HHS Public Access

Author manuscript

J Comp Eff Res. Author manuscript; available in PMC 2016 November 11.

Published in final edited form as:

J Comp Eff Res. 2015 May ; 4(3): 227–238. doi:10.2217/cer.15.11.

Value of a small control group for estimating intervention effectiveness: results from simulations of immunization effectiveness studies

Sonja S Hutchins[*,1], **Cedric Brown**[2], **Robert Mayberry**[3], and **William Sollecito**[4]

[1]Office of Minority Health & Health Equity, Centers for Disease Control & Prevention, 1600 Clifton Road, Atlanta, GA 30333, USA

[2]National Center for Immunization & Respiratory Diseases, Centers for Disease Control & Prevention, GA, USA

[3]Department of Community Health & Preventive Medicine, Morehouse School of Medicine, Atlanta, GA, USA

[4]Public Health Leadership Program, University of North Carolina Gillings School of Global Public Health, NC, USA

Abstract

Aim—To improve evidence for public health practice, the conduct of effectiveness studies by practitioners is needed and may be stimulated if knowledge that smaller than usual samples may provide the same reliability of intervention effect size as larger samples.

Materials & methods—We examined reliability of intervention effect using computerized simulations of 2000 hypothetical immunization effectiveness studies from an actual study population and by small (30 and 60) and larger (100 and 200) control groups compared with an intervention group of 200 participants.

Results & conclusion—Across simulated studies, the mean intervention effect (14%) and effect sizes were equivalent regardless of control group size and equal to the actual study effect. These results are relevant for similarly designed and executed studies and indicate that studies with smaller control groups can generate valid and accurate evidence for effective public health practice in communities.

[*]Author for correspondence: Tel.: +1 770 488 8201, Fax: +1 770 488 8140, ssh1@cdc.gov.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Keywords

community intervention study; effectiveness study; immunization effectiveness study; small control group; small sample size

Background

Widespread interest in comprehensive, scientific evidence for public health practice in the USA that examines community studies for intervention effectiveness is relatively new and earnestly took hold in 1996 with formation of the Community Preventive Services Task Force (CPSTF) and subsequent publication of the Guide to Community Prevention Services (Community Guide) [1–5]. With enactment of the Patient Protection and Affordable Care Act in 2010, the CPSTF has taken on a legislated role [6]. Under the provision to prevent chronic disease and improve health, the CPSTF is to provide recommendations from the scientific review of effectiveness, appropriateness and cost–effectiveness of community preventive interventions such as policies, programs, processes or activities designed to affect health at the population level.

The Community Guide provides a broad array of recommendations for communities. Recommendations include health topics addressing population-level behaviors, diseases, injuries and social factors imposing the greatest health burden and offering the broadest range of health promotion and disease prevention interventions for communities. These recommendations draw on the most current comprehensive assessment of scientific evidence from systematic reviews of population-based effectiveness studies [1,2,7]. Such effectiveness studies typically are conducted in the USA after efficacy studies demonstrate an intervention works under ideal conditions [8]. These effectiveness studies are designed specifically to answer the question of whether an intervention works under real world conditions [8]; and therefore, identify interventions likely to lead to effective and efficient public health programs implemented in the community or at the local level.

The group- or cluster-randomized controlled community trial is a common epidemiologic study for assessing community or population-level effectiveness [1,9–11]. Main advantages of this study design are similar to the randomized controlled trial (RCT) for examining efficacy of clinical practice [9]. Instead of double-blind randomized assignment of individuals to control and treatment groups as in the RCT, in the group-RCT random assignment is of a population at a group or cluster level, while the unit of observation is usually at the individual level [9]. The group or cluster RCT is among the best designs for providing evidence for population-level effectiveness, because it has some advantages of an RCT, providing high levels of internal validity [11]. The group randomized trial also overcomes any impracticality of double blinded randomization of individuals in a population- level effectiveness study. As with the RCT, cost is one major drawback to a large group controlled trial. A large study may be difficult for local public health agencies and communities to conduct because of limited resource availability and competing service priorities for available resources.

In fact, a survey of key public health informants of epidemiologists practicing for more than 20 years in local and state public health agencies identified financial constraints in making comprehensive, evidence-based decisions in public health at the community level because of the perceived need for large studies [12]. Difficult questions may not be addressed or subpopulations may not be evaluated in effectiveness studies because of the concern about the cost of large studies. In a recent review of effectiveness studies published as background articles for recommendations of the Community Guide, many effectiveness studies did not include a control group nor are they able to infer to certain racial and ethnic minority subpopulations because of insufficient sample size [1,7,11]. Ideally, relatively inexpensive, valid and reliable methods are likely of value in improving the evidence base for public health practice and community health.

To improve the amount and quality of available evidence for public health practice and practice-based evidence, an unbalanced design, using a relatively small sample size for a control group with a larger intervention group could be considered. In contrast to studies with no control group, inclusion of a control group improves the internal validity or attribution of intervention effect. As a comparison group, the control group typically implements the standard of care and is not exposed to the intervention but has characteristics similar to those of the intervention group that predict outcome [9,13].

For many researchers, a small control group seems counterintuitive. First, a small sample may lack the precision for providing reliable answers to the question under investigation. In addition, an unbalanced design with a small sample size may not allow for sufficient statistical power to detect a difference between intervention and control groups (i.e., low power) [13]. A fundamental concern is that group-level assignment of control and intervention groups also adds another level of variation (intergroup variance) to consider in analyzing results to account for any intragroup or intra-class correlation among members in the same group [9]. A critical review seems warranted of when and under what circumstances a small control group can be used in group-randomized, effectiveness studies. To answer the question does a small control group size provides valid results, this paper addresses the question using a different approach to describe the reliability in estimating an intervention effect from an effectiveness study (group randomized trial) with a small control group compared with larger control groups. The approach of computerized simulations is more understandable to a broader audience of practitioners. Our objective is to examine whether a small control group yields equivalent size of intervention effects, defined as the simple difference between two proportions [14].

