Published in final edited form as:

Contemp Clin Trials. 2015 March ; 41: 152–159. doi:10.1016/j.cct.2015.01.014.

EVALUATING INTERVENTIONS TO PROMOTE ROUTINE PREVENTIVE SCREENINGS: A COMPARISON OF ANALYTICAL OUTCOMES

Akshay Sharma^a, Brent A Johnson^b, and Patrick S Sullivan^b

^a Department of Epidemiology, Emory University Laney Graduate School and Rollins School of Public Health, 1518 Clifton Road NE, Atlanta GA 30322, United States

^b Department of Biostatistics and Computational Biology, University of Rochester, 601 Elmwood Avenue, Rochester NY 14642, United States

Abstract

Background—Often in public health, we are interested in promoting routine preventive screenings (e.g., blood glucose monitoring, hypertension screening, or mammography). Evaluating novel interventions to encourage frequent screenings using randomized controlled trials can help inform evidence-based health promotion programs. When the desired behavior change is a recurrent event, specifying the most meaningful study outcomes may prove challenging.

Methods—To understand the efficiency of multiple approaches for evaluating an intervention seeking to increase regular health screenings we (a) simulated several replications of a trial with a positive intervention effect under various censoring scenarios, (b) formulated three different analytical outcome definitions (screening a certain number of times during the entire study period versus not, screening at least once within a clinically meaningful time period versus not, “hazard” or instantaneous rate of screening), and (c) compared them with regard to interpreting results and estimating power at different sample sizes.

Results—Approaches which better utilize detailed prospective data, while also accounting for within-participant correlations, are less likely to miss the actual underlying benefits conferred by a new prevention strategy compared to relying on a dichotomous measure derived from aggregating events over the study duration. Such approaches are also more powerful in realistic scenarios wherein some participants are lost to follow-up over time.

Conclusions—Researchers should carefully consider the choice of analytical outcomes and strive to employ more efficient approaches that model comprehensive event-specific information,

© 2015 Published by Elsevier Inc.

This manuscript version is made available under the CC BY-NC-ND 4.0 license.

Corresponding author: Akshay Sharma, MBBS, MPH, Department of Epidemiology, Emory University, 1518 Clifton Road NE, Atlanta GA 30322, United States. Phone: +1 404 857 8460. ashar24@emory.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of Conflicting Interests: The authors declare that there are no conflicts of interest.

rather than summarizing repeated measures into less-informative dichotomous responses, while designing and conducting trials with recurrent preventive screenings.

Keywords

Preventive screenings; analytical outcome choices; randomized trials with recurrent events; rapid home HIV self-testing; men who have sex with men

INTRODUCTION

Preventive screenings are an important component of health promotion efforts which can potentially reduce the significant economic burden of diseases.¹ Subgroup-specific recommendations have been developed to guide the timely identification and treatment for numerous morbidities. For example, the United States Preventive Services Task Force (USPSTF) recommends screening for lipid disorders in all men aged ≥ 35 years and women at an increased risk for coronary heart disease aged ≥ 45 years every five years,² screening for type 2 diabetes in adults with hypertension or hyperlipidemia at 3-year intervals,³ and screening for Hepatitis B infection in pregnant women at the time of their first prenatal visit.⁴

Given the emergence of new screening technologies, it is important to evaluate public health strategies to promote regular health exams. Parallel group randomized controlled trials (RCTs), reported according to established standards,⁵ are considered to be the most rigorous scientific tool for testing new interventions. Despite available guidance for variations in trial design,^{6, 7} intervention content^{8, 9} and mode of delivery,¹⁰ limited discussion exists regarding design and methodological aspects unique to RCTs with recurrent events during follow-up.¹¹⁻¹³ Examples of such events include episodes of healthcare utilization, screening mammography, self-monitoring blood glucose and cholesterol levels, and screening for human immunodeficiency virus (HIV) infection.

