Published in final edited form as:
Am J Prev Med. 2013 May ; 44(5): 513–519. doi:10.1016/j.amepre.2013.01.006.

Measurement in Comparative Effectiveness Research

Jessica Chubak, PhD, **Carolyn M. Rutter, PhD**, **Aruna Kamineni, PhD**, **Eric A. Johnson, MS**, **Natasha K. Stout, PhD**, **Noel S. Weiss, MD, DrPH**, **V. Paul Doria-Rose, DVM, PhD**, **Chyke A. Doubeni, MD, MPH**, and **Diana S.M. Buist, PhD**
Group Health Research Institute (Chubak, Rutter, Kamineni, Johnson, Buist), the Department of Epidemiology (Chubak, Weiss, Buist), the Department of Biostatistics (Rutter), the Department of Health Services (Rutter, Buist), University of Washington School of Public Health, the Program in Epidemiology (Weiss), Fred Hutchinson Cancer Research Center, Seattle, Washington; the Department of Population Medicine (Stout), Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts; Health Services and Economics Branch (Doria-Rose), Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland; the Department of Family Medicine and Community Health (Doubeni), Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

Abstract

Comparative effectiveness research (CER) on preventive services can shape policy and help patients, their providers, and public health practitioners select regimens and programs for disease prevention. Patients and providers need information about the relative effectiveness of various regimens they may choose. Decision makers need information about the relative effectiveness of various programs to offer or recommend. The goal of this paper is to define and differentiate measures of relative effectiveness of regimens and programs for disease prevention. Cancer screening is used to demonstrate how these measures differ in an example of two hypothetic screening regimens and programs.

Conceptually and algebraically defined measures of relative regimen and program effectiveness are also presented. The measures evaluate preventive services that range from individual tests through organized, population-wide prevention programs. Examples illustrate how effective screening regimens may not result in effective screening programs and how measures can vary across subgroups and settings. Both regimen and program relative effectiveness measures assess benefits of prevention services in real-world settings, but each addresses different scientific and policy questions. As the body of CER grows, a common lexicon for various measures of relative effectiveness becomes increasingly important to facilitate communication and shared understanding among researchers, healthcare providers, patients, and policymakers.

Introduction

Comparative effectiveness research (CER) is playing a leading role on the U.S. healthcare stage. The American Recovery and Reinvestment Act (ARRA) of 2009 and the Patient

Address correspondence to: Jessica Chubak, PhD, Group Health Research Institute, 1730 Minor Avenue, Suite 1600, Seattle WA 98101. chubak.j@ghc.org..

No other financial disclosures were reported by the authors of this paper.

Protection and Affordable Care Act (ACA) of 2010—via its establishment of the Patient-Centered Outcomes Research Institute (PCORI)—have invested substantially in CER.[1-4] In requiring coverage of evidence-based preventive services,[2,5] the ACA also highlights the growing relevance of effectiveness research to health policy. Along with information on harms, affordability, patient preferences, and equity,[2,4,6] data on effectiveness are crucial for shaping health policy and enabling patient and provider decision-making on various approaches to achieving a specific health benefit.

Methods for CER are receiving increasing attention[7, 8] and are a focus of the legislation establishing PCORI.[2] However, measures for CER have not been fully characterized. This paper describes measures of relative effectiveness of healthcare interventions for the prevention of disease and seeks to provide a standard framework for presenting CER results for prevention research, decision-making, and policy development. Cancer screening, a central topic in clinical prevention, public health, and CER, is used in the current paper to illustrate the proposed terminology. Previously published frameworks, such as RE-AIM (a model designed to evaluate the public health impact of interventions),[9,10] serve as an example of how shared terminology can advance the study of health services interventions.[11]

The current paper proposes definitions for two measures of relative effectiveness for screening: regimen effectiveness, based on repeated use of one or more types of screening tests; and program effectiveness, based on an organized approach to encouraging and delivering screening. The difference between the effectiveness of regimens versus programs (or guidelines) is an important distinction for clinical care and public health. When two or more screening regimens are available for the same cancer—one regimen that is more effective and another with higher patient participation—which should be recommended? Healthcare providers, public health practitioners, and policymakers may face such questions when deciding, for example, whether to recommend colonoscopy or fecal occult blood screening for colorectal cancer, or human papillomavirus testing versus cytologic screening for cervical cancer.