Materials & methods

Study design

Computerized simulations of hypothetical effectiveness studies were performed to examine equivalence of intervention effects. As an example, simulations were performed with data from an actual effectiveness study of an immunization intervention. The study was conducted in early 1990s by the first author and colleagues. This actual effectiveness study was a group-randomized controlled intervention trial reviewed and rated by the CPSTF as having greatest suitability for assessing effectiveness of immunization interventions [15,16].

The actual effectiveness study included large intervention and control groups (approximately 14,000 children combined).

Actual study population

Study population, design & intervention effectiveness measurements—In the actual study population, change in vaccination levels or coverage among children 13–35 months of age in seven of 48 sites for the Special Supplemental Nutrition Program for Women, Infants and Children (WIC) in Chicago from 1990–1992 was compared between intervention and control groups [16]. Two WIC sites were randomly assigned to a control group and five sites were randomly assigned to three different types of intervention groups with an overall three-part intervention of vaccine assessment, vaccine referral and a voucher incentive (e.g., two sites to one intervention, two sites to second intervention and one site to third intervention). Group assignment was chosen because randomization of individuals in a WIC site was not operationally practical as an intervention. Methods for assessment of vaccination status and implementation of a voucher incentive were the same at all intervention sites; intervention groups only varied by referral to a healthcare provider for immunization either on site in the WIC clinic, down the hall to a pediatric clinic or off site to a provider in the community [16]. The voucher incentive for all interventions included issuing a 3-month supply of food vouchers if vaccinations were up to date instead of a 1-month supply.

The intervention effect was defined as the difference in changes from baseline vaccination coverage (percentage of up-to-date children for vaccination) after the intervention was introduced at a WIC site between intervention and control groups. Vaccination coverage was measured by conducting cross-sectional surveys of 13- to 35-month-old children in a site before the intervention and 1 and 2 years after the intervention began. For the purposes of simulating hypothetical studies, the change from the baseline percentage of up-to-date children for vaccination was only 1 year after the intervention began, and this change in coverage in control sites was subtracted from the change in baseline coverage in the intervention sites (Table 1). For the purpose of generating samples of control and intervention populations in simulated studies, all children in control sites were combined into one control group and all intervention sites were combined into one intervention group. The arithmetic difference between control and intervention groups was chosen as the method for estimating intervention effect from simulated studies [14]. Of five intervention sites, approximately 1000 children were in each of the baseline and follow-up surveys and nearly 400 or more were in these surveys for the two control sites combined. Observed intervention effect in the actual study was 14.4% (95% CI: 13.5–15.2%).

Surveyed children in study sites had similar demographic characteristics, extent of WIC participation and receipt of other federal assistance at baseline as those in seven study sites and all 48 WIC sites in Chicago. Because these characteristics were known to affect vaccination status of children, the survey sample of children was considered representative of all WIC enrolled children in Chicago [16].

Assessment of intragroup correlation—Intraclass correlation (ICC) was assessed for baseline vaccination coverage by intervention and control groups in the actual study to determine if the ICC was significant and should be included in estimating uncertainty of the intervention effect. Using standard computations for ICC [9], approximately 3% of the total variance in the difference in vaccination coverage between the control and intervention groups at baseline was due to correlation among individuals within the same group (control or intervention groups). Because the ICC was very small, it was excluded from simulated studies.

Simulations—Computerized simulations generated hypothetical studies from the actual study population and provided distributions of intervention effects by control group size. From these distributions, intervention effects were examined by precision of mean effects, statistical power and mean square error (MSE). Precision, or the width of the 95% CI, was computed as an epidemiological measure of uncertainty, describing the magnitude of random error in estimating a population parameter from a sample [14,17]. Because the intervention effect was a difference between proportions, precision was computed by subtracting the lower 95% confidence limit from the upper confidence limit [14,17]. Statistical power was chosen as another measure, because it is a fundamental consideration in estimating sample size requirements for a study. Although power is not a direct measure of the degree of uncertainty, power is a critical element in sample size requirements and measures the probability that a study detects a statistically significant difference under the null hypothesis if there is a true difference in the population [14]. MSE was chosen because it represents an index of uncertainty. As a measure of uncertainty that captures both random error and systematic bias, it is an average closeness of an estimator to the population parameter and equals the mean of squared deviations of the observed estimate from the true parameter [18]. The true parameter was the intervention effect of 14%, previously mentioned from the actual study, and the mean squared deviations of the observed effects from simulated studies were each subtracted from the actual study effect of 14% for computing the MSE.

In simulated studies, sample sizes as small as 30 and as large as 200 were examined for the control group. A sample size as small as 30 was chosen because this size is considered large enough to assume that variability of an intervention estimate can be closely approximated by a normal distribution from the central limit theorem [19]. The intervention group included only a large sample size of 200 participants. To select different sizes of a control group, a bootstrap method of random sampling with replacement was performed. The bootstrap macro is a commonly used and well-established macro named Jackboot. It has been updated over the years and we used the latest published version at the time of this study [20,21]. In our study we used the macro to produce estimates with the bias-corrected accelerated methods to produce CIs and to observe both plots and statistics to check if the distribution was approximate normal. These bias-corrected accelerated methods allowed us to correct, if needed, the percentile interval for bias and skewness. To display simulations by control group size, for convenience Minitab 13 software was used because a figure and descriptive statistics could be shown side by side.