Researchers conducting trials to promote preventive screening behaviors need to determine *a priori* what constitutes a meaningful outcome. Imagine a situation where we would like all adults get examined for hypertension annually, but only 20% actually follow this recommendation. Suppose we are studying a new intervention aimed at increasing the frequency of screening. How should we define our outcome? One option is a dichotomous measure, such as checking for high blood pressure ≤ 5 times versus not over a 5-year period. However, this definition might misclassify meaningful behavioral changes (e.g. increasing from one to four screenings) as failures, and a promising intervention could be wrongly described as being ineffective. Further, someone could screen five times in a short period (e.g. within a year), but not again for the remaining 5 years, and still be counted as a success. Potential alternative outcomes are screening at least once within a 1-year interval or the rate of screening. Depending upon the choice of our outcome, different analytical approaches are needed to answer the primary question “Does the intervention work?” directly impacting adequate sample size estimation.¹⁴

Despite the availability of different techniques to analyze recurrent events data,¹⁵⁻¹⁷ researchers often adopt naive approaches which either ignore the existence of multiple

events, their timing during follow-up, or the correlation between repeated measures. For example, a recent systematic review of 83 RCTs evaluating interventions to prevent falls among the elderly indicated that more than half the studies inappropriately employed proportions/odds-ratio (OR) based approaches.¹⁸ Another review of 105 longitudinal studies examining hospitalization data among heart failure patients found that 70% based their analyses on outcomes incorporating only the first admission, and almost one-third compared proportions of individuals experiencing one or more hospitalizations using either a chi-squared test or standard logistic regression (SLR).¹⁹

Given that researchers continue to summarize repeated measures into less-informative dichotomous responses, we sought to demonstrate how different choices of analytical outcomes impact the ability to detect true underlying intervention effects. To understand the efficiency of multiple approaches for evaluating an intervention seeking to increase routine preventive screenings we (a) simulated several replications of a “successful” RCT (i.e. one with a positive intervention effect) under various censoring scenarios, (b) formulated three outcome definitions (screening a certain number of times during the entire study period versus not, screening at least once within a clinically meaningful time period versus not, “hazard” or instantaneous rate of screening) and performed corresponding analyses, and (c) compared them with regard to interpreting results and estimating power at different sample sizes. For demonstration purposes and our own scientific research interests, we are using the rationale of a randomized trial which seeks to determine the effectiveness of rapid HIV self-test kits in increasing testing among men who have sex with men (MSM) in the United States.²⁰

METHODS

Simulation strategy

Consider an RCT among HIV-negative or unknown status MSM prospectively followed for one year. The intervention to be evaluated is one to increase the frequency of HIV screening by distributing rapid HIV self-test kits that can be used at home. Intervention arm participants are given self-test kits and comparison arm participants are provided resources for identifying local HIV testing services. Men can report their test results online at the time of screening or during quarterly surveys. Participants are censored either because they are newly diagnosed as HIV-positive or because they are lost to follow-up.

SAS version 9.3²¹ was used to simulate 360,000 iterations of such a trial under different assumptions and perform all subsequent analyses. Hypothetical participants were assigned demographic characteristics based on a previous study of behavioral risks involving voluntary HIV testing with a home specimen collection kit²² and randomized to either the intervention or comparison arm. In that prospective study, 1% of participants had tested for HIV six times within a year, 1% had tested five times, 3% had tested four times, 8% had tested three times, 17% had tested twice, 31% had tested once and 39% had not tested even once within a year. Screening frequencies for simulated participants were generated using different assumptions for men in either trial arm. Annual HIV testing frequencies for men in the intervention arm were simulated such that the intervention was effective and that participants could screen for a maximum of six times. Testing days were uniformly

generated on the interval 1 to 365, assuming that all days of the year were equally likely to be selected, and the specific days of screening were separated to obtain HIV testing behavior within four 3-month time intervals.

First, two variations of a “successful” RCT (i.e. one with a positive intervention effect) were simulated assuming that 13% of comparison arm men screened (a) 3 times annually (as previously observed²²) and (b) 2 times annually, with the intervention truly doubling these odds (i.e., assuming an OR of 2). Essentially, in the first parameter specification the odds of testing 3 times per year among intervention arm participants were twice the odds of testing 3 times per year among men in the comparison arm, and in the second parameter specification the odds of testing 2 times per year among intervention arm participants were twice the odds of testing 2 times per year among men in the comparison arm. For sensitivity analyses, we considered four additional parameter specifications such that 13% of the comparison arm screened either 3 or 2 times annually with an OR of 2.5, and 5% of the comparison arm screened either 3 or 2 times annually with an OR of 3.

