prior to their diagnosis.[31-34] There is ongoing debate

as to the appropriate testing frequency for MSM, with a CDC workgroup recently

concluding that there was insufficient evidence to recommend testing more frequently

than annually.[29,30,35-41] In spite of this controversy, the US still does much better with HIV

testing than other parts of the world; in the US it is estimated that 18% of the infected

population is unaware of their infection,[42] internationally as many as 50% of the

infected population remains undiagnosed.[1] However, many people with diagnosed

infection do not receive appropriate HIV care, and treatment required to achieve viral

suppression and reduce transmission risk to others.[2,17,18,42]

References:

1) UNAIDS. The Gap Report: The Beginning of the End of the AIDS Epidemic. Geneva; UNAIDS: 2014. Available from: http://www.unaids.org/sites/default/files/media_asset/UNAIDS_Gap_report_en.pdf

2) CDC. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas—2012. *HIV Surveillance Supplemental Report* 2014;19(No.3). Published November 2014. Available from: http://www.cdc.gov/hiv/library/reports/surveillance/

3) CDC. Estimated HIV incidence in the United States, 2007–2010. *HIV Surveillance Supplemental Report* 2012;17(No. 4). Published December 2012. Available from: http://www.cdc.gov/hiv/library/reports/surveillance/index.html#panel1

4) CDC. *HIV Surveillance Report*,2013; vol. 25. Published February 2015. Available from: http://www.cdc.gov/hiv/library/reports/surveillance/2013/surveillance_Report_vol_25.html

5) American Public Health Association. Control of Communicable Disease Manual, 20th Edition: David Heymann, Editor. Washington, DC: American Public Health Association Publications; 2014.

6) Purcell DW, Johnson CH, Lansky A, Prejean J, Stein R, Denning P, Gau Z, Weinstock H, Su J, Crepaz N. Estimating the population size of men who have

sex with men in the United States to obtain HIV and syphilis rates. Open AIDS J. 2012;6:98-107.

7) US Census Bureau. 2010 US Census. Available from:

http://www.census.gov/2010census/data/

8) Friedman MR, Wei C, Klem ML, Silvestre AJ, Markovic N, Stall R. HIV infection and sexual risk among men who have sex with men and women (MSMW): a systematic review and meta-analysis. PLoS One. 2014 Jan 30;9(1):e87139.

9) Paxton KC, Williams JK, Bolden S, Guzman Y, Harawa NT. HIV Risk Behaviors among African American Women with at-Risk Male Partners. J AIDS Clin Res. 2013 Jul 25;4(7):221.

10) Patel P, Borkowf CB, Brooks JT, Lasry A, Lansky A, Mermin J. Estimating per-act HIV transmission risk: a systematic review. AIDS. 2014 Jun 19;28(10):1509-19.

11) Beyrer C, Sullivan PS, Sanchez J, Dowdy D, Altman D, Trapence G, Collins C, Katabira E, Kazatchkine M, Sidibe M, Mayer KH. A call to action for comprehensive HIV services for men who have sex with men.

12) Sullivan PS, Carballo-Diéguez A, Coates T, Goodreau SM, McGowan I, Sanders EJ, Smith A, Goswami P, Sanchez J. Successes and challenges of HIV prevention in men who have sex with men. Lancet. 2012 Jul 28;380(9839):388-99.

13) Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. Lancet. 2009 Jan 3;373(9657):48-57.

14) Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, HPTN 052 Study Team. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med. 2011 Aug 11;365(6):493-505.

15) Cohen MS, Holmes C, Padian N, Wolf M, Hirnschall G, Lo YR, Goosby E. HIV treatment as prevention: how scientific discovery occurred and translated rapidly into policy for the global response. Health Aff (Millwood). 2012 Jul;31(7):1439-49.

16) World Health Organization. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach June 2013. Geneva; World Health Organization: 2013.

17) Bradley H, Hall HI, Wolitski RJ, Van Handel MM, Stone AE, et al. Vital Signs: HIV diagnosis, care, and treatment among persons living with HIV--United States, 2011. MMWR Morb Mortal Wkly Rep. 2014 Nov 28;63(47):1113-7.

18) Rosenberg ES, Millett GA, Sullivan PS, Del Rio C, Curran JW. Understanding the HIV disparities between black and white men who have sex with men in the USA using the HIV care continuum: a modeling study. Lancet HIV. 2014 Dec;1(3):e112-e118.

19) Cohen MS, Muessig KE, Smith MK, Powers KA, Kashuba AD. Antiviral agents and HIV prevention: controversies, conflicts, and consensus. AIDS. 2012 Aug 24;26(13):1585-98.

20) Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med. 2010;363:2587–2599.

21) Thigpen MC, Kebaabetswe PM, Paxton LA, Smith DK, Rose CE; TDF2 Study Group. Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana. N Engl J Med. 2012 Aug 2;367(5):423-34.

22) Baeten JM, Donnell D, Ndase P, Mugo NR, Campbell JD; Partners PrEP Study Team. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. N Engl J Med. 2012 Aug 2;367(5):399-410.

23) Van Damme L, Corneli A, Ahmed K, Agot K, Lombaard J; FEM-PrEP Study Group. Preexposure prophylaxis for HIV infection among African women. N Engl J Med. 2012 Aug 2;367(5):411-22.

24) Centers for Disease Control and Prevention. Pre-exposure prophylaxis for the prevention of HIV infection in the United States—2014: a clinical practice guideline. www.cdc.gov/hiv/pdf/guidelines/PrEPguidelines2014.pdf.

25) Dinh TH, Delaney KP, Goga A, Jackson D, Lombard C, et. al. Impact of maternal HIV seroconversion during pregnancy on early HIV vertical transmission, South Africa 2011-2012. Plos One Accepted January 2015.

26) Moodley, D., et al., High HIV incidence during pregnancy: compelling reason for repeat HIV testing. AIDS, 2009. 23(10): p. 1255-1259

27) Moodley, D., et al., Incident HIV Infection in Pregnant and Lactating Women and Its Effect on Mother-to-Child Transmission in South Africa. Journal of Infectious Diseases, 2011. 203(9): p. 1231-1234.

28) World Health Organization. Guidance on provider-initiated HIV testing and counselling in health facilities. Geneva: World Health Organization; 2007.

29) CDC. Revised recommendations for HIV testing of adults, adolescents and pregnant women in health-care settings. MMWR 2006; 55(No. RR-14).

30) CDC. HIV Testing Among Men Who Have Sex with Men --- 21 Cities, United States, 2008. MMWR 60(21);694-699. Available from: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6021a3.htm

31) Wejnert C, Le B, Rose CE, Oster AM, Smith AJ, Zhu J; Gabriela Paz-Bailey for the NHBS Study Group. HIV infection and awareness among men who have sex with men-20 cities, United States, 2008 and 2011. PLoS One. 2013 Oct 23;8(10):e76878.

32) Cooley LA, Oster AM, Rose CE, Wejnert C, Le BC, Paz-Bailey G; NHBS Study Group. Increases in HIV testing among men who have sex with men--National HIV Behavioral Surveillance System, 20 U.S. Metropolitan Statistical Areas, 2008 and 2011. PLoS One. 2014 Sep 2; 9(9):e104162. doi: 10.1371/journal.pone.0104162. eCollection 2014.

33) Chandra A, Billioux VG, Copen CE, Balaji A, DiNenno E. HIV testing in the U.S. household population aged 15-44: data from the National Survey of Family Growth, 2006-2010. Natl Health Stat Report. 2012 Oct 4;(58):1-26.

34) Hall HI, Song R, Szwarcwald CL, Green T. Time from infection with the human immunodeficiency virus to diagnosis, United States. J Acquir Immune Defic Syndr. 2015 Feb 24.

35) Gray, R. T., et al. Increased HIV Testing Will Modestly Reduce HIV Incidence among Gay Men in NSW and Would Be Acceptable if HIV Testing Becomes Convenient. PLoS ONE 8(2): 2013.

36) Lucas A, Armbruster B. The cost-effectiveness of expanded HIV screening in the United States. AIDS 2013 27(5): 795-801.

37) Wilson DP, Hoare A, Regan DG, Law MG. Importance of promoting HIV testing for preventing secondary transmissions: modelling the Australian HIV epidemic among men who have sex with men. Sexual Health 2009 6(1): 19-33.

38) Long EF HIV Screening via Fourth-Generation Immunoassay or Nucleic Acid Amplification Test in the United States: A Cost-Effectiveness Analysis. PLoS ONE 2011 6(11): e27625. doi:10.1371/journal.pone.0027625

39) Brookmeyer R, Boren D, Baral SD, Bekker LG, Phaswana-Mafuya N, Beyrer C, Sullivan PS. Combination HIV Prevention among MSM in South Africa: Results from Agent-based Modeling. PLoS One. 2014 Nov 14;9(11):e112668. doi: 10.1371/journal.pone.0112668. eCollection 2014.

40) Khanna AS, Goodreau SM, Gorbach PM, Daar E, Little SJ. Modeling the Impact of Post-Diagnosis Behavior Change on HIV Prevalence in Southern California Men Who Have Sex with Men (MSM). AIDS Behav. 2014 Aug;18(8):1523-31.

41) Khanna AS, Goodreau SM, Wohlfeiler D, Daar E, Little SJ, Gorbach PM.

Individualized diagnosis interventions can add significant effectiveness in

reducing human immunodeficiency virus incidence among men who have sex

with men: insights from Southern California. Annals of Epidemiology 2014 epub

ahead of print October 7, 2014. DOI:

http://dx.doi.org/10.1016/j.annepidem.2014.09.012

42) Skarbinski J, Rosenberg E, Paz-Bailey G, Hall HI, Rose CE, Viall AH, Fagan JL,

Lansky A, Mermin JH. Human Immunodeficiency Virus Transmission at Each Step

of the Care Continuum in the United States. JAMA Intern Med. 2015 Feb 23. doi:

10.1001/jamainternmed.2014.8180. [Epub ahead of print]

Chapter 3 - Specific Aim 1

Use the geolocation features of the a social networking application as a novel approach to calculating the population density of a population at high-risk for HIV infection.

Original Paper

Using a geolocation social networking application to calculate the population density of sex-seeking gay men for research and prevention services

Word Count: Abstract 438/450; Main Text 5193/6000

Introduction

In the US HIV epidemic, men who report have sex with men (MSM) have been, and continue to be, the most heavily impacted HIV risk group [1-2]. Although HIV incidence is increasing among MSM overall, there are pronounced disparities in both prevalence and incidence within the United States' MSM HIV epidemic by race/ethnicity. A CDC surveillance study conducted in 2008 [3] found that black non-Hispanic MSM were significantly more likely to be living with HIV than were white non-Hispanic MSM (28% vs 18%), and among those living with HIV, blacks were also significantly more likely to be unaware of their HIV infection (59% vs 26%).   The disparity in HIV prevalence is consistent with a marked difference in estimated incidence of new infections for young minority MSM. From 2006 to 2009, black MSM under age 30 experienced a 47% increase in the estimated annual number of new infections and in 2009, and there were more new infections in black MSM under age 30 than in white MSM under age 39 and more than all Hispanic MSM [4].

As a result, there is renewed emphasis [5] on identifying reasons for these disparities [6-7] and developing and providing interventions specifically for young minority MSM. However, the number of HIV prevention interventions implemented and evaluated with young minority MSM remains relatively low [8-9]. One reason for the lack of interventions specifically targeted to black MSM may be difficulty identifying a sampling frame for this population [6,8]. Stigma experienced by black MSM [10-12] may pose particular challenges in enumerating and accessing these men for provision of services

[11]. A variety of sampling methods have been developed to access hidden or marginalized populations [13-18], with varying degrees of success [17-24].

Social networking websites and applications represent novel means for individual communication. A variety of new social networking tools designed for MSM are now available for most "smart" phones [25-28] and combined these applications have more than 6 million users and 10,000 new users added daily. Many of these applications build their services on the ability to use the geolocation features available on most phones and other communication devices (iPods, iPads, and tablets) to provide location information for other application users, including their geographic proximity (in feet or miles) to the user's location. In this manuscript, we describe methodology for using the geolocation features of one of these applications as a novel approach to calculating the population density of men using the application at given times, and describe how to use this density measure to highlight areas with a high-density of minority and young minority MSM.
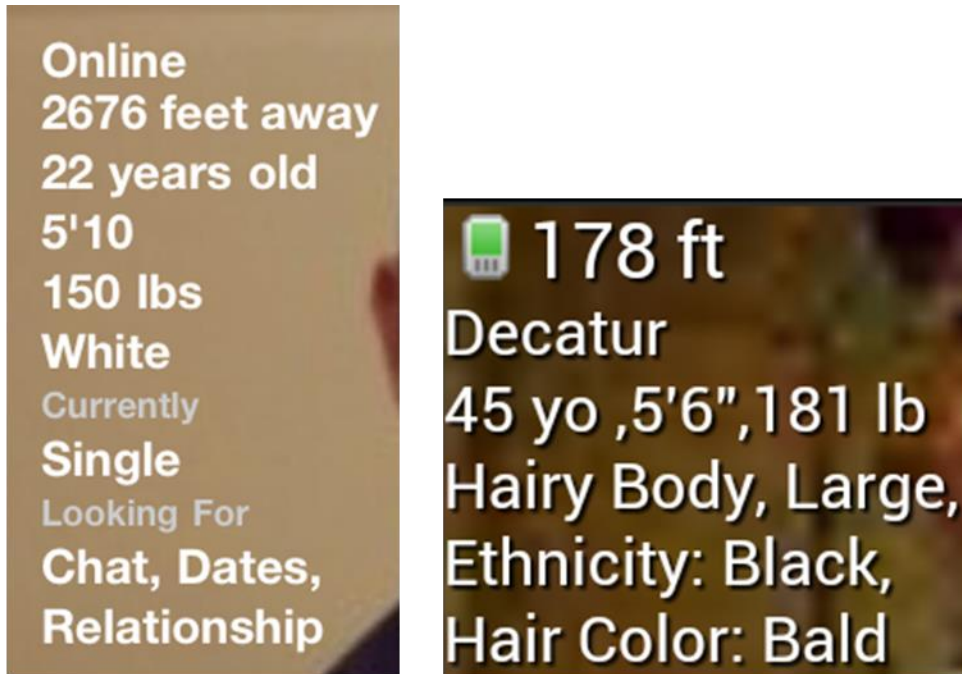
Figure 1: Examples of social networking application profile data provided by application users. For this study we extracted age, race and distance in feet from our location for use in the analysis

Methods

To pilot the study methodology, we chose a sexual networking app, and collected data from publicly available profiles at sampled locations around the City of Atlanta.

Application profiles (See Figure 1 for two examples) include information on the linear distance from the user to each other member, in feet for distances less than one mile, and miles for larger distances. For example, the person whose profile is represented in Figure 1a was 2,676 feet from our sampling location when the profile was viewed.

Although we piloted this approach with several of the available applications [25-28], data generated for this manuscript were from a single application, whose name is not

revealed at the request of the application developer. Application profiles indicate the distance but not the direction of the person in question. In order to develop measures of density of users, we began by establishing a grid over the City of Atlanta, and selecting points within the grid at which to collect information (Figure 2). Points were selected systematically with the
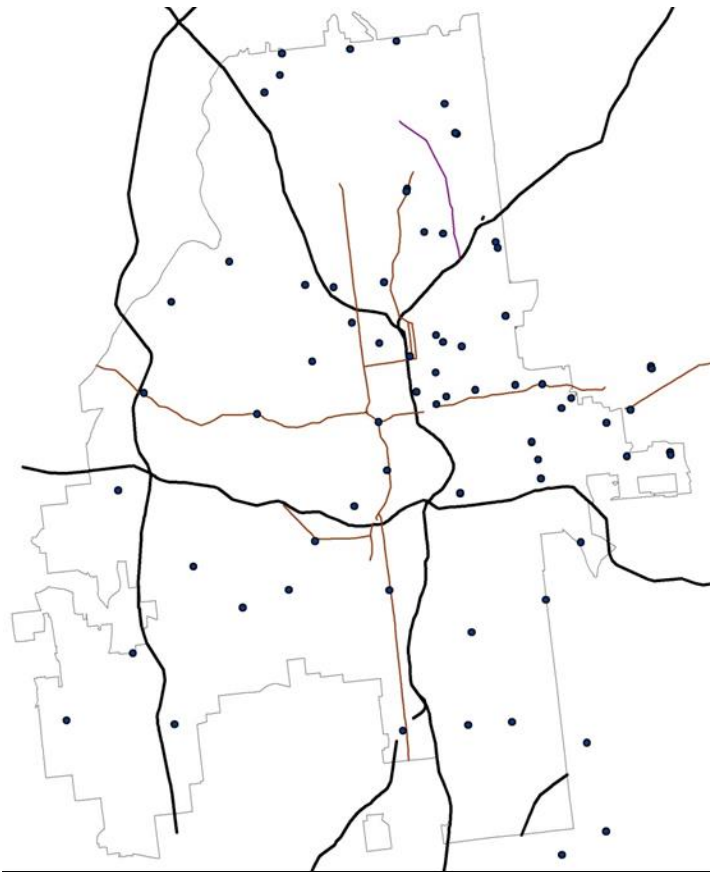
Figure 2: Map of the City of Atlanta (Grey Outline) including major interstates (black lines) and selected major roads (dark red lines). Points in the figure represent the 70 locations at which data were collected.

following protocol: we selected a starting point near KPD's home and drove along major roads to sample at roughly 2 mile intervals through most of the city. In areas with a high density of application users, we used a sampling strategy designed to (see below) collect data more frequently at closer intervals. At each point where profile data were observed, study staff used the "GEOLOCATION" application [29] to pinpoint the location of data collection to latitude and longitude.

Validation of geolocation data: In order to assess the accuracy of the geolocating application, we also recorded the GPS location at a subset of the same points at several

different days/times using both the GeoLocation and a GPS unit (Garmin model GPSmap

60CS [30]). The Geolocation application was found to be consistent with the GPS unit,

with the mean of the difference between the two being 144 feet (Range: 7-344 feet)

over a total of 25 sampled points. The GeoLocation application was also used at the

same 10 locations 6 months apart and found to give consistent results with a mean of

the difference in location coordinates of 76 feet (Range: 0-232). Thus we found it

sufficient to use the free GeoLocation latitude and longitude data available on the same

device as the social networking app for our purposes, rather than using two different

devices for data collection. (See Figure 3 for a screenshot from the GeoLocation iPhone

app, Available from [29]. Similar tools available for Android devices [31] were not

evaluated in this study.

Figure 3: Example of the GeoLocation app used in this study. The app is available from the itunes app store
https://itunes.apple.com/us/app/geolocation/id376832615?mt=8

It relies on cell towers and your internet connection, and provides Latitude and Longitude in decimal degrees.

V

Data collection

At each sampling point study staff collected screen shots of user profiles. These

applications sort profiles based on distance from the user to other users. We collected

profile data for either the 50 closest users or for all users within 2 miles of the sampling

point, whichever was less. Profiles were saved on a password-protected iPod Touch.

These data were entered into a database after field collection of the screenshots. Staff

also recorded the day and time of data collection at each point. We calculated the total

time spent collecting data as a process measure for this pilot study.

For each profile recorded, we extracted self-reported race and age, and the reported

distance from the sampled point (See Figure 1). Race was categorized as "white,"

"black," or "other", and age was recorded as a continuous variable. If a profile included

no information on race or age this was indicated with a missing value in the database.

Because the main objective was to compare the distribution of persons reporting their

race as white to those reporting their race as black, when either race or age were

missing, we recorded missing race as "other" and missing age as missing. Individual

profile data from each sampled point were aggregated as the number of users by self-

reported race (grouped as white, black and other) and self-reported age group (grouped

as 18-24, 25-30, >30 or unknown), and summary measures comparing those reporting

black or white race in their profiles (further described below) were calculated for the

City of Atlanta.

Sampling Strategy

At points where there were greater than 50 users within less than a two mile radius, we
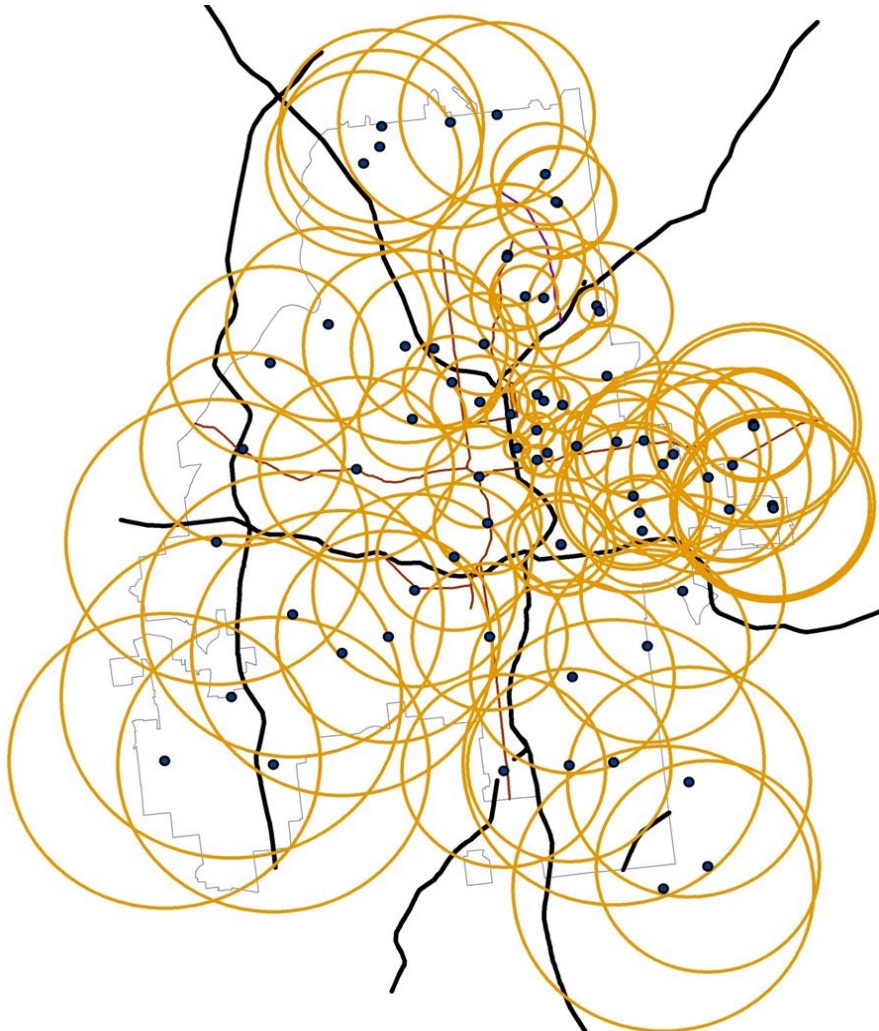
Figure 4: Map of the City of Atlanta showing 79 points at which data were collected from profiles of a sex seeking networking app. Radii of yellow circles represent the distance to the user sample at the maximum distance from the sample point. Overlapping circles completely cover the City of Atlanta, with smaller circular areas used for data collection in areas where there were the largest numbers of application users.

recorded the maximum distance to the 50th closest user (ordered by distance) and

moved this same distance along city streets to establish the next sample point.

Thus smaller radii were utilized in areas with a higher density of users. Figure 4 shows

the sampling radii for each point, the smaller circles represent the areas of Atlanta with

the highest density of users, and thus larger numbers of individual profiles available

within a given (e.g. 2 mile) radius. Because we collected different numbers of users

from circles of different radii we choose to standardize these measures to a common

area, for example, converting each observation into the number of users within 1 mile

of the point (thus describing a circle with a mile radius and/or an area of π square

miles), and stratifying these measures by race and age group.

Analysis
The data in this study provide a somewhat unique challenge to geospatial statistical

methods, because they combine the characteristics of point and area processes [32-37].

Data are collected at points on a grid, but the data at that point represent a density over

an area of sampling in a concentric circle around that point.   Still, the data are more

analogous to point data, with the measure collected at each point representing an area

rather than an individual data point. Thus we chose to treat these densities of users per

square mile as the measure of interest but use point data statistics [32, 38-39] to

summarize over the entire study area. ArcGIS [39] performs kernel smoothing to

estimate the density measured at each sample point where each sample point is

weighted by the observed population density at that point. In our case, the Kernel

Density smoother [34] counts every white and black user observed at that location. For

example a point at which we observed 12 profiles within 2 miles, including 8 white and 4

black users, would be counted 8 times in the white density measure and 4 times in the

black density measure. Next these weighted values for each point are also averaged

with other points within a specified radius [32, 36-37], resulting in a smoothed surface

representing the density of users, by race, in the sample space. The kernel approach

may place non-zero density in areas where no data were collected, but only as a result

of averaging between points separated by the area with no data. We also experimented

with methods for interpolation of spatial data such as kriging [32, 38] and found similar

results. We focus on kernel density estimates here. As noted above, sampling was

conducted at different times and days of the week over a 6 month period (See online

appendix for documentation of days and times sampled). While an in-depth analysis of

time of day and day of week variability is of interest for future research, to illustrate our

approach, we present the kernel densities calculated here as averages over sampled

days and times.

After estimating the population density, we used ArcGIS to compute the mean and

standard deviation for the calculated density measure over the entire sample space.

We compared density surfaces through ratio and difference measures via the Map

Algebra tool in ArcGIS, which solves standard algebraic equations at each point in a grid

across the density surface and creates a new map displaying the results of these

calculations. When comparing the density of users, the difference between surfaces for

different races, e.g. (density of black users – density of white users) has the property

that its null value (no difference) is zero, and if positive, it identifies an area with a

higher density of black users than white users. This represents an absolute difference in the densities of the two groups. When positive, this approach identifies areas where it might be easier to recruit black users because the density of black users is greater in absolute terms (i.e., the number of excess individuals). We note that this example says nothing about the magnitude (size of the density of black and/or white users), only that one number is bigger than the other. To capture areas where there are relatively more black users than white users (i.e., the ratio of black to white users is higher), we also calculated the ratio of the two density surfaces.

As a further exploration of the possibilities with the approach, we also considered a measure to highlight areas with the largest densities for each race, and then compare these areas as follows. First, for each density surface (e.g. the density of black users less than 25 years of age) we identified areas with the highest density values (density value > mean + 2 SD). For example, if the estimated mean density for white users was 14/square mile with standard deviation of 7, we would ask ArcGIS to select points with a density of white users greater than 28. We then used Map Algebra to calculate the difference between the surfaces including these highest density points for each race according to the following formula:

$$I(\text{Density of Black Users} > \text{Mean} + 2SD \text{ of estimated kernel density distribution})$$
$$-$$
$$I(\text{Density of White Users} > \text{Mean} + 2SD \text{ of estimated kernel density distribution})$$

where I(statement) represents an indicator function with value 1 if the statement is true

and zero otherwise. This equation takes only three values: zero when a point is greater

than Mean + 2SD of both distributions or neither is greater than Mean + 2SD, 1 when a

point is greater than the Mean + 2SD for only the first distribution, and -1 when the

point is only greater than the Mean + 2SD of the second distribution. This measure

identifies not only locations with more users of a given race, but also locations with the

highest density areas overall.   Similar measures can be constructed to highlight other

features of interest, e.g. comparing densities by age group or combinations of race and

age.   Finally, to provide some context to our results, we present them in relation to the

location of recruitment sites seeking to enroll MSM for two ongoing HIV prevention

studies in Atlanta.

Results

Over a two-week period we spent a total of 21 hours traversing Atlanta, collecting data

at the 79 sample points (Figure 2) covering 883 square miles of area (Figure 4) in order

to collect overlapping circles of data and cover the entire 132.4 square miles in the city

of Atlanta. The average radius of data collection at each sample point was 1.65 miles,

with smaller radii resulting from the more densely populated areas in Midtown Atlanta

(the area bounded by the rectangle in Figures 2 and 4).

We extracted profile data (race and age) for 2,666 user profiles. Of these 1,563 (59%)

were white, 810 (30%) were black, 146 (5.5%) were some other race, and 147 (5.5%) did

not report a race in their profile. The mean age was 31.5 years, with 591 (22%) between

the ages of 18-25, and 496 (19%) between the ages of 26-30. Age was more likely than

race to be missing from profile information with 593 (22%) of profiles sampled not

providing age information. The remaining 37% of profiles reported ages greater than 30;

whites were more likely to report being > 30 years of age than blacks (46% vs. 25%,

p<0.0001). Black users were younger than white users (median 28 vs. 33 years, *P*<0.001

via the Wilcoxon Sign rank test).

Across the 79 sampled points the mean number of users was 33 per square mile, but the

distribution of users across points was highly skewed with median of 17 and range 0.86-

208 (Figure 5).
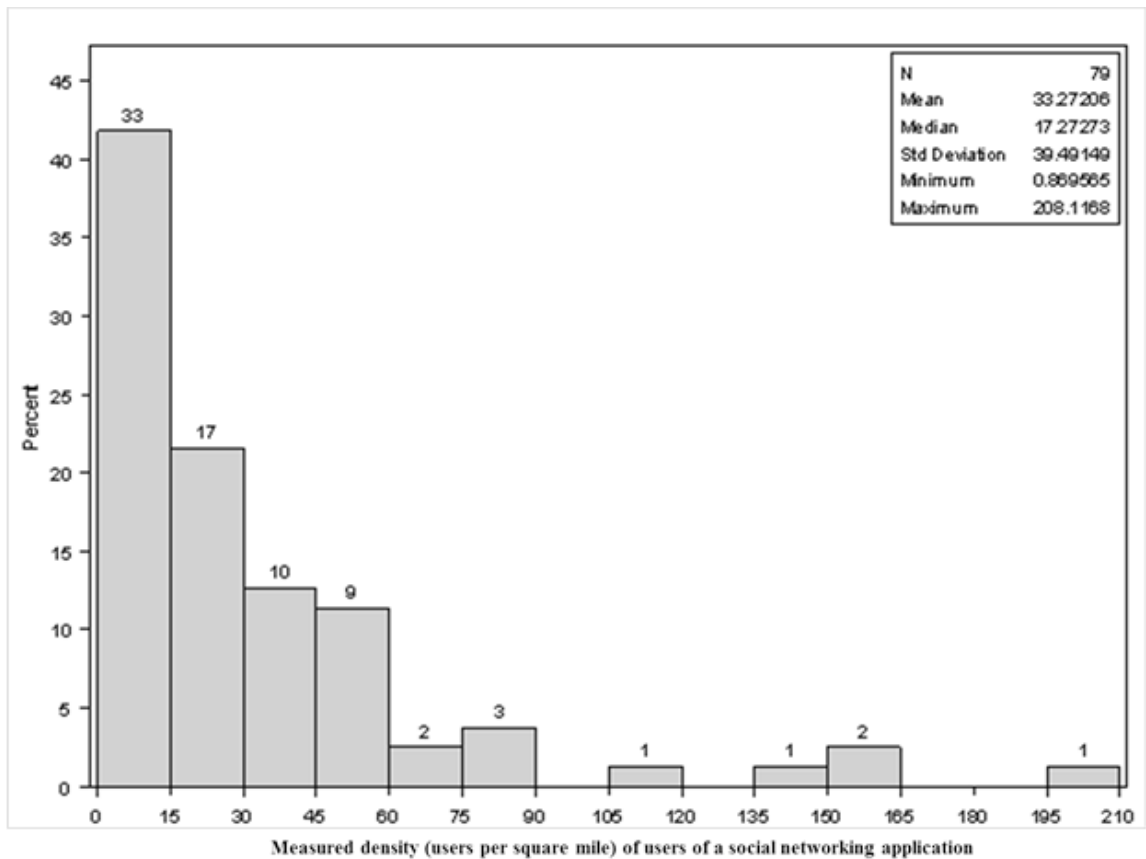
Figure 5: Histogram describing the distribution of observed density of social network application users per 1 mile circle for the 79 sampled locations in the City of Atlanta. Inset includes statistics for the distribution, which is highly skewed with the standard deviation estimated to be larger than the mean. Numbers above the bars are the number of sample points with density along the X axis and the Y axis representing the percent of all points with this density.

Figure 6 shows the density of application users, smoothed using a kernel density

function with a 2 mile radius, for white (A) and black (B) users. A 2 mile radius was

chosen as the smoothing parameter because it was the next largest integer that covered

the average radius of 1.6 miles for in the sampled points, and also was the maximum
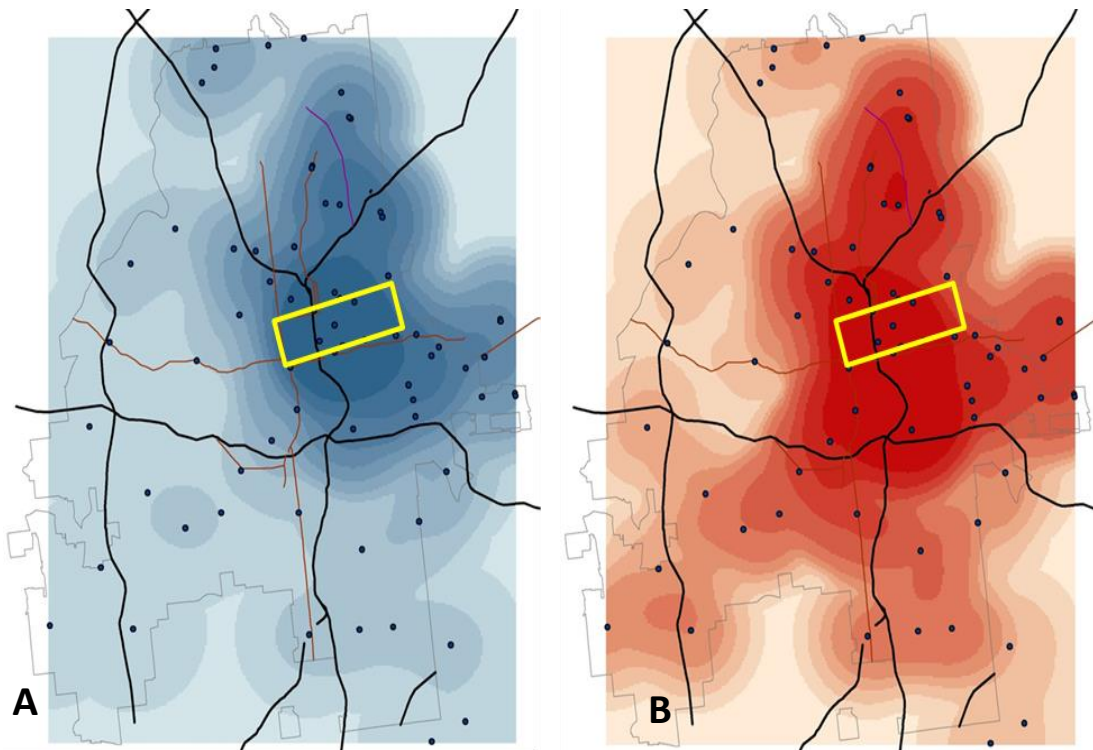
Figure 6: Estimated density of white (A) and black (B) social network application users in the City of Atlanta (grey outline), showing major highways (black lines) and roads (dark red lines) and highlighting the "Midtown" area of Atlanta (yellow rectangle). Kernel Densities were estimated from sample data standardized to 1 mile circular radii, and smoothed to 2 miles using a Gaussian smoother that concentrates the majority of the density at the sample point, and averages over all adjacent data points within the smoothing radius.

distance to which we sampled data when a sample point had fewer than 50 users. The

online appendix shows the analogs of Figures 6 and 7 with a 1-mile kernel density

smoothing parameter for comparison, the results were not qualitatively different. The

highest density of white users (the darkest blues in Figure 6A) concentrates in the

midtown area of Atlanta (roughly bounded by the yellow rectangle on the map). While

much of the highest density of black users also concentrates in this area, it is clear that

there are areas with high densities of black users further south and to the west (to the

lower left) of the midtown area The kernel approach smooths observations according to

a two-dimensional distribution centered at the observed point and declining out to the

radius used to define the search area, essentially "spreading" observations from sample

points across the study area. For example, the density values for white users over the

79 sample points ranged from 0.3 to 154 profiles per square mile, but the range of

values for the smoothed density shown in Figure 6A was 0-57 profiles per square mile.

For the 1-mile smoothed density (supplementary appendix Figure 1) the range (0-138)

was closer to the observed values, but with many more points with density estimates of

zero (i.e., observations were not "spread" as far).

There are several ways to compare surfaces to illustrate local differences between the

densities of white and black users. Figure (7) shows two similar but nonidentical ways

to compare these densities. Figure 7A shows the difference between the two surfaces,

colored so that areas with higher absolute density of white users are blue and areas

with higher density of black users are red. Figure 7B shows the relative difference, with

areas where the ratio of black to white profile densities is higher than one as red and

lower than one as blue.
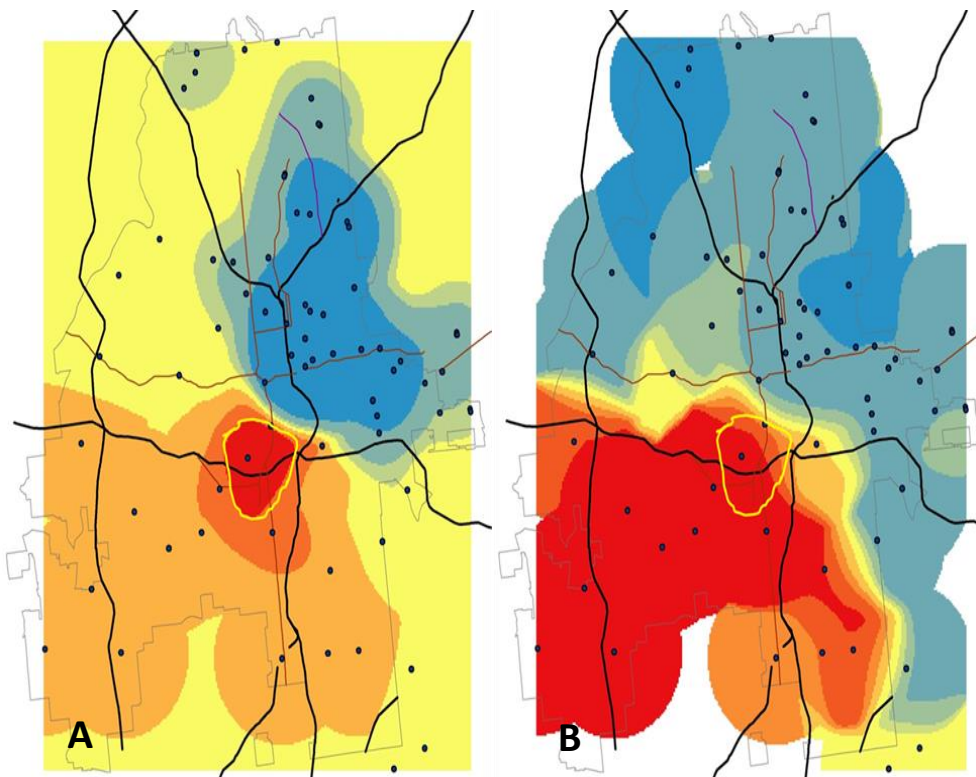
Figure 7: Comparison of the density of black and white social networking application users in the City of Atlanta. Panel A shows the absolute difference in users (Density of black users – Density of white users) color coded so that areas with more black users appear red and those with more white users appear blue. Yellow regions are areas where the two densities are similar. Panel A highlights a small section of the city (the area shaded the darkest red) where there are many more black than white application users. Panel B shows a comparison of the relative size of the densities of black and white users (Density of black users/Density of white users). With this measure, Atlanta is divided nearly in half, with relatively more black users in the southwest and more white users to the North and East. The yellow band in Panel B shows the region with the highest absolute excess of black users for comparison purposes.

The ratio measure shows that most of Southwest Atlanta has relatively more black user

profiles observed than white profiles, but when we compare the map with that of the

overall number of black users, we find a much smaller region in which to focus efforts,

south and west of the midtown area, shown with a yellow band in Figure 7.
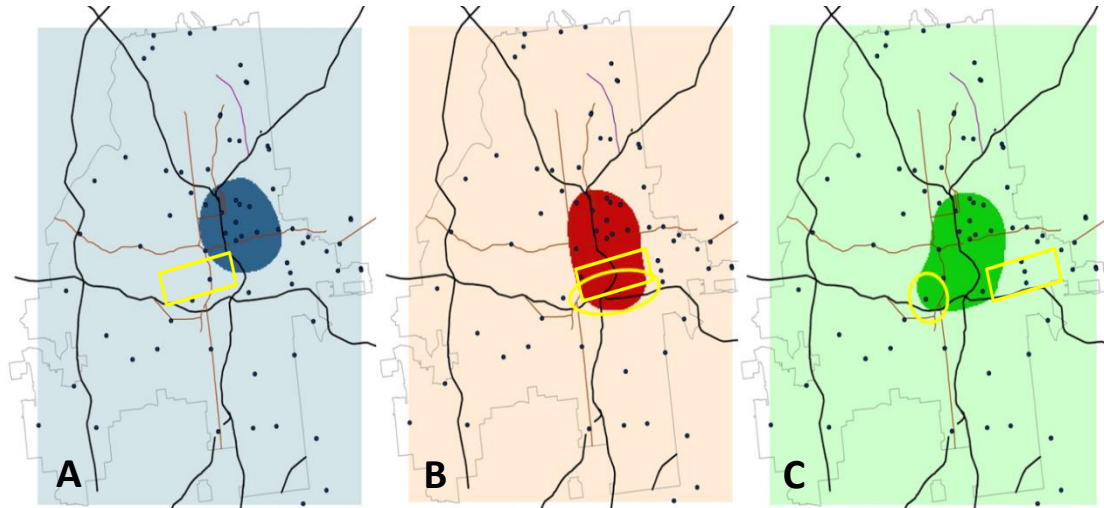
Figure 8: Density of social networking application users in Atlanta, highlighting points with values greater than the 95th percentile of the estimated kernel densities for white (Panel A), black (Panel B) and young black (<25 years of age, Panel C) users. For Panel A points with an estimated density greater than 17.2 users/mile^2 are highlighted dark blue; for Panel B those >5.65/mile^2 are dark red and for Panel C >2.8/mile^2 are dark green. The Yellow rectangle highlights the midtown area of Atlanta for reference. The yellow oval in Panel B highlights an area with high density of black users but not white users. The yellow circle in Panel C highlights an area with a high density of young black users, but not black users overall (i.e. an area highlighted in Panel C but not Panel B).

A third way to visualize differences between the surfaces is to focus on the areas

with extreme values. This provides a within-density comparison: over the entire surface

of the density of black user profiles, where is the density the greatest?   In Figure 8, we

highlight the regions with density greater than the Mean+2 standard deviations over the

entire map, separately for all black (A), all white (B) and young black (<25years old, C)

users based on data in their observed profiles. This approach again highlights the

midtown area of Atlanta (yellow rectangle) as the region with the most users observed

in each graph.

Figure 9, calculates the difference between Figure 8B and Figure 8A, and shows

that black user profiles have high density much further south than white profiles.
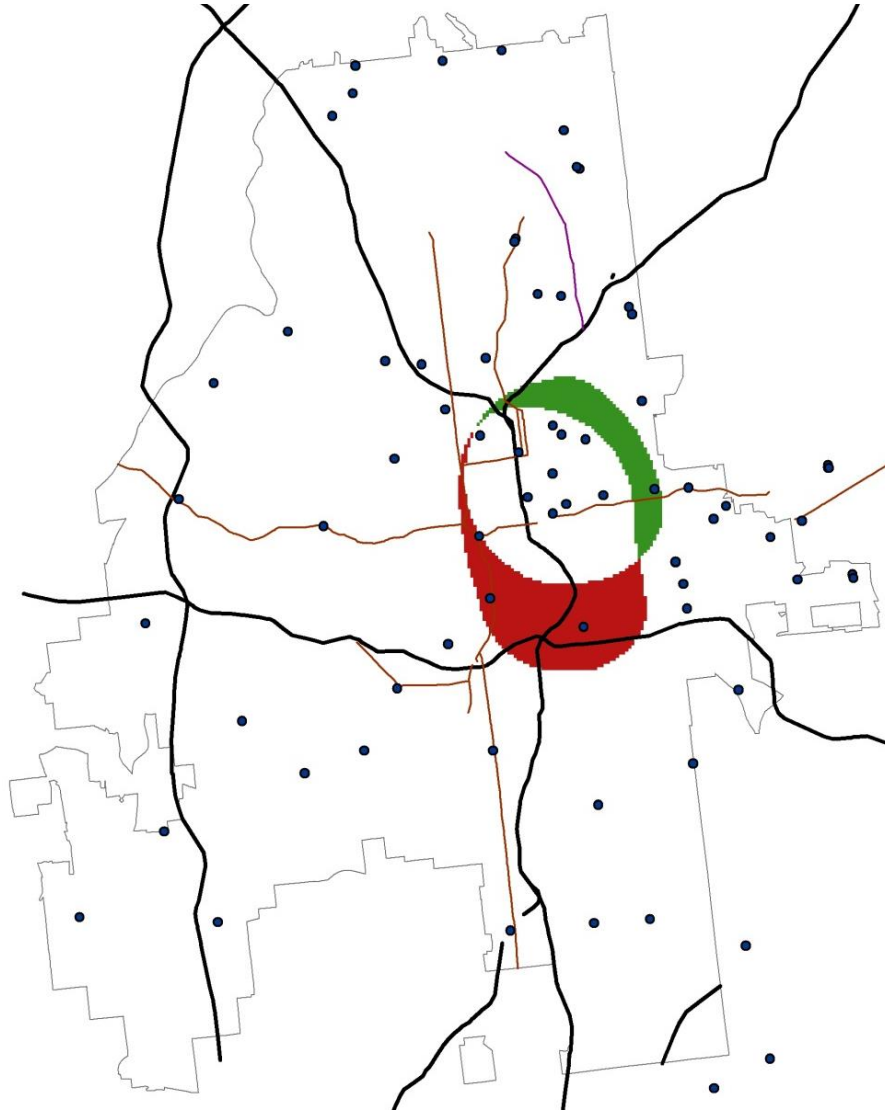
Figure 9: Difference Between Extreme Values of Estimated Kernel Densities of White and Black users of a social networking application.