When the sample size for both the control and intervention groups were 200 participants, the design was considered balanced; otherwise, it was considered an unbalanced design. For each sample size (30, 60, 100 and 200) of the control group, 2000 simulated studies were created; this number was adequate for computing an approximate or asymptotic 95% CI for the mean intervention effect. To obtain the intervention effect for each of the 2000 simulated studies, the bootstrap macro also enabled analysis of the randomly selected children from each of the four cross-sectional surveys (two surveys each (baseline and 1-year follow-up) for the control group and intervention group) [20–22]. Percentage of children aged 13–35 months who were up to date for vaccination was computed from each sample in each of the four surveys.

Statistical analysis—An empirical distribution and asymptotic statistical methods described the intervention effect from the 2000 simulated studies by control group size and reliability of estimation. Asymptotic methods were chosen, because they are most commonly applied in statistical practice [19]. Frequency distribution of estimated intervention effects were, in general, normally distributed [19] and descriptive statistics were computed in MINITAB 13 software for convenience as previously mentioned [23]. Precision of mean intervention effect for simulated studies for each sample size of the control group was derived from the 95% CI in MINITAB (uncorrected for bias). The mean intervention effect was chosen because it was the most common estimate. In MINITAB, the 95% CI was computed by multiplying the approximate standard error by 1.96 [24]. For convenience, mean power was computed by control group size in SAS version 9.3 [25]. A two-sided Type I error (α) of 0.05 was used to estimate power from a study measuring the difference of two means. Although the first difference computed in the actual study at baseline was of two proportions, the second difference (difference of the 1-year changes from baseline vaccination coverage between control and intervention groups) was of two mean proportions, which could have values between -1 to 1. Power was one minus the probability of a Type II error (β), which was derived from a one-sided Z value. From simulated studies by control group size, squared errors from the intervention effect in the actual study were computed as the MSE.

Results

Simulations

Intervention effect—Mean intervention effect for the simulated studies with 200 children in the intervention group and different sample sizes of children in the control group was the same as the observed estimate of 14% in the actual study, regardless of size of control group (Figures 1–4). For instance, 14% was the mean intervention effect from simulated studies with a control group of 30 participants, which was the same for studies with larger control groups. Although the empirical distribution of intervention effects by study design showed that most simulated studies found an elevated intervention effect, the percentage of simulated studies with an elevated effect increased as control group size increased (30 participants: 85%; 60 participants: 92%; 100 participants: 96%; and 200 participants: 99%).

Measures of uncertainty—Precision of mean intervention effect varied by sample size of the control group (Figure 5A). Imprecision was largest (±1.2%) for the study design with a control group of 30 participants and decreased for designs with increasing size of control group. Imprecision was lowest (±0.6%) for the design with a control group of 200, which was one half precision of the design with a small control group. In general, estimated mean intervention effects for simulated studies by control group size based on precision of estimated effect was equivalent regardless of sample size of control group.

Average power in detecting a true difference was high regardless of control group sizes, because the actual intervention effect was a relatively large difference (14%) (Figure 5B). With a two-sided Type I error of 0.05, the mean power for studies with a control size of 30 was approximately 99%. Power was the same for studies with larger sample sizes for control groups.

MSE slightly varied by size of control group. Because study bias was minimal, MSE approximated variance of distribution of intervention effects by size of the control group. The MSE was 1.7×10^{-2} for a small control group of 30 participants, and was lower for larger control groups (60 participants: 0.9×10^{-2}; 100 participants: 0.8×10^{-2}; and 200 participants: 0.7×10^{-2}).

Discussion

In thousands of stimulated studies with control groups of 30, 60, 100 or 200 participants compared with an intervention group of 200 participants the average intervention effect was the same as the effect in the actual study population regardless of sample size of the control group. The empirical distribution suggests that, on average, the mean intervention effect (14%) from stimulated studies would have been the same as the effect in the actual study population if the group-RCT had a control group as small as 30 participants. Our close results, regardless of control group size, may be explained by a well-designed and executed actual study with a representative population of all children enrolled in WIC [16]. It is well known in sampling theory that truly random samples (i.e., not affected by selection bias) result in unbiased estimates of true effects. Since we have randomly sampled from a dataset, it is expected that, regardless of sample size, we should find on average the same effect estimates as the actual study data. Despite this fact from sampling theory, most practitioners still think that very large controls are needed and do not understand or make use of limited power computation tools that are available. This paper confirms that a large control group is unnecessary and will improve the understanding of a broad group of public health practitioners.

In this paper we chose to represent intervention effect as the simple difference between means, because it directly measures the intervention effect most meaningful to public health practitioners, and is consistent with national, state and local goals and objectives. We acknowledge some controversy in the choice of intervention effect presented here, and suggest that this may be worthy of further consideration such as computing multiple definitions of intervention effect size for comparative purposes as for comparative effectiveness research studies using very small sample sizes [26].

In this paper, although a similar mean intervention effect is found in simulated studies by control group size, examination of uncertainty in effect did vary slightly. Of several measures of uncertainty examined, most suggest that a small control group (30 and 60 participants) is likely to provide a reasonable degree of certainty using the binomial distribution. Regardless of control group size, the size of estimated intervention effects was equivalent. In addition, imprecision is approximately 1% using a binomial distribution, mean power of studies is nearly 100% and variability of observed intervention effect from the actual estimate (intervention effect found in actual study population) or MSE is low (approximately 1–2%). The average power in detecting a true difference was high regardless of control group sizes because of the relatively large intervention effect (14%) in the actual study and large size of the intervention group, which was the same size in all simulated studies. Again, the study design and representativeness of WIC-enrolled children in Chicago may also explain similar intervention effect sizes regardless of control group size.