Datasets were simulated by increasing starting sample sizes from 100 to 2000 in increments of 100 for each specification under the following retention and incidence scenarios:

Scenario A (Ideal retention) – No loss to follow-up and no incident HIV diagnoses

Scenario B (Modest retention) – 40% loss to follow-up and 5% incident HIV diagnoses

Scenario C (Poor retention) – 70% loss to follow-up and 10% incident HIV diagnoses

Censoring in Scenarios B and C was implemented by randomly designating some testers from Scenario A as incident positives, followed by selecting men from the remaining pool to be lost over time, and uniformly generating their last day of follow-up on the interval 1 to 365. Consult appendix for further details regarding the simulation strategy.

Analytical outcome definitions

Outcome definition 1 – Trial arms could naively be compared using a dichotomous outcome of screening a certain number of times during the entire study period versus not, e.g. testing 3 times versus fewer within a year, and performing SLR.²³

Outcome definition 2 – Alternatively, one could use a binary outcome of screening at least once within a clinically meaningful time period versus not, e.g. testing at least once versus not within a 3-month interval. Because screening behaviors for each individual may be related across consecutive time intervals, Generalized Estimating Equations (GEE) with an autoregressive first order (AR1) correlation structure could be employed for analyses.²³

Outcome definition 3 – Finally, one could examine whether the intervention increases the “hazard” or instantaneous rate of screening. Stratified Cox proportional hazards (PH) regression for recurrent events could be used to model survival times between screening tests. The total follow-up time approach uses “event times”, i.e. the actual time of two tests from baseline,^{12, 24} whereas the gap time approach uses “gap times”, i.e. the interval length

between subsequent tests.^{12, 25} Both models adjust for the fact that screening events may be correlated within individuals.

Power and sample size estimation

SLR, GEE for correlated data, and stratified Cox PH regression for recurrent events were used to analyze simulated datasets. For SLR, we excluded all those participants who were newly diagnosed but retained available screening data for those lost to follow-up. For GEE and stratified Cox PH regression, we used available information from all participants (including incident positives) until they were censored. Power for different approaches was estimated from 1,000 trial replications for a fixed sample size by determining the proportion for which the intervention effect magnitudes were significantly greater than the null. This process was sequentially repeated with 20 starting sample sizes under each censoring scenario for all six parameter specifications. Consult appendix for data layouts, model specifications and sample SAS code.

RESULTS

Preliminaries

Table 1 illustrates our randomization process with sample data on 1,200 simulated participants from a single trial replication. Study arms are balanced with respect to race/ethnicity, age, education, income, urban or rural residence, and self-identified sexual orientation. Annual screening characteristics under the first set of parameters in Scenario A are presented in Table 2. As specified, men in the intervention arm had higher testing frequencies versus those in the comparison arm, resulting in shorter inter-test intervals. Consult appendix for an illustration of screening patterns under different censoring scenarios (Figures A1, A2 and A3).

Table 3 presents point estimates from different analytical approaches on 10 trial replications under the first set of parameters in Scenario A, each with 1,200 participants. By design, the SLR ORs comparing trial arms with respect to testing 3 times annually are close to 2. Although we present results from different methods in a single table for convenience, their interpretations differ and their magnitudes are not directly comparable. We also performed a subset of analyses wherein we adjusted for demographic covariates, and our findings remained unchanged (data not shown in table). Frequency distributions of estimates from 1,000 replications are presented in the appendix (Figures A5 and A6).

Graphical comparison of analytical outcomes

Figure 1 depicts two sets of power curves generated under the first and second parameter specifications in Scenario A. Each panel includes multiple analytical approaches corresponding to different outcome choices on 20,000 trial replications. Compared to other methods, SLR (using the first outcome definition of testing 3 times versus fewer within a year) was more sensitive to variations in the underlying benefit conferred by the intervention. Despite being powerful when the intervention actually increased the odds of screening 3 times annually (top panel), this approach performed poorly when the intervention increased the odds of screening 2 times annually (bottom panel). Expressed

alternatively, sample size estimation in order to attain a desired level of statistical power was subject to more uncertainty when using the first outcome definition compared to both other choices. For example, depending upon the actual success of the intervention in improving screening behaviors, recruiting either 500 participants would suffice (top panel) or 1,200 would be needed (bottom panel) to detect a difference between trial arms with the conventional 80% power.

Figures 2 and 3 demonstrate that compared to SLR, models for GEE for correlated data (using the second outcome definition of testing at least once versus not within a 3-month interval) and stratified Cox PH regression for recurrent events (using the third outcome definition of the instantaneous rate of testing) performed increasingly better in detecting the intervention effect as greater proportions of participants were censored over time (Scenarios B and C respectively). Similar patterns were observed on comparing results from sensitivity analyses with four additional parameter specifications (Figures A7, A8, A9, A10, A11 and A12 presented in the appendix).