In this figure we use the formula I(Density of black users > mean+2 standard deviations) – I(Density of white users > mean+2 standard deviations). We then present regions where the values of this equation are -1 (green shading, indicating areas with extremes of density for white but not black users), 1 (Red shading, indicating areas with extremes of the density for black but not white users) and 0 (white shading indicating areas which are either not extremes of either density or are extremes for both races).

The supplementary online materials show the kernel density and map algebra

calculations using a 1-mile radius. Overall the results obtained with the smaller radius

are similar (Supplemental Figures 1-3) with the peaks in areas that at 2 miles (Figure 8C)

had previously shown up as having a higher concentration of young black user profiles

rather than black user profiles overall. The Supplemental Figures compare the

difference between the 1 mile smoothed densities for young black and all black users.

Overall the results are similar, but there are a few additional areas (highlighted in the

Supplemental Figures), with extreme densities of young black users that did not appear

in the 2-mile estimates shown in Figure 8C or Figure 9.

Discussion

We sampled 2,666 profiles from a mobile phone-based social networking application at

79 sites in Atlanta, and, under our sampling protocol, observed a mean of 33 application

users per square mile. We also identified areas where there were more black and young

black user profiles observed compared to white user profiles, describing 3 different

summary measures of the density of profiles in a sampling frame. Finally, we showed

the impact of the choice of the kernel radius in construction and interpretation of such

data.

The goal of this study was primarily descriptive, in that we sought to describe a method

for calculating the density of user profiles by race and age in Atlanta, and to compare

and contrast the information provided by different outcome measures that can be

constructed from these data. In addition, the methods described here may have

practical application in HIV prevention research. The results are promising and illustrate

how the use of self-reported location data can provide information on the geographic

distribution of users in time and space. The study methodology could provide a more

efficient way to identify locations for recruitment of MSM in future studies. Significant

time and effort is spent on formative research to develop sampling frames for studies of

MSM [15, 21]. The goal of such formative research is to identify locations for sampling

MSM using time space sampling methods [15]. Our methodology, based on the

geolocation data incorporated into popular social networking applications, allowed us to

quickly describe the density of sex-seeking MSM in Atlanta. Furthermore, we were able

to use profile information to stratify these density measures by race and age. This might

allow for oversampling or exclusive sampling in areas of the city that are expected to

yield a particular subset of the population, for example, young black MSM. As an

example, Figure 10 illustrates how these data can inform study implementation in

practice. Figure 10 shows Figure 7b and a variation of Figure 9, along with recruitment

venues currently in use for two HIV prevention studies in Atlanta (green triangles).

Figure 10a shows that, to date, there have not been very many sampling locations in the

southwestern part of Atlanta, where, based on the ratio of the density of black to white

application users, there are relatively more black users than white users. However,

Figure 10b shows the difference between extremes for the densities of young white and

young black users of the social networking application, using a formula similar to that

used to calculate Figure 9. Looking at this representation of the data, we see that we

have identified recruitment venues in an area of the city where there are the most

young black application users and not that many white users. In this case, while going
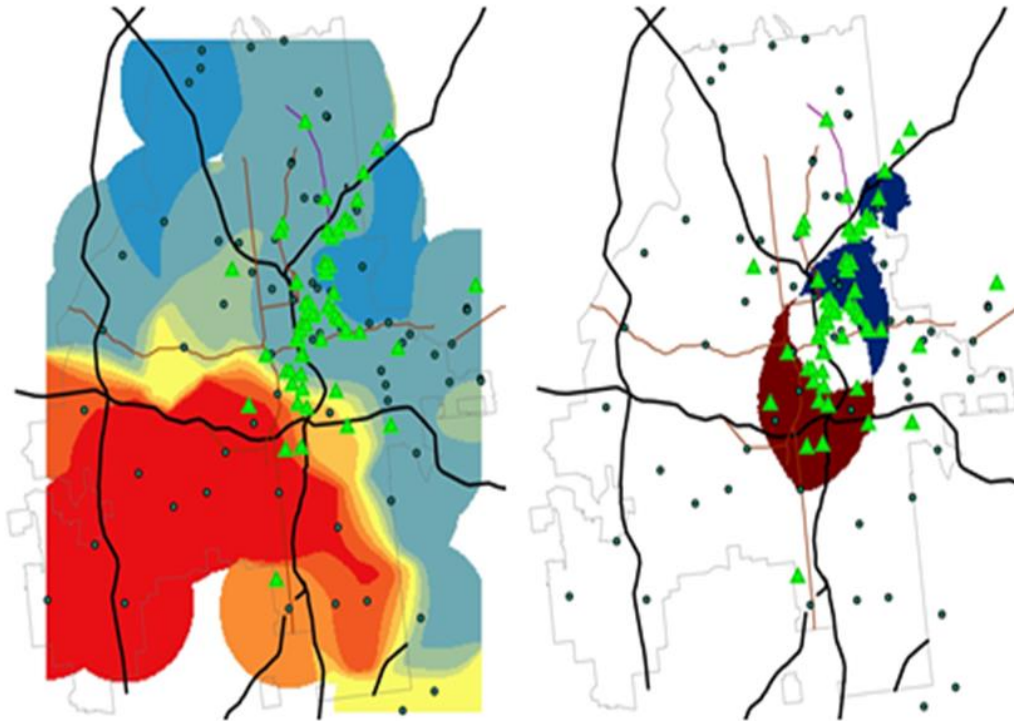
Figure 10: Application of two density metrics to evaluate recruitment for HIV prevention studies in Atlanta, GA.

This figure shows Figure 7b and a variation of Figure 9, along with recruitment venues currently in use for two HIV prevention studies in Atlanta (green triangles). Figure 10a illustrates that there are not very many recruitment locations in the southwestern part of Atlanta, where there are relatively more young black application users than white users. Figure 10b uses the formula (Density of young black users > mean+2 standard deviations) – (Density of young white users > mean+2 standard deviations). Regions where the values of this equation are -1 (blue shading, indicating areas with extremes of density for young white but not young black users), 1(Red shading, indicating areas with extremes of the density for young black but not young white users) and 0 (white shading indicating areas which are either not extremes of either density or are extremes for both races) can then be compared to the locations of current recruitment venues.

further into the areas of higher relative densities of black users might yield additional

recruitment sites, we seem to have covered the areas with the highest number of both

black and white users. Also, we find that there aren't many recruitment sites outside of

the area with the highest densities of white users, black users or both, confirming that

past recruitment sites were located in parts of the city where there are the most

applications users overall. Further potential applications of this methodology include

identification of areas with need for prevention services, e.g. overlaying HIV testing

locations on the density grid to identify local areas with greatest unmet need.

Since the early 2000's, there has been a significant rise in internet usage by MSM [40-

43] and young minorities [44-46]. Three different groups have found that gay men now

report meeting the majority of their sex partners online [40,47-49], and many [43,47-49]

but not all [50] studies of sex behavior have shown increased reports of behaviors

associated with higher HIV risk amongst partners met online compared to offline. The

most popular and well-studied of these location-based social networking applications is

Grindr [51-53], which is currently being used by over 4 million men worldwide [25], and

is likely to continue to grow in popularity. MSM use this application for a variety of

purposes, but a survey of Grindr users in Los Angeles found that 76% have had sex with

someone they met on Grindr [51], suggesting that Grindr users are using the application

to help find sex partners. Many other similar applications exist such as Adam-4-Adam,

Jack'd, and BoyAhoy [26-28], and our methodology can be applied to any such

application that provides data on race and age as well as distance to the user within

member profiles. In our research we have found that users of these applications vary by

race and less so by age, with, for example a greater proportion of white men reporting

using Grindr and more Black men reporting using Jack'd (unpublished Emory University

data). In this study, although we illustrate our approach using only one application (and

have chosen not to identify the specific application used to generate these data) we did

validate the methodology with more than one application. Any of the apps that report race, age and other characteristics of interest (e.g. HIV serostatus) as well as geographic distance from the user's present location could be used to make density maps and calculate summary statistics using the methods we report in this manuscript. In some cases, it may be useful to calculate one or more density measures with more than one app to try to get a better overall picture of the spatial distribution of men using sex-seeking apps in a given location.

Because users of these applications make both their profile information and their location public, it was possible to simply observe these publically available data without contacting the users directly for this research. However, there is still an ethical requirement to protect individually identifying information when the information is collected for research purposes. In this study we used screen captures to record profile information, storing these pictures on a password protected iPod touch until the data of interest (age, race and location information) could be entered into a database with no identifiers. Because we were only recording publically available data from user profiles without identifiers, the Institutional Review Board at Emory University considered the study to be research exempt from IRB review.

More generally, using social networking applications for HIV prevention is likely a key strategy for future research [52-54], but comes with new ethical and methodological questions. Our study only sought to summarize the data publically available within these applications, but social media applications may themselves serve as an important public

health communications tool. Recently, public health agencies have sought to partner with Grindr, and use its built in advertisements as a medium for disseminating prevention information and recruit MSM for research studies [52, 53]. Future research might adapt our methodology further to establish a sampling frame, and then use the density information to sample application users and contact them to either conduct a cross-sectional survey or recruit them into a follow-up study. At that time one would have to develop mechanisms for consenting study participants, as well as a way to keep sensitive information, such as sex and drug use behavior, protected and ideally separate from any identifying online profile information.

Piloting this methodology in Atlanta exposed other challenges as well. Atlanta is geographically large and contains both densely settled neighborhoods in the inner city along with a large amount of semi-urban and even rural areas with less dense populations. Atlanta also exhibits a large degree of geospatial segregation by race, both in the population overall [55] and in the relative measures of the distribution of social-network application users (Figure 7b). However, although the overall black population density is low in the midtown area of Atlanta [55, 56], it still represented the area with the highest concentration of black users of the sex-seeking application (Figure 4). To obtain a picture of the distributions of both black and white application users we therefore had to sample enough points with a sufficiently wide radius to cover the entire city. We also found the density of users to vary widely within the city, and we therefore had to adapt our sampling strategy. We choose to collect either the first 50 profiles and record the distance to the 50th user, or to sample out to a 2-mile radius if

there were less than 50 profiles observed in that area. In areas with large numbers of users, we had to collect data at more closely sampled points, e.g. if there were 50 profiles within a half mile, we only moved that short distance before collecting more data, if there were only 13 users within the 2 mile radius we moved the full 2 miles between sample points. This allowed us to cover the whole city, but despite collecting data at 79 points that represented an area equivalent to 882 square miles, there were still areas of the city where we did not directly sample any users.

This makes the choice of the smoothing parameter (radius) for the kernel smoothing algorithm important, because it provides a balance between too much interpolation of data between sampling points and presuming that the data collected at a particular sampling point only occur at the point and do not represent an area defined by the radius of a circle based on the linear distance to the person whose profile is being observed. Using our sampling plan, we collected data from concentric circles with an average radius of 1.65 miles, and then fit weighted kernel densities smoothed to 1 and 2 miles. Both of these smoothing parameters provided similar interpretations of density of black, white and young black individuals, with the 1-mile radius leaving more areas of the city with no estimates for the density of application users. The 2-mile radius covers the whole city, but as a result it reduces the emphasis of several points which, when using a 1-mile radius are considered to have a particularly high density of black users.

Some questions remain about the precise interpretation of the density of social-network

application users. For example, are users simply a subset of all MSM seeking sex on the internet?   Is the population that uses any one of these applications different by important characteristics (race, age, sex behavior with persons met through a social-network application or with sex partners generally) from the underlying population? Are persons who use specific services, (e.g., Adam-4-Adam, Jack'd, Grindr) different by one or more of these characteristics than those that use other online applications [47-49]?   Future studies [54] will seek to quantify the density and characteristics of men who use each of these applications and compare the characteristics of men who use each of the applications exclusively, while also capturing information about men who use more than one service to describe whether their behaviors vary when using different services.

It would be useful to test the methodology in other cities with significant minority MSM populations (e.g. Washington, DC or Los Angeles, CA) and also to assess the utility of the method in less densely populated areas (e.g. in rural areas of Georgia), to describe the extent to which the utility of the methods vary by characteristics of the geography of the region. We have already identified that Atlanta is a challenging place to conduct this kind of study because of its racial distribution, which was borne out in the social-network application user density data. In areas with sparse numbers of users, our adaptive sampling methodology which sampled a minimum of 50 users or to a 1-mile radius might help to stabilize density estimates, but this needs further testing. Additionally, although we averaged over day and time of sampling in our current analysis, the method could be refined to capture spatiotemporal trends in density. For

example, it would be possible to select points to be sampled multiple times over a grid

of specific times and days [14-15]. This modification could provide a clear description of

how the user profile's population density changes over the course of a week. This last

component may identify trends in the spatial and temporal clustering of application

users, for example on weekend nights, as compared to mid-day during the work week.

We have found that it is possible to use a limited number of sample points to develop a

geospatial density of men using a social-networking application to seek sex in the City of

Atlanta. Such a density could serve as a sampling frame for future cross-sectional or

longitudinal research. We also describe several methods to compare two densities with

a goal of identifying areas with a high density of a particular subset of the population.

We hope that this novel methodology and its further adaptations will prove useful to

future research and prevention efforts that can be tailored to areas of the community

where they will be most effective.

Multimedia Appendix: Additional figures and analysis using 1-mile smoothed kernel

densities and a matrix describing the days and times of sampling at the 79 locations

included in this analysis.

References

1) Centers for Disease Control and Prevention. *HIV Surveillance Report, 2011*; vol. 23. http://www.cdc.gov/hiv/library/reports/surveillance/2011/surveillance_Report_vol_23.html Published February 2013. Archived at: http://www.webcitation.org/6LodIMTIRIR Accessed 12/12/2013

2) CDC. Estimated HIV incidence among adults and adolescents in the United States, 2007–2010. *HIV Surveillance Supplemental Report* 2012;17(4). http://www.cdc.gov/hiv/pdf/statistics_hssr_vol_17_no_4.pdf. Published December 2012. Archived at: http://www.webcitation.org/6LodWskUl Accessed 12/12/2013.

3) CDC. Prevalence and awareness of HIV Infection among men who have sex with men—21 cities, United States, 2008. MMWR 2010;59(37):1201-1207. PMID: 20864920

4) Prejean J, Song R, Hernandez A, Ziebell R, Green T, Walker F, Lin LS, An Q, Mermin J, Lansky A, Hall HI; HIV Incidence Surveillance Group. Estimated HIV incidence in the United States, 2006-2009. PLoS One. 2011;6(8):e17502. doi: 10.1371/journal.pone.0017502. Epub 2011 Aug 3. PMID: 21826193.

5) National HIV/AIDS Strategy. Available from: http://whitehouse.gov/sites/default/files/uploads/NHAS.pdf Archived at: http://www.webcitation.org/6LodbOGzf Accessed: 12/12/2013

6) Millett G, Flores F, Peterson JL, Bakeman R. Explaining disparities in HIV infection among black and white men who have sex with men: a meta-analysis of HIV risk behaviors. AIDS 2007;21:2083–91. PMID: 17885299.

7) Oster AM, Wiegand RE, Sionean C, Miles IJ, Thomas PE, Melendez-Morales L, Le BC, Millett GA. Understanding disparities in HIV infection between black and white MSM in the United States. AIDS. 2011 May 15;25(8):1103-12. PMID: 21505305

8) Maulsby C, Millett G, Lindsey K, Kelley R, Johnson K, Montoya D, Holtgrave D. HIV Among Black Men Who Have Sex with Men (MSM) in the United States: A Review of the Literature. AIDS Behav. 2013 Apr 26. [Epub ahead of print]. PMID: 23620241

9) Bing EG, Bingham T, Millett GA. Research needed to more effectively combat HIV among African-American men who have sex with men. J Natl Med Assoc. 2008 Jan;100(1):52-6. PMID: 18277808

10) Jeffries WL 4th, Marks G, Lauby J, Murrill CS, Millett GA. Homophobia is associated with sexual behavior that increases risk of acquiring and transmitting HIV infection among black men who have sex with men. AIDS Behav. 2013 May;17(4):1442-53. PMID: 22569942

11) Millett GA, Jeffries WL 4th, Peterson JL, Malebranche DJ, Lane T, Flores SA, Fenton KA, Wilson PA, Steiner R, Heilig CM. Common roots: a contextual review of HIV epidemics in black men who have sex with men across the African diaspora. Lancet. 2012 Jul 28;380(9839):411-23. PMID: 22819654.

12) Ayala G, Bingham T, Kim J, Wheeler DP, Millett GA. Modeling the impact of social discrimination and financial hardship on the sexual risk of HIV among Latino and Black men who have sex with men. Am J Public Health. 2012 May;102 Suppl 2:S242-9 PMID: 22401516

13) Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. Soc Probl. 1997;44(2):175–200.

14) MacKellar DL, Vallaroy L, Karon J, Lemp G, Janssen R. The young men's survey: methods for estimating HIV seroprevalence and risk factors among young men who have sex with men. Public Health Rep. 1996;111(Suppl 1):138–44.

15) MacKellar DA, Gallagher KM, Finlayson T, Sanchez T, Lansky A, Sullivan PS. Surveillance of HIV risk and prevention behaviors of men who have sex with men--a national application of venue-based, time-space sampling. Public Health Rep. 2007;122 Suppl 1:39-47. PMID: 17354526

16) Ompad DC, Galea S, Marshall G, Fuller CM, Weiss L, Beard JR, Chan C, Edwards V, Vlahov D. Sampling and recruitment in multilevel studies among marginalized urban populations: the IMPACT studies. J Urban Health. 2008 Mar;85(2):268-80. PMID: 18214686

17) Sanchez T, Smith A, Denson D, Dinenno E, Lansky A. 16. Developing a Web-Based HIV Behavioral Surveillance Pilot Project Among Men Who Have Sex with Men. Open AIDS J. 2012;6:224-31. PMID: 23091580

18) Johnston LG, Prybylski D, Raymond HF, Mirzazadeh A, Manopaiboon C, McFarland W. Incorporating the service multiplier method in respondent-driven sampling

surveys to estimate the size of hidden and hard-to-reach populations: case studies from around the world. Sex Transm Dis. 2013 Apr;40(4):304-10. PMID: 23486495

19) Montealegre JR, Johnston LG, Murrill C, Monterroso E. Respondent Driven Sampling for HIV Biological and Behavioral Surveillance in Latin America and the Caribbean. AIDS Behav. 2013 Apr 9. [Epub ahead of print]. PMID: 23568227

20) Pollack LM, Osmond DH, Paul JP, Catania JA. Evaluation of the Center for Disease Control and Prevention's HIV behavioral surveillance of men who have sex with men: sampling issues. Sex Transm Dis. 2005 Sep;32(9):581-9. PMID: 16118608

21) Lansky A, MacKellar D, Gallagher KM, Lin LS, Sullivan PS, Onorato IM. Evaluation of the new National HIV Behavioral Surveillance System (NHBS) is imperative. Sex Transm Dis. 2006 Apr;33(4):272-3; author reply 274-5.

22) Manopaiboon C, Prybylski D, Subhachaturas W, Tanpradech S, Suksripanich O, Siangphoe U, Johnston LG, Akarasewi P, Anand A, Fox KK, Whitehead SJ. Unexpectedly high HIV prevalence among female sex workers in Bangkok, Thailand in a respondent-driven sampling survey. Int J STD AIDS. 2013 Mar 19. [Epub ahead of print]. PMID: 23512512.

23) Jenkins RA. Recruiting substance-using men who have sex with men into HIV prevention research: current status and future directions. AIDS Behav. 2012 Aug;16(6):1411-9. PMID: 22016329

24) Kendall C, Kerr LR, Gondim RC, Werneck GL, Macena RH, Pontes MK, Johnston LG, Sabin K, McFarland W. An empirical comparison of respondent-driven sampling,

time location sampling, and snowball sampling for behavioral surveillance in men

who have sex with men, Fortaleza, Brazil. AIDS Behav. 2008 Jul;12(4 Suppl):S97-

104. PMID: 18389357

25) www.grindr.com Archived at: http://www.webcitation.org/6LodfWVwq Accessed:

12/12/13.

26) http://www.jackd.mobi/ Archived at: http://www.webcitation.org/6LoduUIDX

Accessed: 12/12/2013

27) www.adam4adam.com Archived at: http://www.webcitation.org/6Loe4v0I0

Accessed: 12/12/2013

28) www.boyahoy.com Archived at: http://www.webcitation.org/6LoePJ6e1

Accessed: 12/12/2013

29) https://itunes.apple.com/us/app/geolocation/id376832615?mt=8 Archived at:

http://www.webcitation.org/6LoeUcVPP Accessed 12/12/2013

30) https://buy.garmin.com/en-US/US/on-the-trail/discontinued/gpsmap-

60cs/prod6447.html Archived at: http://www.webcitation.org/6LoecZCll

Accessed 12/12/13

31) https://play.google.com/store/apps/details?id=com.exposure101.geolocation&fe

ature=search_result#?t=W10. Archived at:

http://www.webcitation.org/6LoejGmQ7 Accessed: 12/12/2013

32) Waller LA, Gotway CA. Applied Spatial Statistics for Public Health. Hoboken, New

Jersey: John Wiley and Sons, Inc.; 2004. ISBN: 0471387711

33) Bivand RS, Pebesma EJ, Gomez-Rubio V. Applied Spatial Data Analysis with R. New York, New York: Springer Science+ Business Media, LLC; 2008. ISBN: 0387781706

34) ESRI ArcGIS Help 10.1 How Kernel Density Works. Available at: http://resources.arcgis.com/en/help/main/10.1/index.html#/How_Kernel_Density_works/009z00000011000000/ Archived at: http://www.webcitation.org/6Loevcer4 Accessed: 12/12/2013

35) Silverman, B.W. *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall; 1986. ISBN: 0412246201

36) Kafadar K. Smoothing geographical data, particularly rates of disease. Statistics in Medicine (1996); 15: 2539-2560 particularly Section 3 starting on page 2543

37) Bithell JF. A classification of disease mapping methods. Statistics in Medicine (2000); 19: 2203-2215.

38) ESRI ArcGIS Help 10.1 Empirical Bayesian Kriging (Geostatistical Analyst) http://resources.arcgis.com/en/help/main/10.1/index.html#//00300000002p000000 Archived at: http://www.webcitation.org/6Lof78o32 Accessed: 12/12/2013

39) http://www.esri.com/products Archived at: http://www.webcitation.org/6LofNmaCJ Accessed: 12/12/2013

40) Bolding G, Davis M, Sherr L, Hart G, Elford J. Use of gay Internet sites and views about online health promotion among men who have sex with men. AIDS Care. 2004 Nov;16(8):993-1001. PMID: 15511731

41) Zhang D, Bi P, Hiller JE, Lv F. Web-based HIV/AIDS behavioral surveillance among

men who have sex with men: potential and challenges. Int J Infect Dis. 2008

Mar;12(2):126-31. Epub 2007 Sep 19. PMID: 17884663

42) Sullivan PS, Khosropour CM, Luisi N, Amsden M, Coggia T, Wingood GM,

DiClemente RJ. Bias in online recruitment and retention of racial and ethnic

minority men who have sex with men. J Med Internet Res. 2011 May

13;13(2):e38. PMID: 21571632

43) Kakietek J, Sullivan PS, Heffelfinger JD. You've got male: internet use, rural

residence, and risky sex in men who have sex with men recruited in 12 U.S. cities.

AIDS Educ Prev. 2011 Apr;23(2):118-27. PMID: 21517661

44) Whiteley LB, Brown LK, Swenson RR, Romer D, DiClemente RJ, Salazar LE, Vanable

PA, Carey MP, Valois RF. African American adolescents and new media:

associations with HIV/STI risk behavior and psychosocial variables. Ethn Dis. 2011

Spring;21(2):216-22. PMID: 21749027

45) Whiteley LB, Brown LK, Swenson RR, Valois RF, Vanable PA, Carey MP,

DiClemente R, Salazar LF, Romer D. African American adolescents meeting sex

partners online: closing the digital research divide in STI/HIV prevention. J Prim

Prev. 2012 Feb;33(1):13-8. PMID: 22293979

46) Cameron KA, Salazar LF, Bernhardt JM, Burgess-Whitman N, Wingood GM,

DiClemente RJ. Adolescents' experience with sex on the web: results from online

focus groups. J Adolesc. 2005 Aug;28(4):535-40. Epub 2005 Jan 20. PMID:

16022887

47) Grov C. HIV risk and substance use in men who have sex with men surveyed in bathhouses, bars/clubs, and on Craigslist.org: venue of recruitment matters. AIDS Behav. 2012 May;16(4):807-17. PMID: 21748276

48) Grov C, Crow T. Attitudes about and HIV risk related to the "most common place" MSM meet their sex partners: comparing men from bathhouses, bars/clubs, and Craigslist.org. AIDS Educ Prev. 2012 Apr;24(2):102-16. PMID: 22468972

49) Mayer KH, Ducharme R, Zaller ND, Chan PA, Case P, Abbott D, Rodriguez II, Cavanaugh T. Unprotected sex, underestimated risk, undiagnosed HIV and sexually transmitted diseases among men who have sex with men accessing testing services in a New England bathhouse. J Acquir Immune Defic Syndr. 2012 Feb 1;59(2):194-8. PMID: 22027871

50) Bolding G, Davis M, Hart G, Sherr L, Elford J. Gay men who look for sex on the Internet: is there more HIV/STI risk with online partners? AIDS. 2005 Jun 10;19(9):961-8. PMID: 15905678

51) Rice E. Sex on Grindr: Young Men Who Have Sex with Men and Geo-Locating Social Networking Technology. ISIS Annual Conference on New Media, Youth and Sexual Health, SexTech 2012; 2012 March 31-April 2; San Francisco, CA.

52) Burrell ER, Pines HA, Robbie E, Coleman L, Murphy RD, Hess KL, Anton P, Gorbach PM. Use of the location-based social networking application GRINDR as a recruitment tool in rectal microbicide development research. AIDS Behav. 2012 Oct;16(7):1816-20. PMID: 22851153

53) Usher D, Frye V, Shinnick J, Greene E, Baez E, Benitez J, Solomon L, Shouse RL, Sobieszczyk ME, Koblin BA. Recruitment by a Geospatial Networking Application for Research and Practice: The New York City Experience. J Acquired Immune Defic Syndr. 2014 Sept 22[Epub ahead of print]. PMID: 25247432

54) Sullivan PS, Grey JA, Simon Rosser BR. Emerging Technologies for HIV Prevention for MSM: What We have Learned, and Ways Forward. J Acquir Immune Defic Syndr. 2013 Jun; 63 Suppl 1:S102-7. PMID: 23673879

55) The New York Times: Mapping the 2010 US Census, Largest racial and ethnic groups in 2010, Atlanta, GA.

http://projects.nytimes.com/census/2010/map?view=MajorGroupView2010&lat=33.7615&lng=-84.285&l=11 Archived at:

http://www.webcitation.org/6Rx49628r Accessed: 9/3/2014

56) New York Times: Mapping the 2010 US Census, Distribution of racial and ethnic groups in 2010, Atlanta, GA.

http://projects.nytimes.com/census/2010/map?view=DotView2010&lat=33.7726&lng=-84.378&l=14 Archived at: http://www.webcitation.org/6Rx44zPyU Accessed: 9/3/2014

Chapter 3: Supplemental Figures:

Figures 6, 7 and 9 from the main text duplicated with Kernel Density calculations set with a 1 mile search radius

Supplemental Figure 1: Estimated density of white (A) and black (B) social network application users in the City of Atlanta (grey outline), showing major highways (black lines) and roads (dark red lines) and highlighting the "Midtown" area of Atlanta (yellow rectangle). Kernel Densities were estimated from sample data standardized to 1 mile circular radii, and smoothed to 2 miles using a Gaussian smoother that concentrates the majority of the density at the sample point, and averages over all adjacent data points within the smoothing radius.
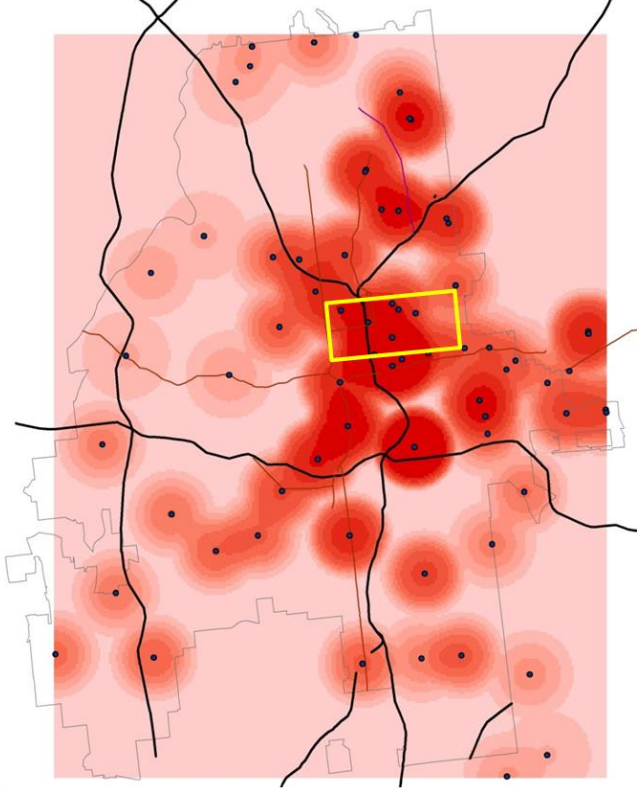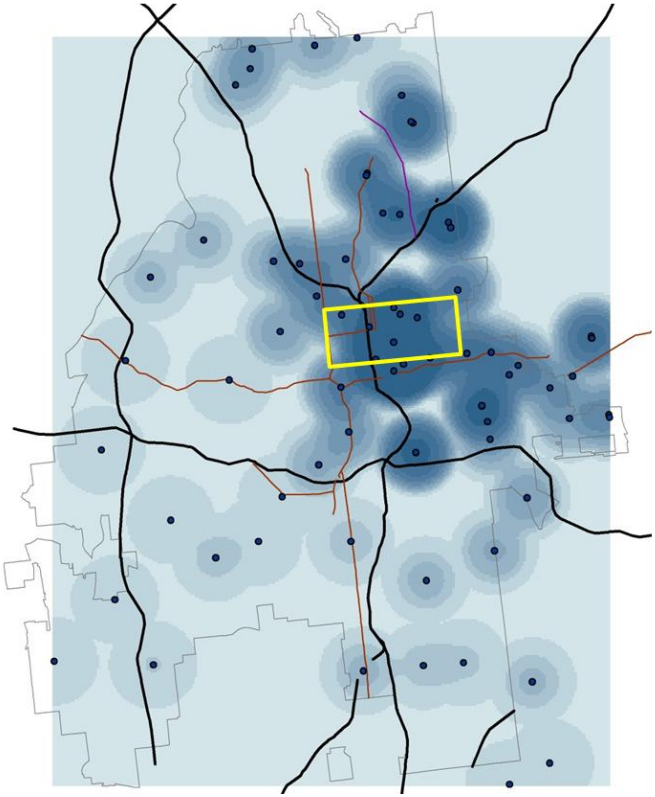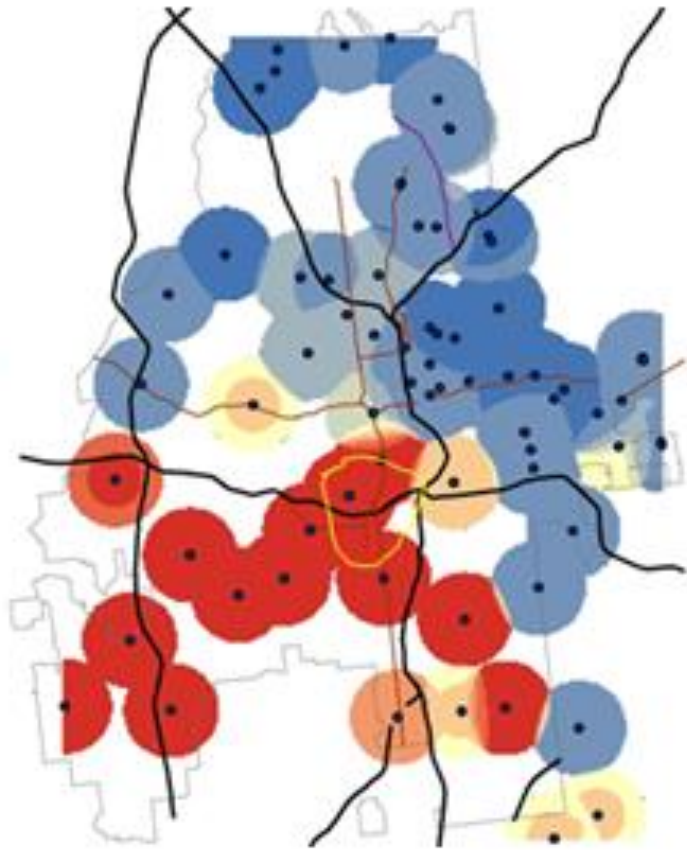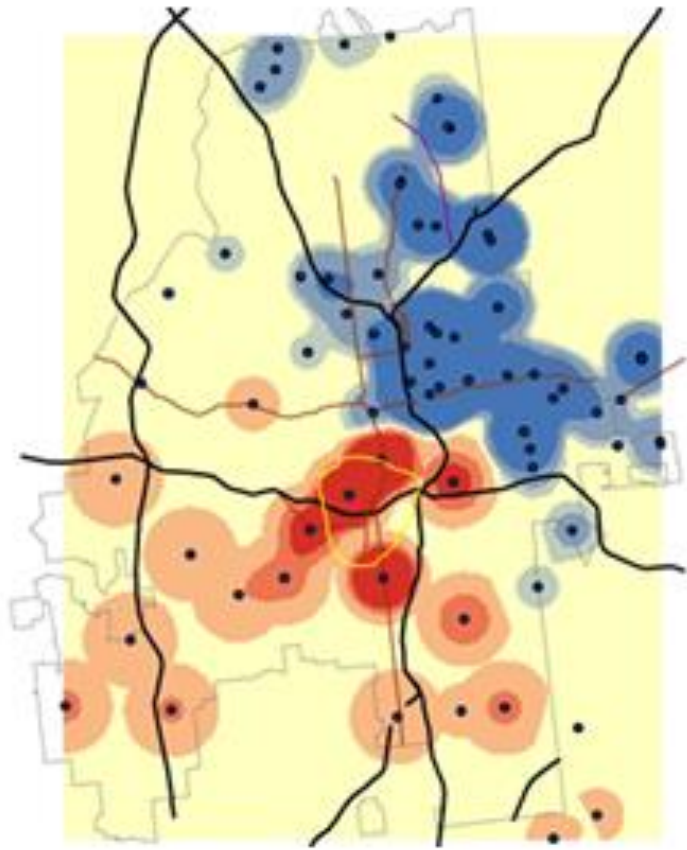
Supplemental Figure 2: Replicate of Figure 7 from the main text with 1 mile search radius for the kernel density smoother.. This figure again shows a comparison of the density of black and white social networking application users in the City of Atlanta. Panel A shows the absolute difference in users (Density of Black users – Density of White users) color coded so that areas with more Black users appear red and those with more white users appear blue. Yellow regions are areas where the density is estimated to be about the same. Panel A highlights a small section of the city (the yellow band around the area shaded the darkest red) where there are many more black than white application users. Panel B shows a comparison of the relative size of the densities of black and white users (Density of Black Users/Density of White Users). With this measure, Atlanta is divided nearly in half, with relatively more black users in the southwest and more white users to the North and East. The yellow band in Panel B shows the region with the highest absolute excess of black users for comparison purposes. Figure 1: Highlighted examples of how outcome measures will change depending on the smoothing radius used for the kernel density. Panel A shows the difference measure (Density of black users – Density of White users) calculated based on densities with a 1 mile smoothing radius. Areas with blue shading indicate points where the absolute value of this difference is negative (indicating more white users); areas with red shading highlight points with higher absolute numbers of black users.   The yellow circle in Panel A highlights an area that was also highlighted at a 2 mile radius, but there are two other points to the south and east of this circle, that did not appear in the analysis using the 2 mile smoothing radius.
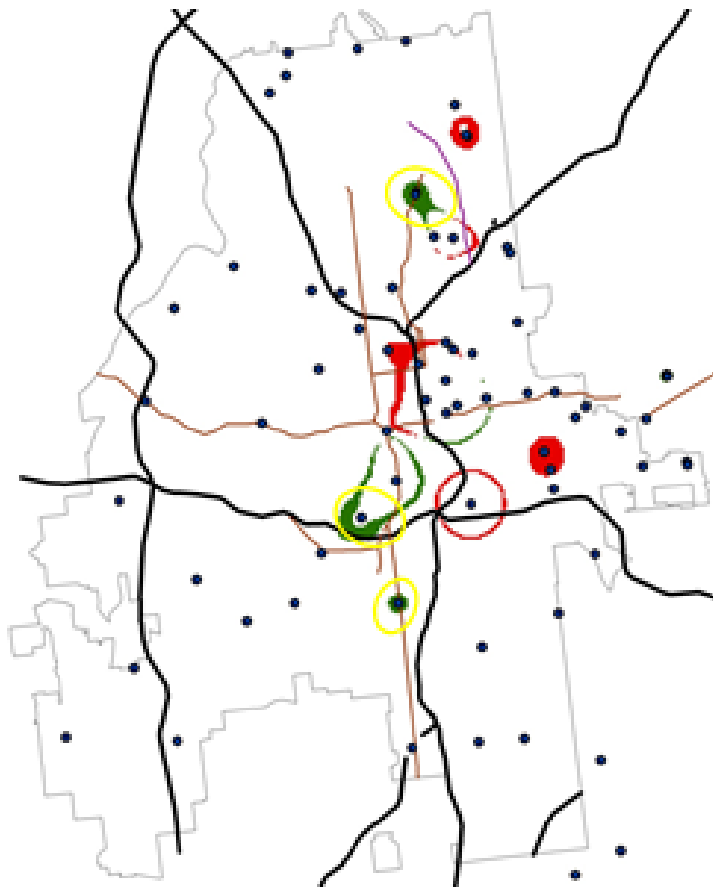
Comparison to figure 9 from the main text, redone with a 1 mile search radius for the kernel density smoother. This version shows the areas where the density of young black users (green) or black users overall (red) exceeds the 95th percentile. White regions of Panel B indicate areas where the density of black and young black users is similar. With the one mile smoothing parameter, there are very few areas with more young black men using a social networking application than black men overall.

Supplemental Table 1: Matrix of day and time of sampling For 79 data collection points included in the analysis

	Time		
	Morning	Afternoon/Evening	Late Night
Day			
Monday	5	4	
Tuesday	6	5	
Thursday			20
Friday	10	18	11

Chapter 4 - Specific Aim 2:

Compare proposed tests for statistical interaction of two exposures using simulated data in which two dichotomous variables interact to have effects that are greater than additive risk differences but less than multiplicative risk ratios, indicative of sufficient component cause interaction

Section 1: Introduction

Sufficient component cause interaction (SCC interaction, sometimes also called biologic interaction,[1-3] causal co-action,[1,2] or synergism[3]) arises in the situation when two or more causes participate in the same sufficient cause for a given disease.[1-3] This particular type of interaction is important to public health because, when present, intervening on either cause will prevent disease. Methods for detecting a mechanism in which two exposures interact in this way to cause disease have seen recent rapid development in the epidemiology literature. Significant work by VanderWeele and Robins[3], further developed by Vanderweele and others,[4-8] provides a theoretical framework and gives conditions sufficient to identify this specific type of interaction. However, this rapid advance comes with many caveats about limited power to identify "interaction" effects in practice.[9-14] The situation is complicated by the fact that the term "interaction" is used loosely by epidemiologists,[11] and there are situations when statistical interaction (also called effect measure modification[1-3,11]), detected as a departure from the assumptions about how risks of two exposures combine in a statistical model, does not indicate the presence of SCC interaction.[1,2,5,8,11]

Vanderweele and Robins[3] defined what they called "definitive interdependence." It is a set of response patterns or types, presented in Appendix 1, which they show definitely imply that SCC interaction is present. They propose[5] statistical tests to detect these response types thereby providing a way to detect SCC interaction. Appendix 1 contains definitions of important terms including statistical interaction, SCC interaction, and definitive interdependence. In this manuscript we use "interaction" to refer to statistical interaction which may or may not be causal, "SCC interaction" to refer the specific type of interaction defined as such by Rothman[1,2] and "definitive interdependence" to refer to the subset of patterns of risk defined by Vanderweele and Robins.[3]

Two recent studies involving the authors of the present manuscript identified significant interaction between two risk factors,[17,18] leading us to seek practical applications of the newly proposed methods as guidance for how to evaluate these interactions in terms of causality. Although some early work on interaction compared two of the many statistical tests,[12] we are not aware of any recent updates directly comparing and contrasting the newer approaches for detection of interaction, particularly those based on definitive interdependence.[19-26]

In the current study, we begin with a detailed description of: the range of ways two dichotomous causal exposures can interact; an outline of several statistical models that accommodate interaction in different ways; and statistical tests of interaction that can be formulated from these models. We then compare performance across 13 variations

of these tests to assess the presence of SCC interaction, including several newly

implemented within existing statistical software for this study (specified in Table 1). We

evaluate test performance in two general areas: 1) the proportion of times the tests

identify interaction when no SCC interaction exists (Type 1 error); and 2) the ability of

each test to detect SCC interaction of varying magnitude when it exists (Power). To do

so, we conduct a series of simulations based on binomial outcomes where we set the

disease risk depending on presence or absence of 2 dichotomous exposures X1 and X2.

We compare the tests across ranges of sample sizes often employed in epidemiologic

studies and a range of risks and risk patterns (detailed below). We conclude with

practical recommendations for how to proceed when the goal is to assess SCC

interaction.

Section 2: A review of proposed methods to assess interaction

When two causes act together

To begin, assume we have two dichotomous exposures X1 and X2 (Notation used

throughout this manuscript is defined and differentiated from that used in other

texts[2,3,5] in Appendix 1.) There are a range of possible effects that could be observed

when two exposures that each has an effect are present. In the presence of SCC

interaction, the combined effects of X1 and X2 might differ from the sum of each

exposure's individual effect measured on the risk difference scale. But, how much

departure from risk difference additivity is observed would depend on the mix of the

sufficient component causes that are present due to interaction of X1 and X2 in the

study population. The range of possible combined effects of X1 and X2 could be

described in many ways, but one is to consider the observed combination of the effects

of X1 and X2 relative to the effects estimated by typical statistical techniques, to see if

and how they depart from assumptions of additive risk differences or multiplicative risk

ratios.

Specifying the Pattern of Disease Risk

To describe the pattern of disease risk in the presence of two exposures, we can specify

the relationship between exposures and risk of disease as additive or multiplicative,

although other specifications have been proposed.[27-28] Equations 1 and 2 provide typical

formulation for additive risk and for multiplicative risk, respectively.

$$R(X1,X2) = P(D|X1,X2) = \alpha_0 + \alpha_1 X1 + \alpha_2 X2 + \alpha_3 X1X2 \qquad \text{(eq 1)}$$

$$R(X1,X2) = P(D|X1,X2) = \exp(B_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X1X2) \qquad \text{(eq 2)}$$

These formulations describe the risk of disease associated with exposures X1 and X2,

using parameters α_3 and β_3 to quantify departures from additive risk differences and

multiplicative risk ratios for X1 and X2, respectively. In Equation 1, when α_3 is equal to

zero, the combined effect of exposures X1 and X2 is exactly what you would expect if

the risk differences associated with X1 and X2 were added together. The risk (the

probability of disease P(D)) for those exposed to both X1 and X2 under additivity is given

by: P(D|X1=1,X2=1)= P(D|X1=1, X2=0) + P(D|X1=0, X2=1) – P(D|X1=0,X2=0). That is,

with an additive pattern the risk for someone exposed to both X1 and X2 equals the sum

of the individual risks, minus the baseline risk for those not exposed to either X1 or X2.

