HEALTH EDUCATION RESEARCH
Theory & Practice

Vol.21 no.5 2006
Pages 688–694
Advance Access publication 31 August 2006

Evaluating the impact of health promotion programs: using the RE-AIM framework to form summary measures for decision making involving complex issues

Russell E. Glasgow[1]*, Lisa M. Klesges[2], David A. Dzewaltowski[3], Paul A. Estabrooks[1] and Thomas M. Vogt[4]

Abstract

Current public health and medical evidence rely heavily on efficacy information to make decisions regarding intervention impact. This evidence base could be enhanced by research studies that evaluate and report multiple indicators of internal and external validity such as Reach, Effectiveness, Adoption, Implementation and Maintenance (RE-AIM) as well as their combined impact. However, indices that summarize the combined impact of, and complex interactions among, intervention outcome dimensions are not currently available. We propose and discuss a series of composite metrics that combine two or more RE-AIM dimensions, and can be used to estimate overall intervention impact. Although speculative and, at this point, there have been limited empirical data on these metrics, they extend current methods and are offered to yield more integrated composite outcomes relevant to public health. Such approaches offer potential to help identify interventions most likely to meaningfully impact population health.

[1]Kaiser Permanente Colorado, Clinical Research Unit, 335 Road Runner Lane Penrose, Denver, CO 81240, USA, [2]Mayo Clinic College of Medicine, Rochester, MN, USA, [3]Kansas State University, Manhattan, KS, USA and [4]Kaiser Permanente Center for Health Research, Honolulu, HI, USA
***Correspondence to: R. E. Glasgow.**
E-mail: russg@ris.net

Introduction

Health promotion and education programs seek to make meaningful improvements in population health, often with limited resources. This is a complex, multilevel challenge [1, 2] and presently, there is little agreement on the criteria necessary to conclude that a program has produced a significant public health impact [3–5]. Standard metrics that accurately summarize complex and multidimensional outcomes would be very helpful.

The Reach, Effectiveness, Adoption, Implementation and Maintenance (RE-AIM) framework offers a comprehensive approach to considering five dimensions important for evaluating the potential public health impact of an intervention [6, 7]. The model includes (i) Reach, the percent and representativeness of individuals willing to participate; (ii) Effectiveness, the impact of the intervention on targeted outcomes and quality of life; (iii) Adoption, the per cent and representativeness of settings and intervention staff that agree to deliver a program; (iv) Implementation, the consistency and skill with which various program elements are delivered by various staff and (v) Maintenance, the extent to which individual participants maintain behavior change long term and, at the setting level, the degree to which the program is sustained over time within the organizations delivering it (www.re-aim.org). RE-AIM builds upon conceptual work by Rogers [8] and Green and Kreuter [2] and focuses attention on these five specific factors.

To date, RE-AIM has only been applied to a single dimension at a time. An overall metric, combining two or more RE-AIM dimensions, would be more useful for making policy decisions

doi:10.1093/her/cyl081

than five separate measures. This paper proposes several combined impact indices, and provides a rationale, calculation and discussion of the advantages and limitations of each. Most indices proposed combine two measures because (i) this is closer to the raw data and easier to understand than more complex indices and (ii) few studies provide data on more than two RE-AIM dimensions.

Individual Level Impact: RE measures (Reach × Effectiveness)

Multiplying Reach and Effectiveness yields a straightforward, composite measure of impact [9, 10]. The basic calculation of R × E is participation rate (number participating/eligible and invited to participate) × effect size (ES) on a primary outcome variable. An RE index can balance the strengths and limitations of programs that reach a wide target audience (but typically have smaller impact per individual) with more intensive interventions that often produce sizable change (but attract a smaller proportion of potential participants).

The RE index also can be expanded to address representativeness of participants. When the denominator of eligible persons is known, sociodemographic and health characteristics of participants can be compared with those who decline participation. When this denominator is not known, participants can be compared with characteristics of persons in that region or nation [11] (www.re-aim.org). Because representativeness comparisons are best made on several characteristics, a 'summary effect size (ES) for differential characteristics' index can be created by using the median ES across the representativeness comparisons calculated to compare characteristics of participants versus those declining participation. Using the median rather than the mean minimizes the influence of outliers. The median ES is then subtracted from the participation rate to provide a summary measure of Reach.