Real-world guidelines should not assume perfect adherence. The U.S. Preventive Services Task Force acknowledged in their colorectal cancer screening guidelines that “adherence will be more important in life-years gained than will the particular regimen selected.”[12] Because the effectiveness of population-based preventive measures depends not only their performance in ideal settings but also their acceptance by the people targeted, a modestly effective regimen with higher patient participation and adherence could be highly effective in community settings. The current paper develops a framework for defining and differentiating the relative effectiveness of regimens and programs. Distinguishing the effectiveness of prevention regimens and prevention programs will enable consistent use of terminology, which is critical for advancing CER.[13,14]

Measures of Relative Effectiveness

Measures of effectiveness differ from those of efficacy: efficacy measures the benefit of an intervention under “ideal conditions,” whereas effectiveness measures the benefit of an intervention “when deployed in the field in routine circumstances.”[15] Researchers are urged to carefully describe the factors that affect the effectiveness of their intervention.[9,16] For example, the effectiveness of cancer screening interventions may be influenced by the setting in which they are adopted, level of participation, test accuracy, skill of the people performing or interpreting the test, adequacy of follow-up of positive test results, adherence to evidence-based intervals, and treatment effectiveness.[17]

Table 1 provides examples of various kinds of relative effectiveness questions in cancer screening. The appropriate measures for answering these questions are described below. Table 2 provides additional detail on how to calculate these measures. Measures that compare rates of outcomes on the relative scale are described, but analogous measures can also be defined on the additive (risk-difference) scale.

Relative Regimen Effectiveness

Patients choosing among competing services may want to know how the effectiveness of one regimen compares to another in the setting in which they receive care. **Relative regimen effectiveness** is defined here as the outcomes in people who participate in and adhere to (i.e., continue in accordance with guideline recommendations) a particular regimen compared to those of people who receive and adhere to an alternative regimen. For cancer screening, the regimens may be different combinations of screening tests (e.g., flexible sigmoidoscopy every 5 years with a mid-interval fecal occult blood test [FOBT] versus annual FOBT alone) or specified screening intervals (e.g., FOBT biennially versus annually). In the extreme, a single test could be considered a regimen (e.g., one-time colonoscopy).

Comparing regimen effectiveness is different from comparing efficacy. Regimen effectiveness may be influenced by real-world implementation (e.g., likelihood of following-up on an abnormal Pap or FOBT). Thus, relative regimen effectiveness goes beyond efficacy to incorporate implementation. Implementation is a key component of the RE-AIM framework. Because regimen effectiveness provides information on outcomes among people who choose specific regimens, this measure is relevant for patients who want to know what to expect if they actually take a test, not just if they are offered one. Estimates of relative regimen effectiveness are highly relevant for public health practitioners and policymakers deciding whether it is worth investing in or designing programs to recommend a particular regimen. Relative regimen effectiveness is also a critical building block for relative program effectiveness, as described below.

Relative Program Effectiveness

Public health practitioners and policymakers interested in the effects of preventive services at population levels should consider another measure: **relative program effectiveness**. A health services “program” is generally an organized effort to engage people in a health-promoting activity; it may be implemented by a variety of entities, ranging from community clinics, to health plans[18,19] to nations.[20-22] Although regimen effectiveness compares the outcome of receipt of tests or combination of tests, relative program effectiveness extends this concept by incorporating differences in individual participation in a regimen that is offered within a program.

Relative program effectiveness is the ratio of rates of outcomes for populations exposed to differing screening programs. For example, a health program that mails FOBT kits to patients directly could be compared to a program that sends reminders for screening colonoscopies to all patients. Program reach[9] (i.e., individual-level participation) is critical for program effectiveness and is most of what distinguishes relative program effectiveness from regimen effectiveness. Relative program effectiveness provides a comprehensive comparison of preventive health services because program success depends on not only the effectiveness of the regimens, but also the reach of a program. For example, a program that invites individuals to receive an effective regimen could be ineffective if few people choose to undergo the test because of cost, inconvenience, or anxiety. The overall impact of a

program will also depend on whether or not a particular organization or setting chooses to adopt and maintain it.[9,11]

Outcomes from an entire population, which would include those for people who undergo an offered/recommended regimen, who choose alternative regimens, and who are not screened at all, form the basis for calculating relative program effectiveness. Participation and adherence contribute to the measured effect of the program. A challenge to estimating relative program effectiveness is that it assumes no unmeasured systematic differences in populations that adopt different programs.