Although assessment of a small versus larger control groups is limited because ICC was not accounted for in total variance in simulated studies as mentioned previously, the ICC at baseline was very low (3%) and adjusted estimates of precision including the ICC in baseline vaccination coverage levels found similar results as unadjusted estimates. Moreover, ICC is a relatively constant value regardless of control group size, and comparisons are made of relative and not absolute measures of uncertainty (precision, power and MSE). Overall, findings of uncertainty (excluding ICC) seem to provide a reasonable assessment of random error, because nearly all of variability in the actual group-randomized trial is between groups and low intragroup correlation is likely acceptable in public health practice for communities in which a large intervention effect has been found. Furthermore, a recent literature review of ICCs in estimating sample size and power in community trials indicate that ICCs vary widely, depend on the outcome measure, depend on adjustment of covariables and should be considered on a study-by-study basis [27–29]. For example, estimation of ICC from 23 variables in the Minnesota Heart Health program ranged from 0.002 to 0.012 [27]. Moreover, immunization effectiveness studies in WIC sites from which the Community Guide has made recommendations on improving vaccination coverage excluded ICCs because they were small. These small ICCs and the small ICC in the present study suggest that careful interpretation of small ICCs from group-randomized controlled community trials relative to intervention effect seems warranted.

As mentioned, to date there are considerable limits to the available evidence base for public health practice [7,11]. Published community effectiveness studies typically have large sample sizes [11,30,31]. However, sample sizes in community-based research typically are small [32,33], and intervention effectiveness seems often undeterminable because of sample size. These small studies are generally not published in the literature (publication bias). In fact when the medical and social science literatures are searched using key words of ‘small community studies or trials,’ case studies in communities are mostly found [11,30,31,34]. If a small sample size is mentioned, it only refers to the number of groups examined and not to the number of participants in a group [34–36]. In the present analysis, we examined if there are circumstances in which a small control group is as reliable in assessing an intervention effect and found that a small sample size for the control group is as reliable as larger sample sizes. This type of effectiveness study may provide adequate internal validity of the study at

an affordable cost. Although an immunization example has been presented, future research with the same or different data sources could include other interventions in which the outcome measure is a mean or count, both control and intervention groups are small or balanced and the intervention effect is expected to be smaller or larger than found in the present study.

The present study could have implications for community-based, effectiveness research, particularly for racial and ethnic minority subpopulations in local communities. If a minority group has small numbers in a community, the control group instead could have been a small population group. Typically, minority populations are oversampled to ensure an adequate number is included in the study [37], which comes at a financial cost that may not be available for all communities and local public health agencies because of competing health or other service priorities. This method may improve evidence-based public health and practice-based evidence in communities by now including valid and reliable evidence applicable to minority populations at a more affordable price.

Currently, public health practitioners are likely concerned about results from small community studies [38]. We found a large probability of detecting elevated intervention effects among simulated studies by control group size (i.e., 85–99% of studies), varying from the smallest to largest control group size based on the empirical distribution. If the empirical distribution instead of the binomial distribution is used to account for uncertainty, public health practitioners may want then to weigh the risk of having an 85 versus 99% likelihood of finding an intervention effect with a small control group versus a large control group against the cost and public health significance of findings. Guidance on public health situations in which small studies provide less or equal certainty as large studies and are cost-effective may help public health agencies and community organizations contribute more to the scientific evidence for public health practice and stimulate a growing literature on practice-based evidence that improves population health. Consideration of public health significance of studies could include community needs for more or less certainty in intervention effect, available funds for study, scope of implementation of findings, immediate use of findings, target population for intervention and consequences of being wrong.

This concept of contextual significance has been applied to findings from small clinical trials. For instance, recommendations from a 2001 Institute of Medicine report on the value of small clinical trials indicate that properly designed trials with small samples can contribute to efficacy in particular situations such as rare diseases, unique study populations, individually tailored therapies, isolated environments, emergency situations, a public health emergency and restricted resources with an important need [39]. These recommendations illustrate which situations clinical significance is considered in accepting results from small trials. Public health practitioners too could consider the public health significance of findings for communities in deciding to apply results from effectiveness studies with a small sample size.

Conclusion

The primary implication of our findings is that an immunization effectiveness study using a group-RCT with a small control group seems to provide an equivalent estimated intervention effect as does a study with a larger control group, particularly for public health practice in which programs could be implemented and estimated to have a substantial impact within the first year of implementation. This is an example of a study with a sizable intervention effect (14%) and small (near zero) background changes in vaccination in the community. Future studies that seek to assess the effectiveness of immunization interventions with a similar study design may consider using a smaller, unbalanced control group to detect changes in mean proportions. This study design may also be used to measure effects from other community interventions. In addition, the concept of public health significance and guidance for application of this significance need further exploration.

Future perspective

To improve public health in the USA, evidence-based recommendations of the CPSTF have and will continue to play an integral role in effective public health practice in communities. The Patient Protection and Affordable Care Act requires the CPSTF to report annually on gaps in topics and populations found in the scientific evidence for community preventive interventions. To close these gaps, public health and community health practitioners are important sources for contributing to the relevant evidence base for effective practice in communities. Knowledge of when smaller sample sizes can reliably identify effective interventions is likely to enable practitioners to initiate and conduct effectiveness research to improve community health.

Acknowledgments

The authors thank M Ibrahim, R Bernier, C Poole, J Kaufman and D Cattelier for their epidemiological and statistical advice during an early draft of this manuscript for a dissertation and P Smith for his statistical advice and assistance during the preparation of this manuscript.

References

Papers of special note have been highlighted as:

• of interest

1. Zaza S, Lawrence RS, Mahan CS, et al. Scope and organization of the guide to community preventive services. Am J Prev Med. 2000; 18(1S):27–34.