In Equation 2, when β_3 equals zero, risk ratios are exactly multiplicative. In terms of risks

(as opposed to risk ratios) a multiplicative pattern would mean that the probability of

disease for those exposed to both X1 and X2 is given by the product of the individual

risks and the inverse of the baseline risk:

P(D|X1=1,X2=1)=P(D|X1=1,X2=0)*P(D|X1=0,X2=1)*(1/P(D|X1=0,X2=0)). Equation 1 is

referred to as a linear binomial model or a binomial model with "identity" link.[23,24]

Multiplicative risk models typically either use a Poisson or Binomial distribution to

model counts of incident disease based on the probability of disease as defined in

equation 2.[13, 29-30] The logistic (logit binomial risk) model is another popular model of

multiplicative risk, a special case of the binomial risk model which uses the logit function

to scale the risks to be between 0 and 1. Other models have also been proposed.

$$R(X1,X2)= P(D|X1,X2) = exp(\beta_0)*(1+ \beta_1X1 + \beta_2X2 + \beta_3X1X2) \quad /$$

$$(1+ (exp(\beta_0)*(1+ \beta_1X1 + \beta_2X2 + \beta_3X1X2))) \qquad (eq\ 3)$$

$$R(X1,X2)= P(D|X1,X2) = Exp(\beta_0+\beta_1X1+ \beta_2X2+ \beta_3X1X2) \qquad / \qquad (eq\ 4)$$

$$(1+ Exp(\beta_0+\beta_1X1+ \beta_2X2+ \beta_3X1X2))$$

Richardson and Kaufman[21] showed how to use equation 3 with logit risks to assess

effect modification on the risk difference scale; with this parameterization β_1, β_2,

correspond to the relative excess risk odds for X1 and X2, and β_3 corresponds to the

relative excess risk odds due to interaction. However, the vast majority of logit models

use the multiplicative form in equation 4.[12,13] To assess interaction as modification of

the risk difference based on equations 2 and 4 requires manipulating model parameters

to estimate risks for each level of exposure.   It is well-known that risk odds ratios, such

as those calculated from parameters of the logistic model, may overestimate

corresponding risk ratios when disease is common.[2,21,29-31]

Measures of departure from additive risk differences

Rothman and Greenland describe two measures of departure from additive risk

differences that can be used, with assumptions, to identify SCC interaction. The

measures are the interaction contrast (IC) and interaction contrast ratio (ICR) also

referred to as the relative excess risk due to interaction (RERI).[2] The IC is the difference

of risk differences

$$IC = (R11 - R10) - (R01 - R00) = R11\text{-}R10\text{-}R01\text{+}R00 \qquad \text{(eq 5)}$$

where, for example, R11 is the $P(D|X1=1,X2=1)$. The risks at each combination of the

levels of the two exposures can be calculated from any statistical model that estimates

these risks; the more difficult piece is calculating the variance of functions of model

parameters such as the IC and ICR.[9] When risk is modeled as if it were additive (equation

1), the IC can be estimated directly as the parameter α_3, with variance and confidence

limits provided by the software used to fit the model. The ICR is calculated by dividing

the IC by the risk in the unexposed group R00, resulting in a difference in the excess

relative risks (RR-1):

$$ICR= R11/R00 \text{ -}R10/R00 \text{ -}R01/R00\text{+}R00/R00$$

$$= RR(11) - RR(10) - RR(01) + 1 \qquad \text{(eq 6)}$$

The IC and ICR are important for the assessment of SCC interaction because when

exposures X1 and X2 are never preventive (the related concept of monotonic exposures

is defined in Appendix 1), and if there is no confounding, then an IC > 0 or ICR >0 implies

the presence of SCC interaction.[2,3,5] Thus, one can test for SCC interaction by testing if

IC or ICR is > 0 (when exposure effects are never preventive).[3,5]

These risk ratios and the ICR can be calculated from the parameters of multiplicative risk

models in several different ways (examples of SAS code in Appendix 2) using software

available to fit these models.

Methods of risk model estimation and tests for interaction using SAS software
The rapid expansion of proposed methods for estimating quantities such as IC and ICR

and for detecting SCC interaction based on them is likely due to the availability of

software packages that provide the elements necessary to calculate them.[21-26] Most of

the published examples use SAS software (SAS Institute, Cary, NC) to implement the

estimation of these parameters, although other software contain similar methods for

estimating risks (e.g. STATA, (Stata Corp, College Station, TX) and R, (www.rproject.org)).

Within SAS there are at least three procedures that can be used to estimate the

parameters in models of disease risk, and we review advantages and disadvantages of

each; example code is referenced in Table 1 and included in Appendix 2. The GENMOD

procedure advocated by Spiegelman[23] and Brumback[24] provides perhaps the most

straightforward methods for estimating the IC using a linear risk model as in equation 1,

and allows for direct estimation of the confidence interval for the α_3 parameter using

both Wald and likelihood ratio methods.[2,12,13,32,33] The main drawback to this approach

is that it does not constrain the individual risks (i.e. R11, R10, R01 and R00) to be

between 0 and 1 and there is potential to obtain estimates that are not possible true

values of risk.[34]

There are similar issues with both Poisson and log-binomial models used to calculate risk ratios, in that they do not constrain the risk to be less than 1. Additionally, in practice log-binomial models frequently fail to converge, while Poisson models will overestimate the standard error of modeled parameters if risks truly follow a binomial distribution.[30,34] GENMOD can also be used to calculate multiplicative risk models, both with model and the empirical standard errors derived from generalized estimating equation fitting methods recommended for Poisson models.[29,30,34] However, GENMOD does not allow one to manipulate the estimated results to enable calculation of the IC or ICR directly from multiplicative model parameters. In contrast, the NLMIXED procedure does allow for custom linear combinations of fitted parameters, and uses the Delta method[35] to approximate their standard errors.[36] For example, to calculate the ICR for the log-binomial model similar to equation 2, one uses an ESTIMATE statement in SAS:

ESTIMATE "ICR" $\exp(\beta_1 + \beta_2 + \beta_3) - \exp(\beta_1) - \exp(\beta_2) + 1$;   (eq 7)

Kuss[25] recommends setting the degrees of freedom for the calculation of the variance for this estimate to a large number (e.g. 10,000), thereby artificially increasing the sample size from that observed in order to approximate a Wald confidence limit. Richardson[21] also provides a SAS macro that utilizes NLMIXED to estimate a likelihood ratio-based confidence limit for a linear odds model (as defined in equation 3 above). Kuss[25] also used a third SAS procedure, PROC NLP (for non-linear programming/optimization), which can directly calculate the likelihood ratio-based confidence interval for the ICR, when the fitted likelihood is coded such that the ICR is a

parameter to be estimated. Note that both Richardson[21] and Kuss[25] provide SAS code based on the logit risk model, and we have extended the latter's approach to use the log-binomial and Poisson risk models in our examples included in Appendix 2.

Tests for SCC interaction when exposure effects are not monotonic

The approaches above identify interactions through tests of ICR>0 or IC>0, which Vanderweele[5] notes are valid indicators of SCC interaction only when exposure is monotonic, specifically when there is never an instance when increasing exposure protects an individual from disease. The rationale for this is presented in detail in Appendix 1. In practice, one often cannot rule out this possibility. In this situation, Vanderweele[5] suggests another test for SCC interaction based on definitive interdependence; this test corresponds to IC− R00 >0 which is equivalent to ICR>1. When considering linear risk models of the form in equation 1, Vanderweele's test for SCC interaction when monotonic effects of exposure cannot be guaranteed is

$$\alpha_3 - \alpha_0 > 0 \qquad\qquad\qquad (eq\ 8)$$

This linear contrast can also be calculated directly using SAS's GENMOD procedure, again with both Wald and likelihood-based confidence limits provided, and it is possible to identify when the lower limit does not include zero. Vanderweele[5] also derives an equivalent test to equation 8 using parameters of a multiplicative risk model:

$$\exp(\beta_1 + \beta_2 + \beta_3) - \exp(\beta_1) - \exp(\beta_2) = (ICR-1) > 0 \qquad\qquad (eq\ 9)$$

and shows this to be equivalent to a joint test of:

$$(\beta_3 + \beta_1 - \log(2) > 0 \text{ and } \beta_3 + \beta_2 - \log(2) > 0) \qquad \text{(eq 10)}$$

However, Vanderweele[5] does not provide an example of how these might be calculated

within statistical software. Because the criterion in equation 9 corresponds to an ICR>1

we can use the confidence bounds around the ICR calculated by any of the

aforementioned SAS procedures to test for situations when the lower limit of the ICR is

greater than 1. The condition in equation 10 can also be coded as a linear contrast

providing a joint test of the two inequalities, including an adjustment for multiple tests

using the SAS PROC PLM procedure. Again, code to assess the criteria in equations 8 –

10 are provided in Appendix 2.

Section 3. Monte Carlo Simulations

We conducted a series of Monte Carlo simulations to evaluate the performance of the

tests to detect SCC interaction based on the IC or ICR as described in Section 2. For all

tables and figures except Supplemental Figure 2 (described in Appendix 1) we assigned

binary exposures X1 and X2 as if exposure was assigned experimentally as in a

randomized controlled trial (RCT) of both X1 and X2; we refer to this scenario as the RCT

design. In this design the number of participants in each of the 4 exposure categories or

"arms" of the trial (see Supplemental Figure 1 for a schematic) is assigned to be equal.

We considered sample sizes (for each arm) of 200, 500, 750, 1,000 and 500,000, where

the last scenario, with a total population size of 2 million, would be realized only in an

extremely large study or perhaps a pooled study. .

After assigning exposure to the population, we next assign a dichotomous disease

outcome as a random variate with the number of disease cases (ones) following a

binomial distribution with density function

$$f(d) = \binom{n}{d} p^d * (1-p)^{n-d},$$

with sample size n as defined in the previous paragraph, d as the count of diseased

cases in n, and p as the underlying disease risk. We next define p as a function of the

baseline probability of disease p0 multiplied by a risk ratio associated with exposures X1

and X2, RR(X1,X2). We assign the baseline risk of disease p0=0.05 and vary the risk ratios

for each exposure RR(X=1,X2=0) and RR(X1=0,X2=1) over a range from 1 to 4. We then

multiply p0 by the risk ratio to obtain the risk p for those exposed to only X1 and only

X2. For the risk associated with both exposures we simulated situations where the risk

was: exactly additive on the risk difference scale, exactly multiplicative on the risk ratio

scale, or the risk for the combined exposure condition was somewhere in between

these two values, as specified below. We define risk as *exactly additive on the risk

difference scale* when the binomial risk parameter p was calculated as

$\quad$ p= (RR(X1=1,X2=0)+RR(X1=0,X2=1)−1)*p0 $\qquad\qquad\qquad$ (eq 11)

which is equivalent to equation 1 with α_3=0 and IC = ICR =0. In this case the β_3

parameter for models defined by equation 2 (log-binomial or Poisson risk models) and

equation 4 (logit binomial risk models) should be negative. In this case the null

hypothesis (IC = 0) is true, although SCC interaction cannot be ruled out for reasons described in Section 5 and Appendix 1.

We define the combined risk as *exactly multiplicative on the risk ratio scale* when:

$$p=RR(X1=1,X2=0)*RR(X1=0,X2=1)*p0 \qquad \text{(eq 12)}$$

Note this is equivalent to equation 2 with $\beta_3=0$. In this case, models such as equation 1 will have positive values for the α_3 parameter, indicating departure from additive risk differences, and the null hypothesis (IC =0, or ICR = 0) is false. IC > 0 or ICR > 0 implies SCC interaction is present because effects of both X1 and X2 are present and monotonic (never preventive).

To set p to be somewhere between additive risk differences (equation 11) and multiplicative risk ratios (equation 12) we define

$$p= \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(eq 13)}$$

$$(RR(X1=1,X2=0)+RR(X1=0,X2=1)-1)*p0 \qquad +$$

$$((RR(X1=1,X2=0)*RR(X1=0,X2=1)*p0 - (RR(X1=1,X2=0)+RR(X1=0,X2=1)-1)*p0)/\kappa)$$

and set κ between 1.25 and 5. The range of magnitudes of risk ratios associated with a single exposure results in a range of possible risk ratios for those with both exposures. Supplemental Table 2 in Appendix 1 shows the expected value of the ICR for values of RR(X1=1,X2=0) and RR(X1=0,X2=1) under the exactly additive and exactly multiplicative

scenarios as well as for κ = 2. For risks in the range defined by equation 13, equation 1 should have an α_3 parameter greater than zero, again indicative of SCC interaction and definitive interdependence when effects of X1 and X2 are monotonic, which is the case here. β_3 in equation 2 and equation 4 should be negative for the range considered, but $\beta_3 < 0$ does not generally indicate SCC interaction, even under monotonicity. For all simulated data sets, because we set the true risk associated with each exposure, we can calculate the true values of the IC and ICR directly.

Two measures of test performance were first compared with 500,000 persons in each exposure category: 1) the proportion of times the measures (incorrectly) "identify" SCC interaction when effects are exactly additive on the risk difference scale, using data generated from equation 11 (Type 1 error); and 2) the ability of each test to detect non-zero values of IC or ICR of varying magnitude when they are positive but less than what would be expected if risk ratios were multiplicative, using data generated from equation 13 (Power). We then compare these measures of test performance across ranges of sample size typically employed in epidemiologic studies. Supplemental Table 1 provides the values for sample size, individual risk ratios for X1 and X2, and interaction effects considered when creating each of the tables and figures included in the manuscript and supplemental results. All simulations were conducted using SAS version 9.2.3. Each scenario outlined in Supplemental Table 1 was repeated for 1,000 replicates. A sample program is included as Appendix 2 and shows both how the simulated dataset (in the RCT scenario) was created and how each test of interaction that we evaluated can be performed. Throughout, we assume no confounding, misclassification or selection bias.

Section 4: Results

Figure 1 shows the proportion of 1,000 simulations in which each of 6 tests for

interaction detected effects that would lead to an IC and ICR > 0, an indicator of SCC

interaction (Power) under the assumption of monotonicity. In this scenario, the IC and

ICR have values that correspond to risks between additive on the risk difference scale

and multiplicative on the risk ratio scale (κ=2 in equation 13), and the effect of each

individual exposure was varied between a risk ratio of 1 and 4, in a sample with 500,000

people in each of four categories of 2 dichotomous exposures. All of the tests

considered detected interaction effects in all 1,000 simulations (100% Power) when the

risk ratio for both individual exposures exceeded 1.25. Figure 2 shows the converse, the

Type 1 error rate when the true risk for those exposed to both X1 and X2 is exactly

additive on the risk difference scale (IC=ICR=0), as defined in equation 11. Contrasting

the top row of Figure 2 with that of Figure 1 illustrates the expected problem with using

multiplicative models to test for interaction. When risk difference modification exists

but is less than what would be observed if risk ratios were multiplicative (Figure 1) the

β_3 term in multiplicative models is often significantly less than zero. In models where the

risks of X1 and X2 have exactly additive risk differences (Figure 2), this is also true. A

"significant" departure from the multiplicative model can be due to a risk pattern that is

exactly additive or greater than additive.

Figure 2 also shows that ICRs calculated from the parameters of a logit binomial risk

(logistic) model as though the risk odds ratio was a risk ratio also have high Type 1 error

rates. This is expected because the estimates of the ICR are incorrectly calculated by

treating risk odds ratios as though they were risk ratios when the outcome is not rare.

Figure 3 displays both the Type 1 error rates (the leftmost bar in each cluster) and

power (the other 8 bars in each cluster) for 8 tests for interaction over 1000 simulations

in which the sample size was set at 1,000 in each strata of exposure, and the risk ratios

for X1 and X2 were set at 2.75 and 2.25, respectively. Figure 3 shows that, even at a

sample size of 1,000 in each of four strata of exposure, power remains low when the

true effects are close to additive (the left side of each cluster of bars), even though the

individual effects of X1 and X2 are strong. However, the power and Type 1 error rates

are generally similar across all the tests based on estimates of the IC from additive risk

models (the first two clusters of columns) or the ICR from multiplicative risk models

(Poisson and log-binomial models). At this sample size, the results are similar within

model type whether we considered Wald or likelihood ratio tests. Results of the macro

suggested by Richardson and Kaufman were nearly identical to the likelihood ratio

results from the logit model estimated using PROC NLP presented in Figure 3(data not

shown). Supplemental Figure 4 in Appendix 1 is similar to Figure 3 except that the

sample size is 200 in each of the 4 combined strata of X1 and X2. This figure along with

Supplemental Figures 2 and 3 shows that power is expected to be too low to

consistently detect interaction effects at this sample size. When baseline risk or the

individual and combined effects of X1 and X2 were smaller, sample sizes required to

detect interaction effects increased, and several of the models (particularly those

estimated using PROC NLP and log-binomial models estimated using NLMIXED) had

convergence issues (data not shown).

Figure 4 shows the simulated power (Row 1) and Type 1 error (Row 2) for the measures

in equations 8 – 10, which indicates SCC interaction when the effects of individual

exposures cannot be assumed to be monotonic. These results can be compared to

Figures 1 and 2, where the test was the less stringent (e.g. null, IC = 0) since IC > 0

indicates SCC interaction under the assumption of monotonic effects. The tests

presented in Figure 4 are based on lower limits of confidence intervals of the linear

contrasts α_3-α_0 and the ICR being greater than 0 and 1 as defined in equations 8 and 9

respectively. We see null hypotheses are rarely rejected when the risk differences are

truly additive (they have low Type 1 error), but hypotheses are also rarely rejected (they

have low power) when the risk ratio for either X1 or X2 is less than 2 and the combined

effects of X1 and X2 are greater than additive on the risk difference scale but less than

multiplicative on the risk ratio scale. Furthermore, as in Figure 3 and Supplemental

Figure 4, the power to detect interaction effects that would lead to an IC or ICR > 0 is

lower when the interaction is closer to additive than multiplicative, even if the risk

ratios for both X1 and X2 are greater than 2 (Figure 5 and Supplemental Figures 6 and

7).

Section 5: Discussion of simulation results and practical recommendations for testing for interaction

We provide a systematic comparison of 13 proposed tests for interaction, considering

situations when risk is in the range between exactly additive on the risk difference scale

and exactly multiplicative on the risk ratio scale. We illustrate how the power to detect

SCC interaction effects of this magnitude depends the strength of the individual effects

of each exposure on disease, and the amount of interaction (reflected in the IC or ICR),

as well as the sample size, particularly for the subset of the population with both exposures. We also showed that only exceptionally large observational studies are likely to have enough participants with both exposures to detect SCC interaction effects in this range. This suggests that studies with a goal of detecting or quantifying SCC interaction should consider a study design such as a prospective cohort or most ideally a randomized controlled trial in order to achieve adequate numbers of individuals with exposure to all combinations of both causes.[37,38]

Vanderweele has methodically illustrated the assumptions necessary to detect what he defines as definitive interdependence along with proposed tests based on the IC and ICR.[3,5] Here we have shown both how to estimate the ICR or other statistics needed for these tests using existing statistical software, when we can and cannot assume monotonic effects. If we cannot assume monotonic effects, detection of SCC interaction when it truly exists is difficult at typically achievable sample sizes.

When the criteria described in equations 8-10 are met, Vanderweele and Robins have shown that SCC interaction must exist. However, the converse is not true, when their proposed criteria are not met SCC interaction may still occur. Vanderweele and Robins have cautioned that the power of these tests would be lower than that of tests of departure from additive risk differences that can be used when effects of X1 and X2 are monotonic. In our simulations, we evaluate scenarios that illustrate the extent to which the power is lowered. It seems unlikely that researchers could presume effects of exposures to always be monotonic, unless the causal mechanism for disease occurrence

is so well understood that a study that would include 1,000 individuals exposed to both exposures is unnecessary, if not unethical. However, we find that the Type 1 error rates are low for both the tests of ICR > 0 and ICR > 1. Thus, differences between the tests for definitive interdependence proposed by Vanderweele and those for SCC interaction proposed by Rothman[2] should be reported and interpreted carefully.

The results of our simulations provide some additional practical guidance for assessment of SCC interaction effects. First, we find tests of IC or ICR > 0 can have sufficient power to detect interaction effects that lie between exactly additive on the risk difference scale and exactly multiplicative on the risk ratio scale, but require large sample sizes. Unfortunately, observational studies generally do not have very large sample sizes in all strata of the exposures of interest. Second, our simulations suggest that interaction patterns that are close to additive on the risk difference scale cannot be detected with sufficient power at sample sizes typical of observational studies. Interaction patterns that are closer to multiplicative on the risk ratio scale were detected with adequate power when the sample size included from 750 to 1,250 individuals in each of the four combinations of exposure to two dichotomous causes of interest. If baseline risk or the individual and combined effects of X1 and X2 were smaller, sample sizes required to detect departure from additive risk differences using the IC or ICR would increase.

Because the effects of each exposure need to be strong to be reliably detected, methods that rely on logistic models to approximate risk ratios via odds ratios are also

not advised. The ICRs calculated from parameters of these models will be biased

estimates of the true ICRs calculated from risk ratios.[2,21,25] We adapted tests using ICRs

calculated from parameters of logistic models offered by Kuss[25] for implementation with

parameters estimated by log-binomial risk models instead and included sample code to

fit these models in Appendix 2.

The results from PROC NLP provide likelihood ratio-based confidence limits directly, and

consensus in the literature is that these limits are generally preferred over the Wald-like

limits produced by NLMIXED. The macro by Richardson and Kaufman[21] could also be

modified to calculate the ICR and its likelihood ratio-based confidence limits using

parameters of a log-binomial model. However, one by-product of requiring large sample

sizes in each strata of exposure is that the Wald-based confidence limits provide nearly

identical results to likelihood ratio-based limits in this setting.

Model fitting algorithms for NLMIXED and NLP do not allow for estimation of empirical

standard errors as recommended when trying to fit Poisson models to data arising from

a log-binomial risk model.[29,30] As a result, ICR estimates that use risk parameters from

log-binomial models (matching the data generation model), when they converge, have

better power than those based on Poisson estimates of risk (which only approximate

the data generation model). The differences between models are largest when sample

sizes and the effect modification on the risk difference scale are small. Programming

log-binomial models to obtain estimates of the ICR is more straightforward using the

NLMIXED procedure compared to the NLP procedure, and using code similar to equation

7, the risks in each strata of exposure, the IC, the ICR, and other related measures[8] (See Appendix 2, example 5) can be calculated directly from the parameters estimated by the model. GENMOD can be used to test for IC > 0 and IC – R00 > 0, and the estimates necessary to formulate these tests may be easier to generate when linear binomial models converge. But, while we have provided the example code to calculate all of the measures we evaluated, for the scenarios considered we feel that performance of tests based on parameters from log-binomial models estimated using NLMIXED as well as its relative ease of use suggest it is the most effective current approach within SAS for testing for IC > 0, ICR > 0 and ICR > 1 within the same procedure, and thus may be preferred in practice.

In our simulations, we focused mainly on RCT analyses and didn't include confounding in our sample size or power calculations. In the presence of a strong confounder it would be necessary to also ensure that there is a sufficient sample size for participants exposed to both X1 and X2 within the strata of the confounder(s).[2,10-11,37,38] There is at least one other approach to modeling risk which we didn't cover here, marginal mean models estimated by back-transforming parameters from a logit binomial model to the risk scale.[39] Software to perform these calculations exists, but, in the presence of confounding, requires assumptions about the distribution of confounders used to transform the log-odds parameters back to log-risk parameters, and the existing software does not calculate the IC or ICR or their confidence intervals directly.[39,40] This represents an area for future research.

In summary, we have provided several examples of new code for tests of for departures from risk difference additivity, designed to detect SCC interaction. We also show through simulations that these tests give similar results when the individual exposures have monotonic effects. However, the power to detect SCC interaction is limited unless exposures have both strong individual effects and their combined effects approach or are greater than what would be expected if risk ratios were multiplicative, within studies based on large sample sizes. When a monotonic effect cannot be assumed, a more stringent test must be used (ICR > 1) and power is reduced. We hope that these results and the compendium of SAS code used to generate them provide practical insight into when and how to detect SCC interaction effects using concepts based on the IC and ICR within epidemiologic studies.

References:

1) Rothman KJ. Causes. Am J Epidemiol. 1976;104:587–592.

2) Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott-Williams & Wilkins; 2012.

3) VanderWeele TJ, Robins JM. The identification of synergism in the sufficient-component-cause framework. Epidemiology. 2007 May;18(3):329-39.

4) Hernán MA, Robins JM. Causal Inference. Chapman & Hall/CRC; 2015.

5) VanderWeele TJ. Sufficient cause interactions and statistical interactions. Epidemiology. 2009 Jan;20(1):6-13.

6) VanderWeele TJ. On the distinction between interaction and effect modification. Epidemiology. 2009 Nov;20(6):863-71.

7) VanderWeele TJ. Richardson TS. General theory for interactions in sufficient cause models with dichotomous exposures. Ann Stat. 2012;40(4):2128-2161.

8) VanderWeele TJ, Tchetgen Tchetgen EJ. Attributing effects to interactions. Epidemiology. 2014 Sep;25(5):711-22.

9) Skrondal A. Commentary: much ado about interactions. Epidemiology. 2014 Sep;25(5):723-6.

10) Greenland S. Interactions in epidemiology: relevance, identification, and estimation. Epidemiology. 2009 Jan;20(1):14-7.

11) Kaufman JS. Interaction reaction. Epidemiology. 2009 Mar;20(2):159-60.

12) Greenland S. Tests for interaction in epidemiologic studies: a review and a study of power. Stat Med. 1983 Apr-Jun;2(2):243-51.

13) Kleinbaum DG, Klein M. Logistic Regression: A Self-Learning Text. 3rd ed. New York, Springer Science and Business Media; 2010.

14) Greenland S. Additive risk versus additive relative risk models. Epidemiology. 1993;4(1):32–36.

15) Darroch JN, Borkent M. Synergism, attributable risk and interaction for two binary exposure factors. Biometrika. 1994;81:259 –270.

16) Skrondal A. Interaction as departure from additivity in case-control studies: a cautionary note. Am J Epidemiol. 2003 Aug 1;158(3):251-8.

17) Delaney KP, Rurangirwa J, Facente S, Dowling T, Janson M, et al. Using a multi-test algorithm to improve the positive predictive value of rapid HIV testing and linkage to HIV care in non-clinical HIV test sites. JAIDS Submitted.

18) Goga AE, Dinh TH, Jackson DJ, Lombard C, Delaney KP, et al for the South Africa PMTCT Evaluation (SAPMCTE) Team. First population-level effectiveness evaluation of a national programme to prevent HIV transmission from mother to child, South Africa. J Epidemiol Community Health. 2014 Nov 4. pii: jech-2014-204535.

19) Knol MJ, Vanderweele TJ, Groenwold RH et al. Estimating measures of interaction on an additive scale for preventive exposures. Eur J Epidemiol 2011;(26(6) 433-438.

20) Knol MJ, Vanderweele TJ. Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol 2012; 41(2): 514-520.

21) Richardson DB, Kaufman JS. Estimation of the relative excess risk due to interaction and associated confidence bounds. Am J Epidemiol. 2009 Mar 15;169(6):756-60.

22) Rose S, van der Laan M. A double robust approach to causal effects in case-control studies. Am J Epidemiol. 2014 Mar 15;179(6):663-9.

23) Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. Am J Epidemiol. 2005 Aug 1;162(3):199-200.

24) Brumback BA, Bouldin ED, Zheng HW, Cannell MB, Andresen EM. Testing and estimating model-adjusted effect-measure modification using marginal structural models and complex survey data. Am J Epidemiol. 2010 Nov 1;172(9):1085-91.

25) Kuss O, Schmidt-Pokrzywniak A, Stang A. Confidence intervals for the interaction contrast ratio. Epidemiology. 2010 Mar;21(2):273-4.

26) Cheung YB. A Modified Least-Squares Regression Approach to the Estimation of Risk Difference. Am. J. Epidemiol. 2007; 166(11): 1337-1344. Greenland S. Tests for interaction in epidemiologic studies: a review and a study of power. Stat Med. 1983 Apr-Jun;2(2):243-51.

27) Ning Y, Woodward M. Relative risk regression models with inverse polynomials. Stat Med. 2013 Aug 30;32(19):3235-46.

28) Greenland S. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. Epidemiology. 1995 Jul;6(4):356-65.

29) McNutt LA, Wu C, Xue X, et al. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am J Epidemiol 2003; 157(10):940-943.

30) Peterson MR. Deddens JA. A comparison of two methods for estimating prevalence ratios. BMC Med Res Methodol. 2008 Feb 28;8:9.

31) Campbell UB, Gatto NM, Schwartz S. Distributional interaction: Interpretational problems when using incidence odds ratios to assess interaction. Epidemiol Perspect Innov. 2005 Mar 3;2(1):1.

32) Kleinbaum DG, Kupper LL, Nizam A and Muller KE. Applied regression analysis and other multivariable methods, 4th Ed. Belmont, Ca: Thomson-Brooks/Cole; 2008

33) Storer BE, Wacholder S, Breslow N. Maximum likelihood fitting of general risk models to stratified data. Appl Stat. 1983;32(2):172–181.

34) Chu H, Cole SR. Estimation of risk ratios in cohort studies with common outcomes: a Bayesian approach. Epidemiology. 2010 Nov;21(6):855-62

35) Cox, C. "Delta Method," Encyclopedia of Biostatistics, Eds. Peter Armitage and Theodore Colton, New York: John Wiley, 1125–1127; 1998

36) SAS/STAT® 9.2 User's Guide, Second Edition. SAS Institute Inc., SAS OnlineDoc 9.2.3, Cary, NC: SAS Institute Inc., 2002-2015.

37) Rothman KJ, Gallacher JEJ, Hatch EE. Why representativeness should be avoided. Int J Epidemiol 2013; 42:1012–14.

38) Rothman KJ, Hatch EE, Gallacher JEJ. Representativeness is not helpful in studying heterogeneity of effects across subgroups Int. J. Epidemiol 2014; 43(2): 633-634

39) Flanders WD, Rhodes PH. Large sample confidence intervals for regression standardized risks, risk ratios, and risk differences. J Chronic Dis. 1987;40(7):697-704.

40) Bieler GS, Brown GG, Williams RL, Brogan D. Estimating model-adjusted risks, risk differences, and risk ratios from complex survey data. Am J Epidemiol. 2010; 171(5): 618-23.

Table 1: Tests of interaction described in Section 2 and tables and figures in which evaluations of each test are presented

Test of Interaction	Equation in main text describing test	Reference	SAS example in Appendix 1	Tables and Figures in which the measure of interaction was evaluated
Tests of $\beta_3 \neq 0$ (departures from multiplicative risk ratios)				
Logit risk model	eq 4	2,13	Examples 3	Figures 1 and 2, supplemental figures 4 and 5
Log-risk model	eq 2	29,30	Examples 2	Figures 1 and 2, supplemental figures 4 and 5
Tests of $\alpha_3 \neq 0$ in linear risk model (departures from additive risk differences, equivalent to the Interaction contrast of eq 5)	eq 1	23,24	Examples 1 and 11	Figures 1,2,3,5, supplemental figures 4-7
Test of $\beta_3 > 0$ in linear odds model	eq 3	21	Example 8	Not presented results equivalent to logit risk model (NLP)
Tests of ICR > 0	eq 6			
Logit risk model (NLMIXED)	eq 7	25	Example 4	Figures 1,2, supplemental figures 4 and 5
Logit risk model (NLP)	eq 7	25	Example 7	Figure 3
Log-binomial model (NLMIXED)	eq 7	a	Example 5	Figures 1,2, 3, supplemental figures 2-7
Log-binomial model (NLP)	b	a	Example 10	Figure 3, supplemental figures 2,3,4,5
Poisson model (NLMIXED)	eq 7	a	Example 6	Figures 1,2, 3, supplemental figures 4 and 5
Poisson model (NLP)	b	a	Example 9	Figure 3, supplemental figures 4 and 5
Tests of $\alpha_3 - \alpha_0 > 0$ (Vanderweele's test for definitive interdependence using a linear risk model parameter when effects cannot be assumed to	eq 8	5, c	Example 11	Figures 4 and 5, supplemental figures 6 and 7

Test of Interaction	Equation in main text describing test	Reference	SAS example in Appendix 1	Tables and Figures in which the measure of interaction was evaluated
be monotonic)				
Test of ICR > 1 (Vanderweele's test for definitive interdependence when effects cannot be assumed to be monotonic using parameters from multiplicative risk models - only presented for NLMIXED log-binomial model)	eq 9	c	Example 5	Figures 4 and 5, supplemental figures 6 and 7
Joint test of: $(\beta_3 + \beta_1 - \log(2) > 0$ and $\beta_3 + \beta_2 - \log(2) > 0)$ Another representation of the test for definitive interdependence when effects cannot be assumed to be monotonic using parameters from multiplicative risk models	eq 10	c	Example 12	Figures 4 and 5, supplemental figures 6 and 7

eq. = equation as referenced in the text, ICR= Interaction contrast ratio

a) Kuss[25] presented only a logit binomial (logistic) model in his examples presented in reference 25, we have adapted both

examples to use both a log-binomial and Poisson risk model estimation

b) Kuss[25] manipulated the linear combination of parameters for a logit binomial (logistic) model such that the ICR was included as a parameter estimated from the model, with the interaction term being a combination of the ICR parameter and the parameters estimating the relative risks for each exposure level: b_interaction=log((ICR+exp(b_X1)+ exp(b_X2)-1)/(exp(b_X1)*exp(b_X2))). He then used this new parameter in the linear estimator of the overall effect: Eta=b_0+b_X1*X1+b_X2*X2+b_interaction*X1*X2. As described above, the examples provided in reference 25 only provided estimation of logit binomial (logistic) models, but we have added examples that use this same manipulation in log-binomial and Poisson based estimations of our risk models.

c) Vanderweele provided a description of these tests for definitive interdependence in reference 5, but no example code to calculate them. We have extended the linear binomial risk model[23,24] example to test whether the lower limit of the confidence interval for the linear contrast $\alpha_3-\alpha_0$ is greater than 0, the NLMIXED examples (i.e. equation 7) to test for the lower limit of the confidence interval of the ICR > 1, and provide example code to test for both inequalities in equation 10 using the SAS PROC PLM procedure. These examples are all included in the code provided in Appendix 2.

Figure Legends:

Figure 1 – Power to detect SCC interaction for six different proposed tests for interaction

This figure presents the proportion of 1,000 simulations in which 500,000 individuals were included in each of four levels of combinations of two dichotomous exposures, X1 and X2. Six tests for interaction are presented: a) the p-value for the test of $\alpha_3 \neq 0$ from a linear risk model[23,24] as in equation 1; b) the p-value for the test of $\beta_3 \neq 0$ in a log-binomial risk model[2] as in equation 2; c) the p-value for the test of $\beta_3 \neq 0$ in a multiplicative logit binomial (logistic) model[2,13] as in equation 4; d) the proportion of times the lower limit of the confidence interval of the ICR (equation 6) calculated from a multiplicative logit binomial (logistic) model did not include 0; e) the proportion of times the lower limit of the confidence interval of the ICR calculated from a log-binomial model of risk did not include 0; and f) the proportion of times the lower limit of the confidence interval of the ICR calculated from a Poisson risk model did not include 0. For d-f all models were calculated using equation 7 and the SAS PROC NLMIXED procedure, which approximates Wald-confidence limits using the delta method.[35,36] The strength of two exposures X1 and X2 was set using risk ratios ranging from 1 to 4; the strength of the interaction was set to be exactly halfway between additive risk differences and multiplicative risk ratios(κ=2 in equation 13) given the two assigned individual risk ratios; and a random draw from a binomial distribution was then used to assign the

disease outcome based on these effects. All methods have an expectation of high power when the risk ratios for both exposures exceeds 1.25.

Figure 2 –Probability of detecting SCC interaction when it doesn't exist (Type 1 error for six different proposed tests of interaction

This figure presents the proportion of 1,000 simulations in which 500,000 individuals were included in each of four levels of combinations of two dichotomous exposures, X1 and X2. Six tests of interaction are presented: a) the p-value for the test of $\alpha_3 \neq 0$ from a linear risk model[23,24] as in equation 1; b) the p-value for the test of $\beta_3 \neq 0$ in a log-binomial risk model[2] as in equation 2; c) the p-value for the test of $\beta_3 \neq 0$ in a multiplicative logit binomial (logistic) model[2,13] as in equation 4; d) the proportion of times the lower limit of the confidence interval of the ICR (equation 6) calculated from a multiplicative logit binomial (logistic) model of risk did not include 0; e) the proportion of times the lower limit of the confidence interval of the ICR calculated from a log-binomial model of risk did not include 0; and f) the proportion of times the lower limit of the confidence interval of the ICR calculated from a Poisson risk model did not include 0. For d-f all models were calculated using equation 7 and the SAS PROC NLMIXED procedure, which approximates Wald-confidence limits using the delta method.[35,36] The strength of two exposures X1 and X2 was set using a risk ratios ranging from 1to 4; and the interaction was set to be exactly additive on the risk difference scale, or equivalently, RR(X1=1,X2=1)= RR(X1=1, X2=0)+RR(X1=0,X2=1)-1 (equation 11); and a random draw from a binomial distribution was then used to assign the disease outcome based on these effects. Methods based on tests to identify departures from

multiplicative risk ratios (parts b and c) identify interaction in these scenarios where SCC

interaction does not exist due to departure from the assumed multiplicative risks. Part

d also shows high type 1 error due to overestimation of the ICR when using parameters

of a multiplicative logit binomial (logistic) model when the outcome is common.

Figure 3 – Probability of detecting an effect indicative of SCC interaction over a range

between exactly additive risk differences and exactly multiplicative risk ratios when

the individual risk ratios for two exposures X1 and X2 are 2.75 and 2.25 and sample

size in each of four strata of exposure is 1000.

This figure presents the proportion of 1,000 simulations in which 1,000 individuals were

included in each of four levels of combinations of two dichotomous exposures, X1 and

X2. Eight tests for interaction are presented as clusters within the figure: a) the

proportion of times when the lower limit of the Wald-based confidence interval of the

α_3 term from a linear risk model did not include zero[23,24]; b) the proportion of times

when the lower limit of the likelihood ratio-based confidence interval of the α_3 term

from a linear risk model did not include zero;[23,24] c) the proportion of times the lower

limit of the confidence interval of the ICR (equation 6) calculated using Wald limits from

a multiplicative logit binomial (logistic)[25] model did not include 0; d) the proportion of

times the lower limit of the confidence interval of the ICR (equation 6) calculated using

likelihood ratio-based limits from a multiplicative logit binomial (logistic)[25] model did not

include 0; e) the proportion of times the lower limit of the confidence interval of the ICR

calculated using Wald limits from a Poisson risk model[29,30] did not include 0; f) the proportion of times the lower limit of the confidence interval of the ICR calculated using Wald limits from a log-binomial model[29] of risk did not include 0; g) the proportion of times the lower limit of the confidence interval of the ICR calculated using likelihood ratio-based limits from a Poisson risk model did not include 0; i) the proportion of times the lower limit of the confidence interval of the ICR calculated using likelihood ratio-based limits from a log-binomial model of risk did not include 0; . For c, e and f Wald-confidence limits were approximated via the delta method[35,36] using the ESTIMATE statement in SAS PROC NLMIXED[25]. For d, g, and h, likelihood ratio-based limits are calculated using SAS PROC NLP using code adapted from Kuss.[25] The strengths of two exposures X1 and X2 were set to risk ratios of 2.75 and 2.25 respectively, relative to a baseline risk of disease of 0.05; the strength of the interaction was set to range between exactly additive risk differences and exactly multiplicative risk ratios given the two assigned individual risks using equations 11-13; and a random draw from a binomial distribution was then used to assign the disease outcome based on these effects. All tests for departures from additive risk differences based on linear binomial (a,b) or multiplicative risk models (e-h) have similar type 1 error rates and power to detect interaction effects in this range. For these tests power exceeds 80% only when the true ICR exceeds 1.25 (κ=1.75 in equation 13), i.e. when risk is closer to multiplicative than additive. Because the outcome is common, the power to detect an ICR based on multiplicative logit binomial risks (c,d) is higher than the corresponding tests of ICRs calculated from models where exponentiated parameters correspond to risk ratios

rather than odds ratios; however the type 1 errors are also higher for tests c and d than for tests e through h.

Figure 4 – Probability of detecting sufficient component cause interaction when it does and doesn't exist for 3 tests of definitive interdependence

This figure presents the proportion of 1,000 simulations in which 500,000 individuals were included in each of four levels of combinations of two dichotomous exposures, X1 and X2. Three additional (relative to Figures 1 and 2) tests[5] hypothesized to be sufficient conditions for detection of definitive interdependence when exposure effects cannot be assumed to be monotonic are presented in columns a-c: a) the proportion of times the lower limit of the confidence interval of the linear combination of parameters α_3-α_0 was greater than zero (equation 8 in the text); b) the proportion of times the lower limit of the confidence interval of the ICR calculated from a log-binomial model of risk did not include 1; and c) the proportion of times the one-sided p-value for both tests of $\beta_3 + \beta_1 > \log(2)$ and $\beta_3 + \beta_2 > \log(2)$ (equation 10) were < 0.05. All tests were calculated using SAS with code included in Supplemental Appendix 2. The strength of two exposures X1 and X2 was set using a risk ratio ranging from 1 to 4. For the top row the strength of the interaction was set to be exactly halfway between exactly additive risk differences and exactly multiplicative risk ratios, given the two assigned individual risk ratios ($\kappa=2$ in equation 13). For the bottom row the interaction was set to be exactly additive on the risk difference scale, that is RR(X1=1,X2=1)= RR(X1=1, X2=0)+RR(X1=0,X2=1)-1 (equation 11). A random draw from a binomial distribution was then used to assign the disease outcome based on these effects. Type 1 error rates for

these tests of interaction (bottom row) are less than 5%. When the true interaction

effects are between those anticipated when risk differences are additive and what

would be expected when risk ratios are multiplicative these three tests only detect this

SCC interaction when the individual risk ratios for exposures X1 and X2 are greater than

2.

Figure 5 – Comparison of the power to detect definitive interdependence when two

dichotomous exposures can and cannot be assumed to have monotonic effects

This figure presents the proportion of 1,000 simulations in which 750 individuals were

included in each of four levels of combinations of two dichotomous exposures, X1 and

X2. Clusters of columns represent 5 different tests for SCC interaction and definitive

interdependence. From right to left: a) the p-value for the test of $\alpha_3 \neq 0$ from a linear

risk model[23,24] as in equation 1; b) the proportion of times the lower limit of the

confidence interval of the linear combination of parameters $\alpha_3 - \alpha_0$ was greater than zero

(equation 8 in the text)[5]; 3) the proportion of times the lower limit of the confidence

interval of the ICR (equation 6) calculated from a log-binomial model of risk did not

include 0; 4) the proportion of times the lower limit of the confidence interval of the ICR

calculated from a log-binomial model of risk did not include 1; and 5) the proportion of

times the one-sided p-value for both tests of $\beta_3 + \beta_1 > \log(2)$ and $\beta_3 + \beta_2 > \log(2)$ (equation

10) were < 0.05. Models were calculated using SAS with code included in Supplemental

Appendix 2. The strength of two exposures X1 and X2 was set as a risk ratio of 2.5

compared to the unexposed who have an underlying risk of 0.05 (5%). In this figure the

strength of the interaction was varied according to equation 13, such that the true ICR

ranged from 0.45 to 1.8; and a random draw from a binomial distribution was then used

to assign the disease outcome within each of four strata of combinations of levels of

exposure based on these effects. The tests that identify definitive interdependence

when the effects of X1 and X2 are monotonic (a and c), have higher power to detect

interaction when it exists than do tests that make no assumption of monotonic effect

(b,d,e). However, none of the tests have very good power to detect interaction except

at the strongest effect sizes presented, with greater than 80% power only when the

ICR=1.8, close to the expected value of multiplicative risk ratios (ICR=2.25) for these

individual exposure risks.