DISCUSSION

Researchers designing trials with recurrent screening events are often challenged with defining appropriate study endpoints. Our goal was to demonstrate how different analytical outcome choices can impact the ability to detect true underlying intervention effects. Therefore, we used simulations of varyingly “successful” RCTs to examine the efficiency of three outcome definition alternatives and their corresponding analyses. We conclude that approaches that model detailed prospective data while accounting for within-participant correlations are better than those which depend upon appropriately specifying a dichotomous outcome measure. In particular, when the dichotomous outcome is defined as number of tests exceeding a certain cut-point, SLR may fail in correctly assessing the effect of the intervention unless the theoretical and user-defined cut-point align closely; GEE for correlated data and survival analysis methods for recurrent events do not suffer this limitation. Besides being less likely to miss the actual benefits of a new prevention strategy, they are also more powerful in realistic scenarios encompassing attrition over time. Although we based our simulations on a trial of rapid HIV self-testing among MSM,²⁰ our findings are more widely applicable to longitudinal studies with repeated health screenings. Depending upon the condition for which someone is getting evaluated, a positive result could lead them to either test more often or to cease screening altogether. For example, pregnant women might increase their frequency of screening for gestational diabetes if their glucose levels are not under control, but someone who receives a positive HIV test result would most likely stop screening, seek confirmatory testing and initiate treatment. This highlights the importance of employing statistical methods which utilize all available information to the fullest extent. Next, we focus on practical issues while choosing relevant outcome measures and some methodological caveats of corresponding analytical approaches.

Suppose an intervention arm participant from our example self-tested three times on the first day of follow-up, and then did not test during the remainder of the study. Using the first outcome definition he would be considered a success at the end of a year, even though the

practical value may only have been that of a single screening test. Contrast this with a comparison arm participant who tested twice annually at regular 6-month intervals (say first with his physician and later at a community based organization). Despite having a more reasonable inter-test interval, he would not be considered a success. Simply aggregating the number of screenings per individual at study conclusion and dichotomizing for SLR while ignoring the key aspect of when they were performed can therefore be problematic. Further, our simulations demonstrated that if the choice of cut-point is not reflective of the actual prevention implications of a novel screening approach, researchers may fail to identify a genuinely superior intervention. Depending upon the disease for which one is screening (e.g. HIV, certain cancers), analyzing data in this manner could also necessitate excluding all new diagnoses, even if they had tested a few times before identification. From our example, someone who was identified as HIV-positive before his third test should not be considered a failure with regard to screening 3 times per year because he might have continued to screen for HIV, and eventually be counted as a success had his previous test result been negative. Yet another statistical decision involves handling missing data due to incomplete follow-up (i.e., should such observations be retained or removed from the dataset?)

Choosing an analytical outcome of any screening within shorter and perhaps more clinically meaningful intervals can help resolve some of these issues. Consider someone from the intervention arm in our example who decided to test because he recently had unprotected sex. Suppose his first test result was negative but he self-tested again on the same day “just to be sure”. He later learned about the approximately 3-month long window period (interval between the time of infection and the production of measurable antibodies to HIV), and decided to test again after 3 months. Using the second outcome definition, his first two tests would be appropriately counted as a single success in one interval and his third test would be counted as a single success in the next interval. Although GEE for correlated data allows for an assessment of the intervention effect in shorter time periods, even this approach does not take into account the actual timing of screening tests. However unlike SLR, this option is less sensitive to variations in the true underlying intervention effects, newly diagnosed individuals need not be excluded and one can better utilize information from individuals who are lost to follow-up.

Researchers who plan on collecting data about the days on which each screening test was performed should consider using the third outcome definition. Although several analytical methods exist to compare instantaneous rates between trial arms,¹² we employed the total follow-up time and gap time models for stratified Cox PH regression for recurrent events. Given the level of detail in our data, we consider these choices to be more appropriate than count models. Both utilize all available information including the order and actual timing of subsequent screening tests. Unlike the previous options considered, one does not have to pre-specify a binary outcome such as testing certain number of times versus not within a particular time period. Power calculations assuming different parameter specifications also showed how both models for this approach were least affected by the manner and extent to which the intervention truly increased preventive screenings in the study population.