Most practical clinical trials and dissemination studies assess multiple outcomes rather than a single dependent variable [12, 13]. A median ES summary measure across key outcomes provides an overall effectiveness index. Another complexity arises when considering potential moderators and consistency of impact across different subgroups. Interventions that produce consistent effects across different population subgroups have greater external validity. We recommend calculating the ES of interactions between patient characteristics (e.g. gender and education) and treatment. For example, the 'differential impact' of an intervention between men and women could be analyzed. If the intervention effect is similar regardless of gender, the $ES_{\text{differential impact}}$ will be zero and indicates robust effectiveness [14].

After calculating the median $ES_{\text{differential impact}}$, Intervention Effectiveness is estimated by calculating the median ES across key outcome measures then subtracting the median ES for negative outcomes and the median ES for differential impact. Finally, the composite estimate of Individual Level Impact or formula RE (1) is calculated by multiplying the composite estimate for Reach by composite Intervention Effectiveness (Table I).

Population impact

Policy makers need to consider not only the impact of intervention on Reach and Effectiveness but also the prevalence of targeted problems. Parallel to the way that epidemiologists combine disease prevalence with risk ratios to produce attributable risk, we recommend multiplying prevalence of a problem by Individual Level Impact [RE (1) above] to produce Attributable Individual Level Impact [RE (2)] of an intervention (Table I).

Economic considerations

Health care decisions are constrained by resources [15]. Other things being equal, decision makers select interventions that most efficiently produce a given level of impact. Thus, RE 'Efficiency' is calculated by dividing the cost of an intervention by its Individual Level Impact [RE (3) in Table I]. We recommend the use of sensitivity analyses in estimating the Cost/Impact for entities that might adopt a given program [16].

Setting Level Indices

The previously described indices provide guidance to organizations that are considering adoption of

Table I. *Proposed RE-AIM summary indices*

Concept	Calculation
RE (Individual Level Impact) measures	
RE (1)	Reach × composite Intervention Effectiveness = (participation rate − median $ES_{\text{differential characteristics}}$) × (median $ES_{\text{key outcomes}}$ − median $ES_{\text{negative outcomes}}$ − median $ES_{\text{differential impact}}$)
RE (2): Attributable Individual Level Impact	Problem prevalence × RE (1) (see above)
RE (3): RE Efficiency	(Incremental cost of treatment − control)/ (incremental RE (1) of treatment − control)
AI (Setting Level Impact) measures	
AI (1)	(Setting adoption rate − median $ES_{\text{differential setting characteristics}}$) × (staff adoption rate − median $ES_{\text{differential staff characteristics}}$) × (median component implementation rate across staff and Tx components − median $ES_{\text{differential implementation}}$)
AI (2): Attributable Setting Level Impact	AI (1) × number of target settings × average no. of persons served per setting
RE-AIM profile	Graph using 0–100 scores of results on all RE-AIM dimensions
RE-AIM average	[Reach (as calculated above) + Effectiveness or Maintenance (see above) + Adoption (see above) + Implementation (see above)]/ 4

$ES_{\text{differential characteristics}}$ = ES for analyses on differences, participants versus non-participants. Note at individual level refers to representativeness of participants; at setting level, refers to representativeness of either settings or staff, as relevant. Low (close to zero) $ES_{\text{differential characteristics}}$ are desirable. $ES_{\text{differential impact}}$ = ES on key outcomes for different patient subgroups. In this case, low or ideally zero $ES_{\text{differential impact}}$ is desirable, since this would indicate little or no difference across subgroups or across different implementation staff. $ES_{\text{differential implementation}}$ = ES for analyses on differences across staff on implementation, low or zero $ES_{\text{differential implemntation}}$ is desirable.

interventions. From a population health perspective, however, there are additional issues. If an intervention is demanding, requires a high level of expertise, a large amount of time to deliver or is extremely costly, it is unlikely that many settings will adopt the program; and thus, its overall societal impact will be limited [15, 17]. Participation and representativeness at the setting level are equally important as at the individual level and we recommend calculation of a Setting Level Impact Index [AI (1), see Table I].

By multiplying Adoption and Implementation, the index yields information that integrates the appeal of a program to potential adopting settings with the extent to which those settings can successfully deliver the intervention. A frequent reason that dissemination studies fail to produce significant impact is that the intervention is not delivered as intended [18].