Figure 1 – Power to detect SCC interaction for six different proposed tests for interaction
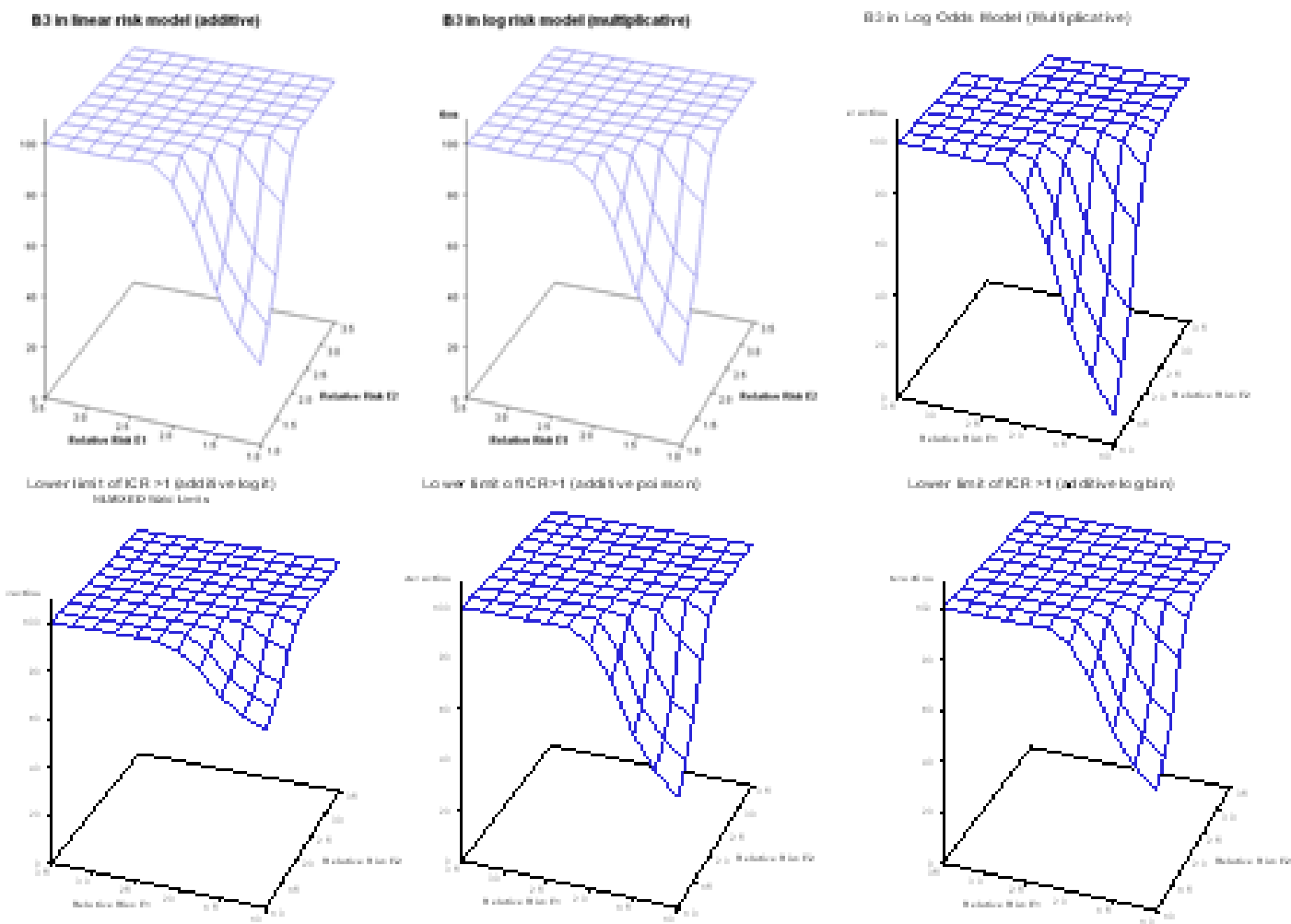
Figure 2 – Probability of detecting SCC interaction when it doesn't exist (Type 1 error) for six different proposed tests for interaction
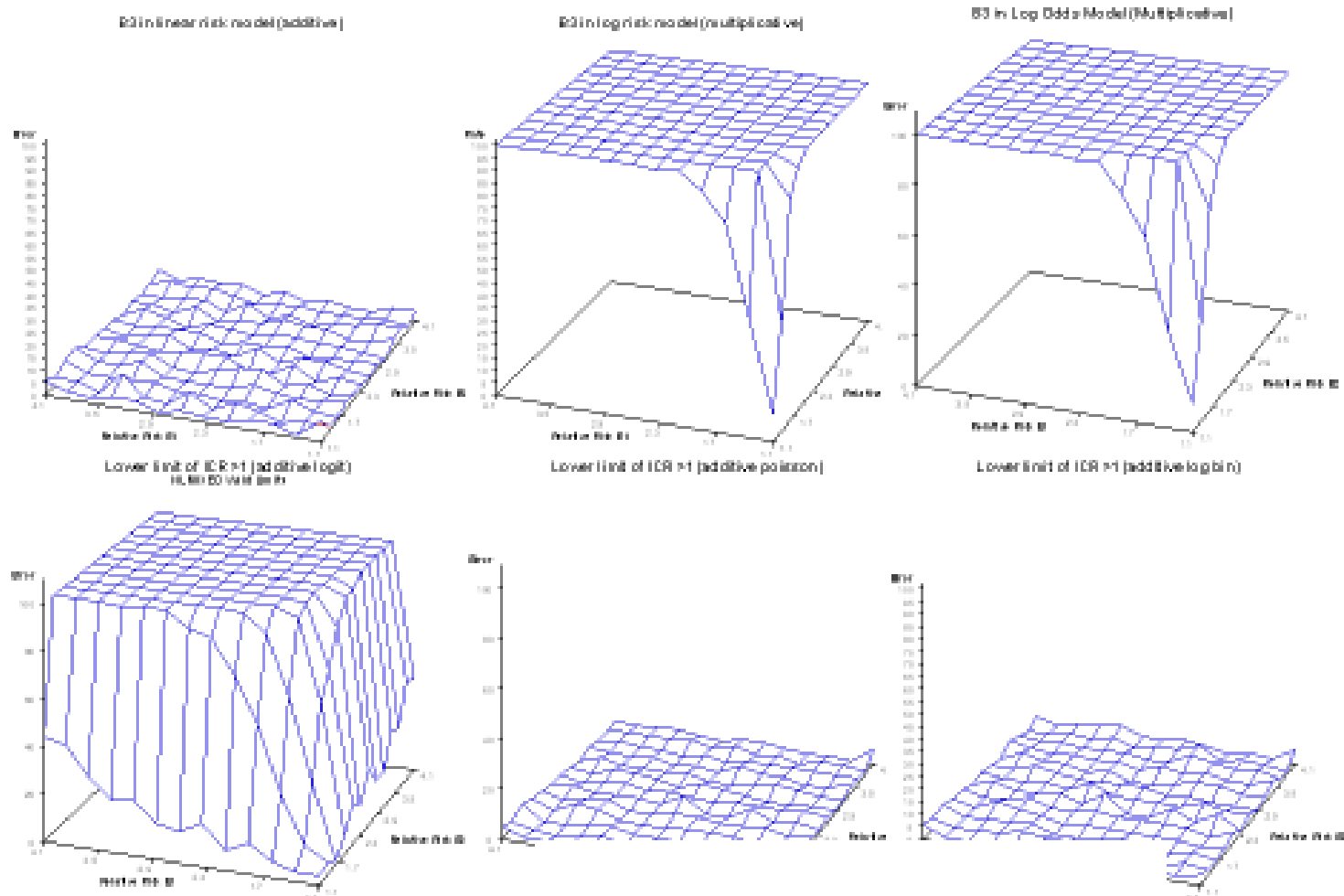
Figure 3 – Probability of detecting an effect indicative of SCC interaction over a range between exactly additive risk differences and

exactly multiplicative risk ratios when the individual risk ratios associated with two exposures X1 and X2 are 2.75 and 2.25 and

sample size in each of four strata of exposure is 1000.
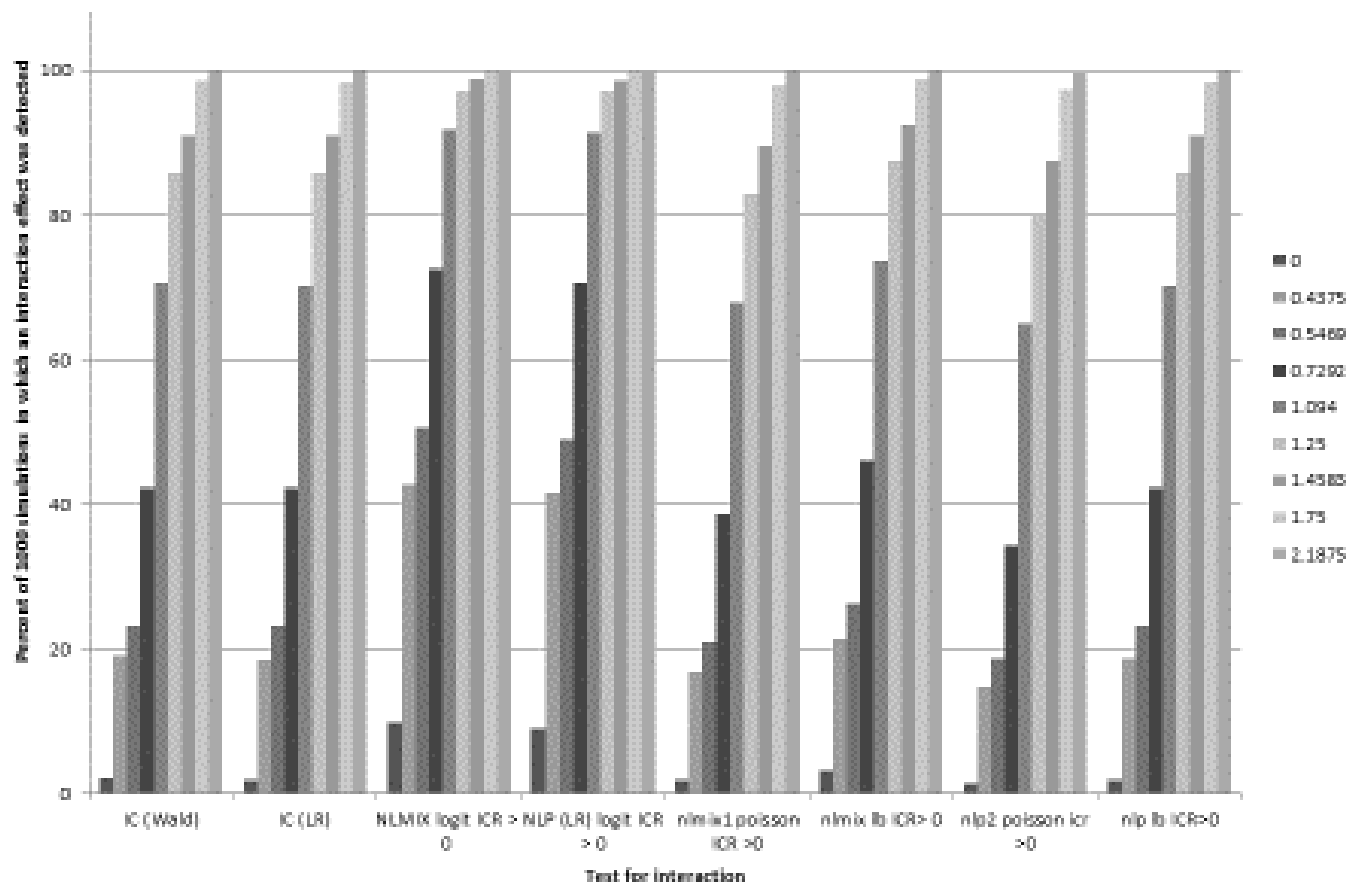
Figure 4 – Probability of detecting sufficient component cause interaction when it does and doesn't exist for 3 tests of definitive interdependence
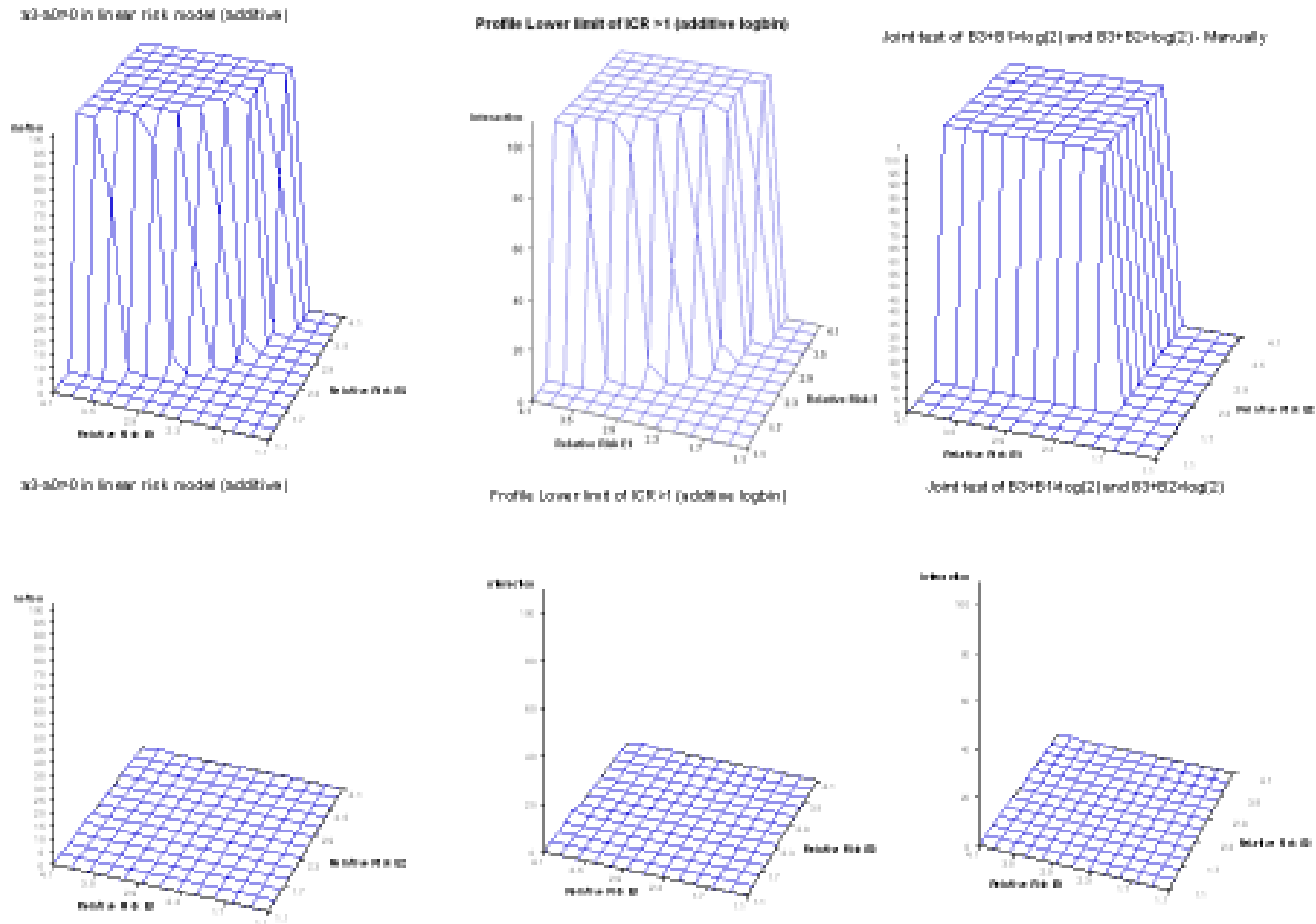
Figure 5 – Comparison of the power to detect definitive interdependence when two dichotomous exposures can and cannot be assumed to have monotonic effects
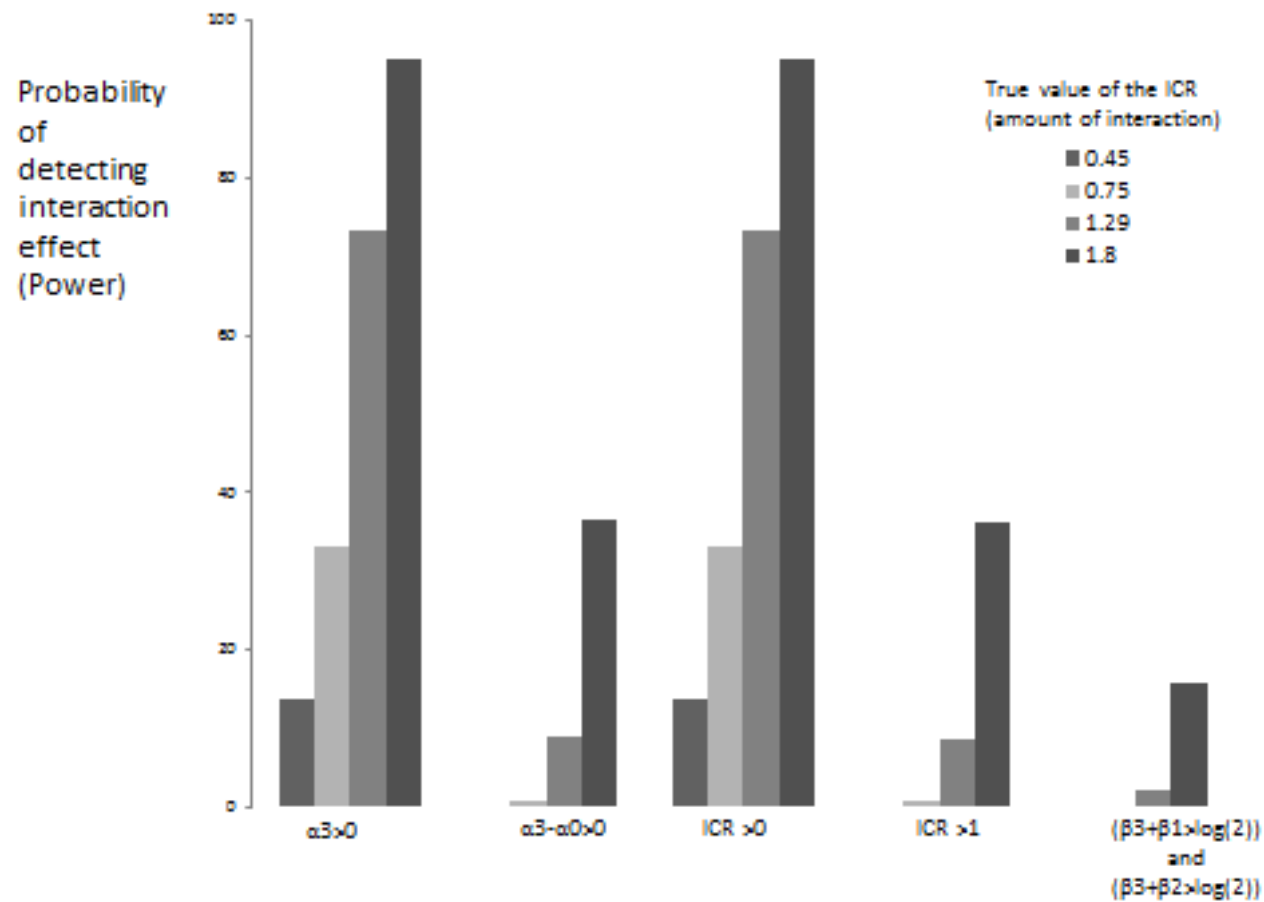
Supplemental materials for Aim 2:

Supplemental Appendix 1: Notation, Definitions and Further description of simulation parameters, distinction between sufficient component cause interaction, definitive interdependence and statistical interaction, and additional results

Notes on Notation

Here we define the notation used in our work and contrast it with notation used in the definitions provided in the next section.

In our examples we have two dichotomous exposures, labeled X1 and X2. In modern epidemiology[1] similar examples use labels X and Z, while Vanderweele and Robins[2,3] used E1 and E2. We indicate the presence of the risk factor X1 via X1=1, and absence of the risk factor X1 via X1=0.

We defined the risk of a dichotomous (0,1) outcome D as the probability of D=1 conditional on the two dichotomous exposures X1 and X2: P(D|X1=x1, X2=x2). For example the risk of disease in the population exposed to X1 but not X2 would be given as P(D|X1=1, X2=0). We also use the shorthand R(X1, X2) to describe risk, the equivalent to P(D|X1=1,X2=0) would be R(10) as in Modern Epidemiology[1], e.g. in Table 5-1 on page 73.

We also use RR(X1,X2) to define the risk ratio for those with exposure to one or both of X1 or X2, relative to the baseline risk in the unexposed (P(D|X1=0,X2=0). For example we define the risk ratio RR(11) to mean RR(X1=1,X2=1) or P(D|X1=1,X2=1) / P(D|X1=0,X2=0).

Finally, we define the excess relative risk as the risk ratio – 1 , such that additivity of risk differences can also be defined in terms of additivity of the excess risk ratios: (RR(11) –

1) = (RR(10) − 1) + (RR(01) − 1) which is algebraically equivalent to the condition when

the Interaction Contrast Ratio (ICR)=0.

Dictionary of terms related to the concept of interaction and how they are used in our manuscript

Sufficient component cause interaction/synergism/causal co-action –This type of interaction is generally defined within the framework of a sufficient component cause model. In this case, of the 9 possible combinations of two exposures X1 and X2 and their other necessary component causes, types F, G, H and I as defined in Figure 5-1 in Modern Epidemiology[1] on pages 80-81 exhibit sufficient component cause interaction in that they require the presence of two different causes in the same sufficient cause for disease to occur. Although these pages explain the concept of sufficient component cause interaction, Vanderweele and Robins[2] define it explicitly, but use the term "synergism" on page 330: "we will say that 2 causes, E1 and E2, for some outcome D, exhibit synergism if E1 and E2 are ever present together in the same sufficient cause. If E1 and $\bar{E}2$ are present together in the same sufficient cause then the 2 causes E1 and E2 are said to exhibit antagonism; in this case it could also be said that E1 and $\bar{E}2$ exhibit synergism. Note that E1 and E2 may exhibit both antagonism and synergism if, for example, E1 and E2 are present together in one sufficient cause and if E1 and $\bar{E}2$ are present together in another sufficient cause. In what follows we will not maintain the distinction between synergism and antagonism in so far as we will refer to a sufficient cause in which both E1 and $\bar{E}2$ are present as synergism between E1 and $\bar{E}2$ rather than as antagonism between E1 and E2." Because others use "synergism" to refer to only a subset of sufficient component cause interaction we use the latter when we refer to this concept in the manuscript, with the abbreviation SCC interaction after the first usage.

Definite Interdependence – This term was coined by Vanderweele and Robins[2], and formally defined as follows on pg. 332 (note that the use of "synergism" in the quote is as defined above, equivalent to SCC interaction): "Suppose that D and 2 of its causes, E1 and E2, are binary. We say that there is definite interdependence between the effect of E1 and E2 on D if there exists an individual ω for whom one of the following holds: $D10(\omega) = D01(\omega) = 0$ and $D11(\omega) = 1$; or $D11(\omega) = D00(\omega) = 0$ and $D01(\omega) = 1$; or $D11(\omega) = D00(\omega) = 0$ and $D10(\omega) = 1$; or $D01(\omega) = D10(\omega) = 0$ and $D00(\omega) = 1$. The definition of definite interdependence is equivalent to the presence within a population of an individual with a counterfactual response pattern of type 7, 8, 10, 12, 14, or 15… although definite interdependence is sufficient for a synergistic relationship, it is not necessary. There may be synergism between E1 and E2 even if they do not exhibit definite interdependence." We will use this term directly throughout the manuscript, particularly in the sections that describe conditions proposed by Vanderweele to test for this type of interaction.

Biologic interaction – This term is defined in two ways within Modern Epidemiology[1], one based on a potential outcomes model, the other based on the sufficient component cause model. In the potential outcomes model it is defined as any of the 16 possible potential outcome response types (See the last section of this appendix for a description of the counterfactual response types) for which the Interaction contrast is not = 0. This is best described in words on page 76 "For an interaction type, the effect of one factor depends on the person's status for the other factor". The definition based on the sufficient component cause model is the same as that used above to define sufficient

component cause interaction, (again, what Vanderweele and Robins[2] refer to as "synergism"), namely that "two or more component causes participate in the same sufficient cause" (Modern epidemiology, page 80.) However, Vanderweele and Robins point out (page 330 at the end of the second to last paragraph) "In contrast with "biologic interaction" which suggests that causes biologically act upon each other in bringing about the outcome, the term "synergism" suggests joint work on the outcome regardless of whether or not the causes act on one another." Also on page 330, they provide an example of "synergism" in which no biologic mechanisms "act upon each other in bringing about the outcome" and state that they prefer not to use this term when they mean what they define as "synergism" and what we refer to as sufficient component cause interaction (see the definition of this term above for why Vanderweele's definition[2] is equivalent to sufficient component cause interaction as defined in Modern Epidemiology.[1]) We agree that the term "biologic interaction" is ambiguous and also refrain from using it in the current work.

Statistical interaction/effect-measure modification – As defined in Modern Epidemiology on page 72, a departure from an additive or multiplicative risk model defined by a statistical test . Effect-measure modification is ambiguous and it is suggested that this term should be replaced with more specific phrasing. In our manuscript when we refer to statistical interaction we are generally referring to risk difference modification. Throughout the text we use risk difference modification or departure from additive risk differences to refer to tests involving interaction contrasts

and interaction contrast ratios and will specifically reserve risk-ratio or odds-ratio

modification for the two tests we evaluate that actually are based on comparisons on

the multiplicative (log or logit additive) scales.

Exactly multiplicative risk ratios– When the combined effect of two exposures X1 and

X2 is exactly what you would expect if the risk ratio comparing risk due to X1 to the

baseline risk, P(D|X1=1,X2=0)/P(D|X1=0, X2=0) and the risk ratio comparing the risk due

to X2 to the baseline risk P(D|X1=0,X2=1)/P(D|X1=0, X2=0) were multiplied together. In

terms of the risks associated with X1 and X2 (as opposed to the risk ratios) this would

mean that the probability of disease P(D) for those exposed to both X1 and X2 is defined

by the product of the individual risks and the inverse of the baseline risk:

P(D|X1=1,X2=1)= P(D|X1=1, X2=0) * P(D|X1=0, X2=1)*(1/P(D|X1=0,X2=0)). For

multiplicative risk models as defined by equation 2 in the main text, exactly

multiplicative risk ratios would lead to β_3=0. Rothman has shown (Modern

Epidemiology[1] pages 82-83, reiterated on pages 299-300) that the condition of exactly

multiplicative risk ratios indicates a departure from additive risk differences and thus

indicates the presence of sufficient component cause interaction.

Exactly additive risk differences – This concept is defined in Modern epidemiology[1] on

page 72: "When both X1 and X2 have effects and the risk difference of one remains

constant across levels of the other, e.g. (P(D|X1=1,X2=1 – P(D|X1=0, X2=1)= R11–R01 =

P(D|X1=1,X2=0) – P(D|X1=0,X2=0)= R10-R00, so there is no modification of risk

differences … the combined effect for X1 and X2 on risk can be computed simply by

adding together the separate risk differences for X1 and X2." Exactly additive risk

differences occur when the combined effect of two exposures X1 and X2 is exactly what

you would expect if the risk differences associated with X1 and X2 were added together.

In terms of risk (the probability of disease P(D)) for those exposed to both X1 and X2 it is

defined by: P(D|X1=1,X2=1)= P(D|X1=1, X2=0) + P(D|X1=0, X2=1) – P(D|X1=0,X2=0). In

this case Rothman has shown (Modern Epidemiology[1] pages 77-78) that while a

departure from additive risk differences indicates the presence of interaction response

types in the potential outcomes model and SCC interaction as defined, because in the

potential outcomes framework interaction response types may cancel each other out,

the condition of exactly additive risk differences does not rule out the possibility of

sufficient component cause interaction occurring in the population under study. In fact

Vanderweele and Robins[2] provide an example in their Appendix 3 to explicitly show why

"There can be synergism without the risk difference condition $P(D=1|E1=1, E2=1, C=c) -$

$P(D=1|E1= 0, E2=1, C=c) > P(D=1|E1=1,E2=0, C=c) -P(D= 1|E1=0, E2=0, C= c)$ holding."

Monotonic effect – Again, this term is defined explicitly in Vandweerle and Robins[2] as:

"We will say that E1 has a positive monotonic effect on D if for all individuals ω we have

$D_{ij}(\omega) \geq D_{i'j}(\omega)$ whenever $i \geq i'$; we will say that E2 has a positive monotonic effect on D if

for all individuals ω we have $D_{ij}(\omega) \geq D_{ij'}(\omega)$ whenever $j \geq j'$. Similarly, we will say that E1

has a negative monotonic effect on D if for all individuals ω we have $D_{ij}(\omega) \leq D_{i'j}(\omega)$

whenever $i \leq i'$ and that E2 has a negative monotonic effect on D if for all individuals ω

we have $D_{ij}(\omega) \leq D_{ij'}(\omega)$ whenever $j \leq j'$. The definition of a monotonic effect essentially

requires that some intervention either increase or decrease some other variable D—not merely on average over the entire population, but rather for every individual in that population, regardless of the other intervention. The requirements for the attribution of a monotonic effect are thus considerable." When an exposure (e.g. X1) is dichotomous this can be translated to the condition that exposure only either causes or prevents disease. In the potential outcomes framework (<u>as described in more detail later in this appendix</u>) requiring two dichotomous exposures to both exhibit monotonic effects eliminates response types 3,5,7,9,10,11,12,13,14,15 such that only interaction response types 2 and 8 remain as potential outcomes that exhibit synergistic (Vanderweele's usage) effects; in the sufficient component cause model this eliminates sufficient cause types D, E G H and I, so that only type F remains among the possible sufficient causes that exhibit sufficient component cause interaction. This concept and the implications of it are defined and utilized to formulate a test for synergism by Vanderweele[2,3] and discussed as a condition in which "we assume that neither factor is ever preventive" but not specifically defined as monotonic effects in Modern Epidemiology[1] on pages 79 and 82.

Description of model parameters used in each output in the main text and this appendix

In Supplemental Appendix 2 we have provided SAS code that was used to make Figure 3 in the main text as well as Supplemental Figures 4 and 5; here we provide a description of the simulations conducted in that code and which factors were varied to make the various Tables and Figures included in our work. In all of our simulations we model two dichotomous exposures X1 and X2, and their relationship with disease outcome D. Supplemental Figure 1 illustrates that the possible combinations of these two dichotomous exposures result in four possible levels of risk for D. Variables included in the simulation and the ways in which these variables were manipulated to create all output presented are listed in Supplemental Table 1. Table 1 in the main text provides a listing of each of the tests of interaction evaluated and shows in which tables and figures of each of these evaluations are presented, as well as the SAS code in appendix 2 that is used for each test.
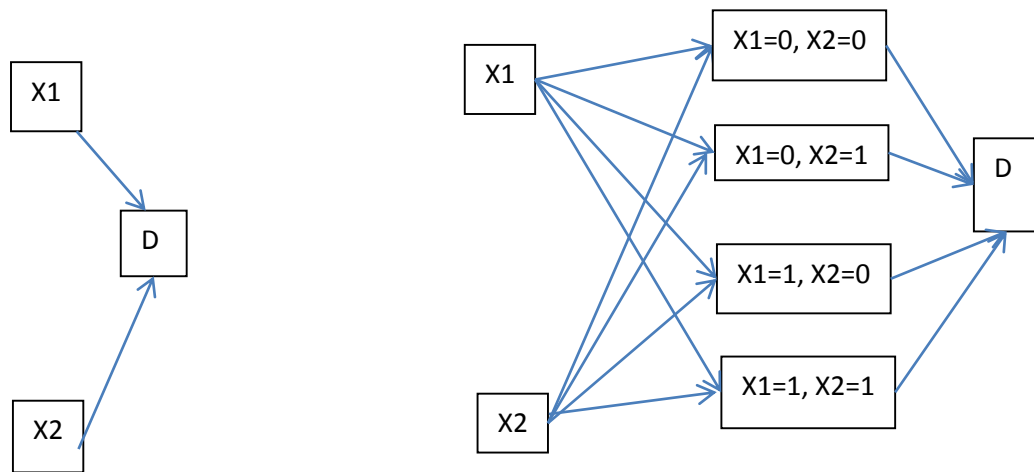
Supplemental Figure 1: Two equivalent causal representations of the relationship between exposures X1 and X2 and disease outcome D. The right side has been adapted to show the sufficient component causes containing different combinations of X1 and X2, similar to the example presented by Vanderweele and Robins.[4]

Supplemental Table 1: Parameter values varied in simulations and corresponding Table and Figure outputs

Sample size in doubly exposed group (X1=1, X2=1)	Scenario		Measure	Output
	Effect Size (Risk ratio) associated with each exposure	Interaction Effect size		
500,000	Range between 1 and 4	Fixed at halfway between additive on the risk difference scale and multiplicative on the risk ratio scale	Ability to detect interaction when present (Power)	Figure 1
500000	Range between 1 and 4	Fixed at exactly additive risk difference	Probability of detecting SCC interaction	Figure 2

Sample size in doubly exposed group (X1=1, X2=1)	Scenario Effect Size (Risk ratio) associated with each exposure	Interaction Effect size	Measure	Output
			when it is not present (Type 1 Error)	
5% of the population independently exposed to X1 and X2, with total population of: 50,100, 500, 750, 1000, 5000, 10000,20000,50000,75000, 100000, and 500000 and expected sample size for the population X1=1	RR(X1=1,X2=0) =2.75 RR(X1=0,X2=1) =2.25	Fixed at RR(X1=1,X2=1) =5.0925 (halfway between additive on the risk difference scale and multiplicative on the risk ratio scale)	Power	Supplemental Figure 2

Sample size in doubly exposed group (X1=1, X2=1)	Scenario Effect Size (Risk ratio) associated with each exposure	Interaction Effect size	Measure	Output
and X2=1 to be 0.25% of these totals				
As in a randomized controlled trial with sample size in each arm (including X1=1 and X2=1) of: 50 to 500 by 50, 1000, 5000, 10000 to 15000 by 1000, 20000, 50000, 75000, 100000 and 500000	RR(X1=1,X2=0) =2.75 RR(X1=0,X2=1) =2.25	Fixed at RR(X1=1,X2=1) =5.09375 (halfway between additive on the risk difference scale and multiplicative on the risk ratio scale)	Power	Supplemental Figure 3
200	RR(X1=1,X2=0) =2.75	Range from exactly additive on the risk difference scale	Power	Supplemental Figure 4

Sample size in doubly exposed group (X1=1, X2=1)	Scenario Effect Size (Risk ratio) associated with each exposure	Interaction Effect size	Measure	Output
	RR(X1=0,X2=1) =2.25	to exactly multiplicative on the risk ratio scale		
1000	RR(X1=1,X2=0) =2.75	Range from exactly additive on the risk difference scale	Power	Figure 3, Supplemental
	RR(X1=0,X2=1) =2.25	to exactly multiplicative on the risk ratio scale		Figure 5
500000	Range 1 to 4	Fixed at halfway between additive on the risk difference scale and	Power	Figure 4

Sample size in doubly exposed group (X1=1, X2=1)	Scenario Effect Size (Risk ratio) associated with each exposure	Interaction Effect size	Measure	Output
		multiplicative on the risk ratio scale		
500000	Range 1 to 4	Fixed at exactly additive on the risk difference scale	Type 1 error	Figure 4
750	RR(X1=1,X2=0)= RR(X1=0,X2=1)=2.5	Range from exactly additive on the risk difference scale to exactly multiplicative on the risk ratio scale	Power	Figure 5
750	RR(X1=1,X2=0)=	Range from exactly additive	Power	Supplemental

Sample size in doubly exposed group (X1=1, X2=1)	Scenario Effect Size (Risk ratio) associated with each exposure	Interaction Effect size	Measure	Output
	RR(X1=0,X2=1)=1.5	on the risk difference scale to exactly multiplicative on the risk ratio scale		Figure 6
750	RR(X1=1,X2=0)= RR(X1=0,X2=1)=3.5	Range from exactly additive on the risk difference scale to exactly multiplicative on the risk ratio scale	Power	Supplemental Figure 7

Description of the range of risks of disease D assigned for those exposed to both X1 and X2 in the simulated data

The methods used to assign exposure and disease status in the simulated population are

described in the main text. Here we elaborate further on the range of values for the ICR

that results from assignment of the combined risk for persons assigned both

dichotomous exposures over a range of risk ratios associated with each exposure X1 and

X2. Supplemental table 2 shows the ICR for values of RR(X1=1,X2=0) and RR(X1=0,X2=1)

under the exactly additive risk difference and exactly multiplicative risk ratio scenarios

as well as for κ = 2 for equation 13 in the main text.

Supplemental Table 2: Example of the magnitude of the Interaction Contrast Ratio (ICR)[a] for two dichotomous exposures wherein the combined risks exhibited exactly additive risk differences, exactly multiplicative risk ratios or the combined risk was somewhere in between these two values.

		RR(X1=1,X2=0)[b]		
		1.5	2.5	3.5
	RR(X1=0, X2=1)[b]			
Risk difference is exactly additive	1.5	0	0	0
RR(X1=1,X2=1)=	2.5		0	0
RR(X1=1,X2=0)+RR(X1=0,X2=1)-1	3.5			0
Risk ratio is exactly multiplicative	1.5	0.25	0.75	1.25
RR(X1=1,X2=1)=	2.5		2.25	3.75
RR(X1=1,X2=0)*RR(X1=0 X2=1)	3.5			6.25
Risk is (exactly halfway) between additive risk differences and multiplicative risk ratios RR(X1=1,X2=1)= (RR(X1=1,X2=0)+RR(X1=0,X2=1)-1) +	1.5	0.125	0.375	0.625
(((RR(X1=1,X2=0)*RR(X1=0,X2=1))-	2.5		1.125	1.875
(RR(X1=1,X2=0)+RR(X1=0,X2=1)-1))/2[c])	3.5			3.125

a) The ICR is calculated as RR(X1=1,X2=1)-RR(X1=1,X2=0)-RR(X1=0,X2=1)+1

b) In simulations the values of each risk ratio were varied between 1 and 4 to

    provide a range of effect sizes. The probability of disease in the unexposed (i.e.

    P(D|X1=0, X2=0)) was fixed at 0.05 and multiplied by these risk ratio values to

    obtain the probability of disease when exposed to one or both of X1 and X2.

c) In simulations the value of the divisor κ was varied between 1.25 and 5 to

    provide a range of true interaction effects (the larger the number the closer the

    effect is to additive risk differences) between additive risk differences and

    multiplicative risk ratios.

In this section we provide additional results that supplement the findings reported in

the main paper.

Power to detect greater than additive risk differences and the need to use study designs where the number of participants exposed to both exposures can be controlled experimentally

We used two different methods to simulate a population exposed to both X1 and X2. In

both cases a binomial distribution was used to generate the data. First, we simulated an

observational epidemiologic study in which the population is not selected based on

exposure risk, or assigned exposure experimentally. This method was used only to

develop Supplemental Figure 2. We assign a probability of exposure to each covariate

(P(X1=1) and P(X2=1)) independently using a draw from a binomial distribution with

probability equal to 0.05, 0.10, and 0.30. This results in approximately 0.0025, 0.01, and

0.09 of the population exposed to both X1 and X2, respectively. Within this scenario we

consider total population sizes between 50 and 500000 people, resulting in an expected

mean of up to 500000*0.0025 = 1250 persons exposed to both X1 and X2 when each

exposure occurs independently in 0.05 of the population. An alternative method was

used for all other tables and figures including Supplemental Figure 3 and is described in

the main text, but the description is repeated here for completeness. For this version of

the simulations, which we call the randomized controlled trial (RCT) scenario, we

assigned X1 and X2 as if exposure was assigned experimentally as in a RCT of both X1

and X2. The population is the same size in each of the 4 exposure categories or "arms"

of the trial depicted in Supplemental Figure 1. We considered sample sizes (for each

arm) of 200, 500, 750, 1000 and 500000, with the last scenario, with a total population

size of 2 million, designed to emulate the performance of the various tests for interaction in expectation.

Supplemental Figures 2 and 3 show the power of each of 6 measures of interaction to detect true interaction effects when both exposures have strong individual effects (RR(X1=1,X2=0)=2.75, RR(X1=0,X2=1)=2.25) and the interaction effect is exactly halfway between what would be observed in the case of additive risk differences and multiplicative risk ratios (κ=2 in equation 13). In this case RR(X1=1, X2=1)=5.0925 which leads to an ICR of 1.0925, indicative of SCC interaction. Figure 2 shows that, when each exposure is rare, it would take hundreds of thousands of subjects to acquire the power necessary to detect SCC interaction. Supplemental Figure 2 shows that, under a study design where the number of persons with exposure to both X1 and X2 is controlled, over 1000 participants are required in each of four strata of exposure to detect interaction of this magnitude. Both Figures show the same trend for the individual measures, with the log-binomial models showing better power than Poisson models and linear-binomial or additive risk models showing similar power to log-binomial models. These figures also show that methods based on likelihood-ratio tests have slightly better power than those based on Wald methods.

Supplemental Figure 2: Power and sample size for 6 tests of SCC interaction when 0.25% of the total population is exposed to both of 2 exposures

Power by sample size for 2 exposures each with an independant prevalence of 5% in the Study Population
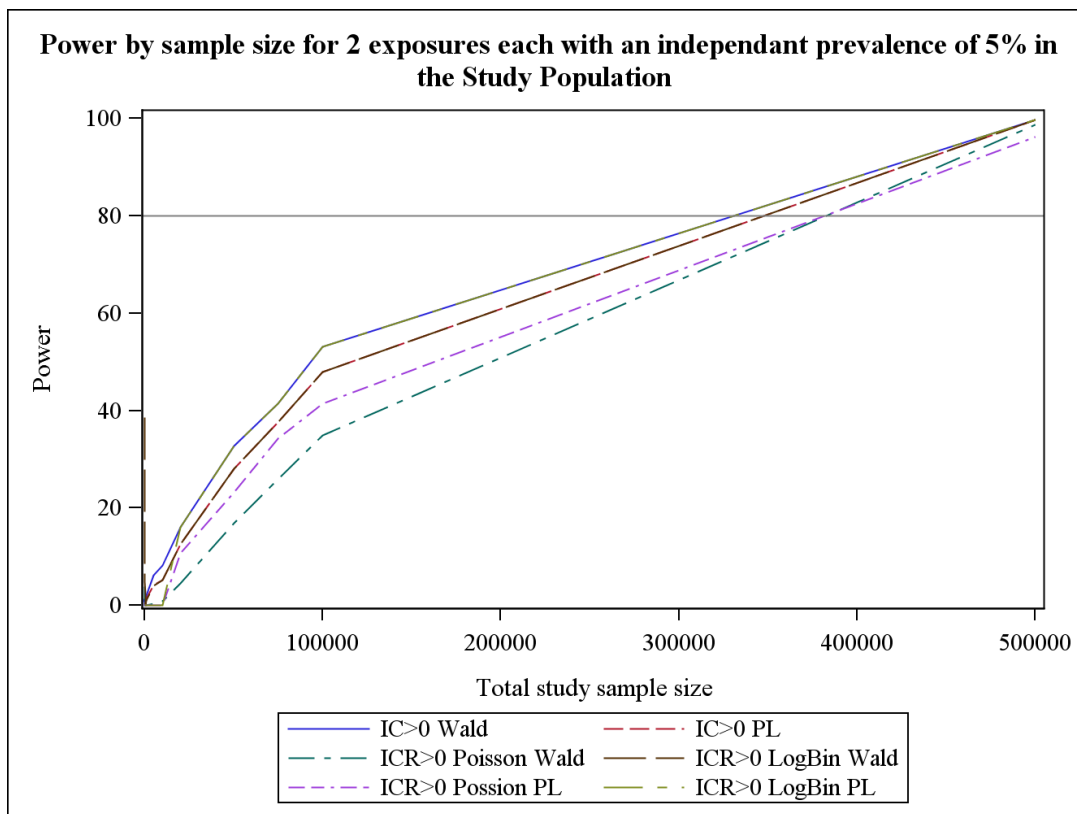
Supplemental Figure 3: Power and sample size for each of four levels of combinations of two dichotomous exposures for 6 tests of SCC interaction

Power by sample size for 2 exposure RCT scenario with N for each of four levels of exposure
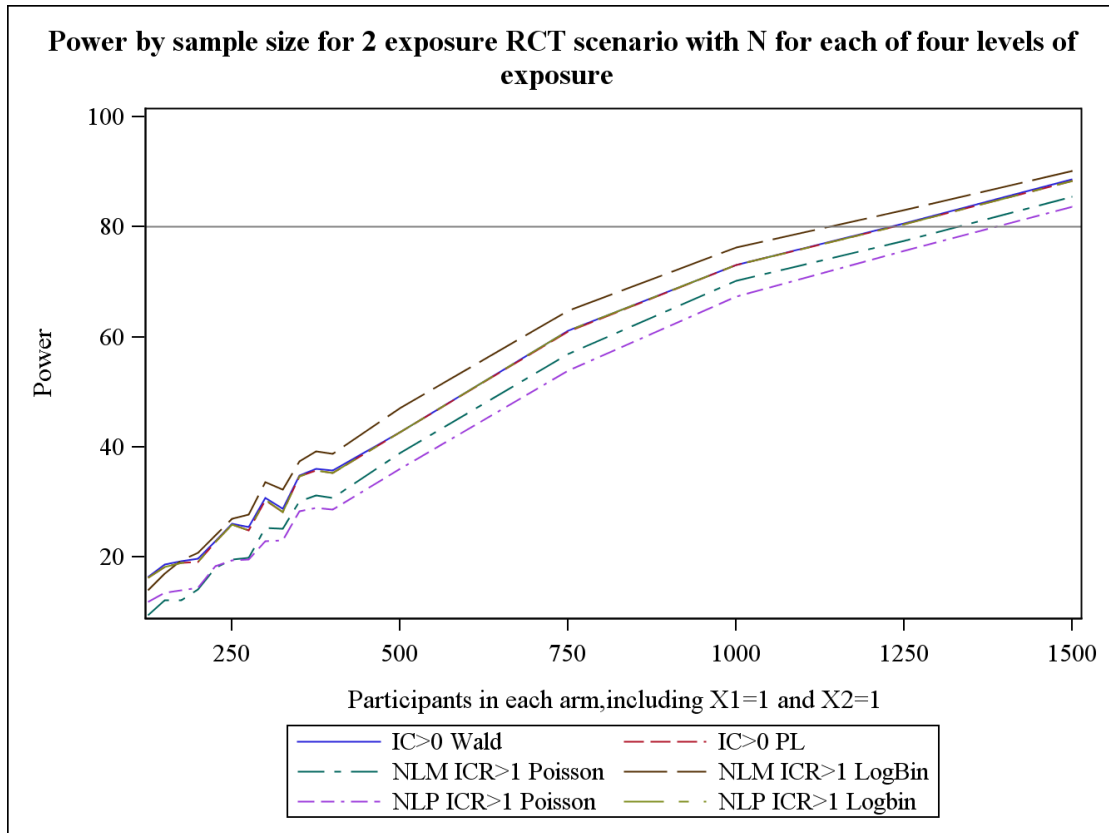
Power to detect SCC interaction is limited to both large sample sizes and interaction effects that are closer to multiplicative on the risk ratio scale than additive on the risk difference scale

Supplemental Figures 4-7 provide additional information on the power to detect SCC

interaction over a range of values between exactly additive on the risk difference scale

and exactly multiplicative on the risk ratio scale, at fixed sample sizes and risk ratios for

each exposure. In figure 4 the sample size in each of four (See supplemental Figure 1)

levels of the two dichotomous exposures is fixed at 200, such that the total sample size

for the simulated study is 800. Supplemental Figure 5 increases the sample size to 1000

in each level or 4000 overall (this figure is similar to Figure 3 in the main text, but

includes results for tests of departure from multiplicative models defined as in equation

2, where the test is whether the interaction term β_3 is different from zero). In both

figures the risk ratios are fixed at RR(X1=1,X2=0)=2.75 and RR(X1=0,X2=1)=2.25,

representing relatively strong individual effects. The underlying prevalence of disease in

the unexposed (X1=0 and X2=0) population is 5%, such that disease occurs in more than

10% of the population exposed to either X1 or X2 and in >20% of the population

exposed to both X1 and X2. Supplemental Figure 4 suggests that no test of interaction

will be able to detect an interaction effect that is greater than what would be expected

if risk differences were additive but less than that expected when risk ratios are

multiplicative reliably when only 200 participants are included in each strata of

combinations of X1 and X2. A maximum of just over 60% of all simulations produced a

significant test of interaction for tests of either IC>0 or ICR>0. Note also that, even at

this moderate sample size, the Wald and likelihood ratio methods for estimating

confidence intervals for both the IC and ICR parameters produce nearly identical results, suggesting that Wald limits may be adequate for testing for interaction when it is present. Further, traditional tests of departure from multiplicative risk ratios (i.e. $\beta_3 \neq 0$ in models defined as equation 2 or 4) have high type 1 error rates when risk differences are exactly additive (ICR=0 in the legend, the leftmost bar in each cluster) and perform poorly relative to tests of effect modification on the risk difference scale at all levels of the ICR. In Supplemental figure 5, when sample size is increased to 4000, all tests of effect modification on the risk difference scale perform adequately when the amount of interaction is closer to multiplicative than additive (an ICR of 1.094 represents the midpoint between exactly additive risk differences and exactly multiplicative risk ratios, $\kappa=2$ in equation 13, when the risk ratios of RR(X1=1,X2=0)=2.75 and RR(X1=0,X2=1)=2.25). None of the tests of effect modification on the risk difference scale perform well when the interaction effect is less than halfway between exactly additive risk differences and exactly multiplicative risk ratios. Again the tests of interaction parameters in multiplicative risk models perform even worse in terms of their ability to detect interaction effects greater than additive risk differences but less than multiplicative risk ratios at this sample size, with large type 1 error rates due to their ability to detect departures from the multiplicative model in the absence of any true SCC interaction, and a general inability to detect interaction as the combined effects approach the expected value under the condition of multiplicative risk ratios.