Our simulations also demonstrated that methods which more fully utilized available data (stratified Cox PH regression and GEE) versus aggregating information (SLR) performed

better in detecting the intervention effect as greater proportions of participants were censored over time. These results were in the expected direction because men who dropped out during later months before their third screening test were coded as failures when using the first outcome definition, and all new diagnoses had to be excluded. Although 40% (Scenario B) and 70% (Scenario C) losses to follow-up by year-end seem unusually high, more than 80% of the sample was retained in each arm at the first follow-up assessment in each scenario (Figure A4 presented in the appendix). This exceeds the 70% retention threshold set by the Centers for Disease Control and Prevention (CDC) to support a designation of “best-evidence” HIV prevention interventions.²⁶ Loss to follow-up might not be problematic in short-term studies, but trials of a longer duration (e.g. an RCT evaluating an intervention to increase routine screenings where the recommended interval is a year) may benefit from analytical approaches that utilize maximum information.

Limitations of our study include making relatively simplistic assumptions while choosing probability distributions and parameters for the simulation process. HIV testing frequencies for comparison arm participants in three of the six parameter specifications were based on an actual behavioral study among MSM,²² but these might not be representative of current trends. However, similar results were observed on performing sensitivity analyses with different specifications in each of the three censoring scenarios and we are confident regarding the validity of our findings. Although the assumption that nobody could test more than six times annually is arbitrary, we believe that counts higher than this would be extremely rare. Because we simulated screening patterns within a year-long trial, the median inter-test intervals of our hypothetical participants were shorter than what have been reported in a recent study following MSM over eight years.²⁷ We also acknowledge we have not considered all analytical options for the third outcome definition (e.g. marginal approach for stratified Cox PH regression,¹⁷ independent-increment model for the rate of recurrent events²⁸). The well-established counting process approach was not employed as this assumes independence between different screening tests on the same individual.^{16, 25} Negative binomial and zero-inflated models were not considered as these necessitate summarizing repeated outcomes on each participant with a single, one-dimensional measure of the total count of screening tests.²⁹ However, if the study design is such that only aggregated data rather than actual event times are collected, these alternatives could be used to estimate intervention effects without losing much efficiency.³⁰

Ultimately, the research question of interest and goals of the RCT should dictate the manner in which recurrent events data are measured and analyzed. Simple and familiar approaches are often appealing, but aggregating and dichotomizing information on repeated screening tests without regard to their timing could miss what are otherwise significant implications of a novel prevention strategy to increase routine health monitoring.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

FUNDING ACKNOWLEDGEMENT

This work was facilitated by the Center for AIDS Research at Emory University (P30 AI050409).

Abbreviations

AR1	Autoregressive first order
CDC	Centers for Disease Control and Prevention
GED	General Educational Development
GEE	Generalized Estimating Equations
HIV	Human Immunodeficiency Virus
HR	Hazard Ratio
IQR	Inter-quartile Range
MSM	Men who have Sex with Men
OR	Odds Ratio
PH	Proportional Hazards
RCT	Randomized Controlled Trial
SLR	Standard Logistic Regression
USPSTF	United States Preventive Services Task Force

REFERENCES

1. NPHPPHC. National Prevention Strategy. National Prevention, Health Promotion and Public Health Council; Rockville, MD: 2011.
2. USPSTF. Final Recommendation Statement: Lipid Disorders in Adults (Cholesterol, Dyslipidemia): Screening. 2013.
3. USPSTF. Final Recommendation Statement: Diabetes Mellitus (Type 2) in Adults: Screening. 2008.
4. USPSTF. Final Recommendation Statement: Hepatitis B in Pregnant Women: Screening. 2014
5. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Medicine*. 2010; 8:18. [PubMed: 20334633]
6. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. *BMJ*. 2012;345.
7. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG, Group C. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA*. 2012; 308:2594–604. [PubMed: 23268518]
8. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Annals of Internal Medicine*. 2008; 148:295–309. [PubMed: 18283207]
9. Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C. Reporting randomized, controlled trials of herbal interventions: an elaborated CONSORT statement. *Annals of Internal Medicine*. 2006; 144:364–7. [PubMed: 16520478]
10. Eysenbach G, Group C-E. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. *Journal of Medical Internet Research*. 2011;13.
11. Cook RJ. The design and analysis of randomized trials with recurrent events. *Statistics in Medicine*. 1995; 14:2081–98. [PubMed: 8552888]
12. Kuramoto L, Sobolev BG, Donaldson MG. On reporting results from randomized controlled trials with recurrent events. *BMC Medical Research Methodology*. 2008; 8:35. [PubMed: 18513418]
13. Cook RJ, Bergeron P-J, Boher J-M, Liu Y. Two-stage design of clinical trials involving recurrent events. *Statistics in Medicine*. 2009; 28:2617–38. [PubMed: 19579215]