There is also the issue of the representativeness of participating settings. We recommend adjusting the setting participation rate by subtracting the median ES for comparisons between participating settings and (i) those settings invited but declining participation or (ii) organizations in that region (or the nation). For example, one might compare participating and non-participating schools on number of students, student:teacher ratio and history of health promotion. Determining the denominator or characteristics of potential settings can usually be estimated with publicly available data (www.re-aim.org). The setting level characteristics most relevant to collect will vary depending on the type of setting. For example, a worksite study might want to conduct representativeness analyses on variables such as type of company; per cent part-, full-time and shift employees; if the site is unionized and history of health promotion. In contrast, a medical office project might want to collect representativeness data on number of physicians and clinical staff, specialty of physicians, type of insurance most patients have, etc. Intervention impact may also be affected by the variety of backgrounds and skill levels of the personnel that deliver an intervention. For example, a hospital smoking cessation program delivered by a trained cessation counselor was highly effective in increasing long-term cessation [19], but when delivered by respiratory therapists, the same program was not effective [20].

Setting Impact also includes the participation rate and representativeness of staff who deliver an intervention. Similar to procedures for Reach, we recommend comparing staff who participate to those who do not on a number of relevant criteria (e.g. gender, age, expertise and experience) and reporting the median ES. Thus,

$$\begin{aligned}\text{Adoption} = {} & (\text{setting level participation rate} \\ & - \text{ES}_{\text{differential characteristics setting}}) \\ & \times (\text{staff level participation rate} \\ & - \text{ES}_{\text{differential characteristics staff}}).\end{aligned}$$

Implementation

Interventions are often inconsistently delivered, so this variability needs to be documented [21]. We recommend evaluating the extent to which various intervention components were delivered compared with protocol or intervention manual recommendations. Because most public health and behavior change interventions consist of multiple components, we recommend reporting the median implementation rate.

Interventions that can be implemented consistently by different staff, and preferably with different levels of training and experience, have greater generalizability [14, 22]. To estimate differential impact of staff, we recommend calculating ES for type of intervention staff on the various Implementation measures, and using the median $\text{ES}_{\text{differential implementation}}$.

Combining setting level factors of Adoption and Implementation, each containing two terms, into a Setting Level Impact Index results in formula AI (1) (see Table I).

Example application

The following hypothetical case study illustrates application of the RE (1) and AI (1) impact measures, used to aid decision making for a state health department deciding between two approaches to tobacco control. Intervention A is a proactive, multicall telephone outreach program designed to reach large numbers of smokers. We assume that it produces a high participation rate (80%) among referred smokers, and that it has consistent appeal across different subgroups of smokers (median $\text{ES}_{\text{differential characteristics}} = 0.05$). However, the ESs on the key outcomes of cessation rate and quality of life are likely to be modest (median = 0.20). Finally, the phone program produces negligible negative outcome (0.01), but is more effective with higher socioeconomic status and female participants (median $\text{ES}_{\text{differential impact}} = 0.15$). The RE (1) composite Individual Impact score for this intervention would then be $(0.75) \times (0.04) = 0.03$.

The alternative program being considered is a more intensive multisession group-based cessation program with pharmacologic aids. We assume that the participation rate (0.25) and differential recruitment indices ($\text{ES}_{\text{differential characteristics}} = 0.12$) for this program are worse than for the phone program. However, the effectiveness of this more intensive intervention among those who participate is likely to be much higher (ES = 0.65); the program should produce less differential results across subgroups (ES = 0.04) and negligible negative outcomes (ES = 0.01). The composite RE (1) index for this more intensive intervention would thus be $(0.13) \times (0.60) = 0.078$; and on the basis of RE (1) scores, the health department would select the intensive in-person smoking cessation program.

Space limitations preclude detailed presentation of setting level results from these programs, but as illustrated in Fig. 1, the phone intervention would likely produce higher adoption scores and more consistent implementation scores than the more intensive program; and thus, result in a substantially higher AI (1) composite Setting Level Impact score—say, 0.22 versus 0.04. Therefore, considering statewide adoption and implementation

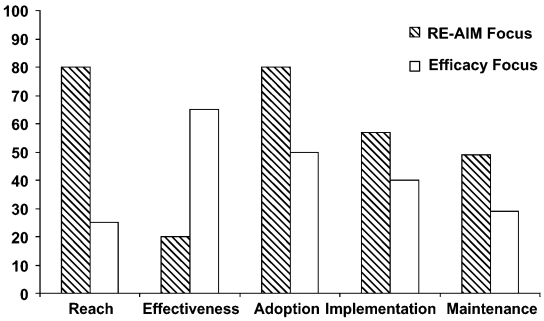

Fig. 1. Visual display of scores of two different types of intervention on RE-AIM dimension.