Supplemental Figure 4: Comparison of the Type 1 error rates and Power to detect SCC interaction effects that are greater than additive on the risk difference scale and less than multiplicative on the risk ratio scale, for 8 tests of interaction over a range of true combined effects of two dichotomous variables
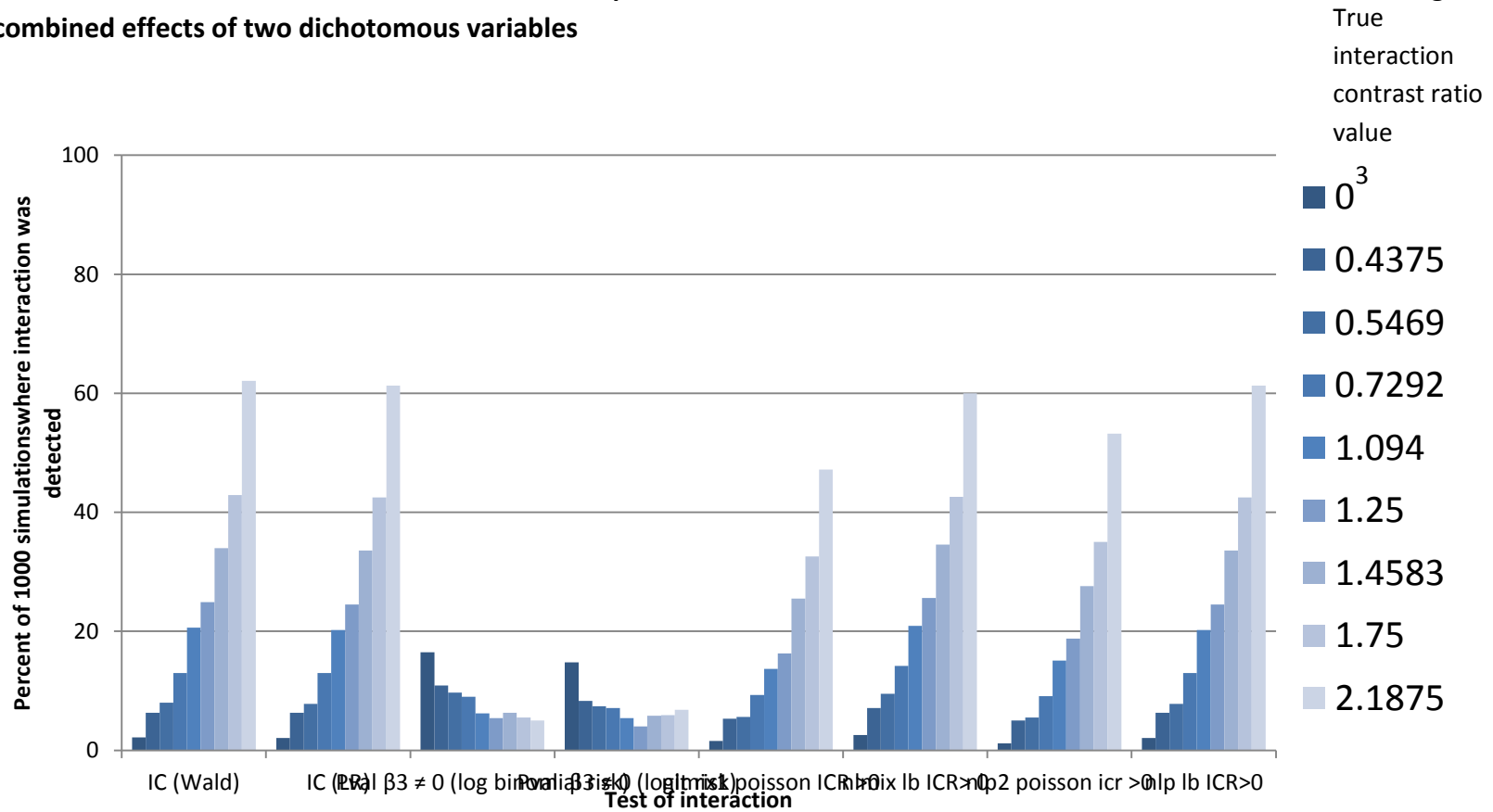
1) NLMIXED calculates confidence limits for linear combinations of parameters through a delta method approximation, in this case the limits approximate Wald limits

2) NLP produces profile likelihood limits for parameters, including the ICR reported here

3) An ICR of zero represents a situation with exactly additive risk differences and any detected interaction represents type 1 error

Supplemental Figure 5: Comparison of the Type 1 error rates and Power to detect SCC interaction for 8 tests over a range of true interaction effects, measured as the proportion of 1000 simulations in which each effect was detected, when the risk ratios for exposures X1 and X2 equal 2.75 and 2.25, respectively, and 1000 participants are included in each of four strata of exposure to X1 and X2
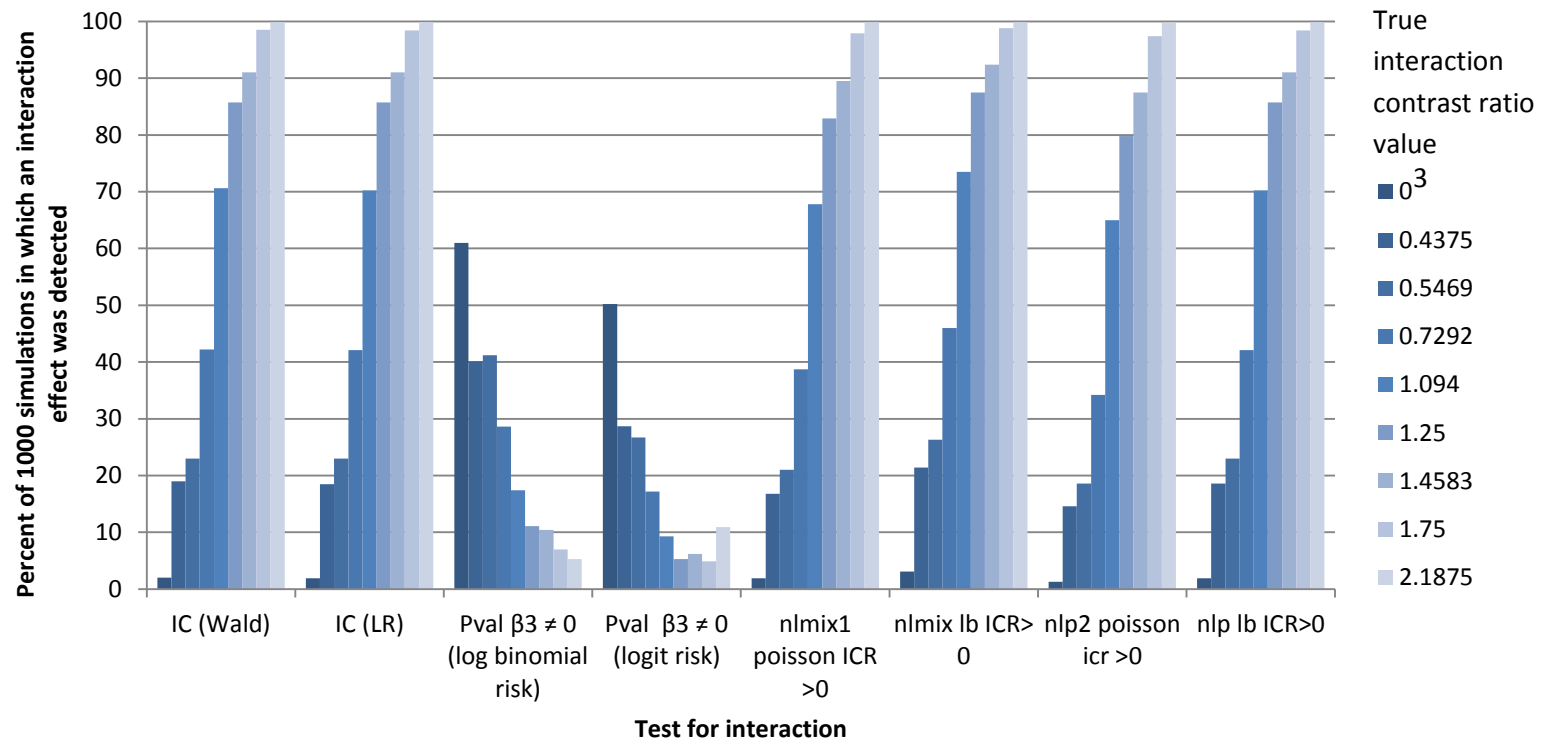
1) NLMIXED calculates confidence limits for linear combinations of parameters through a delta method approximation, in this case the limits approximate Wald limits

2) NLP produces profile likelihood limits for parameters, including the ICR reported here

3) An ICR of zero represents a situation with exactly additive interaction and any detected interaction represents type 1 error

More on the differences between detecting interaction, SCC interaction and definitive interdependence

As defined by Vanderweele and Robins[2] exposure effects are <u>monotonic</u> when they only

cause and never prevent disease, or vice versa, for all individuals in the population. This

is an important assumption that is often not known to be valid in studies being

undertaken to investigate potential causes of disease.

Supplemental Table 5 is adapted from Table 5-2 in Modern Epidemiology,[1] and presents

the individual interaction contrast values for each of 16 possible counterfactual

conditions arising from combinations of potential dichotomous disease outcome D of

two dichotomous exposures X1 and X2. In this framework causal types labeled 3,5,7,8

and 15 show positive interdependence, in that the observed effects of X1 and X2 are

greater when both are present than for one exposure alone and as a result the IC is

greater than zero. Conversely types 2, 9, 10, 12 and 14 show negative interdependence,

with their effects reduced when both exposures are present compared to situations

where only one is present; for these types the IC is less than zero. At the population

level, where p_i represents the proportion of the study population with causal type i, the

IC has been shown[2,3] to be:

$$IC = R_{11} - R_{10} - R_{01} + R_{00}$$

$$= p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + p_8 - (p_1 + p_2 + p_5 + p_6 + p_9 + p_{10} + p_{13} + p_{14}) -$$

$$(p_1 + p_2 + p_3 + p_4 + p_9 + p_{10} + p_{11} + p_{12}) + (p_1 + p_3 + p_5 + p_7 + p_9 + p_{11} + p_{13} + p_{15})$$

$$= (p_3 + p_5 + 2p_7 + p_8 + p_{15}) - (p_2 + p_9 + 2p_{10} + p_{12} + p_{14})$$

which is the difference between the proportion of the population with positive

interdependence and the proportion with negative interdependence types. This is why

departures from zero for the IC indicate that some interaction is present in the

population. However, the IC can be exactly zero in a situation when the proportion with

positive interdependence exactly cancels out with the proportion of the population with

negative interdependence. Note also that causal types 7 and 10 have an IC with an

absolute value of 2 because in both cases the effect of one variable reverses across

strata of the other variable, e.g. for type 7, persons exposed to X2 get disease only when

also exposed to X1, while persons not exposed to X2 only get disease when they are also

not exposed to X1. Inclusion of people with types 7 or 10 in the study cohort is one

reason the IC does not provide a direct estimate of the proportion of the overall effect

due to interaction. The other reason is the potential for types 3, 5, 9, 11, 12, 13, 14, 15,

where exposure can prevent disease. These are the types that lead to a lack of

monotonic effects of X1 and X2 on disease as described by Vanderweele,[2,3] and make it

more difficult to untangle true <u>sufficient component cause interaction</u> from <u>effect</u>

<u>modification</u> on the risk difference scale detected on the basis of an IC or ICR greater

than zero. However, Vanderweele[2,3] showed that if we can exclude types in which one

or both exposures prevent disease (3,5,7,9,10,11,12,13,14,15) then the IC becomes a

comparison of the proportions of the population with type 8 and type 2, and if the IC is

greater than zero then we can conclude that there must be some SCC interaction

because potential outcome type 8 only arises from a sufficient component cause in

which both exposures are component causes. This is the condition that he defines as definitive interdependence. Additionally he showed that even if we can't make assumptions about monotonic effects of X1 and X2, a test of $R_{11} - R_{10} - R_{01} > 0$, parameterized as either equation 8 or equation 9 in the main text is sufficient to conclude definitive interdependence has been detected when the test concludes that it is present. This is because such a test simplifies the calculation of the inequality to be

$$(p7 + p8) - (p1 + p2 + 2p9 + 2p10 + p11 + p12 + p13 + p14) > 0$$

and, if this is true then some people with either types 7 or 8 must be present, and both types achieve disease only through SCC interaction. The implications of this restriction to the tests for interaction based on effect modification on the risk difference scale are presented in Figure 5 in the main text and Supplemental Figures 6 and 7. Supplemental Figure 6 shows the power to detect an interaction when the sample size is fixed at 750 in each of four exposure categories as defined in Supplemental Figure 1, for a total sample size of 3000, with the risk ratios for X1 and X2 set to 1.5, and a range of interaction effects created with equation 13 corresponding to ICRs of 0.05, 0.083, 0.143 and 0.20. Supplemental Figure 6 clearly shows that none of the tests presented is powered to detect interaction when the effects are this small. Supplemental Figure 7 increases the individual risk ratios for X1 and X2 to 3.5, and, again using equation 13 to create a range of effects that are greater than additive on the risk difference scale but less than multiplicative on the risk ratio scale, presents ICRs of 1.25, 2.08, 3.51, and 5.0. Here we see that, as in Figure 5 in the main text, although all of the tests have high

power to detect an ICR of 3.51 or 5.0, there is a large drop in the proportion of times

when the effect is detected when we add the restrictions suggested by Vanderweele.[2,3]

Clusters 2, 4, and 5 show power for tests of definitive interdependence, while clusters 1

and 3 show traditional tests for departures from risk difference additivity, which

Vanderweele shows only test for definitive interdependence when X1 and X2 are

monotonic.

Supplemental Table 3: Listing of 16 potential outcome types for two dichotomous exposures X1 and X2 and a dichotomous outcome, with individual values of the Interaction Contrast (IC) for each type

Causal Type	Individual Risk IF:				IC
	X1=1, X2=1	X1=0, X2=1	X1=1, X2=0	X1=0, X2=0	
1	1	1	1	1	0
2	1	1	1	0	-1
3	1	1	0	1	1
4	1	1	0	0	0
5	1	0	1	1	1
6	1	0	1	0	0
7	1	0	0	1	2
8	1	0	0	0	1
9	0	1	1	1	-1
10	0	1	1	0	-2
11	0	1	0	1	0
12	0	1	0	0	-1

	Individual Risk IF:				
Causal Type	X1=1, X2=1	X1=0, X2=1	X1=1, X2=0	X1=0, X2=0	IC
13	0	0	1	1	0
14	0	0	1	0	-1
15	0	0	0	1	1
16	0	0	0	0	0

Supplemental Figure 6: Comparison of the power to detect SCC interaction when it exists when the monotonicity[1] of effects X1 and X2 can and cannot be assumed: Results of 1000 simulation when X1 and X2 both have a risk ratio of 1.5 relative to the baseline risk of disease of 5%
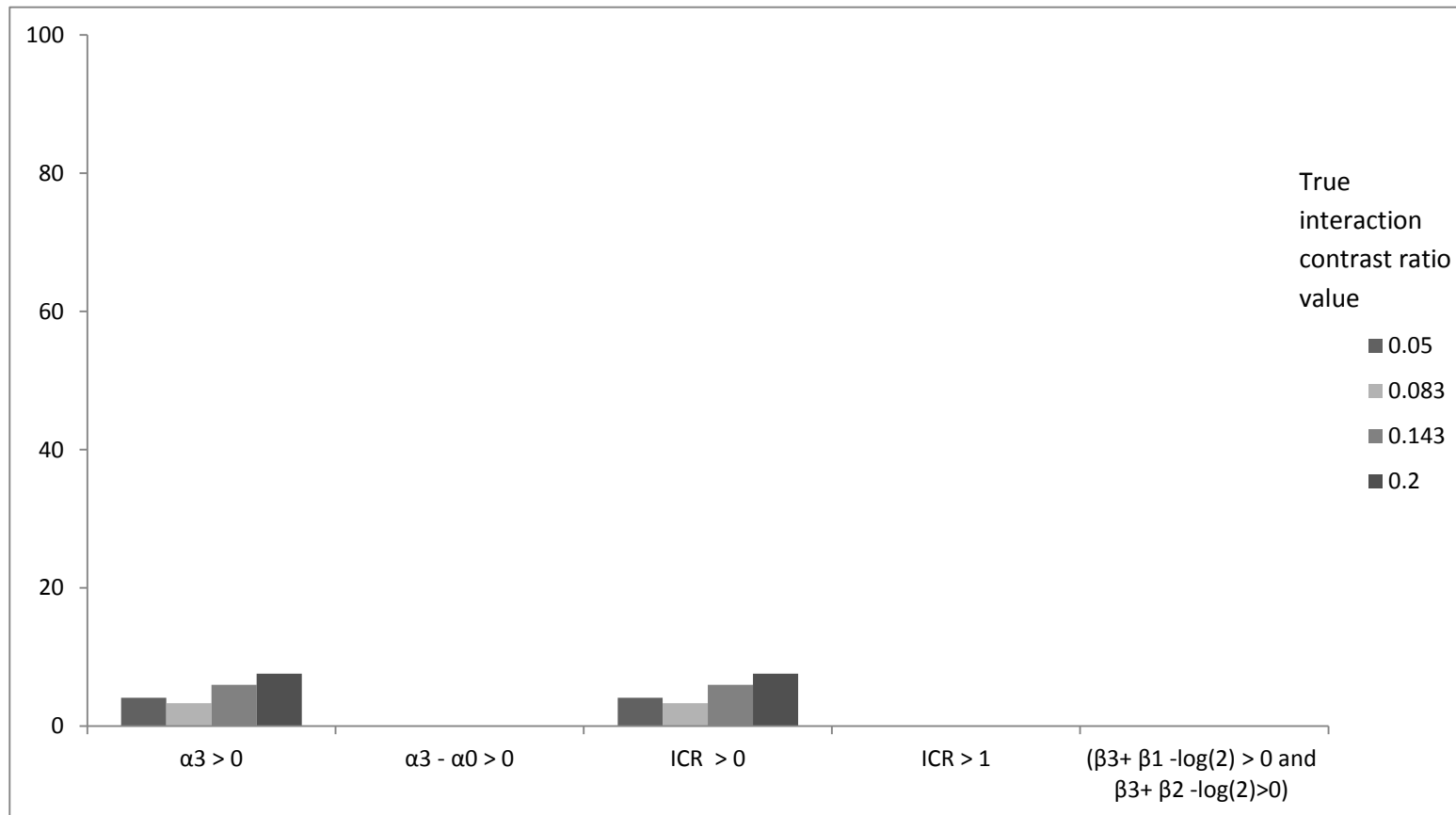
Supplemental Figure 7: Comparison of the power to detect SCC interaction when it exists when the monotonicity[1] of effects X1 and X2 can and cannot be assumed: Results of 1000 simulation when X1 and X2 both have a risk ratio of 3.5 relative to the baseline risk of disease of 5%

True interaction contrast ratio value: ▪ 1.25 ▪ 2.08 ▪ 3.57 ▪ 5

X-axis categories: $\alpha 3 > 0$, $\alpha 3 - \alpha 0 > 0$, ICR > 0, ICR > 1, ($\beta 3 + \beta 1 - \log(2) > 0$ and $\beta 3 + \beta 2 - \log(2) > 0$)
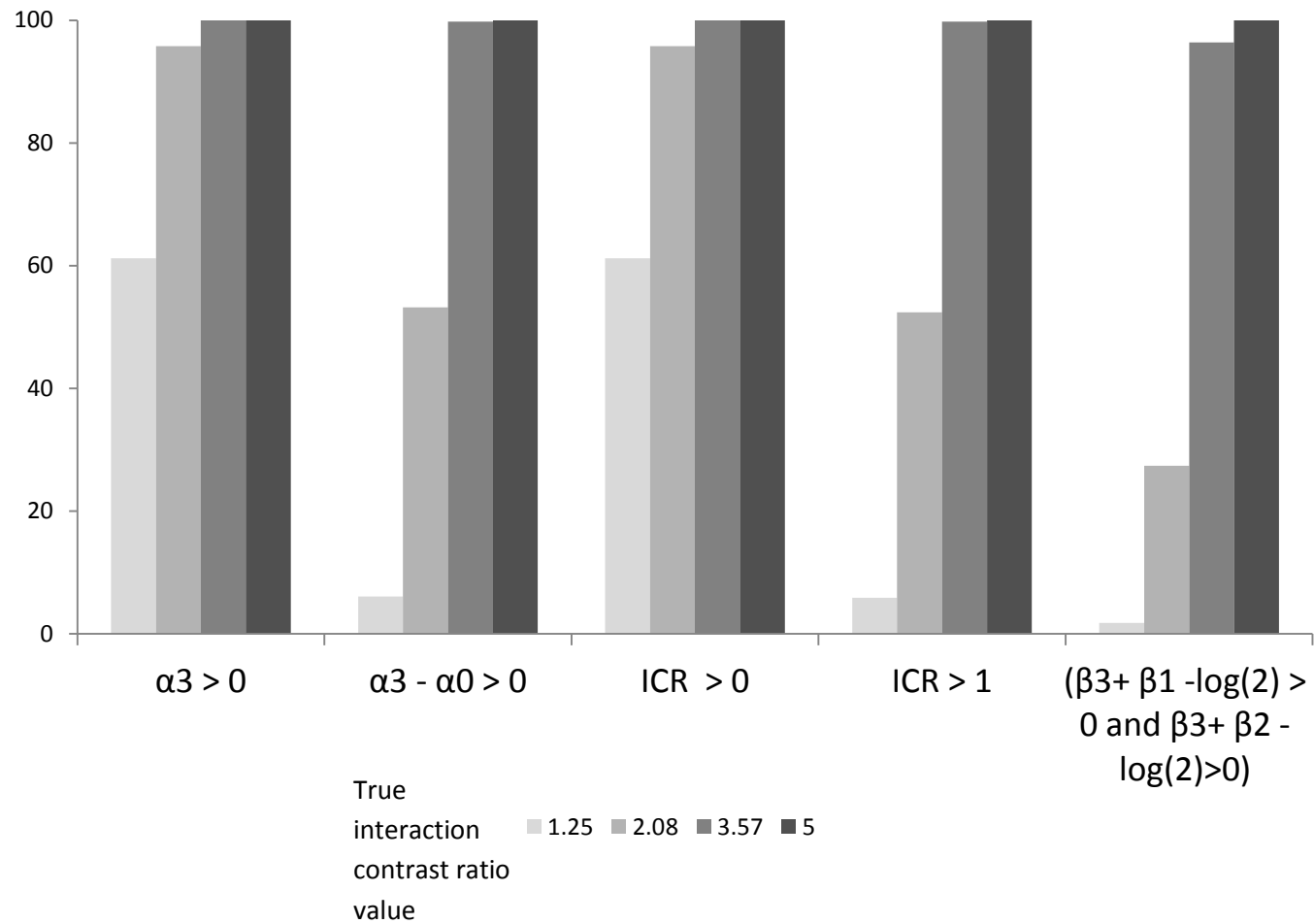
References

1) Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott-Williams & Wilkins; 2012.

2) VanderWeele TJ, Robins JM. The identification of synergism in the sufficient-component-cause framework. Epidemiology. 2007 May;18(3):329-39.

3) VanderWeele TJ. Sufficient cause interactions and statistical interactions. Epidemiology. 2009 Jan;20(1):6-13.

4) VanderWeele TJ, Robins JM. Directed acyclic graphs, sufficient causes and the properties of conditioning on a common effect. Am J Epidemiol. 2007; 166:1096-1104.

Supplemental Appendix 2: Example SAS code for all tests of interaction evaluated

```
ods html close;
ods listing;

proc format; value power 0-0.05="Detected" 0.05<-high="Not Detected";
                Value ICR_power low-0="Not Detected" 0<-
high="Detected";
                value icr_powerb low-1="Not Detected" 1<-
high="Detected";
                run;

ods listing close;
ods trace off;

data from_RRs3;
array RR (9) RR1-RR9;
do n=200,500,750,1000;*I am surprised (even though I know the outcome
is very rare, how wide the CI's still are;
call streaminit(1234);
do rep=1 to 1000;
Do RR10=1.5, 2.5,3.5;
do RR01=1.5, 2.5,3.5;
do type=1,2,3,5,7,9;
if type=1 then do;
rr1=.;
rr2=.;
rr3=.;
rr4=.;
rr5=.;
rr6=.;
rr7=.;
rr8=.;
rr9=.;
end;
if type=1 then RR(type)=RR10+RR01-1;*Exactly additive, used to assess
type 1 error;
if type=2 then RR(type)=RR10*RR01;*Exactly multiplicative;
*A Range of values between additive and multiplicative, used to assess
power
over a range of effects;
if type=3 then RR(type)=RR(1)+((RR(2)-RR(1))/5);
if type=4 then RR(type)=RR(1)+((RR(2)-RR(1))/4);
if type=5 then RR(type)=RR(1)+((RR(2)-RR(1))/3);
if type=6 then RR(type)=RR(1)+((RR(2)-RR(1))/2);
if type=7 then RR(type)=RR(1)+((RR(2)-RR(1))/1.75);
if type=8 then RR(type)=RR(1)+((RR(2)-RR(1))/1.5);
if type=9 then RR(type)=RR(1)+((RR(2)-RR(1))/1.25);
*if type=10 then RR(type)=R10*RR01*1.2;
do p0=.05 ;*This and RR10 act to determine the number of events;
true_ICR=RR(type)-RR10-RR01+1;
true_ic=true_icr*p0;
R10=RR10*p0;
R01=RR01*p0;
R11=RR(type)*p0;
```

```
*Assume that there is a fixed population (variable n above) in each
risk group;
*This is a way to represent the counterfactual risks and the "ideal"
RCT;

count=rand("Binomial",p0,n);
e1=0; e2=0;
output;
count=rand("Binomial",R10,n);
e1=1; e2=0;
output;
count=rand("Binomial",R01,n);
e1=0; e2=1;
output;
count=rand("Binomial",R11,n);
e1=1; e2=1;
output;
end;
end;
end;
end;
end;
end;
run;
options notes;
proc sort data=from_RRs3 out=for_figures;
by rep type R11 RR10 RR01 ;
*suggest you only use one sample size to limit output;
Where n=1000 and type in(3,9) and RR10=1.5;
*Can use the commented statement above (remove first semi-colon and
asterisk (;*)) to limit the range of interaction effects included in
analysis;
run;


*This code is necessary because of the large number of simulation runs;
*In practice, with one dataset with which to implement these tests of
interaction this code
is unnecessary and the notes/output may be informative;
ods trace off;
ods html close;
ods listing close;
options nonotes;
ods results off;



*Example 1: Linear risk model to produce estimates of the interaction
contrast (IC);
ods output estimates=ICsRR;
proc genmod data=for_figures descending;
by rep type R11 RR10 RR01;
*where rr11=&RR11_2;
class e1(param=ref ref="0") e2(param=ref ref="0");
model count/n=e1 e2 e1*e2/link=identity dist=binomial lrci;
*Comments below calculate risks at each of four levels of exposure;
*estimate "R10" intercept 1 e1 1 ;
```

```
*estimate "R01" intercept 1 e2 1 ;
*estimate "R11" intercept 1 e1 1 e2 1 e1*e2 1 ;
*estimate "R00" intercept 1;
estimate "IC"  e1*e2 1;
*Commented code produces estimates of risk differences;
*estimate "RD RR10-RR00" e1 1;
*estimate "RD RR01-RR00" e2 1;
*estimate "RD RR11-RR00" e1 1 e2 1 e1*e2 1;
run;




ods results off;
ods listing close;
*Example 2: Log-binomial risk model, Interaction contrast ratio cannot
be calculated directly in GENMOD,
but a test of multiplicative interaction provided by default;
ods output
parameterestimates=B3_LogRiskRR2(where=(trim(left(upcase(parameter)))="
E1*E2"));
proc genmod data=for_figures descending;
by rep type R11 RR10 RR01;
*where rr11=&rr11_2;
class e1(param=ref ref="0") e2(param=ref ref="0");
model count/n=e1 e2 e1*e2/link=log dist=binomial lrci;
run;




ods results off;
ods listing close;
options nonotes;
ods output
parameterestimates=B3_ORRR2(where=(trim(left(upcase(parameter)))="E1*E2
"));
*Example 3: Logit risk model, will produce overestimates of risk when
outcome is common;
proc genmod data=for_figures descending;
by rep type R11 RR10 RR01;
class e1(param=ref ref="0") e2(param=ref ref="0");
model count/n=e1 e2 e1*e2/link=logit dist=binomial;
run;




*Examples above show B3 is significant in multiplicative models with
exactly additive risk... but what happens when we calculate the ICR?;

ods listing close;
options nonotes;
ods results off;
ods trace off;

ods output additionalestimates=NLM_ICR_RR1;
*Example 4: Logit risk model in NLMIXED to estimate ICR directly, with
Wald-like confidence limits;
*Title2 "PROC NLMIXED Wald CI for ICR";
```

```
*This is a logit risk model, so estimates of ICR do not approximate
true estimates well when outcome is common;
PROC NLMIXED DATA=for_figures DF=10000 ;
by rep type R11 RR10 RR01;
PARMS beta0=0 b_e1=.4055 b_e2=.4055
          b_interaction=0;

     eta=(beta0+b_e1*e1+b_e2*e2+b_interaction*e1*e2);
       p= exp(eta)/(1+exp(eta));
     MODEL count~Binomial(n,p);
*Use NLMIXED to calculate the ICR directly, with degrees of freedom
(df=) very large to approximate Wald confidence intervals for the ICR
estimate;

     ESTIMATE "ICR" exp(b_e1+b_e2+b_Interaction)-
                    exp(b_e1)-exp(b_e2)+1;



RUN;




ods listing close;
ods results off;
options nonotes;
ods output additionalestimates=NLM_ICR_RR2;

PROC NLMIXED DATA=for_figures df=10000;
by rep type R11 RR10 RR01;
*Example 5: Log-binomial risk model, calculates the ICR directly with
Wald-like Confidence intervals when df=10000 (see Kuss 2008);
PARMS beta0=-4 b_e1=.4055 b_e2=.4055
          b_interaction=0;*A large negative starting value for
intercept helps convergence in log-binomial models;

     eta=(beta0+b_e1*e1+b_e2*e2+b_interaction*e1*e2);
       mu=exp(eta);
      * p= exp(eta);
       ll = count*log(mu)- mu - lgamma(count+1);
     MODEL count~general(ll);
     ESTIMATE "ICR" exp(b_e1+b_e2+b_Interaction)-
                    exp(b_e1)-exp(b_e2)+1;
*Can also calculate the Estimate for ICR>1 proposed by Vanderweele
(2009);
      ESTIMATE "TVW2009" exp(b_e1+b_e2+b_Interaction)-
                    exp(b_e1)-exp(b_e2);
*And the newest estimate for the proportion of total effect due to
interaction (pX1=pX2=0.5 in our RCT design...);
     estimate "TVW2014" (exp(b_e1+b_e2+b_Interaction)-exp(b_e1)-
exp(b_e2)+1)*.5
                                        /
                              (exp(b_e1)-
1+(exp(b_e2+b_e1+b_Interaction)-exp(b_e2)-exp(b_e1)+1)*.5);

RUN;
```

```
ods listing close;
ods results off;
options nonotes;
ods output additionalestimates=NLM_ICR_RR3;
PROC NLMIXED DATA=for_figures df=10000;
by rep type R11 RR10 RR01;
PARMS beta0=-4 b_e1=.4055 b_e2=.4055
          b_interaction=-1;


*Example 6: Poission risk model, slightly worse power than the log-
binomial model, probably due to inability to estimate robust standard
errors;
     eta=(beta0+b_e1*e1+b_e2*e2+b_interaction*e1*e2);
       p= exp(eta);
       ll =  count*log(p) + (n-count)*log(1-p);
     MODEL count~general(ll);
     ESTIMATE "ICR" exp(b_e1+b_e2+b_Interaction)-
                  exp(b_e1)-exp(b_e2)+1;

RUN;


*PROC NLP can be manipulated to calculate the ICR directly, but
estimating other version proposed by
Vanderweele is not as simple;
*NLP provides both Wald and Profile likelihood limits, at the sample
sizes required to show an effect these
are nearly the same!;
Options nonotes;
ods listing close;
ods results off;

ods output WaldPLLimits=NLP_PL_Limits_ICR1;
*title2 "PROC NLP Profile and WALD CI for the ICR";
PROC NLP DATA=for_figures VARDEF=N cov=2 pstderr maxit=200 all;
by rep type R11 RR10 RR01;
      *Bounds -500 <= ICR <= 500;
      PARMS beta0=-3, b_e1=1.4, b_e2=1.4,
          ICR=1;
    b_interaction=log((ICR+exp(b_e1)+
                      exp(b_e2)-1)/
                      (exp(b_e1)*exp(b_e2)));
       Eta=beta0+b_e1*e1+b_e2*e2+b_interaction*e1*e2 ;
*Example 7: Logit risk model, again a poor estimate of the true ICR
because the outcome is common;
p=    exp(eta)/
       (1+ exp(eta));

     loglike=((count)*log(p)) + ((n-count)*log(1-p));

     MAX loglike;


     PROFILE ICR /alpha=.05 ;
RUN;


*Example 8: Richardson and Kaufman demonstrated a linear odds model in
which the B3 parameter of NLMIXED is manipulated to estimate the ICR
directly (on the logit scale);
```

```
*As with Examples 4 and 7 this overestimates the true ICR because
interaction can only be detected when the effects are strong and
therefore the outcome is common;

*their macro used data in the "counting" format of input, so let's
create a dataset in that form for our example;
data for_figures_freq;
set for_figures;
freq=count;
outcome=1;
output;
freq=n-count;
outcome=0;
output;
run;


%macro bounds (data= , outcome= , odds= , param= , replicate=);

ods output fitstatistics = fitstatistics
ParameterEstimates=ParameterEstimates;
ods listing close;
ods results off;
proc nlmixed data=&data ;
by rep n type R11 RR10 RR01 ;
odds = &odds;
where rep=&rep and n=&n and type=&type;
model &outcome ~ binary( odds/(1+odds) );
%if %scan(&replicate,1) ne %str() %then %do;
replicate &replicate;
%end;
run;


data fitstatistics;
set fitstatistics;
     if (Descr = "-2 Log Likelihood") then do;
     call symput ('LL', put(value,best16.));
     call symput("refLL",put(value,best16.));
     end;
run;

data ParameterEstimates;
set ParameterEstimates;
if (PARAMETER = "&param") then do;
call symput("istep",put(STANDARDERROR,best16.));
call symput("ibeta",put(ESTIMATE,best16.));
end;
if (PARAMETER = "b0") then call symput("ibeta0",put(ESTIMATE,best16.));
if (PARAMETER = "b1") then call symput("ibeta1",put(ESTIMATE,best16.));
if (PARAMETER = "b2") then call symput("ibeta2",put(ESTIMATE,best16.));
run;


data lci uci;
set _null_;
     null= 0;
```

```
      neglogl=0;
      difference=0;
      param=0;
      step=0;
%DO I = 1 %TO 2;
%* I=1 is Lower Bound ;
      %let conv=0;
      %let step=&istep;
      %let beta=&ibeta;
%DO %WHILE (&CONV=0);
      ods output fitstatistics = fitstatistics
ParameterEstimates=ParameterEstimates;
      proc nlmixed data=&data ;
      by rep n type R11 RR10 RR01 ;
      where rep=&rep and type=&type and n=&n;
      parms b0=&ibeta0 b1=&ibeta1 b2=&ibeta2 ;
      &param=&beta;
      odds = &odds;
      model &outcome ~ binary( odds/(1+odds) );
      title1 "beta is &beta " ;
      %if %scan(&replicate,1) ne %str() %then %do;
      replicate &replicate;
      %end;
      Run;

      data fitstatistics;
      set fitstatistics;
      format value best16.;
      if (Descr = "-2 Log Likelihood") then call symput ('LL',
put(value,best16.));
      run;

      data ParameterEstimates;
      set ParameterEstimates;
      if (PARAMETER = "b0") then call
symput("ibeta0",put(ESTIMATE,best16.));
      if (PARAMETER = "b1") then call
symput("ibeta1",put(ESTIMATE,best16.));
      if (PARAMETER = "b2") then call
symput("ibeta2",put(ESTIMATE,best16.));
      run;

      %let diff=%sysevalf(&LL-&refLL);
      %if %sysevalf(3.8413 <= &diff) %then %do;
            %if %sysevalf( &diff <= 3.8415) %then %do;
                  %let CONV=1;
            %end;
      %end;
      %if %sysevalf( &diff > 3.8415 ) %then %do;
            %IF &I=1 %then %let beta=%sysevalf(&beta+&step);
            %IF &I=2 %then %let beta=%sysevalf(&beta-&step);
            %let step=%sysevalf(&step*0.5);
      %end;
      data tmp;
      null= &refLL;
      neglogl=&LL;
      difference=&diff;
```

```
      param=&beta;
      step=&step;
      run;
      %IF &I=1 %then %do;
      proc append base =lci data = tmp;
      run;
      %end;

      %IF &I=2 %then %do;
      proc append base =uci data = tmp;
      run;
      %end;

      %IF &I=1 %then %let beta=%sysevalf(&beta-&step);
      %IF &I=2 %then %let beta=%sysevalf(&beta+&step);
      %end;
      %end;
      /*
      dm 'out; clear; pgm';
      proc print data=lci;
      title1 "95% Lower Confidence Bound - Iterations";
      run;

      proc print data=uci;
      title1 "95% Upper Confidence Bound - Iterations";
      run;
      */
      data lb (keep=PARAM);
      set lci end=eof;
      if eof then output lb;
      run;

      data ub (keep=PARAM);
      set uci end=eof;
      if eof then output ub;
      run;
*the code below was added to the original Kaufman macro to allow this
to run the simulated data with multiple repetitions and across a range
of parameter values;

proc sql;
create table output as
select a.param as lb, b.param as ub,
input(trim(left(symget("rep"))),best.) as rep,
        input(trim(left(symget("n"))),best.) as n,
        input(trim(left(symget("type"))),best.)as type,
        input(trim(left(symget("ibeta"))),best.) as estimate
from lb as a, ub as b;
quit;
%if %sysfunc(exist(out)) %then %do;
      data out;
      set out output;
      run;
      %end;
%else %do;
        data out;
        set output;
```

```
        run;
      %end;


/*data bd (rename=(param=Bound));
      merge lb ub;
      by _n;
      rep=&rep;
      type=&type;
      n=&n;
      run;

ods listing;
      proc print data=bd NOOBS;
      title1 "Likelihood-Based 95% Lower and Upper Confidence Bounds
for Paramater &param";
      title2 "Point Estimate is &ibeta";
      run;
*/
%mend bounds;



options nomprint nomlogic nosymbolgen nonotes;
ods results off;
ods listing close;
*this macro is added to call the Richardson and Kaufman code many times
over the simulated repetitions and a range of interaction effects;
%macro wrapper ;
%do rep=1 %to 1000;*Change as needed;
%let typea=1 2 4 6 7 9;
%let n=1000;*again limited to one sample size at a time;
%do kd=6 %to 6;
%let type=%scan(&typea,&kd);*Limit number of types, here only type 9;
%bounds (data=from_rrs_freq, outcome=outcome, odds=exp(b0)*(1+ b1*e1 +
b2*e2 + b3*e1*e2), param=b3,replicate=freq );
%end;
%end;
%mend wrapper;

%wrapper

ods listing close;
options nonotes;
ods results off;
ods output WaldPLLimits=NLP_PL_Limits_ICR2;
title2 "PROC NLP Profile and WALD CI for the ICR";
PROC NLP DATA=for_figures VARDEF=N cov=2 pstderr maxit=200 all;
by rep type R11 RR10 RR01;
*where rep in (1,2);
*where rr10=2.75 and rr01=2.25 ;

      *Bounds -500 <= ICR <= 500;
      PARMS beta0=-3, b_e1=1.4, b_e2=1.4,
          ICR=1;
    b_interaction=log((ICR+exp(b_e1)+
                       exp(b_e2)-1)/
                      (exp(b_e1)*exp(b_e2)));
      Eta=beta0+b_e1*e1+b_e2*e2+b_interaction*e1*e2 ;
```

```
        p=     exp(eta);
          *Example 9: Poission risk model, virtually identical to NLMIXED
output, i.e. worse than the log-binomial in terms of type 1 error and
power;
          ll = count*log(p)- p - lgamma(count+1);



        MAX ll;

        PROFILE ICR /alpha=.05 ;
RUN;



ods listing close;
options nonotes;
ods results off;
ods output WaldPLLimits=NLP_PL_Limits_ICR3;
ods trace off;;
PROC NLP DATA=for_figures VARDEF=N cov=2 pstderr maxit=200 all;
by rep type R11 RR10 RR01;
*where rep =1 ;*and RR10=1.2 and RR01=3.4;
*where rr10=2.75 and rr01=2.25 ;

        *Bounds -500 <= ICR <= 500;
         PARMS beta0=-4, b_e1=1.4, b_e2=1.4,
              ICR=1;
       b_interaction=log((ICR+exp(b_e1)+
                           exp(b_e2)-1)/
                           (exp(b_e1)*exp(b_e2)));
         Eta=beta0+b_e1*e1+b_e2*e2+b_interaction*e1*e2 ;
       p=     exp(eta);
          *Example 10: Log-binomial model, nearly identical to NLMIXED
output;
          ll = count*log(p)+(n-count)*log(1-p);



        MAX ll;

        PROFILE ICR /alpha=.05 ;
RUN;



*Below this line tests will be for causal interaction per Vanderweele
2007 and 2009, rather than any interaction;
*Example 11: Test of a3-a0>0 from linear risk model;
ods output estimates=ICsRR2;
ods output contrasts=TVWLRtest;
ods output
parameterestimates=a3_LogRiske1_e2a2(where=(trim(left(upcase(parameter)
))="E1*E2"));
ods output
parameterestimates=a0_LogRiske1a2(where=(trim(left(upcase(parameter)))=
"INTERCEPT"));
proc genmod data=for_figures descending;
by rep type R11 RR10 RR01;
*where rr11=&RR11_2;
class e1(param=ref ref="0") e2(param=ref ref="0");
```

```
model count/n=e1 e2 e1*e2/link=identity dist=binomial lrci;


*estimate "R10" intercept 1 e1 1 ;
*estimate "R01" intercept 1 e2 1 ;
*estimate "R11" intercept 1 e1 1 e2 1 e1*e2 1 ;
*estimate "R00" intercept 1;
estimate "IC"  e1*e2 1;
*Estimate produces the value of a3-a0 and a likelihood ratio based
confidence interval by default;
*Vanderweele proposes that a test of a3-a0>0 represents a test for
causal interaction;
estimate "TVW" e1*e2 1 intercept -1/;
*the contrast statement provides an actual likelihood ratio test of the
signficance of this term, although it is a two-sided test;
contrast "TVW" e1*e2 1 intercept -1;
*estimate "RD RR10-RR00" e1 1;
*estimate "RD RR01-RR00" e2 1;
*estimate "RD RR11-RR00" e1 1 e2 1 e1*e2 1;
run;



data tvw;
*Limit output to data of interest;
set icsrr2;
where label="TVW";
run;


proc sql;
*combine estimates into a summary table;
create table tvw_Ic as
select a.type,a.rr10,
a.rr01,trim(left(put(a.type,best8.)))||"_"||trim(left(put(a.rr10,best8.
)))||"_"||trim(left(put(a.RR01,best8.))) as indicator,
        a.estimate as a0, b.estimate as a3, a3-a0 as tvw_ic,
        (c.meanlowercl>0) as test_ICgt0, ((d.probchisq<.05) and
((b.lowerlrcl-a.lowerlrcl)>0)) as test_JOINT_MANUAL
        from a0_logriske1a2 as a, a3_logriske1_e2a2 as b,tvw as
c,tvwlrtest as d
        where a.rep=b.rep=c.rep=d.rep and a.type=b.type=c.type=d.type
and a.RR10=b.RR10=c.RR10=d.rr10 and a.RR01=B.RR01=c.rr01=d.rr01;
        quit;


*We can also develop these tests from within a log-binomial risk model,
as in eq 10 in the main text;
ods output
parameterestimates=B3_LogRiske1_e2(where=(trim(left(upcase(parameter)))
="E1*E2"));
ods output
parameterestimates=B3_LogRiske1(where=(trim(left(upcase(parameter)))="E
1"));
ods output
parameterestimates=B3_LogRiske2(where=(trim(left(upcase(parameter)))="E
2"));
proc genmod data=for_figures descending;
by rep type R11 RR10 RR01;
class e1(param=ref ref="0") e2(param=ref ref="0");
model count/n=e1 e2 e1*e2/link=log dist=binomial lrci;
```

```
store work.genmod2;
run;


*You can calculate the ICR by hand, and could program your own delta
method solution to the Confidence intervals if you were so inclined;
proc sql;
create table byhand as
select
a.rr10,a.rr01,a.type,trim(left(put(a.type,best8.)))||"_"||trim(left(put
(a.rr10,best8.)))||"_"||trim(left(put(a.RR01,best8.))) as indicator,
        a.estimate as b1, b.estimate as b2, c.estimate as b3,
(exp(b1 + b2 + b3) -
exp(b1) - exp(b2) +1) as ICR
from B3_logriske1 as a, b3_logriske2 as b, b3_logriske1_e2 as c
where a.rep=b.rep=c.rep and a.type=b.type=c.type and
a.rr10=b.rr10=c.rr10 and a.rr01=b.rr01=c.rr01;
run;
quit;


*But PROC PLM will conduct a JOINT test of Vanderweele's requirements
for you;
data _null_;
testval=put(log(2),16.15);
call symput("testval",testval);
run;
%put &testval;


ods listing close;
ods results off;
options nonotes;
ods output
plm.estimates=KDTV_est_testval1(where=(trim(left(upcase(label)))="TESTV
ALA"));
ods output
plm.estimates=KDTV_est_testval2(where=(trim(left(upcase(label)))="TESTV
ALB"));
*Example 12: THE PLM procedure produces ChiBar square tests of the
joint, one-sided inequality, but these do not perform well with
dictomous outcomes;
ods output plm.contrasts=kdtv_Pvals_testval2;
*Testval makes the macro run more easily and is probably more
straightforward in general;
*Note that with 1000 repetitions this takes a long time (much longer
than examples above) but it does eventually finish,
and for a single dataset this is not a problem;
proc plm source=work.genmod2;
estimate "testvala"  e1*e2 1 e1 1 ,"testvalb"
                    e1*e2 1 e2 1 /e  upper joint testvalue=&testval
adjust=T cl ;


run;
*this dataset manipulates the output from PROC PLM to test for the
joint effects described in eq 10 of the main text manually;
*ChiBar square output can also be considered, but it behaves strangely
when the strength of one effect is <RR=2 and the other has an effect
RR>2;
proc sql;
```

```
create table byhand2 as
select a.rep,a.type, a.rr10,a.rr01,
trim(left(put(a.type,best8.)))||"_"||trim(left(put(a.rr10,best8.)))||"_
"||trim(left(put(a.RR01,best8.))) as indicator,
        a.estimate as T1, b.estimate as T2,
((put(a.probZ,power.)="Detected") and (put(b.probz,power.)="Detected"))
as kdtest,c.probchibarsq
from kdtv_est_testval1 as a, kdtv_est_testval2 as b,kdtv_pvals_testval2
as c
where a.rep=b.rep=c.rep and a.type=b.type=c.type and
a.rr10=b.rr10=c.rr10 and a.rr01=b.rr01=c.rr01;
run;
quit;
```