14. Delucchi KL. Sample size estimation in research with dependent measures and dichotomous outcomes. *American Journal of Public Health*. 2004; 94:372–7. [PubMed: 14998798]
15. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika*. 1981; 68:373–9.
16. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*. 1982;1100–20.
17. Wei L-J, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*. 1989; 84:1065–73.
18. Donaldson MG, Sobolev B, Cook WL, Janssen PA, Khan KM. Analysis of recurrent events: a systematic review of randomised controlled trials of interventions to prevent falls. *Age and Ageing*. 2009; 38:151–5. [PubMed: 19106254]
19. Metcalfe C, Thompson S, Cowie M, Sharples L. The use of hospital admission data as a measure of outcome in clinical studies of heart failure. *European Heart Journal*. 2003; 24:105–12. [PubMed: 12559942]
20. ClinicalTrials.gov. Evaluation of Rapid HIV Self-testing Among MSM (eSTAMP). 2014.
21. SAS 9.3 Macro Language: Reference. SAS Institute; Cary, NC: 2011.
22. 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: randomized controlled trial. *Journal of Medical Internet Research*. 2013;15.
23. Kleinbaum, D.; Klein, M. *Logistic Regression: A Self-Learning Text*. 3rd. Springer Science + Business Media, LLC; New York: 2010.
24. Twisk JW, Smidt N, de Vente W. Applied analysis of recurrent events: a practical overview. *Journal of Epidemiology and Community Health*. 2005; 59:706–10. [PubMed: 16020650]
25. Kleinbaum, D.; Klein, M. *Survival Analysis: A Self-Learning Text*. 3rd. Springer Science + Business Media, LLC; New York: 2012.
26. CDC. PRS Efficacy Criteria for Best-Evidence (Tier I) Risk Reduction (RR) Individual- level, Group-level, and Couple-level Interventions (ILIs/GLIs/CPLs). 2014.
27. Katz DA, Dombrowski JC, Swanson F, Buskin SE, Golden MR, Stekler JD. HIV intertest interval among MSM in King County, Washington. *Sexually Transmitted Infections*. 2013; 89:32–7. [PubMed: 22563016]
28. Lin D, Wei L, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2000; 62:711–30.
29. Cook RJ, Lawless JF. Analysis of repeated events. *Statistical Methods in Medical Research*. 2002; 11:141–66. [PubMed: 12040694]
30. Dean CB, Balshaw R. Efficiency lost by analyzing counts rather than event times in Poisson and overdispersed Poisson regression models. *Journal of the American Statistical Association*. 1997; 92:1387–98.

FIGURE 1.

Power curves from 1,000 trial replications assuming the first and second parameter specifications under Scenario A.

FIGURE 2.

Power curves from 1,000 trial replications assuming the first and second parameter specifications under Scenario B.

FIGURE 3.

Power curves from 1,000 trial replications assuming the first and second parameter specifications under Scenario C.

TABLE 1

Illustrative demographic characteristics of 1,200 simulated participants from a single replication of a randomized trial.

Characteristic	Intervention n	(N=570) (%)	Comparison n	(N=630) (%)
Race/Ethnicity:				
White, non-Hispanic	368	(65)	414	(66)
Black, non-Hispanic	88	(15)	93	(15)
Hispanic	114	(20)	123	(20)
Age group (years):				
18-24	229	(40)	246	(39)
25-34	199	(35)	219	(35)
35-44	79	(14)	106	(17)
45	63	(11)	59	(9)
Education:				
College or Post graduate	195	(34)	234	(37)
Some college or Associate's degree	271	(48)	290	(46)
High school or GED	94	(16)	91	(14)
Less than high school or unknown	10	(2)	15	(2)
Annual income:				
< \$14,999	183	(32)	200	(32)
\$15,000 – \$39,999	176	(31)	214	(34)
\$40,000 – \$74,999	119	(21)	117	(19)
> \$75,000	68	(12)	65	(10)
Unknown	24	(4)	34	(5)
Residential location:				
Urban	345	(61)	407	(65)
Rural	191	(34)	191	(30)
Unknown	34	(6)	32	(5)
Self-identified sexual orientation:				
Homosexual	485	(85)	529	(84)
Bisexual	72	(13)	84	(13)
Heterosexual or unknown	13	(2)	17	(3)

Abbreviations: GED, General educational development.