2•. Truman BI, Smith-Akin CK, Hinman AR, et al. Developing the guide to community preventive services – overview and rationale. Am J Prev Med. 2000; 18(1S):18–26. This foundational article on the Community Guide describes the purpose and rationale.

3•. US Preventive Services Task Force. Guide to Clinical Preventive Services. Williams and Wilkins; PA, USA: 1989. http://wonder.cdc.gov For WIC studies also see www.thecommunityguide.org/vaccines/RRWICsettings.html

4. Heffner JE. Does evidence-based medicine help the development of clinical practice guidelines? Chest. 1998; 113:172S–178S. [PubMed: 9515888]

5. Agency for Healthcare Research and Quality. Clinical Practice Guidelines. www.ahrq.gov

6•. Patient Protection and Affordable Care Act. 111th Congress of the United States. US Congress; Washington, DC, USA. 2010; This is the legislation that solidifies the role of the Community Preventive Services Task Force in providing evidence-based recommendations for population health in the USA

7. Anderson LM, Brownson RC, Fullilove MT, et al. Evidence-based public health policy and practice: promises and limits. Am J Prev Med. 2005; 28(5S):226–230. [PubMed: 15894157]

8. Haddix AC, Teutsch SM, Shaffer PA. , et al. Prevention Effectiveness: a Guide to Decision Analysis and Economic Evaluation. Oxford University Press; NY, USA: 1996.

9•. Murray DM. Design and Analysis of Group-Randomized Trials. Oxford University Press; NY, USA: 1998. This is a major textbook on the design, implementation and analysis of group-randomized trials, a commonly conducted research study in communities

10. Shadish W, Cook T, Campbell D. Experimental and Quasi-Experimental Designs for Generalized Casual Inference. Houghton Mifflin; Boston, MA, USA: 2002.

11. Briss PA, Zaza S, Pappaioaniou M, et al. Developing an evidence-based guide to community preventive services-methods. Am J Prev Med. 2000; 18(1S):35–43. [PubMed: 10806978]

12. Hutchins S. Improving effectiveness studies of immunization interventions for public health policy and practice. Dissertation abstract. 2002. www.nsf.gov/sbe/srs/ssed/start.htm

13. Rothman KJ, Greenland S. Accuracy consideration in study design. In: Rothman KJ, Greenland S, editorsModern Epidemiology. 2. Lippincott Williams & Wilkins; PA, USA: 1998.

14. Fleiss JL, Levin B, Cho Palik M. Statistical Methods for Rates and Proportions. 3. John Wiley, Inc; 2003. 161–162. (eBook)http://onlinelibrary.wiley.com/book/10.1002/0471445428

15. Task Force on Community Preventive Services. Recommendations regarding interventions to improve vaccination coverage in children, adolescents, and adults. Am J Prev Med. 2000; 18(1S): 92–96. [PubMed: 10806981]

16. Hutchins SS, Rosenthal J, Eason P. Effectiveness and cost-effectiveness of linking the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) and immunization activities. J Public Health Policy. 1999; 20:408–426. [PubMed: 10643168]

17. Poole C. Low p-values or narrow confidence intervals: which are more durable? Epidemiology. 2001; 12:291–294. [PubMed: 11337599]

18. Efron B, Morris C. Stein's paradox in statistics. Scientific Am. 1977; 236(5):119–127.

19. Mendenhall W, Scheaffer RL, Wackerly DD. Mathematical Statistics with Applications. Duxbury Press; CA, USA: 2001.

20. Efron B, Tibshirani R. Bootstrap methods for standard error, confidence intervals, and other measures of statistical accuracy (with comments and rejoinder). Statistical Science. 1986; 1(1):54–77.

21. Efron B, Tibshirani R. An introduction to the bootstrap. Chapman and Hall Monographs on Statistics and Applied Probability. Chapman and Hall; NY, USA: 1993.

22•. SAS Institute. Jackknife and Bootstrap Analyses. 2000. www.sas.comIn this reference, the in-depth description of the Jackboot macro is provided. It specifically describes how simulated studies were conducted from randomly selecting control groups with replacement and computing outcomes

23. MINITAB. Minitab 13 more powerful and easier to use: release 13. 2002. www.minitab.com

24. Fleiss JL. Statistical Methods for Rates and Proportions. 2. John Wiley & Sons; NJ, USA: 1981. 101–104.

25. SAS Institute. SAS, version 9.3. 2013. www.sas.com

26•. Sullivan GM, Fenin R. Using effect size – or why the p value is not enough. J Grad Med Educ. 2012; 4(3):279–282. Describes the different definitions used to measure intervention effect. [PubMed: 23997866]

27•. Hannan PJ, Murray DM, Jacobs DR, et al. Parameters to aid in the design and analysis of community trials: intraclass correlations from the Minnesota Heart Health Program. Epidemiology. 1994; 5(1):88–95. Illustrates a large community trial with low intraclass correlations for many measures of intervention effect. [PubMed: 8117787]

28. Hade EM, Murray DM, Pennell ML, et al. Intraclass correlation estimates for cancer screening outcomes: estimates and applications in the design of group-randomized cancer screening studies. J Natl Cancer Inst Monogr. 2010; 40:97–103.

29. Baldwin SA, Murray DM, Shadish QR. Intraclass correlation associated with therapists: estimates and applications in planning psychotherapy research. Cogn Behav Ther. 2011; 40:15–33. [PubMed: 21337212]

30. Hopkins DP, Briss PA, Ricard CJ, et al. Reviews of evidence regarding interventions to reduce tobacco use and exposure to environmental tobacco smoke. Am J Prev Med. 2001; 20(2S):16–66. [PubMed: 11173215]

31. Hohmann AA, Shear MK. Community-based intervention research: coping with the “noise” of real life in study design. Am J Psychiatry. 2002; 159:201–207. [PubMed: 11823259]

32. Mayberry RM, Daniels P, Yancey EM, et al. Enhancing community-based organizations’ capacity For HIV/AIDS education and prevention. Eval Program Plann. 2009; 32:213–220. [PubMed: 19376579]

33. Richter D, Prince M, Potts L, et al. Assessing the HIV prevention capacity building needs of community-based organizations. J Public Health Manag Pract. 2000; 6:86–97.