[as well as likely cost implications when considering RE (3) Efficiency scores], the health department would likely opt for the phone-based program.

This hypothetical example illustrates that the use of RE-AIM metrics will not always result in clear-cut decisions. They will, however, facilitate more informed and comprehensive consideration of all relevant factors and make explicit the values and priorities (e.g. Adoption versus Effectiveness versus Cost) on which decisions are based.

Impact of settings

Different intervention settings have different levels of penetration into the community. To consider population-wide impact of programs conducted in different settings, we recommend multiplying the Setting Level Impact AI (1) by the number of such settings in the geographic area and by the average number of individuals served per setting to produce an estimate of Attributable Setting Level Impact AI (2). For example, to compare the impact of an after-school physical activity program with that of a faith-based program, one should consider the number of such facilities as well as the average number of children served by each type of organization (Table I).

Often program developers do not consider all potential individuals or settings (e.g. all worksites) for inclusion. In such cases, the exclusion rate needs to be taken into account in calculating Attributable Individual Level Impact and Attributable Setting Level Impact. For example, if only medical practices having electronic medical records are selected for participation, the multiplication factor used for prevalence in AI (2) should be adjusted. Because not all medical clinics are eligible, there will be a corresponding reduction in population impact.

Long-term maintenance is an additional important issue. Maintenance is critically important for individual behavior change, and possibly, even more important as program sustainability at the setting level. Using long-term data, we recommend that a maintenance score be substituted for Effectiveness in the Individual Level Impact Score.

Finally, attrition should be accounted for in Reach and Effectiveness estimates. At the setting level, intervention sites may discontinue an intervention or close during a study, and alternatives for imputing setting level results and estimating the impact of such attrition are needed.

Graphical display

The calculations described involve several assumptions and procedures for combining RE-AIM scores. Although necessary to produce composite indices, these manipulations involve value judgments and assume factors (e.g. participation rate and representativeness) are of equal importance. This is often defensible [23], but may not be applicable in all situations. There is no way to 'prove' that multiplying Reach by Effectiveness is a better method of summarizing impact than would be adding scores, using a weighted average, a quadratic model, etc. Also, summary scores can sometimes hide or obfuscate important differences.

A more 'transparent' method of summarizing results along RE-AIM dimensions is to plot the various RE-AIM dimensions using a 0–100 scale (Fig. 1) to provide a visual display [24]. Visual displays are useful in comparing relative strengths and weaknesses of two or more alternative interventions [12] since, at present, an insufficient number of studies have reported data along multiple RE-AIM dimensions to interpret absolute scores. Fig. 1 presents a hypothetical comparison of an intensive intervention ('Efficacy Focus') to a low-intensity treatment program ('RE-AIM Focus').

A final approach involves collapsing the RE-AIM dimensions into a single overall index using methods developed for summarizing prevention quality among care systems [25]. Using the data in Fig. 1, each dimension is scored on a scale of 0–100 as in Healthplan Employer Data Information ratings [26]. Scores on the five (or four, since at a given time point, data are only used on either Effectiveness or Maintenance) RE-AIM dimensions would be summed and divided by 5 (or 4) to produce the overall measure of intervention impact (Table I).

Summary

The proposed summary indices are speculative. However, a metric representation of impact is timely since many programs of proven efficacy fail when implemented in real-world settings, resulting in wasted resources and unmet needs. Discussion of impact estimation is necessary before consensus can be reached on optimal methods for summarizing treatment outcomes. The options presented extend discussion to issues like Reach or Adoption that move beyond a restricted focus on one primary outcome or over reliance on cost-effectiveness indices.

Consistent with the recent Transparent Reporting of Evaluations with Non-Randomized Designs statement [3], we propose the formulas and methods in this paper to promote discussion and invite comments and suggestions for refinement. An implicit assumption that needs to be experimentally confirmed is that multilevel interventions should produce more lasting impact on RE-AIM summary scores than single interventions.

Limitations related to the assumptions involved in combining RE-AIM dimensions are recognized. Identifying optimal ways to form impact measures would be aided by more consistent reporting on all RE-AIM dimensions. Then, adequate data would be available to provide norms on individual dimensions, understand relationships among dimensions and document decisions that would be made using different calculations.

Significant improvements in population health depend on developing ways to help policy makers select health promotion and education programs. The RE-AIM framework helps to understand the broad array of issues that an effective program must address. A RE-AIM summary impact index should help decision makers to make more informed judgments and effective use of scarce resources.