Chapter 5 - Specific Aim 3

Additional background on Infectious Disease modeling

Epidemiologic considerations

Quantifying the effect of an intervention designed to reduce the transmission of an infectious disease is complicated by the lack of independence between the disease in one individual and the requirement that that individual also interact with someone susceptible to disease (not yet infected) in order to cause a new case of disease.[1] This complicates the definition of a counterfactual effect for an intervention, because the effects of multiple individuals can be seen to "interfere" with each other leading to multiple potential outcomes with the same set of exposures depending on how individual members of the population and population level factors such as disease prevalence interact.[1-3] However, this also means there is an opportunity for the effects of an intervention be be greater than the expected effect, particularly when interventions can be combined, this concept is also referred to in the infectious disease modeling literature as "synergy".[4] Vanderweele (ref) discusses the detection of this "interference" and its relation to detection of sufficient cause interaction, defining interference as a population average effect greater than the individual direct effects of the intervention – interaction of effects in the sense of Chapter 5, but at the population level.[3] However, the examples he uses involve no more than 2 individuals (one treated, one untreated), and as described in the overall background for this proposal, scaling a directed acyclic graph (DAG) or potential outcomes model to the population level would be a challenge.[1,2,5]

Compartmental Models

However, methods have been developed to model transmission dynamics of many

infectious diseases, including sexually transmitted infections. The simplest model for

infectious disease transmission that adequately accounts for these dependencies is the

Kermack-McKendrick model of population transmission dynamics.[6-7] Because persons
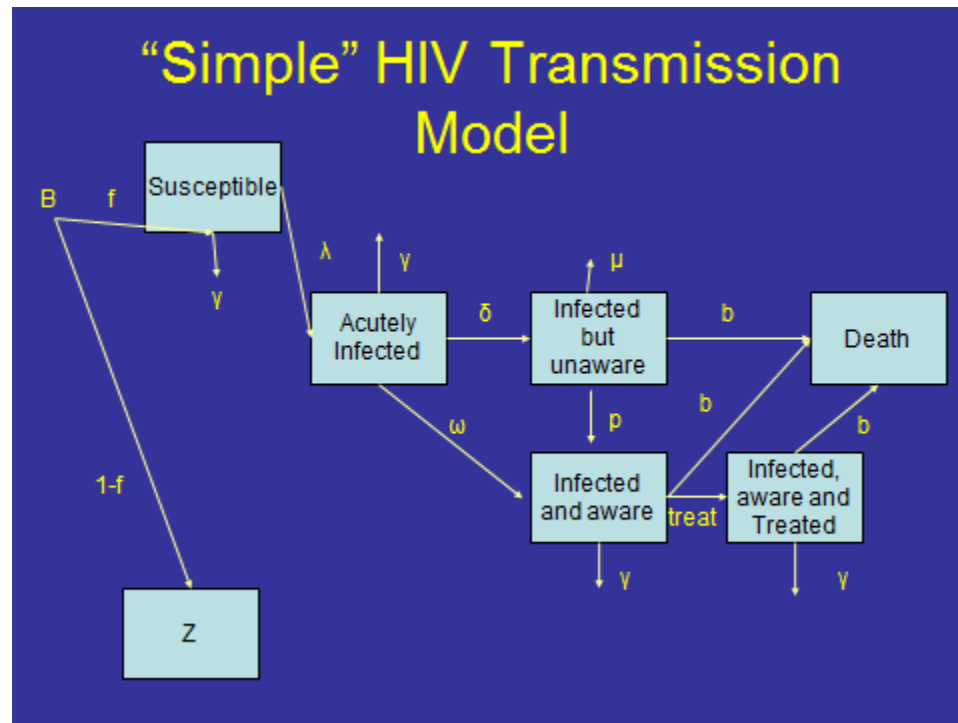
Figure 1: Compartmental model of HIV transmission

infected with HIV do not develop immunity and recover, the simplest model of HIV

transmission would contain only two compartments, one for those susceptible to

infection, and another with those infected with HIV. This "SI" model would be

completely defined by only two differential equations:

$$\frac{\partial S}{\partial t} = B - \beta SI - \mu(N - S)$$

$$\frac{\partial I}{\partial t} = \beta SI - (\mu + \epsilon)I$$

Where B= the number of new births in time t, S= the number of persons in the population that is susceptible to disease, I=the number infected at time t. The greek letters β, μ and ε correspond to rates of infection, death due to causes other than disease and death due to disease respectively. The model of Granich[8] and that shown as Figure 1 are examples of how this simple compartmental model can be expanded to allow for heterogeneity in population risk(through the addition of more compartments with different rates of risk), and changes in risk over time(through either additional compartments or partial differential equations). Figure 1 is even simpler than the Granich model as it does not distinguish HIV serostatus knowledge or treatment in the acute or late stages from treatment in other stages, and limits the population at risk of HIV infection to a proportion f that ever engage in risk. As illustrated by Granich,[8] when a compartmental model is adequately developed, simulations can be performed in both the presence and the absence of an intervention, to quantify the counterfactual average effects at the population level.

These models have gained popularity because they allow for the derivation of an analytic solution to the model through a system of ordinary differential (or partial differential) equations. However, the use of "compartments" leads to only population average effects, because we model the average movements of groups of individuals over time.[6-7] This approach makes sense for a variety of pathogens, e.g. respiratory viruses such as influenza, where persons are exposed to an airborne pathogen and their

interaction with infected individuals is not necessary for disease to occur. Furthermore,

for rare diseases (in the general population HIV occurs in less than 5% of the population

even in sub-Saharan countries such as Nigeria,[9] and in less than 2% of the population in

areas such as China and the US[9-10]), with very low probability of transmission during a

given exposure (a per sex act risk <<1/100 for all but receptive anal sex[11]) stochastic, or

random chance effects can mean the difference between the extinction of an epidemic

and an outbreak in a closed population. The deterministic Kermack-McKendrick model

can be modified to capture stochasticity, by substituting a distribution of values for each

parameter (greek letter in Figure 1) and sampling from that distribution. However, this

leads to a loss of the simple analytic solution for the model.[7]

For a sexually transmitted disease, the interaction with two individuals leading to

exposure is explicitly defined (pun intended).   It is possible, through the use of

increasing complex ODE models with more and more compartments, to make the

interaction between populations more explicit. However, for a sexually transmitted

disease the timing of interactions between groups defines transmission risk as well.[10-14]

Figures 2 and 3 show two examples of sequencing of sexual partnerships.

A1:
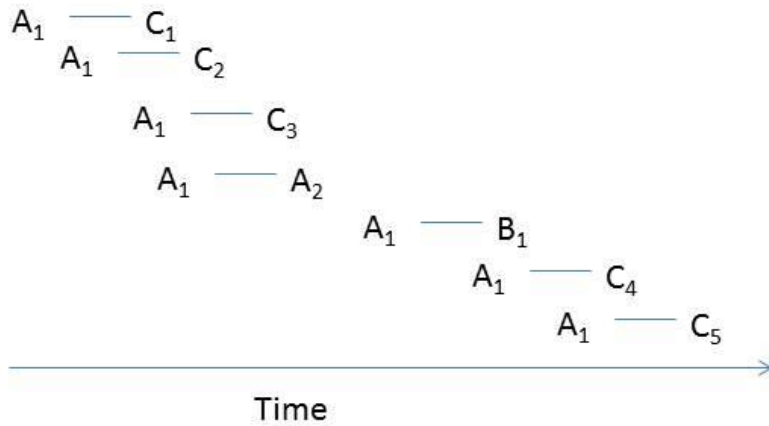
A_1 —— C_1
A_1 —— C_2
A_1 —— C_3
A_1 —— A_2
A_1 —— B_1
A_1 —— C_4
A_1 —— C_5

Time

C4:

C_4 —— C_6
C_4 —— A_1
C_4 —— C_3

Time

Group A – High risk HIV-uninfected
Group B – HIV-infected
Group C – Low risk HIV-uninfected

Who's actually at risk for HIV infection here??

169
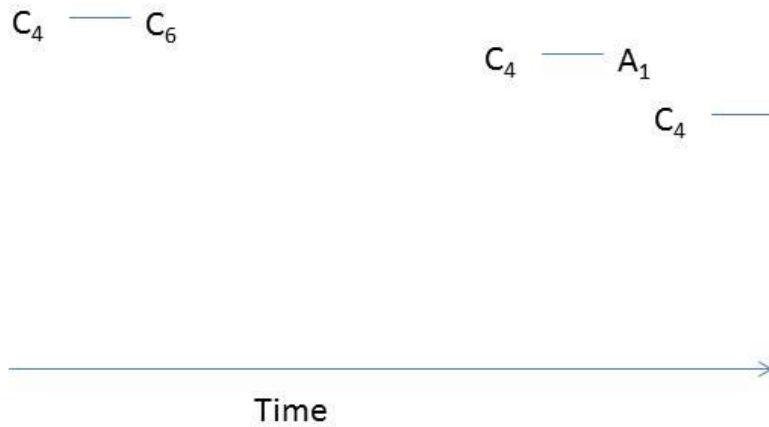
Figure 2: Schematics of sexual partnerships over time from the perspective of two individuals A1 and C4, with no concurrency

(overlap of partnerships) for person C4.

If group A individuals do not interact with an infected person in group B and subsequently become infected before or while engaging

in a sexual relationship with group C then there is no risk from the group A person or the group B person to the group C person. In

Figure 2, only persons C4, C5, and C6 are at risk of HIV infection. However, if there is overlap between partnerships the risk of

infection can increase by to the power of the number of overlapping partnerships (i.e. exponentially). In Figure 3 person C6 is also at

risk due to the overlap between C4's relationship with C6 and A1.
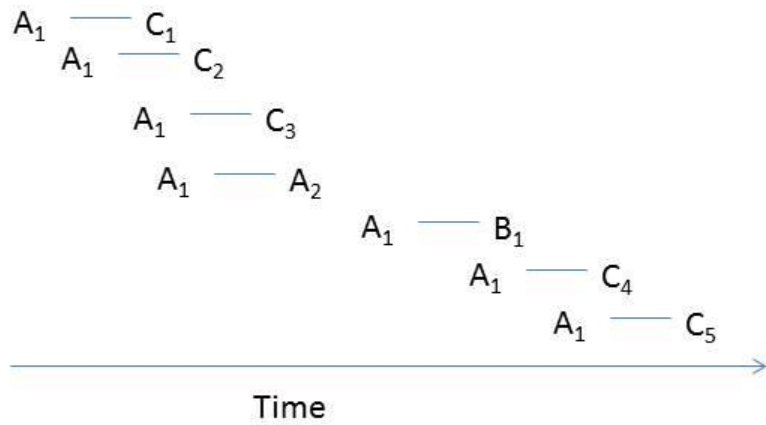
A_1 —— C_1
A_1 —— C_2
A_1 —— C_3
A_1 —— A_2
A_1 —— B_1
A_1 —— C_4
A_1 —— C_5

Time

C_4 —— C_6
C_4 —— A_1
C_4 —— C_3

Time

Group A – High risk HIV-uninfected
Group B – HIV-infected
Group C – Low risk HIV-uninfected

Who's actually at risk for HIV infection here??
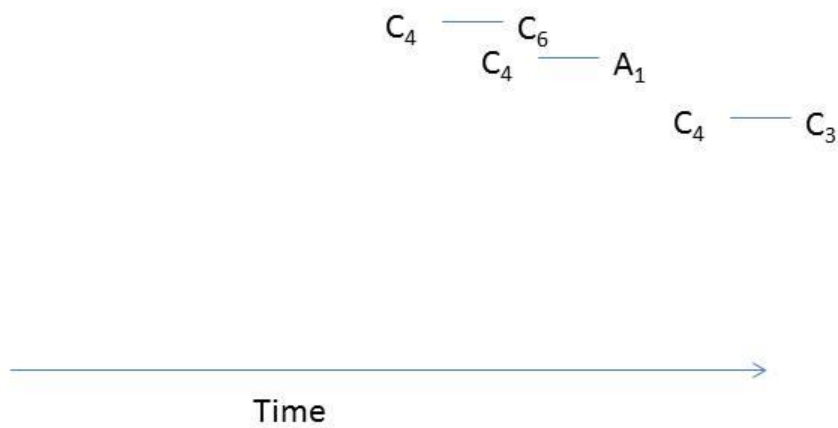
Figure 3: Sexual partnerships with overlap for person C4.

So the timing of the interactions of groups matters, and this can be captured through the use of partial differential equations. When using a system of partial differential equations, a closed form solution becomes far more complicated, and depending on the parameters available to be held fixed, a closed form solution may not be estimable. Coupled with the desire to add complexity through stochasticity for our model for HIV transmission we end up in a situation where our model becomes one of simulated interactions amongst population groups over time.

Agent-based models

If we take pains to add stochasticity and time-dependant interactions to our model, we can do so by explicitly modeling individuals, rather than groups of individuals, and their interactions over time. Such models are called individual or agent-based models.[15-16] These models create a population of individuals with given characteristics (agents) and then model their interactions explicitly through a series of tests of probabilistic equations designed to mimic how persons come into contact, in this case, have sex, and then move on to the next interaction. They allow us to form and dissolve relationships based on described distributions of partner types, and relationship duration, defined by sampling from these distributions over the course of time. In our case, we can then add an intervention that affects some or all of the population at a given point in time, and see what would have happened in the presence and absence of the intervention.

Exponential Random Graph Models

In the current work we developed a separable temporal exponential random graph model[17] (STERGM) for sexual partnership formation. STERGMs are the dynamic, stochastic and agent-based extension of cross-sectional exponential random graph models (ERGMs) and as such allow for simulation of stochastically evolving sexual partnership networks.[18,19] These models accommodate the statistical dependence among partnerships described in Figures 2 and 3, and model parameters describe factors associated with both partnership formation and dissolution over time. The ability to include statistical dependence among partnerships (e.g. as in the case of two

or more partnerships overlapping in time, see figure) is a major strength of the

applicability of the ERGM framework to the modeling of sexual partnerships, since

classical compartmental methods for modeling HIV transmission cannot capture such

dependence.[12-13] Additionally ERGMs allow us to define a set of network features, and

the STERGM portion of the model allows simulation of partnership formation and

dissolution over time while maintaining the underlying network structure (within

stochastic variation). This give statistical rigor to partnership formation and dissolution,

making STERGM models reproducible within stochastic variation and mathematically

tractable in ways that agent-based models are not able to be.[15-17]

References:

1) Hudgens MG, Halloran ME. Towards causal inference with interference. J Am Stat Assoc. 2008;103:832–842.

2) Halloran ME, Struchiner CJ. Causal inference for infectious diseases. Epidemiology. 1995;6:142–151.

3) VanderWeele TJ, Vandenbroucke JP, Tchetgen EJ, Robins JM. 26. A mapping between interactions and interference: implications for vaccine trials. Epidemiology. 2012 Mar;23(2):285-92.

4) Dodd PJ, White PJ, Garnett GP. Notions of synergy for combinations of interventions against infectious diseases in heterogeneously mixing populations. Math Biosci. 2010 Oct;227(2):94-104.

5) Lauritzen SL, Richardson TS. Chain graph models and their causal interpretations. J R Stat Soc Ser B 2002; 64: 321-348.

6) Anderson RM, May RM. Infectious Diseases in Humans: Dynamics and Control. New York, New York: Oxford University Press; 1991.

7) Keeling MJ, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton, NJ: Princeton University Press; 2008.

8) Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. Lancet. 2009 Jan 3;373(9657):48-57.

9) UNAIDS. The Gap Report: The Beginning of the End of the AIDS Epidemic. Geneva; UNAIDS: 2014. Available from:

http://www.unaids.org/sites/default/files/media_asset/UNAIDS_Gap_report_en.pdf

10) CDC. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas—2012. HIV Surveillance Supplemental Report 2014;19(No.3). Published November 2014. Available from: http://www.cdc.gov/hiv/library/reports/surveillance/

11) Patel P, Borkowf CB, Brooks JT, Lasry A, Lansky A, Mermin J. Estimating per-act HIV transmission risk: a systematic review. AIDS. 2014 Jun 19;28(10):1509-19.

12) Morris M, Epstein H. Role of concurrency in generalised HIV epidemics. Lancet. 2011 Nov 26;378(9806):1843-4; author reply 1845-6.

13) Goodreau SM, Cassels S, Kasprzyk D, Montaño DE, Greek A, Morris M. Concurrent partnerships, acute infection and HIV epidemic dynamics among young adults in Zimbabwe. AIDS Behav. 2012 Feb;16(2):312-22.

14) Rosenberg ES, Rothenberg RB, Kleinbaum DG, Stephenson RB, Sullivan PS. The implications of respondent concurrency on sex partner risk in a national, web-based study of men who have sex with men in the United States. J Acquir Immune Defic Syndr. 2013 Aug 1;63(4):514-21.

15) Railsback SF, Grimm V. Agent-based and Individual-based Modeling: A Practical Introduction. Princeton, NJ: Princeton University Press; 2012.

16) Van Dyke Parunak H, Savit R, Rialo RL. Agent-based modeling vs. equation-based modeling: A case study and users' guide. Proceedings of the First International

Workshop on Multi-Agent Systems and Agent-Based Simulation. London, UK:

Springer-Verlag; 1998.

17) Krivitsky PN, Handcock MS. A separable model for dynamic networks. Journal of

the Royal Statistical Society: Series B (Statistical Methodology). 2014; 76.1: 29-

46.

18) Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. statnet: Software

tools for the Statistical Modeling of Network Data. Seattle, WA; 2003. Version

2.0. Available from: http://statnetproject.org.

19) Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M (2008). statnet:

Software Tools for the Representation, Visualization, Analysis and Simulation of

Network Data. Journal of Statistical Software, 24(1). Available from:

http://www.jstatsoft.org/v24/i01/paper Accessed: 11/20/14

Aim 3 - Manuscript to be submitted to Lancet HIV

Introduction

In the US HIV epidemic, men who have sex with men (MSM) have been, and continue to be, the most heavily impacted HIV risk group.[1-2] Recent estimates of HIV incidence among young black MSM have led to calls[3,4] for immediate action to improve HIV prevention in the United States. HIV-infected persons who are not aware of their HIV infection are more likely to engage in behaviors that place their partners at risk of HIV transmission, and there is meta-analytic evidence that most persons who learn of their HIV-positive status take steps to reduce the risk of HIV transmission to others.[5]

The US Centers for Disease Control (CDC) recommends that MSM should test for HIV at least annually.[6,7] However, recent reports suggest that most men are not testing for HIV this frequently.[4,7-8] How many US MSM have ever tested, how frequently they test, and if MSM should test more frequently than once a year are still under debate.[9-15] In addition to behavioral changes that result from an HIV diagnosis, this diagnosis serves as a necessary but insufficient first step to receiving HIV medical care and antiretroviral treatment, which can result in reduced viral load and associated decreases in infectiousness.[16] In the United States, a minority of those diagnosed with HIV actually achieve viral suppression.[17-19] The proportion of all infections that are diagnosed and the testing frequency among MSM are increasing[7] but there has been relatively little change in the proportion of all diagnosed persons achieving viral suppression[17-19], with some

estimates[18] suggesting that this proportion has recently decreased. It is unclear if the potential behavioral changes associated with increased awareness of HIV infection resulting from increased testing frequency are sufficient to result in reduced incidence in the absence of increased viral suppression in the population.

We developed interaction and transmission models, parameterized using data from a national online survey of MSM[20] to assess the impact of increases in testing frequency on 3-year HIV incidence, and to determine how testing frequency interacts with viral suppression, test sensitivity and the proportion of the population seeking testing.

Methods

Modeling strategy and source of parameter values

We developed individual based models (IBMs) using the time-varying extensions of exponential random graph models (ERGMs), called Separable Temporal ERGMs (STERGMs).[21] STERGMs have the ability to capture statistical properties of sexual networks, a feature that cannot be described in either IBMs alone or within compartmental (e.g., susceptible-infected) models for HIV transmission.[22] We used this methodology to define partnership formation and dissolution and simulate HIV transmission in a MSM sexual network of 5250 men. STERGM models were fit using the Statnet[21] suite of packages in R, and we use customized extensions to the R package EpiModel[23] to control testing intervention parameters within our simulations.

A national online survey of MSM provided information on behaviors with up to five partners in the 6 months prior to interview,[20,24] data that were used to define

partnership formation and dissolution parameters of the model. Additional detail about the structure of our STERGM model and other model parameters that were held constant in all simulations is included in the Supplemental appendix. The survey also collected: partnership type (main, casual, and one-time), current partners (number at the time of interview) and in the last 6 months, partnership duration, and self-reported HIV serostatus of the participant and each reported partner. In addition, a 735-person subset of study participants was tested for HIV at baseline, and, if HIV-negative, enrolled in follow-up with an HIV-test performed 12 months after completion of the baseline questionnaire. This longitudinal follow-up provided a measure of HIV incidence against which to validate our model.[25]

Statistical analysis of outcomes from model simulations

The model was utilized to assess the impact of hypothetical manipulations of HIV testing on HIV incidence, circulating virus, and time to diagnosis. These simulated interventions, including changes in testing frequency distribution and other testing intervention parameters were varied as described in Panel 1 and the Supplemental Appendix, and complete cohorts were simulated 20 times to capture the stochastic variation inherent in the model and develop a sample of populations in which three year incidence of HIV infection could be summarized. Supplemental Figures S4 and S5 describe the rationale for presenting data on 20 simulations of each scenario. We calculated the median and interquartile range, used these observed incidence measures to calculate incidence rates, then compared these values for the baseline model and models with modified

testing frequency distributions or other key factors related to HIV testing as described in Panel 1.

Investigation of the effects of changes in testing frequency on transmission dynamics

Like individual based models, our STERGM approach allow us to: track each MSM throughout the 3-year simulation; calculate summary measures of the number of infected men, their total viral load, and the number of onward transmissions each man contributes; and stratify each of these measures by serostatus awareness, and testing status. Calculation details appear in the Supplemental Appendix. The model records how often each man tests for HIV and calculates the time between infection and diagnosis. To assess the impact on HIV incidence, we compare the distributions of these intermediate factors and their impact on HIV transmission dynamics in the population across levels of testing frequency and other testing interventions.

Investigation of population level interactions of testing intervention components

To address how the effect of increasing testing frequency varied depending on the values for parameters of other testing interventions under our control in the simulations, we evaluated the effect of testing frequency at higher levels of viral suppression, with more sensitive tests, and with more complete coverage of testing interventions (i.e. with fewer MSM who never test for HIV). We modified these variables to represent typical US situations as well as an ideal situation, and assess the effect of increasing testing frequency under each of these counterfactual scenarios. By varying one facet of a testing intervention at a time and then combining interventions we can

describe any synergy across interventions, where the effect of interventions in combination is more (or less) than when each intervention was implemented on its own.

Results

Baseline model and the effects of increases in testing frequency
Our baseline scenario, in which 80% of the MSM population test on average annually with 43·4% of the diagnosed population achieving viral suppression (Panel 1), resulted in a median of 422·5 infections over three years, corresponding to an incidence rate of 2·87 per 100 person-years (Table). In this scenario, over 11,000 HIV tests were performed over the 3-year study period. Both the total number of tests performed (Figure 1) and the mean number of tests per person (Supplemental Appendix, Table S3) increased as testing frequency increased. Under the scenario where participants tested every 90 days, we observed a greater than 3-fold increase in testing compared to baseline. Both the median time from infection to diagnosis and the variance around that time decreased with more frequent testing (Figure 2). HIV incidence decreased slightly (Figure 1) but not meaningfully with increasing testing frequency.

Effects of testing interventions other than increases in testing frequency
In simulations that reflected other possible interventions related to HIV testing, incidence did not change from baseline with more sensitive HIV tests (detection at 22 or even 0 days), or with decreasing proportions of the population that does not ever test (Table). In contrast, increasing from the estimated US national average[17] of 43.4% achieving viral suppression to 68.5% (corresponding to a scenario where all those

currently estimated to be receiving HIV care[17] are suppressed) reduced the median

incidence by 24% to 2·19 per 100 person-years (IQR: 2·03–2·47). Increasing the

proportion virally suppressed to an ideal goal of 100% of those diagnosed reduced

incidence to a median of 1·43 per 100 person-years (IQR: 1·32–1·57), a reduction of 50%

relative to baseline.

Effects of increases in testing frequency under a scenario where viral suppression is improved
To investigate why increasing the frequency of testing had little to no impact on HIV

incidence, whereas HIV care had a large impact, we estimated the proportion of the

total infections and circulating viral load in each stage of serostatus awareness and

infection, under varied testing and care scenarios. Figures 3 and 4, and Figure S3 in the

supplemental appendix show how increasing testing frequency can only have a limited

impact on the total circulating viral load (Figure 3) and, as a result, on transmitted

infections (Figure 4), because diagnosis alone only moves those previously undiagnosed

to the diagnosed group and has a small impact on circulating viral load overall.

However, if treatment coverage can be increased, changes in testing frequency begin to

have a greater impact on HIV transmission. Figures 3 and 4 show that increasing viral

suppression among those whose infection is diagnosed reduces circulating viral load

(Figure 3) and onward transmission (Figure 4) substantially within both the baseline

testing scenario and a scenario where MSM test every 90 days. In Figure 3, when the

diagnosed population no longer contributes to the total circulating viral load, increasing

testing frequency to once every 90 days and using a test that detects HIV within 22 days

of infection removes an additional 26% of the circulating viral load, reducing the total viral load to only 35% of the baseline estimate. In Figure 4, under the baseline testing scenario, a change to a more sensitive test alone also does not have any impact on incidence. However, when treatment coverage improves to 100% of the diagnosed, changing to a more sensitive test appears to decrease incidence even further than increasing treatment coverage or testing frequency alone. Figure 5 shows the proportion of all infections averted for each component of this last scenario. With three testing interventions combined, there is an additional 7% reduction in incidence over the 3-year simulation, over and above what was observed when each intervention was applied to the population individually. A similar synergistic effect was observed when viral suppression among the diagnosed population was 100% and more (93% compared to 80% at baseline) MSM test for HIV at least annually (See supplemental appendix).

Discussion

We developed a simulation model for MSM sexual partnerships parameterized in part by data from an online survey of MSM, and evaluated 4 hypothetical interventions related to HIV testing in terms of their effects on resulting HIV incidence in our simulated population over a short (3-year) period. We found that, at current rates of HIV viral suppression in the US, increasing the frequency of HIV testing by MSM increased the numbers of test performed, but did little to affect 3-year incidence of HIV. The only scenario that led to reduced HIV incidence was increasing the proportion of MSM who proceed from diagnosis to achieve HIV viral suppression. Under the

assumption that 100% of those who are diagnosed with HIV achieve viral suppression, circulating viral load and, ultimately, HIV transmissions could be reduced if MSM tested every 90 days instead of testing annually. Thus, increased HIV testing frequency for MSM has a role to play as part of a comprehensive package of services that culminates in effective linkage and care, but, according to our data, it would not lead to reductions in HIV incidence if implemented without improvements in downstream continuum steps.

Our baseline scenario was parameterized using behavioral data from a prospective HIV incidence cohort of US MSM,[20,25] and produced an annualized estimate of HIV incidence of 2·87/100 person years). The cohort study observed a very similar HIV incidence (2·4/100 person years; 95% CI 1·4–4·1),[25] which suggests our model was reasonably calibrated. The model estimates of incidence and the cohort estimate are consistent with meta-analysis results describing the annual HIV incidence among MSM recruited from community-based studies in the US (mean 2·25, 95% CI: 2·05–2·45).[26] In addition, we achieved a baseline distribution of HIV testing frequency that was consistent with our survey data and which produced relatively stable estimates of HIV testing over time, but at equilibrium values that may be more realistic than those used in other models of HIV testing behavior.[9-15]

Our finding that testing frequency had little to no impact on 3-year HIV incidence is consistent with several of these other studies. Gray[9] found that increasing testing to twice or four times annually resulted in a non-significant trend in infections averted,

with reductions in 10 year incidence of 8·5% (range -5·7–20%) and 13·8% (-4·2–20·6%)

respectively. For the baseline scenario in her deterministic model of HIV transmission,

Long[12] assumed that only 23% of high-risk individuals (including MSM) test annually; the

counterfactual scenarios were to deploy high sensitivity testing strategies (i.e., able to

identify persons at 22, 17 and 11 days after infection). She found a large reduction in

HIV incidence if more high risk people tested annually, but only small marginal benefits

from testing every 6 months. Semi-annual testing of MSM with a test with a 17 day

window period was estimated to provide a 1·9% reduction in incidence at a cost of $4·9

billion dollars for additional testing.[12] Likewise, Lucas[10] found that testing MSM every 3

months would be the "optimal" strategy, being cost effective ($45,074/QALY), but at a

cost of $8·1 billion per year. It is unclear whether the publically and privately funded

healthcare systems responsible for such testing are prepared to absorb these costs.

Khanna[15] studied a scenario where testing frequency was tailored to personal risk

(based on number of non-main sex partners with whom condomless anal intercourse

occurs), and found this strategy to significantly reduce incidence compared to both

annual and less frequent (testing every 2 and 10 years on average) testing. CDC

recommends that highest risk MSM, which would include men with more than 3 casual

sex partners in 3 months as modeled by Khanna, consider pre-exposure prophylaxis,

which also requires frequent re-testing for HIV.[27] Because the average partnership

duration reported by our survey participants was greater than 365 days, annual testing

corresponds roughly to testing with every new partner. Strategies that can encourage

testing with each new partner[28-29] may be an alternative that is appealing to lower risk MSM.

We further previous reports by examining the intermediate impacts of HIV testing on transmission. Given our current situation, with large gaps in viral suppression for those MSM with diagnosed HIV infection,[18-19] the interaction between increasing testing frequency and achieving viral suppression given a positive test result is important. Even without additional testing, in our model increasing the proportion of HIV-infected men on treatment reduced incidence dramatically. In the absence of treatment and viral suppression for those who are aware of their infection, the only mechanism for reducing incidence after diagnosis is through behavioral changes due to serostatus awareness. When only 43% of the diagnosed population achieves viral suppression (reported by CDC in early 2014,[17] but recently updated estimates are even lower[18,19]) the majority of testing results in moving small percentages of men from undiagnosed to "diagnosed but not suppressed", leading to only small reductions in the circulating viral load in the population overall. In contrast, if universal and prompt suppression of viral load after HIV diagnosis were achieved, increasing testing frequency and increasing test sensitivity could further reduce the remaining circulating viral load by a relatively larger percentage, ultimately reducing transmission. This interaction of the impacts of testing programs and treatment efforts has been observed in other models[10,13,15] but not always identified as such. For example, Lucas et al[10] accounted for this concept by assuming immediate access to therapy for all those who were diagnosed when determining that the "optimal" strategy was for MSM to test every 90 days, but did not

present a counterfactual situation in which testing frequency was increased without perfect follow-up care. CDC estimates that proportion of all infected individuals who are aware of their infection has increased in the last 3 years, but will likely not reach the NHAS 2015 goal of 90%.[4,7] The current estimates of the proportion of MSM that have achieved viral suppression fall even shorter of the 2015 NHAS target.[18-19] Efforts to improve both outcomes simultaneously are needed.

Our analysis has several potential limitations. In our model, although we parameterized test sensitivity to allow for false-negative results to occur, we did not consider the impact of false-positive results. With a low prevalence of HIV in the population overall, increasing the frequency of testing among the uninfected population will likely increase the number of false-positive results given to MSM annually, which could impact future testing behavior. We assume testing behaviors are periodic, and do not account for risk-based episodic HIV testing (for example, testing based on recent exposure or symptoms of HIV seroconversion illness). The extent to which such testing could change the transmission dynamics of HIV will again be limited by how quickly men newly identified as being infected are able to access treatment and suppress their viral load. As in Lucas,[10] in our model treatment and viral suppression are started by MSM immediately after diagnosis. This is in contrast to other recent models[9,11-15] which instead assumed that treatment will commence based on time since infection and disease progression. However, in the US it is recommended that all HIV-infected persons be offered treatment as soon as possible after diagnosis.[30] Even if this is not happening currently

throughout the US, for our most optimistic scenario we wanted to model viral

suppression as being available to all as soon as possible after diagnosis.

According to our data more frequent HIV testing by US MSM will not result in reduced

HIV incidence unless combined with improvements in effective HIV care. With limited

resources available for reducing onward transmission of HIV, efforts focused on

increasing the proportion of infected men who achieve and maintain viral suppression

would have a larger impact on HIV transmission. Only once viral suppression becomes

the normative outcome of HIV diagnosis does additional focus on increasing HIV testing

as the gateway to this outcome become warranted.

References

1. CDC. Estimated HIV incidence in the United States, 2007–2010. HIV surveillance supplemental report 2012; Vol. 17(No. 4). Atlanta, GA: US Department of Health and Human Services, CDC; 2012. Available at http://www.cdc.gov/hiv/library/reports/surveillance.

2. CDC. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data: United States and 6 U.S. dependent areas—2011. HIV surveillance supplemental report, 2013. Vol. 18(No. 5). Atlanta, GA: US Department of Health and Human Services, CDC; 2013. Available at

   http://www.cdc.gov/hiv/pdf/2011_monitoring_hiv_indicators_hssr_final.pdf   .

3. Rosenberg ES, Sullivan PS, Kelley CF, Sanchez TH, Luisi N, et al. Race and age disparities in HIV incidence and prevalence among MSM in Atlanta GA. CROI 2014: March 3-6, 2014; Boston, Masssachusetts. Abstract 38. Available from: http://www.croiconference.org/sessions/race-and-age-disparities-hiv-incidence-and-prevalence-among-msm-atlanta-ga Accessed: 2/9/2015

4. Office of National AIDS Policy. National HIV/AIDS strategy for the United States. Washington, DC: Office of National AIDS Policy; 2010. Available at http://aids.gov/federal-resources/national-hiv-aids-strategy/nhas.pdf Accessed:

5. WHO Guidelines: prevention and treatment of HIV and other sexually transmitted infections among men who have sex with men and transgender people: recommendations for a public health approach http://www.who.int/hiv/pub/guidelines/msm_guidelines2011/en/ (accessed Nov 20 2014)

6. CDC. Revised recommendations for HIV testing of adults, adolescents and pregnant women in health-care settings. MMWR 2006; 55(No. RR-14).

7. Cooley LA, Oster AM, Rose CE, Wejnert C, Le BC, Paz-Bailey G; NHBS Study Group. Increases in HIV testing among men who have sex with men--National HIV Behavioral Surveillance System, 20 U.S. Metropolitan Statistical Areas, 2008 and 2011. PLoS One. 2014 Sep 2; 9(9):e104162. doi: 10.1371/journal.pone.0104162. eCollection 2014.

8. Chandra A, Billioux VG, Copen CE, Balaji A, DiNenno E. HIV testing in the U.S. household population aged 15-44: data from the National Survey of Family Growth, 2006-2010. Natl Health Stat Report. 2012 Oct 4;(58):1-26.

9. Gray, R. T., et al. Increased HIV Testing Will Modestly Reduce HIV Incidence among Gay Men in NSW and Would Be Acceptable if HIV Testing Becomes Convenient. PLoS ONE 8(2): 2013.

10. Lucas A, Armbruster B. The cost-effectiveness of expanded HIV screening in the United States. AIDS 2013 27(5): 795-801.

11. Wilson DP, Hoare A, Regan DG, Law MG. Importance of promoting HIV testing for preventing secondary transmissions: modelling the Australian HIV epidemic among men who have sex with men. Sexual Health 2009 6(1): 19-33.

12. Long EF HIV Screening via Fourth-Generation Immunoassay or Nucleic Acid Amplification Test in the United States: A Cost-Effectiveness Analysis. PLoS ONE 2011 6(11): e27625. doi:10.1371/journal.pone.0027625

13. Brookmeyer R, Boren D, Baral SD, Bekker LG, Phaswana-Mafuya N, Beyrer C, Sullivan PS. Combination HIV Prevention among MSM in South Africa: Results from Agent-based Modeling. PLoS One. 2014 Nov 14;9(11):e112668. doi: 10.1371/journal.pone.0112668. eCollection 2014.

14. Khanna AS, Goodreau SM, Gorbach PM, Daar E, Little SJ. Modeling the Impact of Post-Diagnosis Behavior Change on HIV Prevalence in Southern California Men Who Have Sex with Men (MSM). AIDS Behav. 2014 Aug;18(8):1523-31.

15. Khanna AS, Goodreau SM, Wohlfeiler D, Daar E, Little SJ, Gorbach PM. Individualized diagnosis interventions can add significant effectiveness in reducing human immunodeficiency virus incidence among men who have sex with men: insights from Southern California. Annals of Epidemiology 2014 epub ahead of print October 7, 2014. DOI: http://dx.doi.org/10.1016/j.annepidem.2014.09.012

16. Sullivan PS, Carballo-Dieguez A, Coates T, Goodreau SM, McGowan I, Sanders EJ, et al. Successes and challenges of HIV prevention in men who have sex with men. Lancet. 2012; 380(9839):388–99.

17. Gray KM, Cohen SM, Hu X, Li J, Mermin J, Hall HI. Jurisdiction level differences in HIV diagnosis, retention in care, and viral suppression in the United States. J Acquir Immune Defic Syndr. 2014 Feb 1;65(2):129-32.

18. Bradley H, Hall HI, Wolitski RJ, Van Handel MM, Stone AE, et al. Vital Signs: HIV diagnosis, care, and treatment among persons living with HIV--United States, 2011. MMWR Morb Mortal Wkly Rep. 2014 Nov 28;63(47):1113-7.

19. Rosenberg ES, Millett GA, Sullivan PS, Del Rio C, Curran JW. Understanding the HIV disparities between black and white men who have sex with men in the USA using the HIV care continuum: a modeling study. Lancet HIV. 2014 Dec;1(3):e112-e118.