34. Brookmeyer R, Chen YQ. Person-time analysis of paired community intervention trials when the number of communities is small. Stat Med. 1998; 30(17):2121–2132.

35. Donner A, Klar N. Statistical considerations in the design and analysis of community intervention trials. J Clin Epidemiol. 1996; 49:436–439.

36. Kerry SM, Bland JM. Statistical notes: sample size in cluster randomization. BMJ. 1998; 316:549. [PubMed: 9501723]

37. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey. Note on 2007–2010 Sampling Methodology. 2009. www.cdc.gov

38. Barker ND. Basic concepts of statistics. In: Stroup D, Teutsch S, editorsStatistics in Public Health: Quantitative Approaches to Public Health Problems. Oxford University Press; NY, USA: 1998. 30

39•. Institute of Medicine. Small Clinical Trials: Issues and Challenges. National Academy Sciences; Washington, DC, USA: 2001. Highlights the clinical context in which small clinical trials are useful and calls on additional research for alternative methods such as simulations for testing the efficacy of small trials

Executive summary

- Findings from this paper that sometimes smaller than usual samples may provide the same reliability of intervention effect size as larger samples may encourage public health and community health practitioners to conduct effectiveness research and strengthen the evidence base for the Guide to Community Preventive Services (Community Guide), which seeks to improve evidence-based public health in the USA.
- To show the value of a small control group in a group-randomized trial, we performed computerized simulations of hypothetical immunization effectiveness studies using data from an actual study population and found results likely to interest practitioners in initiating effectiveness research in their communities.
- In 2000 simulations of hypothetical group-randomized controlled trials examining immunization effectiveness by comparing either a small (30 and 60 children) control group or larger (100 and 200) control groups with the same large intervention group of 200 children, an equivalent estimated intervention effect was found for trials regardless of control group size when there is a sizable intervention effect (14%) in 1 year and small (near zero) background changes in vaccination in the community.
- Although the intraclass (ICC) or intragroup correlation was excluded from our analyses because baseline ICC was small (3% of total variance) relative to total variance in the actual study, our reliability findings of intervention effects provide a reasonable assessment of random error because comparisons between trials with small and larger control groups are relative and not absolute measures of uncertainty and unadjusted and adjusted ICC precisions are similar.
- Furthermore, low intragroup correlations are likely acceptable in public health practice in communities when there is a large intervention effect and consideration of public health significance could complement decision making as has been discussed by the Institute of Medicine in a report on the issues and challenges of small clinical trials.
- Potential gains from a small control group compared with no control group are improved internal validity at an affordable price for practitioners.
- Although we may not generalize to all studies, results may be relevant to similarly well-designed and executed effectiveness studies measuring a difference between mean proportions and those seeking to include minority populations.
- As the role of the Community Preventive Services Task Force matures in generating the evidence base for public health practice for the nation, relevant effectiveness research initiated and conducted by public health and community health practitioners will be increasingly important as a critical

resource for identifying what works in improving population-level health in communities.

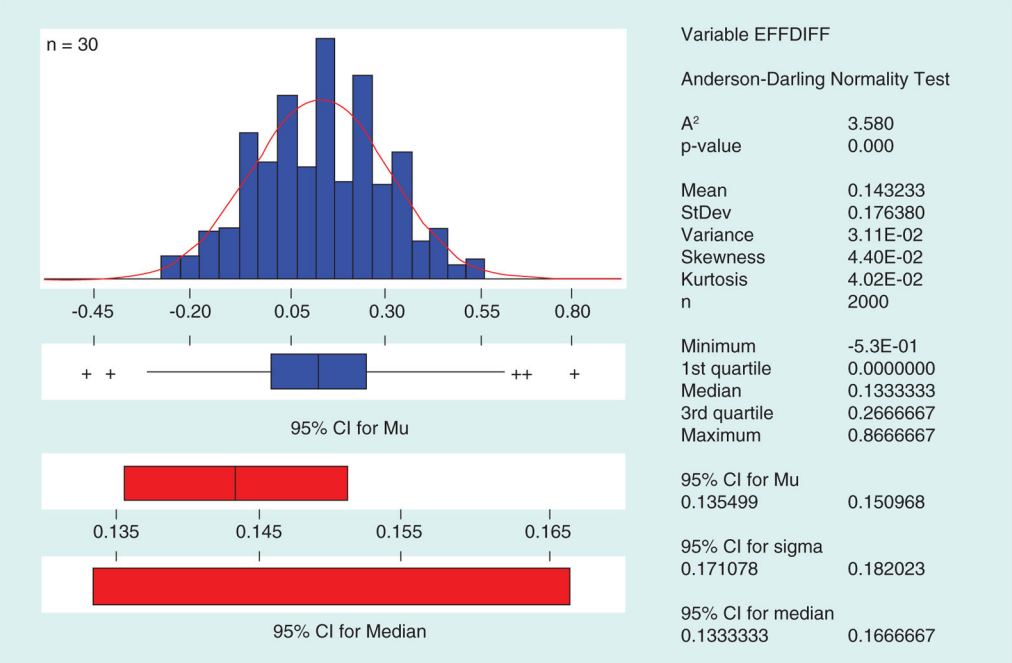

Figure 1. Descriptive statistics of intervention effects from 2000 simulated studies with a control group of 30 children and an intervention group of 200 children

EFFDIFF: Intervention effect; n: Number of simulated studies; Mu: Mean; StDev: Standard deviation.