This assumption is valid when patients are randomized to be invited into one of two screening programs. However, randomizing entire communities can be difficult.[23-25] Relative program effectiveness is usually assessed in observational studies by comparing organizations or settings that have adopted different programs, such as health plans that use different approaches to increase screening. The potential for bias in these observational studies may be strong. For example, if underlying cancer incidence or cancer mortality rates differ across populations, this difference could be mistakenly attributed to the relative effectiveness of a screening program. Thus, valid comparison across populations requires careful consideration of confounding at multiple levels.[26-28]

Relationship Among Measures: An Example

Measures of relative effectiveness are related: relative regimen effectiveness influences relative program effectiveness. However, these measures can differ in magnitude and even direction. A screening regimen with low effectiveness and high participation and adherence may result in a screening program with better results than one employing a test with high effectiveness but low participation or adherence.

In the example in Table 3, Regimen A is more effective than Regimen B because the cancer mortality rate in people screened by A (11.0 per 100,000) is lower than that in people screened by B (13.0 per 100,000). This lower mortality rate in **people** exposed to A compared to B is reflected in a relative **regimen** effectiveness of 0.85, favoring A compared to B. However, even though Regimen A is more effective than Regimen B at reducing cancer mortality, differences in participation result in more people being exposed to Regimen B. As a result, the program that invites people to receive Regimen B is more effective than a program that invites people to receive Regimen A.

This difference in effectiveness is seen in the relative program effectiveness of 1.10 when outcomes are compared for **populations** receiving recommendations for Regimen A versus B. For example, a colonoscopy every 10 years might be more effective than FOBT annually. But if people are more willing to undergo FOBT than colonoscopy an FOBT screening program could be more effective than a colonoscopy screening program in reducing mortality. This might be the case if FOBT can reach more people because tests can be mailed, or if access to colonoscopy is limited by clinical capacity.

Interpreting Differences in Relative Effectiveness Estimates Across Studies

Prevention researchers, healthcare providers, policymakers, payers, and patients often need to compare and synthesize results from multiple CER studies. This fact highlights the importance of understanding why findings might differ across studies in the absence of chance, confounding, or other biases. At least two potential reasons exist: (1) different measures answer different questions; and (2) the relative effectiveness of various regimens and programs may truly differ across subgroups, populations, and settings.

Different Measures Answer Different Questions

Estimates of relative regimen effectiveness and relative program effectiveness may differ in magnitude and even direction, because they compare different aspects of health service effectiveness (Table 3). Syntheses and meta-analyses of study results should report which measure or measures of relative effectiveness are being summarized, as well as provide information that can explain the relative effectiveness, such as—in the case of screening—test performance within the population, participation in and adherence to screening and follow-up tests, and treatment effectiveness.

Relative Effectiveness May Differ Across Subgroups, Populations, and Settings

Caution should be taken when extrapolating CER results from one subgroup, population, or setting to another. Relative regimen or program effectiveness may vary across subgroups or settings because of differences in natural history of the disease, test accuracy, clinical follow-up of a positive test, or treatment effectiveness. The impact of an intervention depends, in part, on which settings adopt it. It is not always possible to predict how results from one setting will differ from those in another. Uncertainty about the transferability of results across populations and settings may limit their generalizability.

Computing relative effectiveness in various groups and settings is necessary if participation, adherence, test accuracy, follow-up care, or treatment effectiveness are expected to vary substantially. This is synergistic with the patient-centered CER goals of elucidating the relative harms and benefits of multiple approaches across diverse populations. Studies evaluating relative effectiveness should provide information on the influential variables that might differ among populations. When presenting findings, estimates of relative effectiveness measures for various groups should be presented separately for regimens and programs to facilitate the tailoring of recommendations and policies.

Examples of differences across subgroups or populations

Mammography provides an example of these potential differences across subgroups or populations. Digital mammography is more accurate than film mammography for women with dense breasts, but may be less accurate for women with fatty breasts.[29,30] Because breast density declines with age, digital mammography may be more effective than film mammography in younger women, while film mammography may be more effective in older women. Here, the relative effectiveness of the two regimens could differ in the two subgroups because the accuracy of these two tests differs between subgroups.

