Contemporary Clinical Trials 41 (2015) 152–159

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Evaluating interventions to promote routine preventive screenings: A comparison of analytical outcomes

Akshay Sharma [a,*], Brent A. Johnson [b], Patrick S. Sullivan [a]

[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

ARTICLE INFO

Article history:
Received 19 November 2014
Received in revised form 20 January 2015
Accepted 21 January 2015
Available online 29 January 2015

Keywords:
Preventive screenings
Analytical outcome choices
Randomized trials with recurrent events
Rapid home HIV self-testing
Men who have sex with men

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

1. 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

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.

* Corresponding author at: Department of Epidemiology, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322, United States. Tel.: +1 404 857 8460.
E-mail address: ashar24@emory.edu (A. Sharma).

http://dx.doi.org/10.1016/j.cct.2015.01.014
1551-7144/

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].

2. Methods

2.1. 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 9.3 [21] 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 kit [22] 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 observed [22]) 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.