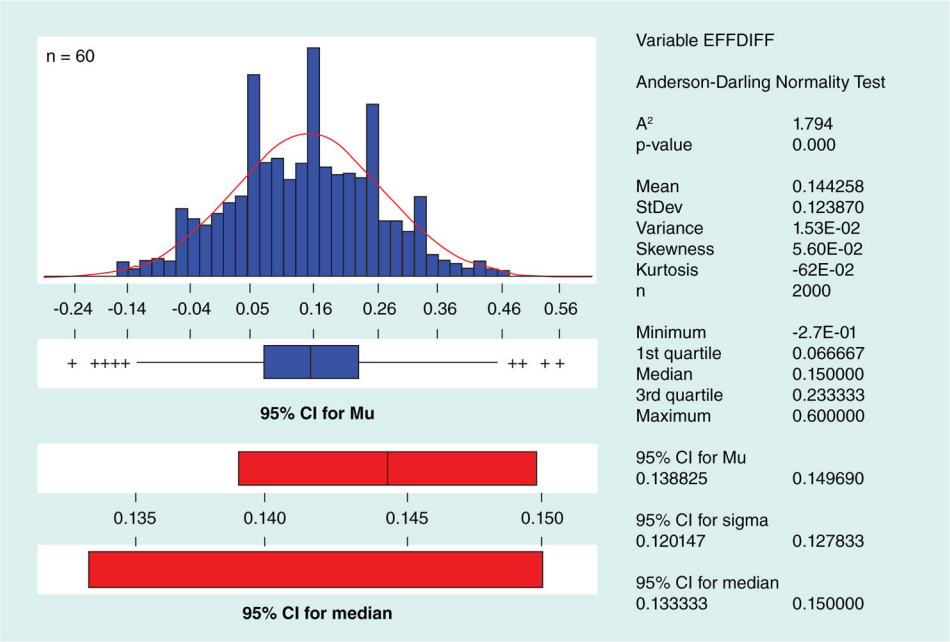

Figure 2. Descriptive statistics of intervention effects from 2000 simulated studies with a control group of 60 children and an intervention group of 200 children

EFFDIFF: Intervention effect; n: Number of simulated studies; Mu: Mean; StDev: Standard deviation.

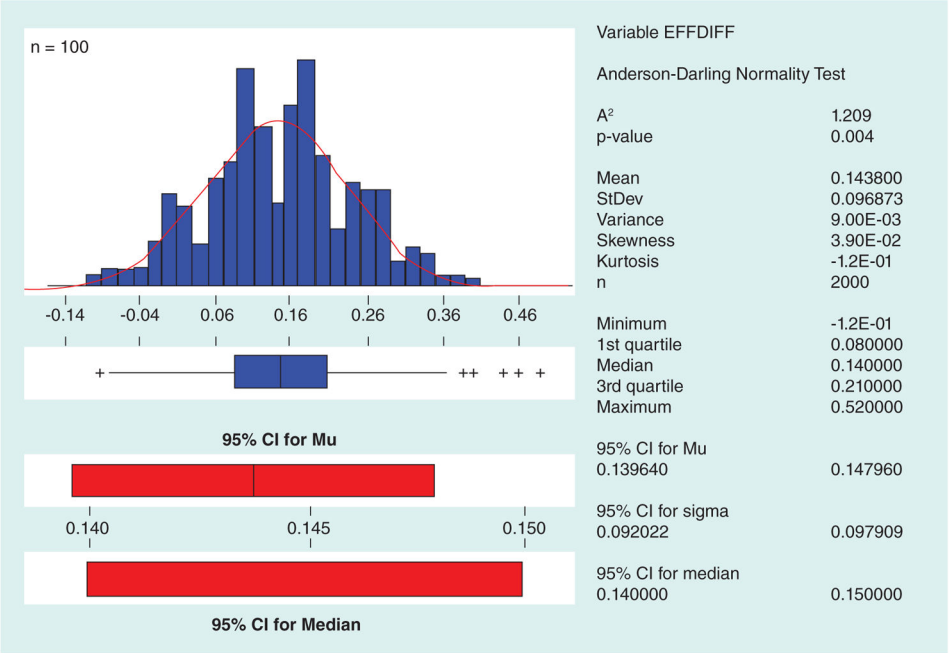

Figure 3. Descriptive statistics of intervention effects from 2000 simulated studies with a control group of 100 children and an intervention group of 200 children

EFFDIFF: Intervention effect; n: Number of simulated studies; Mu: Mean; StDev: Standard deviation.