This example illustrates that results from relative regimen effectiveness studies in one group may not be generalizable to another group. As with relative regimen effectiveness, relative program effectiveness may also differ across settings or populations. A strategy that works in one setting may not work in another,[31] especially if reach is likely to differ. For example, relative program effectiveness may differ across racial or socioeconomic groups if beliefs and preferences for screening tests differ.[32,33]

Examples of differences across settings

Relative regimen effectiveness may also differ across settings due to differences in follow-up care or treatment. This fact is illustrated by cervical cytology and visual inspection of the cervix using acetic acid (VIA), which are two cervical cancer screening methods. VIA is slightly more sensitive but less specific than cervical cytology.[34] It can be performed in a primary care clinic and results are immediate, so follow-up for abnormal findings can be

performed without delay. Cervical cytology requires laboratory facilities and trained cytotechnologists, so follow-up tests for positive results require additional visits.

In developed countries, cervical cytology is currently the most common screening method for cervical cancer. Most patients receive follow-up care after a positive exam, and they return for screening on a routine basis. However, the effectiveness of cervical cytology may be lower in countries that lack healthcare infrastructure and resources, and have poor compliance with follow-up visits,[35,36] even if the regimens' sensitivities and specificities do not vary across settings. Therefore, depending on the setting, a VIA regimen may have a higher relative regimen effectiveness than cervical cytology in reducing cervical cancer mortality.[37,38] This example illustrates the need for caution when extrapolating relative effectiveness study results from one setting to another.

As with regimen effectiveness, relative program effectiveness may vary across settings. For example, a recommendation to receive colonoscopy may be more effective than a recommendation to receive FOBT in urban settings; conversely, an FOBT program may be more effective in rural areas where endoscopists are scarce or difficult to visit because of geographic barriers. This example underscores the importance of considering all aspects that might influence program effectiveness when formulating health policy based on findings from a particular population or setting and recognizing that if reach or other factors change, so might the estimate of relative program effectiveness.

Conclusion

The impact of a preventive healthcare service or program depends on many factors. Increasingly, the importance of patient preference, equity, affordability, ease of adoption, implementation, and maintenance are being recognized.[2,4,6] Effectiveness remains critical to decision-making at the population and individual levels. Previous frameworks, such as RE-AIM, have illustrated how shared terminology can advance the study of health services interventions. To this end, this paper offers a framework for thinking about effectiveness at multiple levels.

Although this paper uses a specific preventive service, cancer screening, to distinguish relative regimen effectiveness and relative program effectiveness and highlight the factors that influence them, most of the considerations and definitions are broadly applicable to CER. Each measure provides different information. Relative regimen effectiveness assumes perfect adherence to a regimen, and provides important information for individuals and providers making decisions about preventive care. Relative program effectiveness compares intervention effects in populations and is influenced by participation, and so is generally more useful for policy formation. As the body of comparative effectiveness research grows, a common lexicon for various measures of relative effectiveness becomes increasingly important to facilitate communication and shared understanding among researchers, healthcare providers, patients, and policymakers.

Acknowledgments

Research reported in this publication was supported by the National Cancer Institute of the NIH under Award Number UC2CA148576 (DSMB and CAD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The authors thank Dr. Michael Von Korff, Dr. Rebecca Hubbard, and Dr. Chris Tachibana for comments on earlier versions of the paper. This work resulted from a collaboration generated within the HMO Cancer Research Network (CRN), a collaboration of 14 HMOs that conducts research to determine the effectiveness of preventive, curative, and supportive interventions for major cancers that span the natural history of those cancers among diverse populations and health systems.