Acknowledgements

Preparation of this manuscript was supported in part from a grant from the Robert Wood Johnson Foundation.

Conflict of interest statement

None declared.

References

1. Brownson RC, Gurney JG, Land GH. Evidence-based decision making in public health. *J Public Health Manag Pract* 1999; **5**: 86–97.
2. Green LW, Kreuter MW. *Health Promotion Planning: An Educational and Ecological Approach*: New York: Mayfield Publishing Co., 2005.
3. Des Jarlais DC, Lyles C, Crepaz N *et al.* Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health* 2004; **94**: 361–6.
4. Dzewaltowski DA, Estabrooks PA, Klesges LM *et al.* TREND: an important step, but not enough. *Am J Public Health* 2004; **94**: 1474.
5. Glasgow RE. Translating research to practice: lessons learned, areas for improvement, and future directions. *Diabetes Care* 2003; **26**: 2451–6.
6. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health* 1999; **89**: 1322–7.
7. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy to effectiveness transition. *Am J Public Health* 2003; **93**: 1261–7.
8. Rogers EM. *Diffusion of Innovations*, 5th edn. New York: Free Press, 2003.
9. Abrams DB, Emmons KM, Linnan LA. Health behavior and health education: the past, present, and future. In: Glanz K, Lewis RM, Rimer BK (eds). *Health Behavior and Health Education: Theory, Research, and Practice*. San Francisco, CA: Jossey-Bass, 1997, 453–78.

10. Prochaska JO, Velicer WF, Fava JL *et al*. Evaluating a population-based recruitment approach and a stage-based expert system intervention for smoking cessation. *Addict Behav* 2001; **26**: 583–602.
11. Hughes JR. Data to estimate the similarity of tobacco research samples to intended populations. *Nicotine Tob Res* 2004; **6**: 177–9.
12. Tunis SR, Stryer DB, Clancey CM. Practical clinical trials. Increasing the value of clinical research for decision making in clinical and health policy. *J Am Med Assoc* 2003; **290**: 1624–32.
13. Glasgow RE, Magid DJ, Beck A *et al*. Practical clinical trials for translating research to practice: design and measurement recommendations. *Med Care* 2005; **43(6)**: 551–557.
14. Leviton L. International encyclopedia of the behavioral and social sciences. In: Smelser NJ, Baltes PB (eds). *International Encyclopedia of the Behavioral and Social Sciences*. Burlington, MA: Elsevier Science and Technology Books, 2001, 5195–200.
15. Lamm RD. *The Brave New World of Health Care*. Golden, CO: Fulcrum Publishing, 2004.
16. Meenan RT, Stevens VJ, Hornbrook MC *et al*. Cost-effectiveness of a hospital-based smoking cessation intervention. *Med Care* 1998; **36**: 670–8.
17. Lenfant C. Clinical research to clinical practice—lost in translation? *N Engl J Med* 2003; **349**: 868–74.
18. Basch CE, Sliepcevich EM, Gold RS. Avoiding Type III errors in health education program evaluations. *Health Educ Q* 1985; **12**: 315–31.
19. Stevens VJ, Glasgow RE, Hollis JF *et al*. A smoking cessation intervention for hospitalized patients. *Med Care* 1993; **31**: 65–72.
20. Stevens VJ, Glasgow RE, Hollis JF *et al*. Implementation and effectiveness of a brief smoking cessation intervention for hospital patients. *Med Care* 2000; **38**: 451–9.
21. Bellg AJ, Borrelli B, Resnick B *et al*. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the Behavior Change Consortium. *Health Psychol* 2004; **23**: 443–51.
22. Cronbach LH, Glesser GC, Nanda H *et al*. *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: John Wiley & Sons, 1972.
23. Cohen J, Cohen P, West SG *et al*. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. London: Lawrence Erlbaum, 2003.
24. Tufte ER. *Beautiful Evidence*. Cheshire, CT: Graphics Press LLC, 2004.
25. Vogt TM, Aickin M, Ahmed F *et al*. The prevention index: using technology to improve quality assessment. *Health Serv Res* 2004; **39**: 511–30.
26. Schneider E, Riehl V, Courte-Wiencke S *et al*. Enhancing performance measurement: NCQA's road map for health information framework. *J Am Med Assoc* 1999; **282**: 1184–90.

Received on February 28, 2004; accepted on May 9, 2005