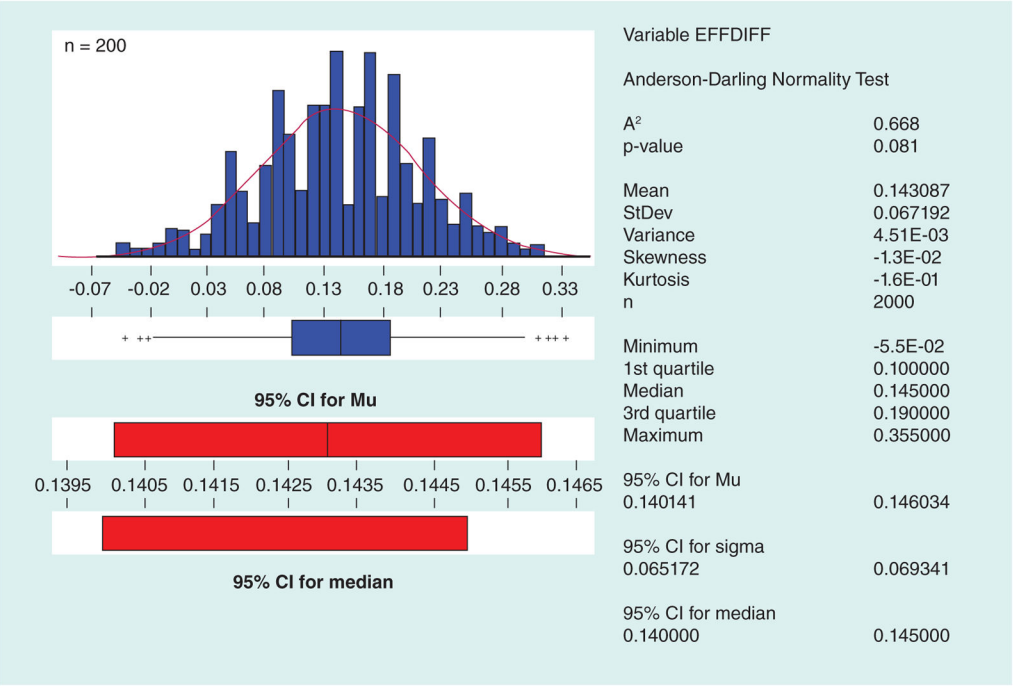

Figure 4. Descriptive statistics of intervention effects from 2000 simulated studies with a control group of 200 children and an intervention group of 200 children

EFFDIFF: Intervention effect; n: Number of simulated studies; Mu: Mean; StDev: Standard deviation.

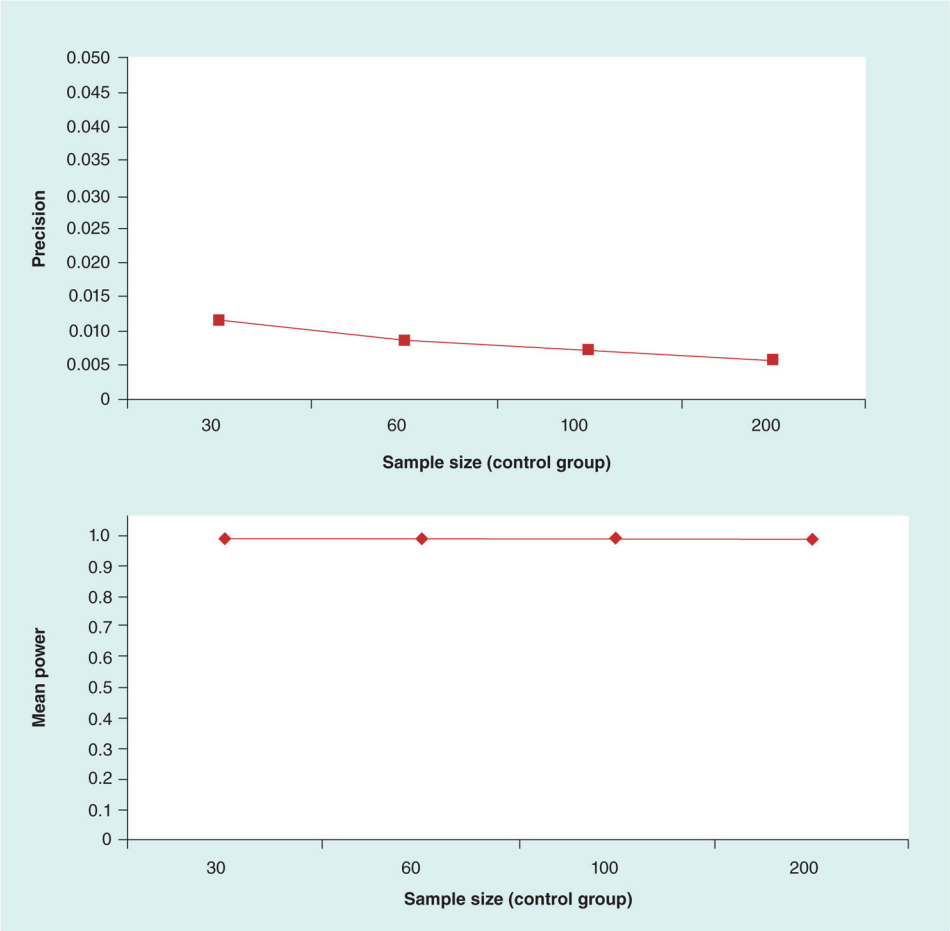

Figure 5. Measures of uncertainty and power by sample size in control group in 2000 simulated studies

Precision is the 95% CI and mean power is average statistical power of simulated studies in detecting a difference using a two-sided type I error of 0.05.

Table 1

Effect of immunization intervention in women, infants and children program sites, Chicago, 1990–1991.

Group	Baseline		Post intervention		Change from baseline
	n	% (95% CI)	n	% (95% CI)	
Control	576	69 (65–72)	389	65 (60–70)	−4 (−10–3)
Site 1	347	73 (69–78)	224	69 (63–75)	−4 (−12–2)
Site 2	229	61 (55–67)	165	60 (53–67)	−1 (−11–9)
Intervention	1015	59 (56–62)	993	69 (67–72)	11 (7–15)
Site 1	255	69 (64–75)	154	70 (63–77)	1 (−8–10)
Site 2	210	58 (51–64)	216	67 (61–73)	10 (4–19)
Site 3	176	59 (52–66)	192	74 (68–80)	15 (5–24)
Site 4	168	48 (41–56)	198	67 (61–74)	19 (9–29)
Site 5	206	55 (48–62)	233	70 (67–72)	15 (6–24)

Intervention effect = 14.4% (95% CI: 13.5–15.2%). Intervention effect is defined as the difference in the change from baseline percentage of up-to-date children between intervention and control groups 1 year after intervention began.

Percentage of up-to-date children 13–35 months of age.