20. Khosropour, C. M., B. A. Johnson, A. V. Ricca and P. S. Sullivan (2013). "Enhancing Retention of an Internet-Based Cohort Study of Men Who Have Sex With Men (MSM) via Text Messaging: Randomized Controlled Trial." J Med Internet Res 15(8): e194.

21. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M (2008). statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. Journal of Statistical Software, 24(1). Available from: http://www.jstatsoft.org/v24/i01/paper Accessed: 11/20/14

22. Morris M, Kurth AE, Hamilton DT, Moody J, Wakefield S. Concurrent partnerships and HIV prevalence disparities by race: linking science and public health practice. Am J Public Health. 2009;99(6):1023–31.

23. Jenness SM, Goodreau SM, Morris M. EpiModel – Mathematical Modeling of Infectious Disease Version 1.1.1, A Tutorial. Available from: www.epimodel.org Accessed 11/20/14

24. Sullivan PS, Peterson J, Rosenberg ES, Kelley CF, Cooper H, Vaughan A, Salazar LF, Frew P, Wingood G, Diclemente R, del Rio C, Mulligan M, Sanchez TH. Understanding racial HIV/STI disparities in black and white men who have sex with men: a multilevel approach. PLoS One. 2014 Mar 7;9(3):e90514

25. Ricca A, Khosropour C, Sullivan PS. HIV incidence among men who have sex with men enrolled in an online behavioral risk study. XIX International AIDS

Conference. Washington, DC, USA July 22-27, 2011. Abstract: TUPE205 Available from: http://pag.aids2012.org/abstracts.aspx?aid=11790

26. Stall R, Duran L, Wisniewski SR, Friedman MS, Marshal MP, McFarland W, Guadamuz TE, Mills TC. Running in place: implications of HIV incidence estimates among urban men who have sex with men in the United States and other industrialized countries. AIDS Behav. 2009 Aug;13(4):615-29. doi: 10.1007/s10461-008-9509-7. Epub 2009 Feb 10.

27. Centers for Disease Control and Prevention. Pre-exposure prophylaxis for the prevention of HIV infection in the United States—2014: a clinical practice guideline. www.cdc.gov/hiv/pdf/guidelines/PrEPguidelines2014.pdf.

28. Chakravarty, D., et al. (2012). Rates of testing for HIV in the presence of serodiscordant UAI among HIV-negative gay men in committed relationships. AIDS and Behavior(7): 1944-1948.

   http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3429654/

29. Sullivan PS, Wall KM, O'Hara B, Jones J, Barnes J, DiClemente R, Hoff C, Scales L, Salazar LF, Sanchez T, White D, Wingood G, Allen S, Stephenson R. The prevalence of undiagnosed HIV serodiscordance among male couples presenting for HIV testing. Arch Sex Behav. 2014 Jan;43(1):173-80

30. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1–infected adults and adolescents. Dept Health Hum Serv. 2012;1–166. Available at:

http://www.aidsinfo.nih.gov/contentfiles/lvguidelines/adultandadolescentgl.pdf.

Accessed:

Panel 1: Description of Testing Interventions

Increase the frequency of testing

Our modeling of HIV testing frequency is fully described in the Supplemental Appendix. At Baseline, on average MSM test approximately annually, but some test more or less frequently (9.3% test every 90 days, 14.3% test only once every 3 years). We modified this so that all men actually test annually, and also more (every 180 days, every 90 days, every 5 days) or less (every 3 years) frequently.

Improve the proportion that ever test

To consider the effects of a strategy which would get more MSM to test at least once (such as a social media campaign, routine testing in a hospital or other clinical setting, or other outreach to the MSM population that is not currently testing), we increased the proportion of the population that ever tests for HIV from the baseline value of 80% to 93% and 99%.

Changing the type of test used for HIV screening

We implemented a test sensitivity parameter by varying the number of days after infection when a test, if conducted on the infected individual, would return a (correct) HIV-positive result. We used 22, 45 and 70 days to correspond to estimates of sensitivity for the newest lab tests, standard blood tests, and oral

fluid rapid tests that are also available for home use. We also include a

hypothetical "perfect" HIV test which can detect infection the day it occurs.

Improve HIV treatment access and viral suppression

In the US there is a large proportion of the HIV-infected population that is aware

of their infection but not receiving effective therapy to suppress their virus and

thereby reduce the risk of transmission.[17-19] However, although the national

average in the US is low (43.4%), there are some cities with higher proportions of

the population achieving viral suppression, as high as the 68.5% of all HIV-

infected persons estimated to be in care for HIV[17], and guidelines for HIV

treatment in the US recommend universal treatment for all HIV-infected

persons.[30] Thus we varied the proportion of the population suppressed, setting

this parameter to 43.4%, 68.5% and 100% to assess the effect of increased

testing frequency in situations where a higher proportion of the population

follows their diagnosis with effective treatment.

Panel 2: Research in context

Evidence before this study

Although the US Centers for Disease Control and prevention currently recommends that Men who have sex with Men (MSM) test for HIV at least annually[6], recent reports suggest that most men are not testing for HIV this frequently[7-8] and how many MSM have ever tested, how frequently they test, and how frequently they should test is still being debated.[9-15] In October 2014, we searched PubMed, Medline, and Embase for studies that included the terms , "HIV Seropositivity", "HIV Infections ", "AIDS Serodiagnosis " , "Men who have sex with men" and "MSM", "high risk", "Test" , and/or "screen." We limited the results to studies published after January 2005, and limited the review to studies conducted in the United States, Europe and Australia, and reported in English, with an endpoint of incident HIV infection that compared annual testing with more frequent testing among MSM. We identified 14 studies that compared the HIV incidence in populations with different HIV testing frequencies, 12 of 14 studies were mathematical models with different assumptions about HIV transmission in the MSM population. Some models suggested trends toward a reduction in HIV incidence when MSM test more frequently[10-12], and at least two concluded that more frequent testing of MSM in the United States could be cost-effective.[11-12] However, those models also suggested that the costs associated with more frequent testing would be substantial (>$8billion dollars per year[12],

nearly $5 billion annually to increase from annual testing to testing every 180 days[11]).

Added value of this study

Unique aspects of our modeling approach include: acknowledging that an unknown and perhaps large segment of the US MSM population is not currently testing for HIV at all, and that the duration of sexual partnerships varies by the perceived HIV infection status of the MSM partners. While other studies[10,13,15] have alluded to the potential for synergistic effects when considering the level of HIV treatment coverage and viral suppression in the population and any strategy aimed at increasing the uptake of testing (for example increasing testing frequency among those MSM already testing or increasing the percentage that ever test), we believe this is the first study to describe how this synergy impacts the effects of these interventions on the incidence of HIV infection for US MSM. We found that, under current conditions for HIV therapy in the US, an intervention that increased the frequency with which MSM test to more than the annual testing currently recommended by CDC did little to affect overall incidence of HIV, many more HIV tests having been performed. Increasing the proportion of the infected population that successfully achieves HIV viral suppression was the only intervention that, on its own, was able to significantly reduce incidence compared to the baseline scenario. However, if those who are aware of their infection achieve viral suppression at much higher rates than

currently observed in the US, increasing the testing frequency of MSM and using more sensitive HIV tests to conduct this testing may be able to impact HIV incidence. In an optimal scenario where 100% of those who are aware of their infection achieve viral suppression, HIV transmissions could be reduced by an additional 17% if MSM tested every 90 days instead of testing annually.

Implications of all available evidence

It appears that, for MSM that test at all, testing more than annually will have little impact on HIV incidence until the proportion of MSM with diagnosed infection who go on to achieve viral suppression can be improved. The limited benefits of more frequent testing occur secondarily to the reduction in incidence that arises from viral suppression of the diagnosed population. However, if we could combine improvements in the proportion of the diagnosed population who achieve and sustain viral suppression, with increasing frequency of HIV testing in the at-risk but undiagnosed population we could reduce HIV incidence more than what was observed as a results of each intervention individually.

Table: Median and Inter-quartile range of incidence observed in 20 simulations of 3 years of follow-up in a simulated population of 5250 MSM under Baseline and 12 alternative scenarios for HIV testing interventions in the United States.

Model Scenario	Value	New infections/ 100 person years	
		Median	IQR
Baseline*		2·87	2·76–3·06
Increasing test frequency (Distribution's for each scenario reported in Supplemental Table S3)	Test every		
	5 days	2·63	2·32–2·75
	90 days	2·57	2·34–2·78
	180 days	2·80	2·50–2·92
	365 days	2·98	2·64–3·07
	1095 days	3·27	3·10–3·57
Reduce the proportion of MSM who have never tested for HIV (Baseline 21%)	7%	2·67	2·58–2·86
	0.2%	2·60	2·44–2·86
Increasing viral suppression through linkage to care and treatment (Baseline 43%)	68.5%	2·19	2·03–2·47
	100%	1·43	1·32–1·57
Using tests with different sensitivity for early infection (Baseline 45 days)	70 days	2·89	2·83–3·12
	22 days	2·68	2·53–3·02
	0 days	2·84	2·64–3·17

*Baseline testing frequency is assigned based on data collected from an online survey; under this scenario MSM test almost annually on average, but many test more (9.7% test every 90 days) and less (14.7% test once every 3 years). A complete description of the baseline testing frequency distribution and how testing was implemented is included in the supplemental appendix. Baseline values for other testing intervention parameters include 21% of MSM never testing, 43% of the population with diagnosed infection achieving viral suppression, and a test capable of detecting infection 45 days after it occurs.

Figure 1: Population level impact of increasing testing frequency on both total tests performed and 3-year HIV incidence in a simulated population of 5250 men who have sex with men (MSM) in the United States

Scenario	Baseline	Test every 3 years	Test every 365 days	Test every 180 days	Test every 90 days	Test weekly
Median number of HIV tests performed	11,279	3,457	9,879	18,880	32,537	88,002
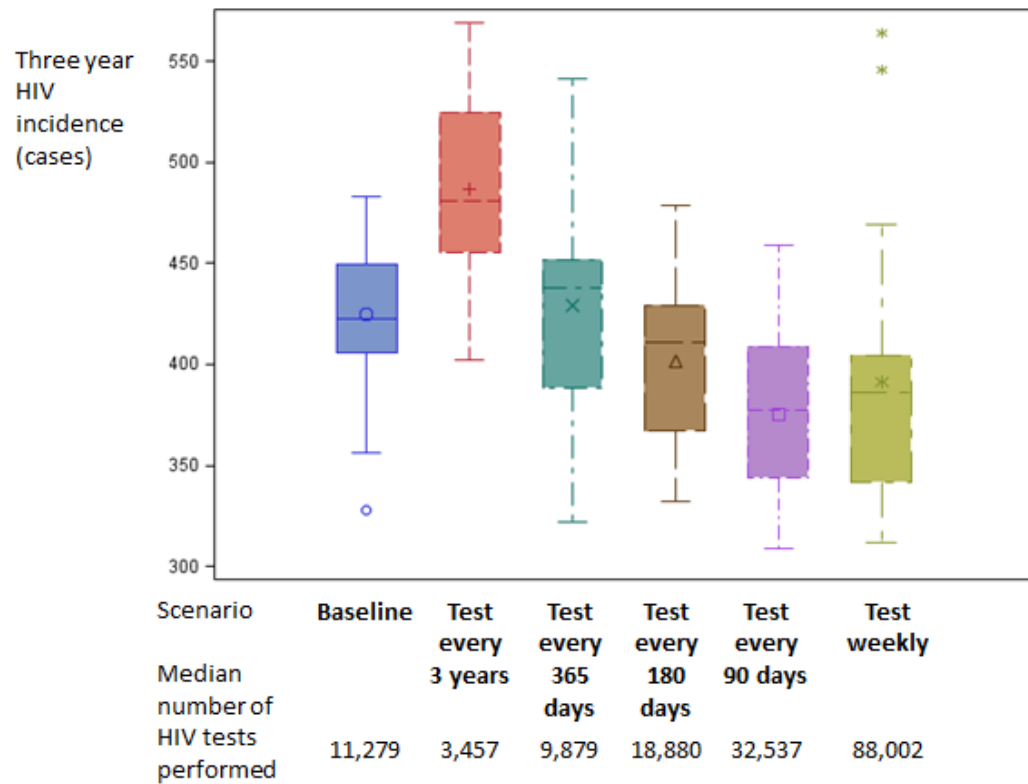
Figure 2: Individual level impact of increasing testing frequency on time to HIV diagnosis over 20 3-year simulations of a population of 5250 men who have sex with men (MSM) in the United States
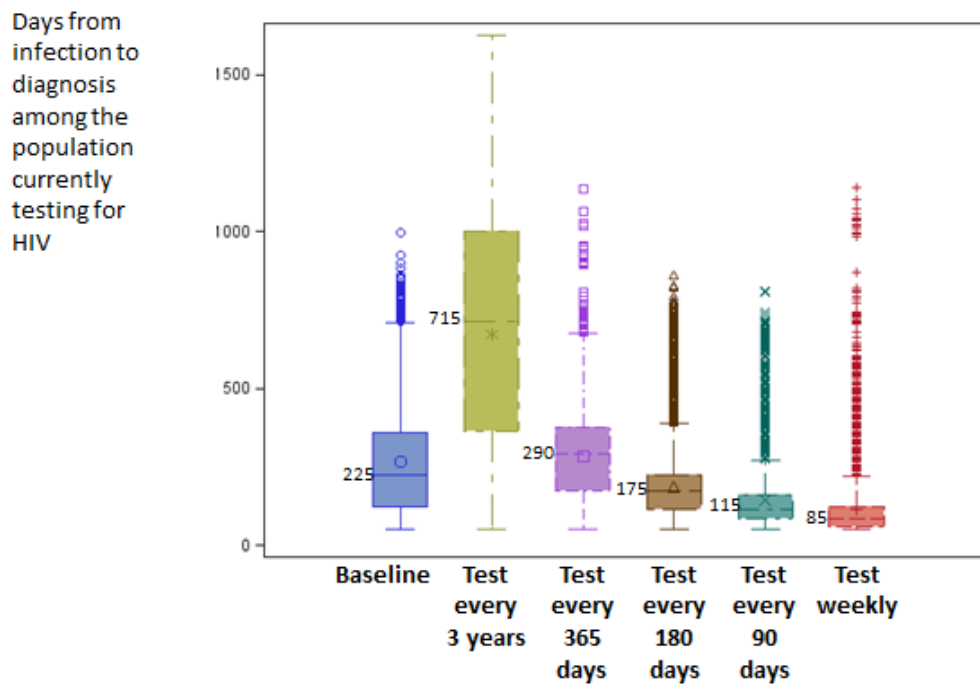
Figure 3: Synergistic effects of HIV testing frequency, test sensitivity and viral suppression among the diagnosed population on the median 3-year total of circulating viral load, by HIV testing and diagnosis group of MSM in a simulated population
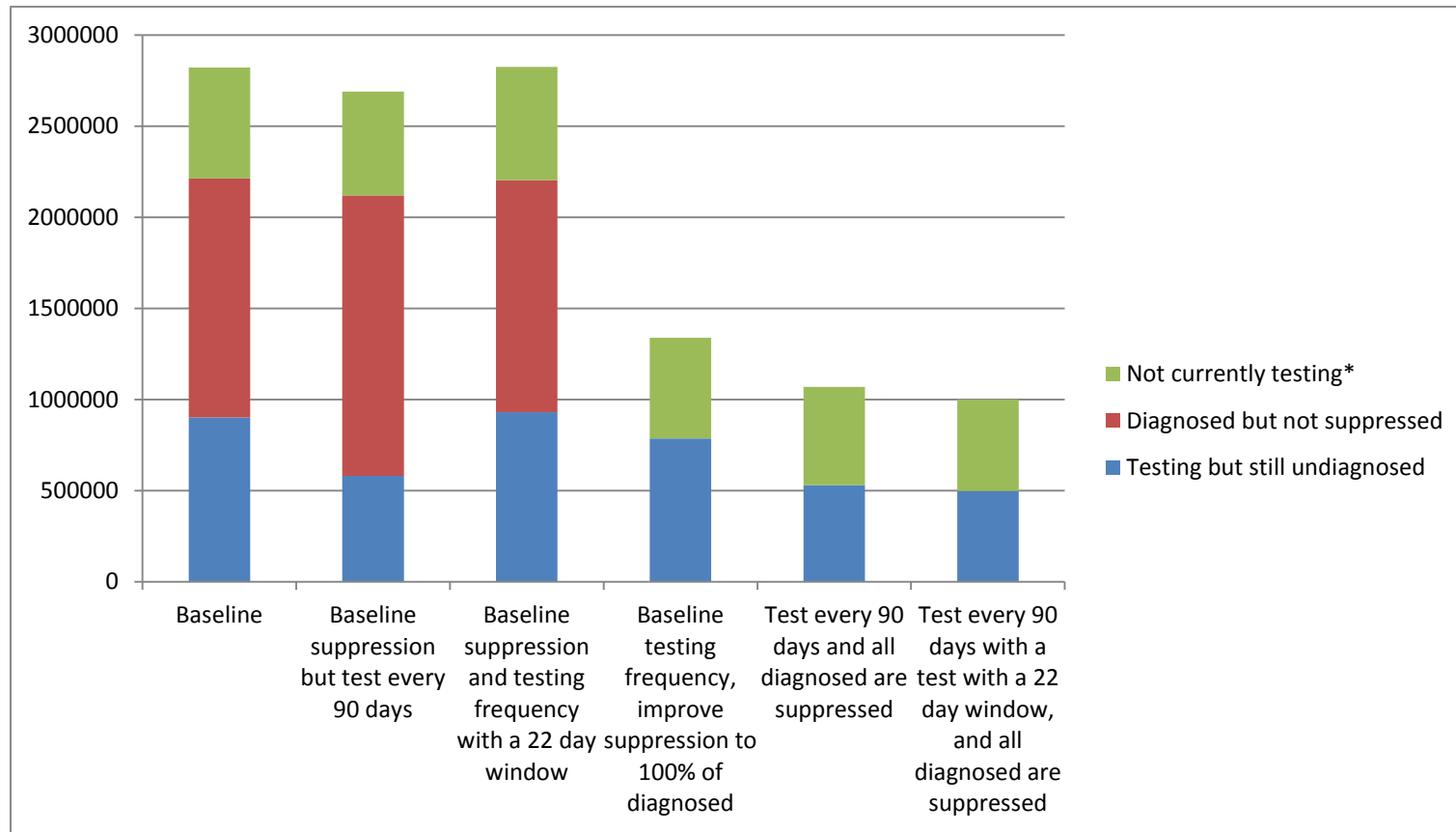
Figure 4: Synergistic effects of HIV testing frequency, test sensitivity and viral suppression among the diagnosed population on 3-year HIV incidence across 20 simulations of a population of 5250 men who have sex with men (MSM) in the United States
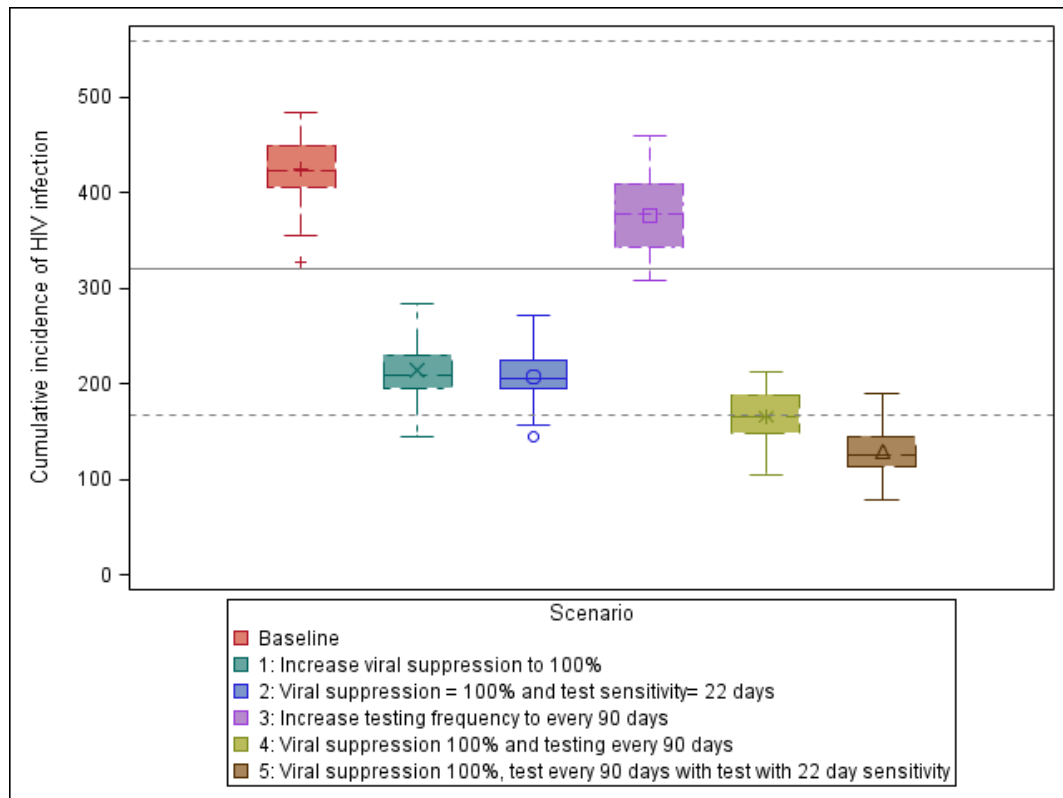
Figure 5: Breakdown of the synergistic effects on the median of 3-year HIV incidence estimated from 20 simulations of a population of men who have sex with men (MSM) in the United States in which MSM test every 90 days with a test that detects HIV 22 days after infection and all diagnosed MSM suppress their HIV virus through HIV care and treatment.
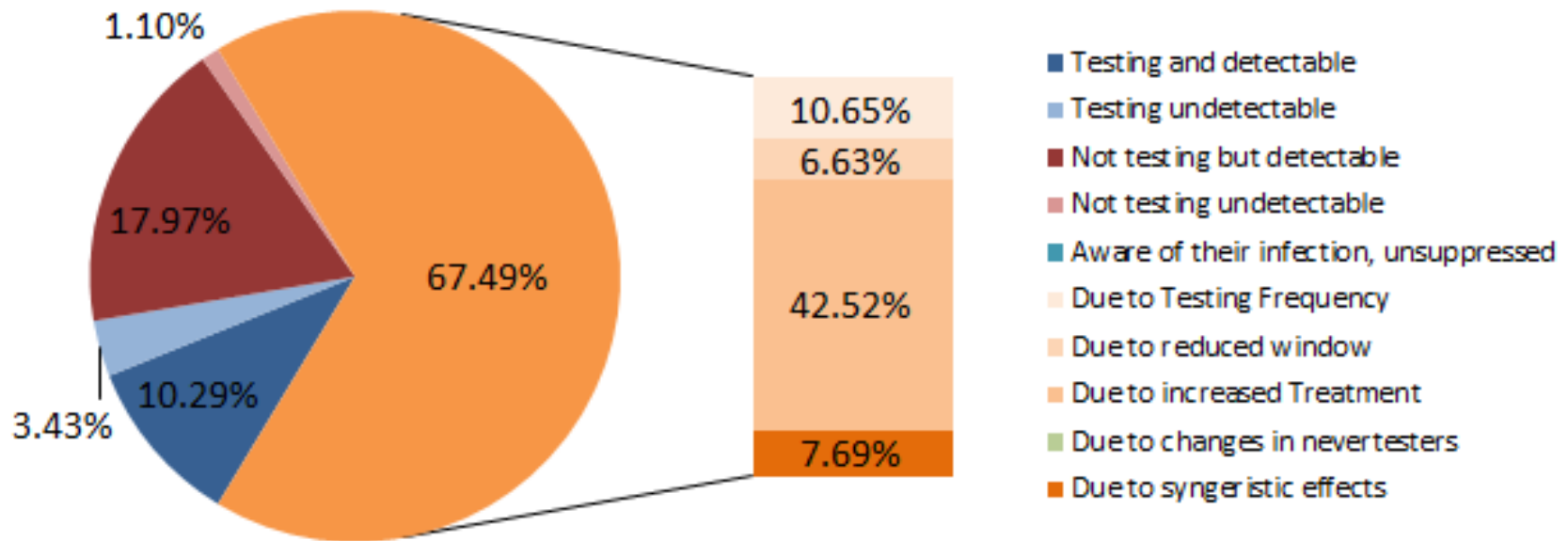
Supplemental Materials for Aim 3:

In the current work we developed a separable temporal exponential random graph

model[1] (STERGM) for sexual partnership formation. STERGMs are the dynamic,

stochastic extension of cross-sectional exponential random graph models (ERGMs) and

as such allow for simulation of stochastically evolving sexual partnership networks[2,3].

These models accommodate statistical dependence among partnerships, and model

parameters describe factors associated with both partnership formation and dissolution

over time. Dissolution of partnerships is approximated using the method derived by

Carnegie et al[4], and dynamic extensions to the R *statnet* packages[5-7]. Once the

partnership network statistics are estimated we use customized extensions to the R

package EpiModel[8] to control all of the testing intervention parameters within our

dynamic stochastic model of HIV transmission. Because the time horizon of interest in

changes in HIV incidence is limited to 3 years, we do not model either entrance to or

exit from the study population of 5250 MSM.

**Parameterization of the separable temporal random graph model for partnership
formation and other aspects of model for transmission dynamics.**

As far as we know our model is unique to the literature in that partnership formation

probabilities differ based on the perceived HIV serostatus of the two men interested in

forming a partnership. This assumption is based on data collected in an online survey[9]

which we have used to parameterize the following STERGM model.

STERGM Model parameters

In defining the model, $y_{i,j,t}$ is a variable describing whether MSM i and j are in a sexual partnership at time t taking the value 1 when they are and 0 otherwise. The variable $Y^e_{i,j,t}$ describes the rest of the network (i.e. all other sexual partnerships among the 5250 total men in our simulation, EXCLUDING the partnership information for MSM i and j.

Partnership formation is described by:

$Logit(P(y_{i,j,t}=1|y_{i,j,t-1}=0, Y^e_{i,j,t})) =$

β_0*edges +

β_1*concordant_Unknown_status + β_2*concordant_both negative +

β_3*concordant_both_positive + β_4*Degree0_unknown_status + β_5*Degree0_negative +

β_6*Degree0_positive + β_7*concurrent_unknown_status +

β_8*concurrent_negative + β_9*concurrent_positive

This form of the model is written such that the log odds of a partnership forming between MSM i and MSM j at time t, given both that i and j were not in a partnership together at the last time step and the composition of the rest of the network under study, is a function of 10 parameters. The terms included in our model are: a) the total number of partnerships in the network (edges), b) the selective mixing by perceived HIV status, so that the number of ties between individuals of the same perceived HIV status is greater than would be expected by chance, and different for each perceived status, c) the number of individuals of each serostatus who do not currently have a partner (degree 0) or d) who have 2 or more partners (concurrent partnerships).

These network statistics are calculated for each of three groups of perceived serostatus: unknown serostatus for those never tested; negative serostatus for those whose most recent test result was reported as negative; and positive for those who have received a positive HIV test result. For those with either an unknown or negative perceived status, a proportion of them are actually infected with HIV, they just don't know it. Note that, as in traditional multivariable generalized linear models (GLMs), the edges term (which describes the total number of partnerships not defined by any other parameter, and is equivalent to an intercept term in a GLM) in this captures the log odds of forming a partnership for those with discordant perceived HIV status.

Parameters for the model are averaged over data reported for both main and causal partnerships in the survey that provides the source of these data. Supplemental table 1 provides network characteristics that were used to parameterize the initial model. There are subtle differences in both the number of partners (expressed as mean degree in Table S1) by serostatus awareness, but the most interesting finding in terms of transmission is that HIV-positive participants reported more partners overall and more perceived discordant partnerships than HIV-negative participants. However, in our survey HIV-infected participants were also more likely to form partnerships with other HIV-infected participants (31% of all partnerships were between 2 HIV-infected men, compared with 5% that would have been expected to be observed by chance); as a result many of the sex partners of HIV-infected men are not at risk of acquiring HIV infection.

Table S1 Characteristics of individual respondents collected from an internet-based survey of men-who have sex with men, used to parameterize our model

	Perceived HIV status		
	Don't know	Negative	Positive
Degree	0.88	1.11	1.14
Homophily	0.41	0.70	0.31
Expected Homophily	0.32	0.6	0.05

Baseline prevalence and serostatus awareness

The population includes 5250 men who have sex with men, and we simulate both network formation and dissolution and HIV transmission dynamics over a three year period. Baseline perceived HIV serostatus is based on the information collected in our survey sample[9], while prevalence of diagnosed and undiagnosed infection are set to be consistent with data from most sites participating in CDC surveillance of MSM[10-12] and population based surveys of this risk group[13-15].

Table S2: Distribution of HIV infection and perceived serostatus at simulation initiation

Network model characteristics			N (%)
Sample size			5250
HIV Status	Truth	Perceived	
	HIV-infected	HIV-Positive	350
		HIV-negative	105
		Unknown	140
	HIV-Uninfected	HIV-negative	3395
		Unknown	1260

At the start of our simulation, 585 of the 5250 men (11%) are infected with HIV, with

350 (60%) of these aware of their infection. A total of 3500 believe they are uninfected

(e.g. based on a test result in the past 4 years), including 105 (3%) who are actually

infected. The remaining 1400 (26%), including 140 (24% of all HIV-infected individuals)

who are infected, are considered unaware of their HIV status.

The model is incremented in 5 day time steps for efficiency, based on two assumptions

that should make this simplification valid:

1) The average partnership duration is an order of magnitude larger than this time

increment

2) MSM have sex on average once in this time period[16]

We model HIV transmission as occurring through anal intercourse acts without a

condom within an ongoing partnership, with an allowance for additional one-time acts

with other HIV-infected individuals occurring outside this partnership as a probability of

a one-time act per five days as described below. We consider only an average rate of

transmission per sex act in which no condom was used, and do not attempt to account

for differences in risk associated with insertive, receptive or both types of anal

intercourse occurring within a given sexual encounter. Similar to a model considered by

Goodreau[2], we model the risk of transmission as a daily per act probability of

transmission that changes over the infected lifetime of each individual based on

changes in their HIV viral load. For those aware that they are infected with HIV, we also

include a parameter for the overall probability of accessing HIV anti-retroviral therapy

and achieving full suppression of detectable HIV virus.

Entry into and exit from the population

Although migration into and out of a given community of MSM undoubtedly occurs, the

majority of entries into a given population of MSM are through aging into sexual debut

and exiting due to death either from HIV or other causes. Others[2,3,17] have used rates of

birth and death based on US population dynamics, but these effects should be minimal

in our 3 year time scale. However, by not allowing for births and deaths our population

will not settle into equilibrium values for HIV prevalence, as no one with infection will

die and be removed from the simulation. Thus we actually capture the short term

dynamics in a closed system, rather than long-range dynamics of a system in

equilibrium. It should be pointed out that both incidence and prevalence are increasing

rapidly in some subsets of the MSM population in the US, and insights from a short term

evaluation of the impact of an intervention on a system that is not at equilibrium[18-19]

may help to guide expectations for short term outcomes of interventions better than

estimating effects on longer timescales[17]. In future work we will expand the model to

include entry to and exit from the system to see how our interventions might affect HIV

incidence once the system achieves equilibrium.

**Parameters for partnership durations, and one-time partnerships and comparison of
our approach to others in the literature**

After partnerships are formed based on the STERGM model described above, they

persist for an average duration based on serostatus of the men in each partnership, with

durations based on information collected in our surveys. Recently other research

groups have reported STERGM models that capture the nature of overlapping

partnerships[2,3,17], but they model only main partnerships, considering all casual

partnerships to be one-time occurrences, which our data suggest is not the case. Others

have chosen to model partnership formation within agent or individual based models,

capturing main, casual and one-time partnerships as a function of partner

characteristics, but without considering the overall structure of the sexual network in

their model[20-22]. Still others have used deterministic models with groups of high and

lower risk MSM, but these capture neither the overlap we have found to be common in

the sexual partnerships of MSM (both men with perceived HIV-negative and perceived

HIV-positive status have >1 partnership ongoing at the same time on average, Table S1),

nor the stochastic nature of HIV transmission[23-25]. In this analysis, we take a different

approach than others because we have found that many non-main partnerships are

recurring, if not ongoing, relationships, so we do not treat them as one-time events.

Instead we capture the duration of all partnerships and use these data to describe the

duration of partnerships using differing rates of partnership dissolution by perceived

serostatus. Thus, our partnership durations tend to be shorter on average than those

reported in other similar work that employs STERGM models[2,17].

Table S2 shows the distribution of partnership durations used in our initial

parameterization of the model, in which we average partnership duration over main and

casual partnership types, for: serodiscordant partnerships, partnerships where the

study participant reported that both he and the partner did not know their HIV status,

partnerships where both were reported as negative and partnerships where both were

reported as HIV-infected.

Table S3: Mean duration of partnerships used to parameterize the dissolution model

Model 1	Perceived HIV status of participant and partner			
	Discordant*	Neither knows their serostatus	Both Negative	Both Positive
Partnership duration				
Exponential Mean	369	254	436	570
Weibull Mean	332	217	465	646

*This category combines survey respondents who reported being HIV-negative with an HIV-positive partner, and those who reported being HIV-positive with an HIV-negative partner. Reported partnership durations were similar for these two groups

Persons in concordant HIV-positive partnerships have the longest duration, those in

perceived concordant negative partnerships are nearly ½ year shorter on average, and

those for persons with discordant and unknown partners are the shortest, with average

partnership lengths less than 1 year. In[2,3,17] average partnership duration was 1248 days,

or between 2 and 6 times longer than the durations used in our model. As a sensitivity

analysis to attempt to quantify the importance of partnership duration to our main

findings about the effects of testing interventions on HIV incidence, we reran the

simulations with partnership durations set to one quarter of the lengths reported in

Table S2.

Discussion of the importance of partnership duration as a model assumption

When we decreased the average duration of all partnerships modeled within our STERGM to 1/4 of median length of the duration observed in our data, incidence increased to a median of 3.84 (IQR: 3.61-4.44) per 100 person years under the assumptions of the baseline model related to testing interventions. Under this scenario in which partnerships were on average much shorter, increasing testing frequency from the baseline distribution to once every 90 days still had little effect (median 3.87/100 person years) on 3 year HIV incidence.

Thus, we found that partnership duration can impact overall HIV incidence, suggesting that structural interventions aimed at reducing the number of partners and encouraging stable partnerships could be important areas for future research[26-28]. The drastic differences in mean partnership duration that we found, with short partnership durations for serodiscordant men and the population that did not know their serostatus, lead to more frequent opportunities for the infected men in these groups to expose new partners to the virus. As has been reported by others[2,17,19] the mean degree of HIV-infected men who were aware of their infection is higher than that of HIV-negative men, this is in part the reason they are now infected. However, in our data, infected men who are aware of their infection were observed to be more likely to be in a relationship with other infected men, and for those relationships to be more stable than other combinations of HIV serostatus. Among those who know they are infected, the average durations for relationships with men who are of negative or unknown status was 332 days compared to 646 days for relationships with other HIV-infected men. The lack of

effect of increasing testing frequency to more than annually is partially due to the average relationship lasting longer than one year.

Probability of having a one-time partner

Because we attempt to capture the duration of casual partnerships through shorter average partnership duration, the only portion of our sample of partnership distributions that remains poorly explained is the number of partnerships that are one-time events. In our data these account for 40% of all reported anal sex partners in the last 6 months. The proportion of all partnerships reported as one-time was substantially lower when participants reported knowing both their HIV status and that of their partner (35%) than when reporting unknown serostatus for themselves and or their partners (45%), suggesting the anonymous nature of a one-time sexual encounter precluded discussion of HIV serostatus. However, from the survey we were unable to ascertain the proportion of all one-time partnerships that were relevant for transmission, namely those in which one partner is HIV-infected, the other is not, and a condom is not used for the entire anal sex encounter. To keep the parameterization of these episodes simple, but still attempt to account for them in our model, we added an additional unprotected anal sex act in each simulated 5-day time step for serodiscordant partnerships with probability based on an exponential distribution. The probabilities were based those above; when one or both partners did not know their serostatus the 5-day probability of having an additional one-time act in addition to all ongoing

relationships was 0.03, when both partners perceived themselves to be either HIV-

negative or HIV-infected the 5-day probability of a one-time act was set to 0.014.

Evaluation of the underlying rate of HIV testing and its impact on our transmission models

The main goal of this analysis is to assess the impact of increasing average testing

frequency from the baseline scenario, which results in approximately annual testing, to

every 6 months and every 3 months. Changes in testing frequency result in more

frequent updating of the percieved serostatus of the men in our simulated population,

providing more accurate serostatus information leading to changes in partnership

formation based on this perceived serostatus over time.

We used data from our online survey of MSM[9] to define the baseline probability of HIV

testing.

Table S4: Testing frequency at Baseline and under various intervention scenarios that increase the frequency of testing

Testing Frequency	Baseline N	Test Every 3 years N	Annual testing N	Twice annual testing N	Every 90 days N	Every 5 days N
Never	1400	1400	1400	1400	1400	1400
Every 3 years	700	3850				
Every 2 years	1050					
Annually	1050		3850			
Twice Annually	595			3850		
Every 90 days or less	455				3850	
Every 5 days or less						3850
Median number of tests	2.93	0.89	2.57	4.9	8.45	22.86
Expected Number of tests		1	3	6	12	219

If testing occurred at a constant rate (which we know is not true) and there were an infinite amount of resources to conduct testing daily (also not true) testing 0.01425 percent of the population every 5 days should achieve a population mean of testing once every 351 days, and 64% of the population should be tested annually. However, we also wanted to control the inter-test interval for men, such that once they test they will not consider themselves eligible to retest for 365, 180 or 90 days after the date of their most recent test. We choose to increase the daily probability of testing to 0.1308; at this level we would expect 99.996% of the population to test annually (every 365.25 days), and 92% of the population to be able to test every 90 days. We then vary a

parameter for when MSM would seek retesting from 1095, to 365, 180, 90 and finally to every 5 days after the most recent test. At this rate, the average number of tests recorded by those MSM testing in the simulation in each scenario was 2.93 at baseline, and increased from 0.89 in our scenario where men test every 3 years to 22.86 tests when men test every 5 days. While we would have expected 1 and 219 tests if we performed testing deterministically on every individual in the simulations of three year and 5 day testing frequencies, the actual testing frequencies were somewhat lower than this (See Table S3) and become more divergent from deterministic testing as testing becomes more frequent. Although we don't achieve the number of annual tests that would be expected if testing occurred deterministically, in our model we do a) observe a distribution of simulated testing frequency that corresponds roughly with what we observed in an online survey and b) are able to dramatically shift this distribution by instituting interventions that increase the frequency of testing.

Modeling the impact of men who never test on HIV incidence

Another difference between our STERGM model and others that have examined this question is that we have explicitly included a proportion of the MSM population that never tests for HIV, and can manipulate this proportion along with other aspects of testing interventions. Khanna et al[3,17] attempted to account for the large proportion of MSM who never test for HIV through a scenario in which MSM test only once every 10 years on average, and found much higher incidence in this scenario compared to their baseline scenario where MSM test annually. In their model based on data from MSM in

Australia[21], Gray found that an intervention targeting the estimated 5% of MSM that have never tested and getting them to test annually would have a modest impact on HIV incidence. In our baseline model, 20% of the population has never been tested for HIV, higher than observed in venue-based samples of MSM[10-12] but consistent with or lower than population based estimates[13-15]. To consider the effects of a strategy which would get more MSM to test at least once (such as a social media campaign, routine testing in a hospital or other clinical setting, or other outreach to the MSM population that is not currently testing, we increased the proportion of the population that ever tests for HIV from the baseline value of 80% to 93% and 99%. We found (See Table in the main text) that, at baseline levels of testing frequency, test sensitivity, and treatment coverage, increasing the proportion of this group that received at least one HIV test had a small but linear relationship with HIV incidence.

Figure S1 shows that, when combined with a test capable of detecting infection within 22 days and 100% viral suppression for those whose infection is diagnosed, synergistic effects similar to that shown for increasing the frequency of testing to once every 90 days can be achieved by reducing the proportion who of MSM who never test from 20% to 7%. If social media or other campaigns could be targeted to this largely undescribed population, such a campaign would require a 20% increase in annual testing compared to baseline, and thus might be a more efficient strategy than one that would require 4 times as much testing as is currently performed on MSM in the US.

Figure S1: Synergistic effects of accessing more of the population for testing while using a highly sensitive test and achieving 100% Viral suppression among the population with diagnosed infection

Proportion of all observed HIV transmissions over three years of simulation, including (Orange section) transmissions averted due to changes in HIV testing intervention parameters compared to baseline

Proportion of averted transmissions (relative to baseline model) attributable to changes in HIV testing

0.24%
5.09%
3.91%
22.84%
68.05%

6.63%
42.52%
6.98%
11.91%

- Testing and detectable
- Testing undetectable
- Not testing but detectable
- Not testing undetectable
- Aware of their infection, unsuppressed
- Due to Testing Frequency *
- Due to reduced window
- Due to increased Treatment
- Due to changes in nevertesters
- Due to syngeristic effects

*In this model HIV testing frequency was held constant at the baseline distribution observed in our online survey and thus does not contribute to infections averted. This can be compared to the effects reported in Figure 5 in the main text in which the proportion

of MSM that never test was held constant while testing frequency, treatment coverage for the diagnosed population, and HIV test sensitivity were varied.

Future research is needed to better describe this group of MSM that never test for HIV despite recommendations to test at least annually, and to design interventions that would motivate them to join the population of MSM that do seek out testing for HIV.

Calculating the total circulating HIV viral load over time

We also calculate and report the total circulating viral load, by diagnosis status of the men in the model. Similar to Goodreau and Khanna[2,3] we model viral load over time using a function with 5 parameters for $\log_{10}$ viral load:

(a) Days 0-21: rises linearly from 0 to 6.886

(b) Days 21-100 declines linearly from 6.886 to 4.5

(c) Days 100-3370: assumes a set point of 4.5 that lasts until the onset of AIDS approximately 9 years post-infection [17]

(d) Days 3371-3650: linear rise from 4.5 to 7.0

(e) Day 3650: death

Figure S2: Distribution of log$_{10}$ HIV-1 viral load over time of infection
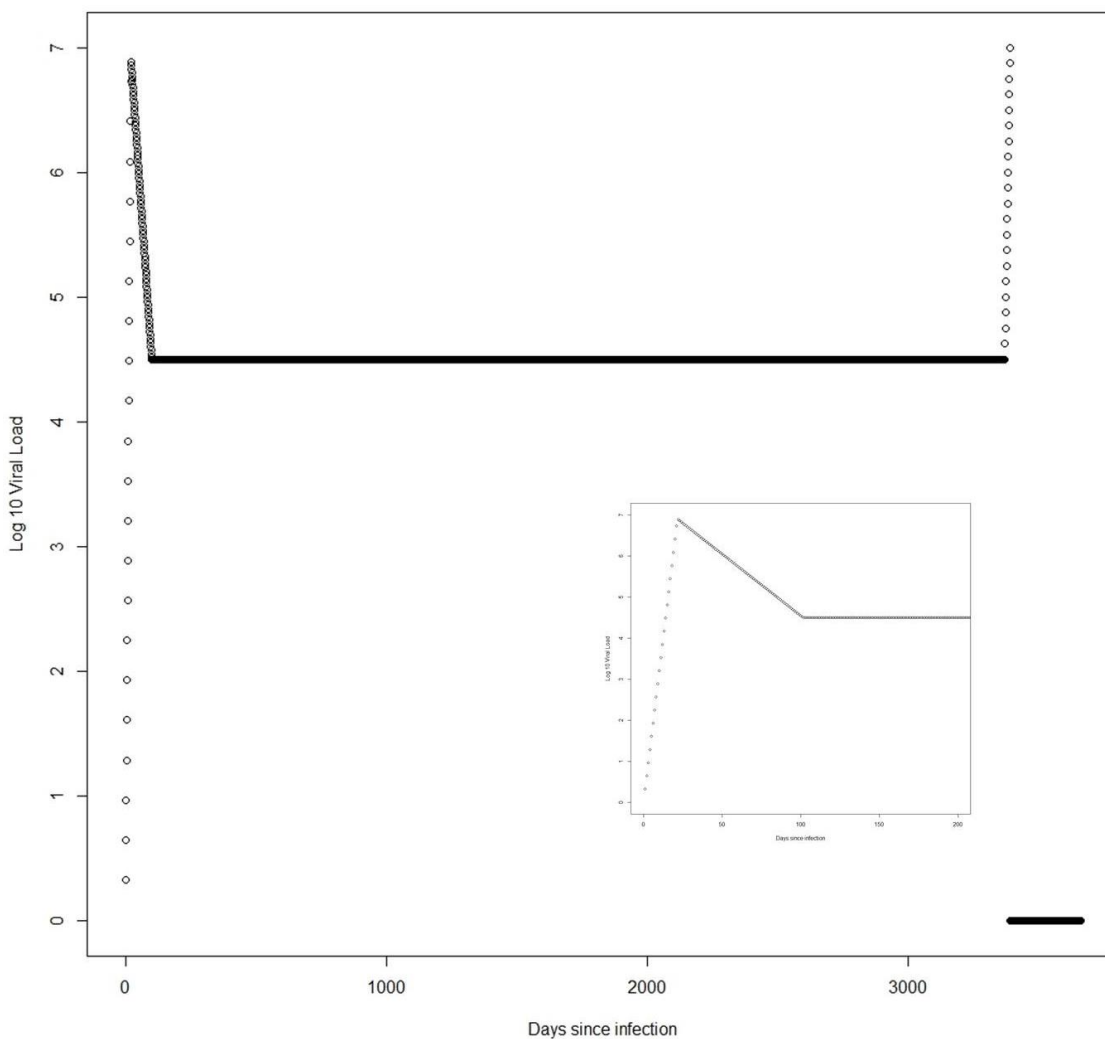
This is a simplistic function (Figure S2) of viral load over time, which we can then

integrate for each individual after they become infected, summing their viral load for all

observed person time for each infected man in the simulation. To illustrate the impact

(or lack thereof) of diagnosis alone on circulating viral load we can then divide each

man's "Cumulative Viral load" by other factors, including their perceived HIV status. For

Figure S3, we categorize perceived HIV status as diagnosed as HIV-infected, perceived to

be HIV-negative even though they are infected and currently testing for HIV, or currently

infected but with unknown serostatus for the group of men who do not seek HIV testing. For those who are diagnosed and access treatment we assume their viral load will be reduced to an undetectable level, and that at that point it does not contribute to the calculation of the total circulating viral load. We calculated the median of the total viral load in each category (diagnosed HIV-infection, undiagnosed but not testing, undiagnosed and currently testing for HIV) across 20 simulations for each of our model scenarios. Then, we compare the resulting proportions of circulating viral load over time in each category across model scenarios with different testing frequency distributions for the MSM who test.