TABLE 2

Screening characteristics from a single replication of a randomized trial with 1,200 simulated participants assuming the first parameter specification under Scenario A.

Characteristic	Intervention		Comparison	
	(N=570)		(N=630)	
Categorical	n	(%)	n	(%)
Annual testing frequency:				
0	171	(30)	256	(41)
1	141	(25)	203	(32)
2	108	(19)	87	(14)
3	92	(16)	42	(7)
4	34	(6)	29	(5)
5	8	(1)	6	(1)
6	16	(3)	7	(1)
	(N=399 with a total of 905 screening tests)		(N=374 with a total of 691 screening tests)	
Continuous	Median	IQR	Median	IQR
Inter-test interval ^a (days)	79	(36-160)	97	(43-191)

Abbreviations: IQR, Inter-quartile range.

^aDefined as time between consecutive screenings within a year among participants who tested at least once, with time to the first test being recorded from study entry.

TABLE 3

Point estimates from different analytical approaches on 10 replications of a randomized trial each with 1,200 simulated participants assuming the first parameter specification under Scenario A.

Replication # (Intervention N, Comparison N)	SLR		GEE for correlated data ^a				Stratified Cox PH regression for recurrent events			
	Odds of testing 3 times within a year among:		Average 3-month interval-specific odds of testing at least once among:		Average rate of testing (per 1,000 person-days) among:		Total follow-up time model ^b		Gap time model ^c	
	Intervention	Comparison	OR ^d	Intervention	Comparison	OR ^e	Intervention	Comparison	HR ^f	HR ^f
1 (614, 586)	0.36	0.14	2.62	0.43	0.29	1.49	4.19	2.91	1.36	1.36
2 (613, 587)	0.32	0.16	1.97	0.40	0.31	1.31	3.94	3.16	1.18	1.19
3 (582, 618)	0.31	0.20	1.55	0.37	0.35	1.06	3.90	3.55	1.08	1.09
4 (588, 612)	0.29	0.17	1.70	0.42	0.33	1.26	4.14	3.24	1.23	1.23
5 (611, 589)	0.29	0.17	1.69	0.38	0.34	1.13	3.86	3.30	1.13	1.14
6 (609, 591)	0.33	0.17	1.96	0.39	0.34	1.15	3.93	3.32	1.16	1.15
7 (597, 603)	0.41	0.19	2.19	0.48	0.33	1.44	4.59	3.34	1.29	1.30
8 (621, 579)	0.34	0.12	2.75	0.42	0.31	1.36	4.13	3.05	1.30	1.30
9 (596, 604)	0.32	0.14	2.28	0.40	0.31	1.27	3.96	3.10	1.23	1.24
10 (599, 601)	0.32	0.18	1.75	0.40	0.30	1.33	3.95	3.13	1.20	1.19

Note: Results from different approaches presented in a single table for convenience. Magnitudes of the ratio effect estimates are not directly comparable. Consult appendix for frequency distributions of point estimates from 1,000 replications (Figures A5 and A6).

Abbreviations: AR1, Autoregressive first order; GEE, Generalized Estimating Equations; HR, Hazard Ratio; OR, Odds Ratio; PH, Proportional Hazards; SLR, Standard Logistic Regression.

^a AR1 working correlation structure assumed for describing the relationship between 3-month interval-specific testing behaviors within each participant.

^b Modeling repeated screening test patterns over a year.

^c Modeling time between each of the recurrent screening test events.

^d Interpreted as the odds of testing 3 times within a year among intervention arm participants are x times the odds of testing 3 times within a year among comparison arm participants. Estimates in bold represent a statistically significant effect ($P < 0.05$).

^e Interpreted as the odds of testing at least once in a 3-month interval among intervention arm participants are x times the odds of testing at least once in a 3-month interval among comparison arm participants after accounting for correlation from repeated measures. Estimates in bold represent a statistically significant effect ($P < 0.05$).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Interpreted as the instantaneous rate of testing among intervention arm participants is λ times the instantaneous rate of testing among comparison arm participants after accounting for recurrent screening events. Estimates in bold represent a statistically significant effect ($P < 0.05$).