In Figure S3 the rows represent the distribution of the: 1) Diagnosis status awareness of infected MSM, 2) the total circulating viral load and 3) the total number of HIV transmissions. The columns are different scenarios for HIV testing frequency. In the baseline scenario for column 1, on average MSM test approximately annually, but some test more or less frequently (9.3% test every 90 days, 14.3% test only once every 3 years). We modified this so that all men test closer to every 6 months, every 90 days, every 5 days (weekly), with the results of these scenarios presented in columns 2-4 respectively.

In the first row, light blue is the proportion of all infections undiagnosed at the end of follow-up, light purple represents the proportion undiagnosed at the start of the simulation but diagnosed by the end(i.e. over 3 years of follow-up), dark purple is the proportion diagnosed at the start and red the proportion who never test. As testing

frequency increases the proportion of all infections that remain undiagnosed decreases, but this is the smallest grouping of HIV-infected menat baseline.

In the second row the color coding is retained, with light blue corresponding to the proportion of the total viral load contributed by men while infected but undiagnosed (including during acute infection), purple is the proportion of the total circulating viral load contributed by those with diagnosed infection

Figure S3: Impact of testing frequency on transmission dynamics in a population where only 43% of those with diagnosed HIV infection achieve viral suppression
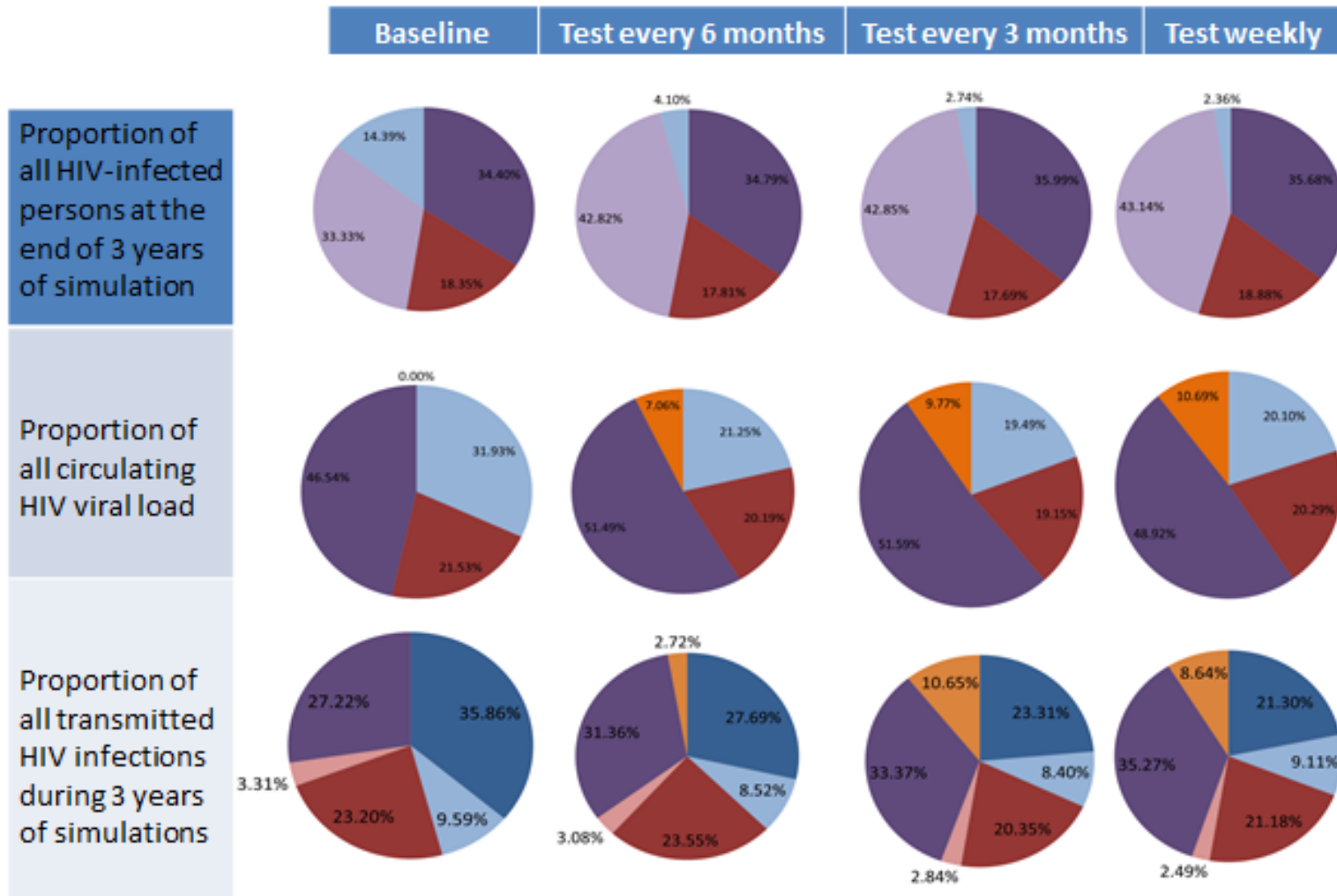
(under conditions where only 43.4% of the diagnosed population has their virus suppressed by therapy) and red corresponds to the viral load contributed by the MSM that never test. Columns 2-4 have orange sections corresponding to the percent of circulating virus removed compared to baseline, in this case through increases in testing frequency.

In the third row, information about the effect of test sensitivity on diagnosis category has been added. In this row, dark blue represents the proportion of transmissions that arise from the population that is currently infected and testing but undiagnosed, including those testing false-negative based on the sensitivity of the test, and those testing infrequently such that they have been infected since their last test. The light blue is the transmissions in this group that occur within the first 22 days of infection and are currently deemed "unstoppable" through testing alone. Likewise the dark red shading represents transmissions arising from the population that never tests, and the pink corresponds to those transmissions occurring within the first 22 days for this subset. The purple again represents the proportion of transmissions from the diagnosed population. The orange section now represents the proportion of transmissions removed compared to baseline, i.e. present in the baseline scenario but which did not occur in the counterfactual scenarios where testing frequency increased.. For all three rows the data presented represent the changes in median values across testing scenarios, the interquartile ranges for these estimates overlapped for all scenarios at baseline values of treatment coverage.

Taken together Figure S3 shows that at baseline only 14% of the HIV-infected

population is "up for grabs" from a testing frequency intervention's perspective. While

this group does represent a disproportionate amount of the circulating viral load (row

2), testing alone only moves them to the diagnosed group and has a small impact on

circulating viral load overall. The impact on onward transmission is therefore also small,

with the confidence interval around the 10.65% of infections averted when comparing

the scenario where MSM test every 90 days to baseline testing frequency including zero

(-8.6%,26.8%). The non-significance is also shown in a lack of a trend for increasing

testing frequency and percent of infections averted, with testing weekly having slightly

fewer infections averted on average.

Figure 3 in the main text shows that the median values of total circulating viral load

decreased substantially when we parameterize the model such that the entire

diagnosed population achieves viral suppression, and in Figure 4 in the main text this is

also shown to lead to reductions in 3 year HIV incidence.

Stochastic variation in models and implications for our main outcomes
Despite the computational intensity of building these models (each simulation takes on

average 2 hours to run) we report 340 total simulations with large variability in

incidence within a given model. In another recently published individual-based

stochastic model, of HIV transmission among South African MSM, Brookmeyer et. al.[22]

varied the number of simulations based on a pre-defined standard error threshold of

0.01 (1%) for the mean of the proportion of the population infected (prevalence) over 5

years after each replication using all replications performed up to that point. If the

standard error was above 0.01 they proceeded and performed an additional replication,

stopping when the standard error fell below this threshold. The mean number of

replications performed for their 163 distinct combinations of 4 interventions was 13

with a minimum of at least 5 replications performed for each combination of

interventions. This should be contrasted with deterministic models that have examined

the impact of testing frequency on HIV incidence, which by definition only have one

mathematical solution and therefore can only observe variations in outcome measures

through sensitivity analyses of uncertain parameter values. In our work we observed a

large range in HIV incidence within a given model scenario across the 20 simulations

reported in the main paper. Additional simulations of a given scenario might have

provided tighter fInterquartile ranges of around our median values, but due to the

stochastic nature of the models and the effect of network substructure on model

incidence, it is unlikely that the distribution of the values we report here would have

changed substantially beyond the interquartile ranges we observed, even if we ran

additional simulations. Supplemental Figures S4 and S5 illustrate this point for the main

comparison of interest. The panels for both figures show the distribution of observed 3

year HIV incidence (Figure S4) and the standard error of incidence per 100 person years

(Figure S5) for between 20 and 100 simulations of the baseline and 90 day retesting

scenarios.

Figure S4: Impact of the number of simulations on the primary outcomes of 3 year HIV incidence for populations of 5250 MSM, comparing baseline testing frequency to testing every 90 days.
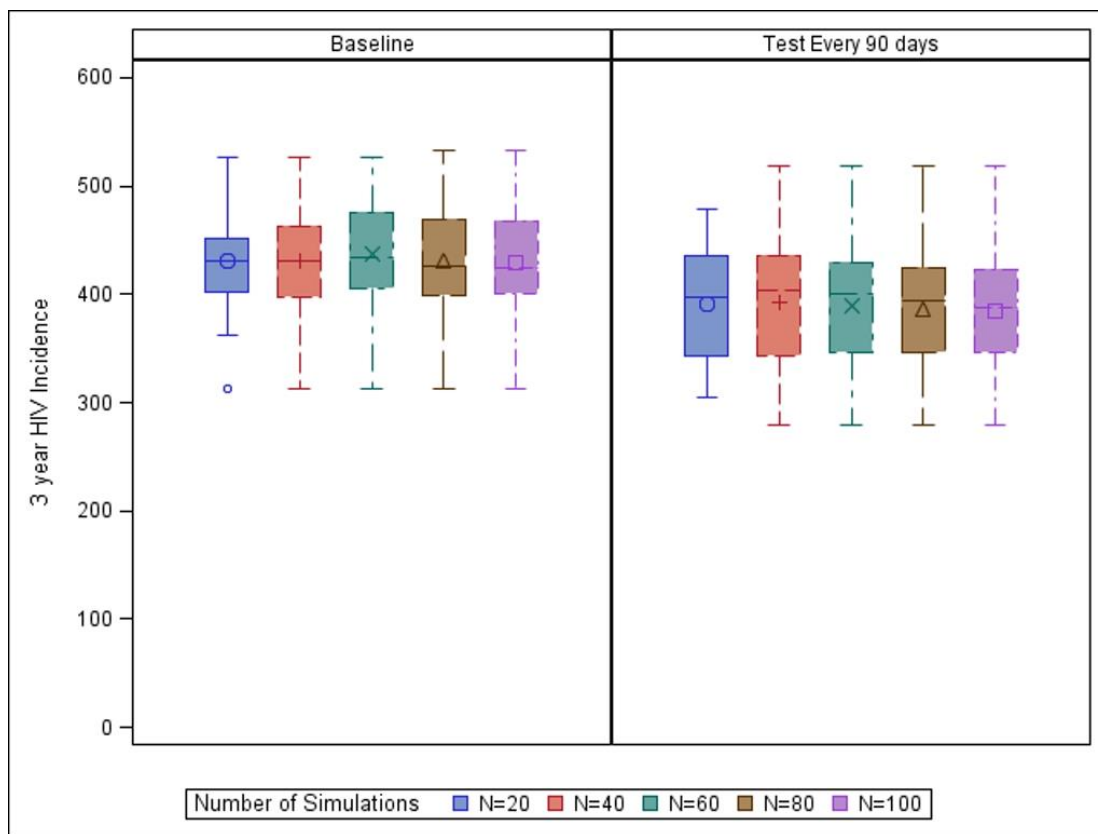
Figure S5: Impact of the number of simulations on the standard error of annualized incidence for populations of 5250 MSM, comparing baseline testing frequency to testing every 90 days.
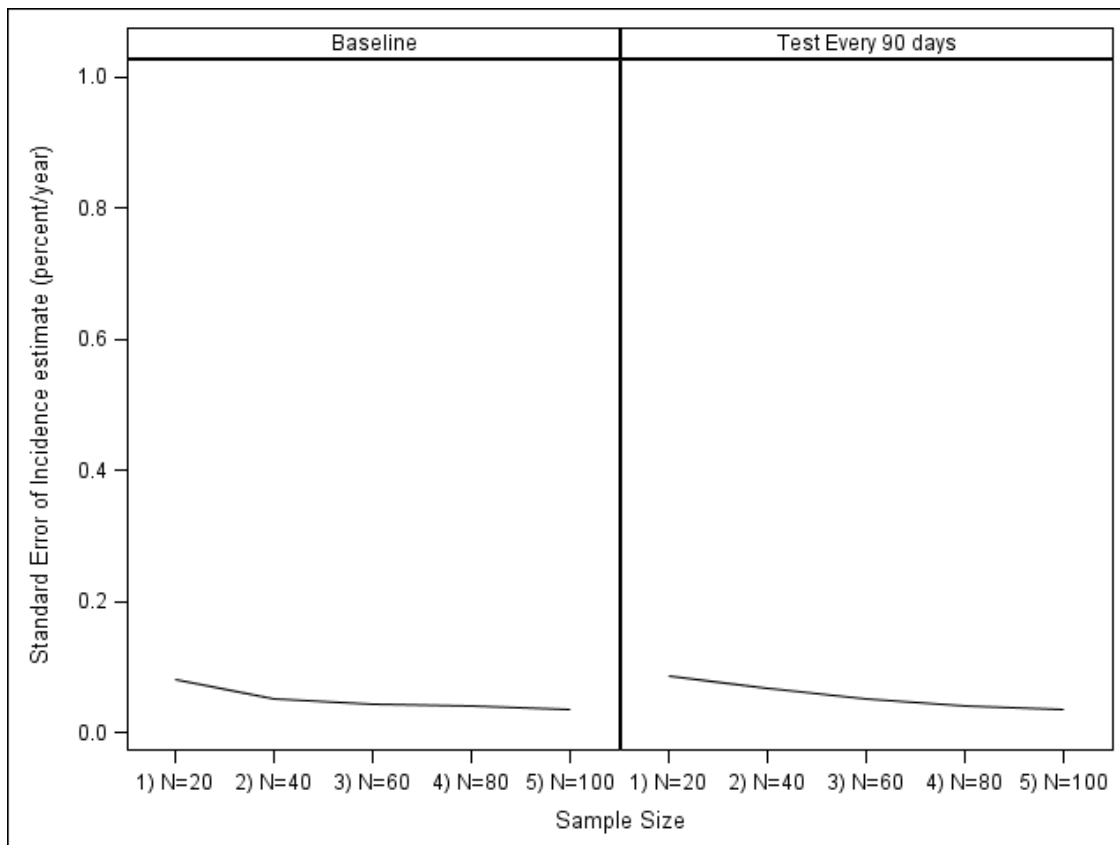
In each scenario, the minimum, maximum and interquartile range (IQR) vary little with increasing number of simulations, suggesting that by 20 simulations we have described most of the expected range in incidence for our model. The range of the IQR and extreme values for the scenario where men test every 90 days overlap those from the baseline scenario even after 100 simulations. Although the standard error of incidence per 100 person-years decreases slightly as the number of simulations increases, this value is driven more by the sample size of the population (set at 5250) and number of events (observed incidence), which are not affected by increasing the number of

simulations. Even if simulations with a larger population (e.g. a simulation of the entire US population of MSM) or an extremely large number of simulations were to indicate that a difference in incidence could be considered 'significant' under traditional statistical inference , we only observed a 10% reduction in median incidence comparing baseline (near annual) testing with testing every 90 days, and the total volume and therefore cost of testing was so different for quarterly and annual testing that it is unlikely to be clinically meaningful or cost effective to decrease testing interval to less than annually, relative to interventions that increase the proportion of the infected population that achieves viral suppression.

References

1. Krivitsky PN, Handcock MS. A separable model for dynamic networks. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2014; 76.1: 29-46.

2. Goodreau SM, Carnegie NB, Vittinghoff E, Lama JR, Sanchez J, Grinsztejn B, et al. What drives the US and Peruvian HIV epidemics in men who have sex with men (MSM)? PLoS One. 2012; 7(11):e50522. http://dx.doi.org/10.1371. Accessed 23 Oct 2013.

3. Khanna AS, Goodreau SM, Gorbach PM, Daar E, Little SJ. Modeling the Impact of Post-Diagnosis Behavior Change on HIV Prevalence in Southern California Men Who Have Sex with Men (MSM). AIDS Behav. 2014 Aug;18(8):1523-31. doi: 10.1007/s10461-013-0646-2.

4. Carnegie NB, Krivitsky PN, Hunter DR, Goodreau SM. An Approximation Method for Improving Dynamic Network Model Fitting. Technical Reports and Reprints Number 12-04: October 2012, Department of Statistics, Pennsylvania State University, University Park, PA, USA. Available from: http://stat.psu.edu/research-old/technical-reports/2012-technical-reports/TR1204.pdf Accessed: 11/20/14.

5. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. statnet: Software tools for the Statistical Modeling of Network Data. Seattle, WA; 2003. Version 2.0. Available from: http://statnetproject.org.

6. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M (2008). statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. Journal of Statistical Software, 24(1). Available from: http://www.jstatsoft.org/v24/i01/paper Accessed: 11/20/14

7. Goodreau SM, Handcock MS, Hunter DR, Butts CT, and Morris M. A statnet Tutorial. Journal of Statistical Software. 24(9). Available from: http://www.jstatsoft.org/v24/i09/paper Accessed 11/20/2014

8. Jenness SM, Goodreau SM, Morris M. EpiModel – Mathematical Modeling of Infectious Disease Version 1.1.1, A Tutorial. Available from: www.epimodel.org Accessed 11/20/14

9. Khosropour, CM, Johnson BA, Ricca AV, Sullivan PS. Enhancing Retention of an Internet-Based Cohort Study of Men Who Have Sex With Men (MSM) via Text

Messaging: A Randomized Controlled Trial. J Med Internet Res (2013); 15(8): e194.

10. CDC. HIV testing and risk behaviors among gay, bisexual, and other men who have sex with men—United States. MMWR 2013;62:958–62.

11. Wejnert C, Le B, Rose CE, et al. HIV infection and awareness among men who have sex with men—20 cities, United States, 2008 and 2011. PLoS ONE 2013;8:e76878.

12. Cooley LA, Oster AM, Rose CE, Wejnert C, Le BC, Paz-Bailey G; NHBS Study Group. Increases in HIV testing among men who have sex with men--National HIV Behavioral Surveillance System, 20 U.S. Metropolitan Statistical Areas, 2008 and 2011. PLoS One. 2014 Sep 2;9(9):e104162. doi: 10.1371/journal.pone.0104162. eCollection 2014

13. Helms DJ Weinstock HS, Mahle KC, Bernstein KT, Furness BW, Kent CK, Rietmeijer CA, Shahkolahi AM, Hughes JP, Golden MR. HIV Testing Frequency Among Men Who Have Sex With Men Attending Sexually Transmitted Disease Clinics: Implications for HIV Prevention and Surveillance. J Acquir Immune Defic Syndr   Volume 50, Number 3, March 1, 2009

14. Hamel L, Firth J, Hoff T, Kates J, Levine S, Dawson L. HIV/AIDS In The Lives Of Gay And Bisexual Men In The United States. Kaiser Family Foundation. 25 September 2014. Available from: http://kff.org/hivaids/report/hivaids-in-the-lives-of-gay-and-bisexual-men-in-the-united-states Accessed: 11/20/2014.

15. Chandra A, Billioux VG, Copen CE, Balaji A, DiNenno E. HIV testing in the U.S. household population aged 15-44: data from the National Survey of Family Growth, 2006-2010. Natl Health Stat Report. 2012 Oct 4;(58):1-26.

16. Wall KM, Stephenson R, Sullivan PS. Frequency of sexual activity with most recent male partner among young, Internet-using men who have sex with men in the United States. J Homosex. 2013;60(10):1520-38. doi: 10.1080/00918369.2013.819256.

17. Khanna AS, Goodreau SM, Wohlfeiler D, Daar E, Little SJ, Gorbach PM. Individualized diagnosis interventions can add significant effectiveness in reducing human immunodeficiency virus incidence among men who have sex with men: insights from Southern California. Annals of Epidemiology 2014 epub ahead of print October 7, 2014. DOI: http://dx.doi.org/10.1016/j.annepidem.2014.09.012

18. White PJ, Fox J, Weber J, Fidler S, Ward H. How Many HIV Infections May Be Averted by Targeting Primary Infection in Men Who Have Sex With Men? Quantification of Changes in Transmission-Risk Behavior, Using an Individual-Based Model. J Infect Dis. 2014 Dec 1;210 Suppl 2:S594-9. doi: 10.1093/infdis/jiu470.

19. Eaton JW, Hallett TB. Why the proportion of transmission during early-stage HIV infection does not predict the long-term impact of treatment on HIV incidence. Proc Natl Acad Sci U S A. 2014 Nov 11;111(45):16202-7. doi: 10.1073/pnas.1323007111. Epub 2014 Oct 13.

20. Wilson DP, Hoare A, Regan DG, Law MG. Importance of promoting HIV testing for preventing secondary transmissions: modelling the Australian HIV epidemic among men who have sex with men. Sexual Health 2009 6(1): 19-33.

21. Gray, R. T., et al. Increased HIV Testing Will Modestly Reduce HIV Incidence among Gay Men in NSW and Would Be Acceptable if HIV Testing Becomes Convenient. PLoS ONE 8(2): 2013.

22. Brookmeyer R, Boren D, Baral SD, Bekker LG, Phaswana-Mafuya N, Beyrer C, Sullivan PS. Combination HIV Prevention among MSM in South Africa: Results from Agent-based Modeling. PLoS One. 2014 Nov 14;9(11):e112668. doi: 10.1371/journal.pone.0112668. eCollection 2014.

23. Lucas A, Armbruster B. The cost-effectiveness of expanded HIV screening in the United States. AIDS 2013 27(5): 795-801.

24. Long EF HIV Screening via Fourth-Generation Immunoassay or Nucleic Acid Amplification Test in the United States: A Cost-Effectiveness Analysis. PLoS ONE 2011 6(11): e27625. doi:10.1371/journal.pone.0027625

25. Long EF, Brandeau ML, Owens DK. The cost-effectiveness and population outcomes of expanded HIV screening and antiretroviral treatment in the United States. Annals of internal medicine 153.12: 2010 778-789 .

26. Sullivan PS, Carballo-Dieguez A, Coates T, Goodreau SM, McGowan I, Sanders EJ, et al. Successes and challenges of HIV prevention in men who have sex with men. Lancet. 2012; 380(9839):388–99.

27. Sullivan PS, Stephenson R, Grazter B, Wingood G, Diclemente R, Allen S, Hoff C, Salazar L, Scales L, Montgomery J, Schwartz A, Barnes J, Grabbe K.   Adaptation of the African couples HIV testing and counseling model for men who have sex with men in the United States: an application of the ADAPT-ITT framework. Springerplus. 2014 May 16;3:249. doi: 10.1186/2193-1801-3-249. eCollection 2014.

28. Sullivan PS, Wall KM, O'Hara B, Jones J, Barnes J, DiClemente R, Hoff C, Scales L, Salazar LF, Sanchez T, White D, Wingood G, Allen S, Stephenson R. The prevalence of undiagnosed HIV serodiscordance among male couples presenting for HIV testing. Arch Sex Behav. 2014 Jan;43(1):173-80.

Chapter 6: Summary of findings and next steps

In this dissertation we used three different studies to look at different aspects of interaction. In this chapter we conclude by first summarizing the findings of the three studies and how they interrelate and contribute collectively to advance HIV prevention research. . We then suggest directions for future research in each of the areas studied, with both specific questions and research trajectories that arise from the results decribed in Chapters 3 through 5 and also some opportunities to expand infectious disease epidemiology methods and HIV prevention research more broadly.

Contributions to HIV Prevention Research

In the first study we demonstrated a new way to collect and apply data from smartphone applications being used for social interaction for HIV Prevention research. We demonstrated that, in less than 24 hours of effort only one research staff (KD) could sample 2666 application user profiles. We then used spatial statistics to describe a method for calculating the density of user profiles by race and age in Atlanta, and compared and contrasted the information provided by different outcome measures that can be constructed from these data. The methods described here may have practical application in HIV prevention research. The results are promising and illustrate how the use of self-reported location data can provide information on the geographic distribution of users in time and space. The study methodology could provide a more efficient way to identify locations for recruitment of MSM in future studies. Significant time and effort is spent on formative research to develop sampling frames for studies of

MSM.[1-2] The goal of such formative research is to identify locations for sampling MSM using time space sampling methods.[1] Our methodology, based on the geolocation data incorporated into popular social networking applications, allowed us to quickly describe the density of sex-seeking MSM in Atlanta. Furthermore, we were able to use profile information to stratify these density measures by race and age. This might allow for oversampling or exclusive sampling in areas of the city that are expected to yield a particular subset of the population, for example, young black MSM.

In the second study we provided a systematic evaluation of 13 proposed tests for interaction of two dichotomous exposure effects on an individual. The rapid proliferation of tests for interaction suggests that there is much interest in this area.[3-17] Unfortunately we have shown that only exceptionally large observational studies will likely have enough participants with both exposures to observe a true interaction effect that is greater than additive but less than multiplicative. This is however very practical information for the design of randomized trials of combinations of HIV prevention interventions.[18-20] It implies that studies with a goal of quantifying or describing the significance of interaction effects should consider a randomized controlled design in which there are at least 4 different arms of the study, where patients are assigned one intervention, the other intervention and both together and compared to a control condition. We have also provided context to another study we recently authored.[21] In this study we detected an effect that met the criteria for a test for detection of any interaction, but not for causal interaction, in that the interaction contrast ratio was greater than zero but less than 1. We also observed that the effects in different strata of

our exposure were not monotonic. The results presented in Chapter 4 suggest that, while this effect could possibly be causal interaction, it requires more study to explain what we observed. In addition to providing information that is useful to us currently, we hope that Aim 2 has also provided practical guidance for others seeking to quantify causal interaction of two exposures, with SAS code, sample size calculations and practical advice on how to proceed with such an analysis.

In our third aim we used a mathematical model to study the impact of various changes to HIV testing policies for MSM in the United States. We parameterized the model to consider perceived HIV serostatus as part of sexual partnership formation and duration, something that we believe to be unique in the HIV prevention literature. We were able to validate our baseline model against an external estimate of incidence obtained from the population that was surveyed for the model parameters, something that can also not normally be done in practice. With this new model we were able to show that there is little benefit to increased frequency, coverage or sensitivity of HIV testing among MSM unless men diagnosed through enhanced testing programs are also engaged in effective HIV care. This has implications both for HIV testing guidelines and for HIV prevention in the United States more broadly. It suggests that the White House's National Strategy for HIV and AIDS rightly focuses on the poor rates of viral suppression in the US and that more focus is needed to improve this outcome of HIV diagnosis.[22] But we also showed that, if HIV care and viral suppression are improved to higher rates than currently observed, that changes in HIV testing frequency can have additional impact. This helps to explain conflicting results of both mathematical models and other HIV

prevention studies. For example, studies from Seattle routinely show poorer performance for HIV tests than studies conducted in other settings.[23-25] However, Seattle seems to have some of the highest rates of viral suppression among the diagnosed population and highest frequency of repeat testing in the US.[25] Insights into the interactions of viral suppression, HIV testing frequency and test sensitivity are important when establishing HIV testing policy in the US. Our results suggest that the focus of HIV testing and prevention programs more broadly should likely vary based on the current proportion of all infections that remain undiagnosed and rates of viral suppression among the diagnosed, a finding that we do not believe has been previously described.

Areas for Future Research

In addition to the opportunity for immediate impact of this work, there are also ample opportunities for future growth and funded research that build on our current findings. Each of our studies represents a unique area of research, but we also present a general body of work that advances the methodology for infectious disease epidemiology. As such there are specific foci that arise from each individual study and a general overarching path to growth and opportunity that emanate from this work.

There are specific questions that could be answered to further the work of aim 1. For example, are users of a sexual networking smartphone application simply a subset of all MSM seeking sex on the internet? In order to evaluate this question we need to collect data from a larger group of MSM and compare the group that uses these types of sex

seeking applications to those that don't use them, or that use the internet but not the apps with geolocation services. A similar question is whether the population that uses specific services, (e.g., Adam-4-Adam, Jack'd, Grindr) different by one or more characteristics than those that use other online applications?[26-28] There is really two parts to this question. First, would we find different results for the densities of application users in Atlanta or elsewhere if we used our methods with multiple services. Then there are questions about whether the types of men who use a particular service are different from those who use another service. There is anecdotal evidence that this is true. In future studies we will seek to quantify the density and characteristics of men who use each of these applications and compare the characteristics of men who use each of the applications exclusively, while also capturing information about men who use more than one service to describe whether their behaviors vary when using different services.

It would also be useful to test the methodology in other cities with significant minority MSM populations (e.g. Washington, DC or Los Angeles, CA) and also to assess the utility of the method in less densely populated areas (e.g. in rural areas of Georgia), to describe the extent to which the utility of the methods vary by characteristics of the geography of the region. Additionally, although we averaged over day and time of sampling in our current analysis, the method could be refined to capture spatiotemporal trends in density. For example, it would be possible to select points to be sampled multiple times over a grid of specific times and days.[1-2] This modification could provide a clear description of how the user profile's population density changes over the course

of a week. This last component may identify trends in the spatial and temporal

clustering of application users, for example on weekend nights, as compared to mid-day

during the work week.

More generally, using social networking applications for HIV prevention is likely a key

strategy for future research, but comes with new ethical and methodological

questions.[29-31] Our study only sought to summarize the data publically available within

these applications, but social media applications may themselves serve as an important

public health communications tool. Recently, public health agencies have sought to

partner with Grindr, and use its built in advertisements as a medium for disseminating

prevention information and recruit MSM for research studies.[29,30] This is one example

of an area where Aim 1 and Aim 3 overlap. One recent study offered MSM using a

smartphone app to meet sex partners access to HIV tests through the internet.[32] Based

on findings from Aim 3 this could be an additional useful strategy for accessing the

population of MSM who currently never test for HIV, as well as those MSM who test less

than annually. The authors of this study did point out that this strategy, in which there is

no confirmation that the men receiving the test kits via the internet use the kits and, if

they test positive, access care and treatment, poses additional challenges.[32] Given the

importance of ensure linkage to care, treatment and viral suppression that we found in

Aim 3, future research will need to find ways to combine the access to high-risk men

through smartphone applications with linkage to care for those with HIV. Future

research might adapt our methodology further to establish a sampling frame, and then

use the density information to sample application users and contact them to either

conduct a cross-sectional survey or recruit them into a follow-up study. At that time one

would have to develop mechanisms for consenting study participants, as well as a way

to keep sensitive information, such as sex and drug use behavior, protected and ideally

separate from any identifying online profile information. Another approach would be to

again use the spatial data as the sampling frame, and weight data from surveys that use

banner ads or other methods within the app by the density distribution data to adjust

for selection bias in the resulting sample. When weighted samples are reported for

geographic areas of interest it might be easier to direct needed HIV prevention, care and

treatment services to the areas highlighted as those with the highest density of

application users.

Aim 2 represents a different area for future research entirely. I am very interested in

developing manuscripts similar to this one that take complicated concepts and provide

practical epidemiologic advice based in simulated and practical examples. An obvious

extension of our findings is to author a paper describing modeling strategy for studies

that seek to assess whether two exposures interact, with examples in which the

exposures are and are not monotonic, and when the results of tests for additive

interaction give conflicting results. Such a manuscript would complement the

recommendations for modeling strategy already in the literature and being taught in the

Emory Department of Epidemiology,[33,34] and add newer methods and modeling

techniques to this strategy. This is just one example of work in the area of practical

application of epidemiologic methods that I would like to pursue further. Another

example of this type of work would be providing practical guidance on when collinearity

is important in epidemiologic modeling, how to test for it and how to evaluate the results of those tests for different types of models (e.g. linear, log-linear, logistic, log-binomial).

The type of research conducted in Aim 3 is likely to have the most opportunity for growth and future work. In both the US and sub-Saharan Africa there is growing awareness of the need for "combination prevention" the idea that one intervention will not be enough to control the HIV epidemic, and that there are synergies like the ones we observed when interventions can be combined effectively.[18-20] The beauty of building a model for infectious disease transmission is that it can then be adapted to address many different new questions. One area of immediate need is to assess the interaction between the factors we have not yet included in our model, and factors associated with HIV testing. The most obvious prevention intervention to include in future modeling work is Pre-exposure prophlylaxis (PreP).[35] This intervention has been shown to significantly reduce the incidence of HIV in the highest risk MSM.[36] Furthermore, the clinical regimen for PreP that has been recommended by the CDC includes testing quarterly to ensure that infection hasn't occurred and reduce the risk of developing resistance by taking sub-optimal HIV therapy after infection.[35] In a future collaboration between Emory and CDC we will propose to look at the interaction between PreP guidelines and HIV testing guidelines to assess the optimal mix of these two interventions and whether there is any additional benefit of more frequent HIV testing for high risk MSM who choose not to initiate PreP.

Another way to expand the model would be to look at the overlap between the MSM

epidemic and the heterosexual epidemic. To our knowledge to date there has not been

an ERGM model developed that models both MSM and heterosexuals and the mixing

between the two groups. This risk factor may be particularly important in minority MSM

and as a bridge to the black female population in the United States.[37] Developing a

heterosexual ERGM model would also be useful to answer research questions specific to

sub-Saharan Africa.[38] For example, in countries such as Nigeria, where the epidemic is

large but is still concentrated in high-risk groups such as sex workers and MSM, we

could model the impact of targeted interventions such as PreP, testing, and treatment

to see how each would impact the epidemic overall. We can also answer questions

related to the optimal coverage of multiple interventions for epidemics such as that in

South Africa where a much higher proportion of the population is infected. In summary,

there are a wide variety of additional applications of infectious disease modeling,

particularly separable-temporal exponential random graph models such as the ones I

employed in my dissertation.

Through the three research aims of this dissertation I have explored various aspects of

interaction, and how interaction can be leveraged to improve HIV prevention efforts. I

have demonstrated a novel method for quantifying the amount of social and sexual

interaction occurring in a defined geographic area using social networking applications

and spatial statistics. I have also described the performance of a wide variety of

statistical tests appropriate for assessing the interaction of two or more disease causing

exposures in individuals. Finally, I have built a model of HIV transmission parameterized

based on the US MSM population and described how different public health

interventions can interact at the population level to prevent disease. I hope to continue

to study different aspects of interaction and build on this work in future research over

the next several years.

References:

1) MacKellar DA, Gallagher KM, Finlayson T, Sanchez T, Lansky A, Sullivan PS. Surveillance of HIV risk and prevention behaviors of men who have sex with men--a national application of venue-based, time-space sampling. Public Health Rep. 2007;122 Suppl 1:39-47. PMID: 1735452

2) Lansky A, MacKellar D, Gallagher KM, Lin LS, Sullivan PS, Onorato IM. Evaluation of the new National HIV Behavioral Surveillance System (NHBS) is imperative. Sex Transm Dis. 2006 Apr;33(4):272-3; author reply 274-5.

3) VanderWeele TJ, Robins JM. The identification of synergism in the sufficient-component-cause framework. Epidemiology. 2007 May;18(3):329-39.

4) Hernán MA, Robins JM. Causal Inference. Chapman & Hall/CRC; 2015.

5) VanderWeele TJ. Sufficient cause interactions and statistical interactions. Epidemiology. 2009 Jan;20(1):6-13.

6) VanderWeele TJ. On the distinction between interaction and effect modification. Epidemiology. 2009 Nov;20(6):863-71.

7) VanderWeele TJ. Richardson TS. General theory for interactions in sufficient cause models with dichotomous exposures. Ann Stat. 2012;40(4):2128-2161.

8) VanderWeele TJ, Tchetgen Tchetgen EJ. Attributing effects to interactions. Epidemiology. 2014 Sep;25(5):711-22.

9) Skrondal A. Commentary: much ado about interactions. Epidemiology. 2014 Sep;25(5):723-6.

10) Greenland S. Interactions in epidemiology: relevance, identification, and estimation. Epidemiology. 2009 Jan;20(1):14-7.

11) Kaufman JS. Interaction reaction. Epidemiology. 2009 Mar;20(2):159-60.

12) Richardson DB, Kaufman JS. Estimation of the relative excess risk due to interaction and associated confidence bounds. Am J Epidemiol. 2009 Mar 15;169(6):756-60.

13) Rose S, van der Laan M. A double robust approach to causal effects in case-control studies. Am J Epidemiol. 2014 Mar 15;179(6):663-9.

14) Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. Am J Epidemiol. 2005 Aug 1;162(3):199-200.

15) Brumback BA, Bouldin ED, Zheng HW, Cannell MB, Andresen EM. Testing and estimating model-adjusted effect-measure modification using marginal structural models and complex survey data. Am J Epidemiol. 2010 Nov 1;172(9):1085-91.

16) Kuss O, Schmidt-Pokrzywniak A, Stang A. Confidence intervals for the interaction contrast ratio. Epidemiology. 2010 Mar;21(2):273-4.

17) Cheung YB. A Modified Least-Squares Regression Approach to the Estimation of Risk Difference. Am. J. Epidemiol. 2007; 166(11): 1337-1344. Greenland S. Tests for interaction in epidemiologic studies: a review and a study of power. Stat Med. 1983 Apr-Jun;2(2):243-51.

18) Beyrer C, Sullivan PS, Sanchez J, Dowdy D, Altman D, Trapence G, Collins C, Katabira E, Kazatchkine M, Sidibe M, Mayer KH. A call to action for comprehensive HIV services for men who have sex with men. Lancet. 2012 Jul 28;380(9839):424-38.

19) McNaghten A, Kearns R, Siegler AJ, Phaswana-Mafuya N, Bekker LG, Stephenson R, Baral SD, Brookmeyer R, Yah CS, Lambert AJ, Brown B, Rosenberg E, Blalock Tharp M, de Voux A, Beyrer C, Sullivan PS. Sibanye Methods for Prevention Packages Program Project Protocol: Pilot Study of HIV Prevention Interventions for Men Who Have Sex With Men in South Africa. JMIR Res Protoc. 2014 Oct 16;3(4):e55.

20) Brookmeyer R, Boren D, Baral SD, Bekker LG, Phaswana-Mafuya N, Beyrer C, Sullivan PS. Combination HIV prevention among MSM in South Africa: results from agent-based modeling. PLoS One. 2014 Nov 14;9(11):e112668.

21) Delaney KP, Rurangirwa J, Facente S, Dowling T, Janson M, et al. Using a multi-test algorithm to improve the positive predictive value of rapid HIV testing and linkage to HIV care in non-clinical HIV test sites. JAIDS Submitted.

22) Office of National AIDS Policy. National HIV/AIDS strategy for the United States. Washington, DC: Office of National AIDS Policy; 2010. Available at http://aids.gov/federal-resources/national-hiv-aids-strategy/nhas.pdf

23) Stekler JD, O'Neal JD, Lane A, Swanson F, Maenza J, Stevens CE, Coombs RW, Dragavon JA, Swenson PD, Golden MR, Branson BM. Relative accuracy of serum, whole blood, and oral fluid HIV tests among Seattle men who have sex with men. J Clin Virol. 2013 Dec;58 Suppl 1:e119-22.

24) Katz DA, Dombrowski JC, Swanson F, Buskin SE, Golden MR, Stekler JD. HIV intertest interval among MSM in King County, Washington. Sex Transm Infect. 2013 Feb;89(1):32-7.

25) Katz DA, Cassels SL, Stekler JD. Replacing clinic-based tests with home-use tests may increase HIV prevalence among Seattle men who have sex with men: evidence from a mathematical model. Sex Transm Dis. 2014 Jan;41(1):2-9.

26) Grov C. HIV risk and substance use in men who have sex with men surveyed in bathhouses, bars/clubs, and on Craigslist.org: venue of recruitment matters. AIDS Behav. 2012 May;16(4):807-17. PMID: 21748276

27) Grov C, Crow T. Attitudes about and HIV risk related to the "most common place" MSM meet their sex partners: comparing men from bathhouses, bars/clubs, and Craigslist.org. AIDS Educ Prev. 2012 Apr;24(2):102-16. PMID: 22468972

28) Mayer KH, Ducharme R, Zaller ND, Chan PA, Case P, Abbott D, Rodriguez II, Cavanaugh T. Unprotected sex, underestimated risk, undiagnosed HIV and sexually transmitted diseases among men who have sex with men accessing testing services in a New England bathhouse. J Acquir Immune Defic Syndr. 2012 Feb 1;59(2):194-8. PMID: 22027871

29) Burrell ER, Pines HA, Robbie E, Coleman L, Murphy RD, Hess KL, Anton P, Gorbach PM. Use of the location-based social networking application GRINDR as a recruitment tool in rectal microbicide development research. AIDS Behav. 2012 Oct;16(7):1816-20. PMID: 22851153

30) Usher D, Frye V, Shinnick J, Greene E, Baez E, Benitez J, Solomon L, Shouse RL, Sobieszczyk ME, Koblin BA. Recruitment by a Geospatial Networking Application for Research and Practice: The New York City Experience. J Acquired Immune Defic Syndr. 2014 Sept 22[Epub ahead of print]. PMID: 25247432

31) Sullivan PS, Grey JA, Simon Rosser BR. Emerging Technologies for HIV Prevention for MSM: What We have Learned, and Ways Forward. J Acquir Immune Defic Syndr. 2013 Jun; 63 Suppl 1:S102-7. PMID: 23673879

32) Medline A, Huang E, Marlin R, Young SD, Kwok J, Klausner J. Using Grindr™, a Social-Media–Based Application, to Increase HIV Self Testing Among High-Risk Men Who Have Sex With Men in Los Angeles, California, 2014. Presented at CROI 2015 as Abstract 1098. Seattle, Washington; February 23-26,2015.

33) Kleinbaum DG, Klein M. Logistic Regression: A Self-Learning Text. 3rd ed. New York, Springer Science and Business Media; 2010.

34) Kleinbaum DG, Kupper LL, Nizam A and Muller KE. Applied regression analysis and other multivariable methods, 4th Ed. Belmont, Ca: Thomson-Brooks/Cole; 2008

35) Centers for Disease Control and Prevention. Pre-exposure prophylaxis for the prevention of HIV infection in the United States—2014: a clinical practice guideline. www.cdc.gov/hiv/pdf/guidelines/PrEPguidelines2014.pdf.

36) Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med. 2010;363:2587–2599.

37) Friedman MR, Wei C, Klem ML, Silvestre AJ, Markovic N, Stall R. HIV infection and sexual risk among men who have sex with men and women (MSMW): a systematic review and meta-analysis. PLoS One. 2014 Jan 30;9(1):e87139.

38) UNAIDS. The Gap Report: The Beginning of the End of the AIDS Epidemic. Geneva; UNAIDS: 2014. Available from:

http://www.unaids.org/sites/default/files/media_asset/UNAIDS_Gap_report_en.pdf