References

1. Federal Coordinating Council for Comparative Effectiveness Research. Report to the President and the Congress. Jun 30. 2009 2009
2. Patient Protection and Affordable Care Act. US. 2010. Public Law 111-148, section 6301. www.pcori.org/images/PCORI_EstablishingLeg.pdf
3. Sox HC. Comparative effectiveness research: a progress report. Ann Intern Med. 2010; 153(7):469–72. [PubMed: 20679544]
4. Selby JV, Beal AC, Frank L. The patient-centered outcomes research institute (PCORI) national priorities for research and initial research agenda. JAMA. 2012; 307(15):1583–1584. [PubMed: 22511682]
5. Laiteerapong N, Huang ES. Health care reform and chronic diseases: Anticipating the health consequences. JAMA. 2010; 304(8):899–900. [PubMed: 20736475]
6. IOM. Crossing the quality chasm: A new health system for the 21st century. National Academies Press; Washington, D.C.: 2001.
7. Sox HC, Goodman SN. The methods of comparative effectiveness research. Annu Rev Public Health. 2012; 33:425–45. [PubMed: 22224891]
8. Gabriel SE, Normand S-LT. Getting the Methods Right -- The Foundation of Patient-Centered Outcomes Research. New England Journal Of Medicine. 367(9):787–790.
9. 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(9):1322–7. [PubMed: 10474547]
10. Kessler RS, Purcell EP, Glasgow RE, Klesges LM, Benkeser RM, Peek CJ. What Does It Mean to "Employ" the RE-AIM Model? Evaluation Health Professions. 2012
11. Rabin, BA.; Brownson, RC. Developing the terminology for dissemination and implementation research. In: Brownson, RC.; Colditz, GA.; Proctor, EK., editors. Dissemination and implementation research in health: translating science to practice. Oxford University Press; New York: 2012. p. 23-51.
12. U.S. Preventive Services Task Force. Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2008; 149(9):627–637. [PubMed: 18838716]
13. Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. Ann Intern Med. 2009; 151(3):203–5. [PubMed: 19567618]
14. Sox HC. Defining comparative effectiveness research: the importance of getting it right. Med Care. 2010; 48(6 Suppl):S7–8. [PubMed: 20473202]
15. Last, JM., editor. A Dictionary of Epidemiology. 4th ed. Oxford University Press; New York: 2001.
16. Taplin SH, Anhang Price R, Edwards HM, et al. Introduction: Understanding and Influencing Multilevel Factors Across the Cancer Care Continuum. JNCI Monographs. 2012; 2012(44):2–10.
17. Gatsonis C. The promise and realities of comparative effectiveness research. Stat Med. 2010; 29(19):1977–81. discussion 1996-7. [PubMed: 20683886]
18. Levin TR. Editorial: It's time to make organized colorectal cancer screening convenient and easy for patients. Am J Gastroenterol. 2009; 104(4):939–41. [PubMed: 19293790]
19. Taplin SH, Thompson RS, Schnitzer F, Anderman C, Immanuel V. Revisions in the risk-based Breast Cancer Screening Program at Group Health Cooperative. Cancer. 1990; 66(4):812–8. [PubMed: 2386908]
20. Verbeek ALM, Broeders MJM. Evaluation of The Netherlands breast cancer screening programme. Ann Oncol. 2003; 14(8):1203–1205. [PubMed: 12881378]
21. Flitcroft KL, Salkeld GP, Gillespie JA, Trevena LJ, Irwig LM. Fifteen years of bowel cancer screening policy in Australia: putting evidence into practice? Med J Aust. 2010; (1):193, 37–42.
22. International Cancer Screening Network. Inventory of colorectal cancer screening activities in ICSN counties. 2011. May 2008 May 5appliedresearch.cancer.gov/icsn/colorectal/screening.html

23. Mazor KM, Sabin JE, Boudreau D, et al. Cluster randomized trials: opportunities and barriers identified by leaders of eight health plans. Med Care. 2007; 45(10 Supl 2):S29–37. [PubMed: 17909379]
24. Murray DM, Pals SL, Blitstein JL, Alfano CM, Lehman J. Design and analysis of group-randomized trials in cancer: a review of current practices. J Natl Cancer Inst. 2008; 100(7):483–491. [PubMed: 18364501]
25. Mazor KM, Sabin JE, Goff SL, et al. Cluster randomized trials to study the comparative effectiveness of therapeutics: stakeholders' concerns and recommendations. Pharmacoepidemiol Drug Saf. 2009; 18(7):554–561. [PubMed: 19402030]
26. Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part III. Value Health. 2009; 12(8):1062–1073. [PubMed: 19793071]
27. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: The International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II. Value Health. 2009; 12(8):1053–1061. [PubMed: 19744292]
28. Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report-Part I. Value Health. 2009; 12(8):1044–52. [PubMed: 19793072]
29. Pisano ED, Hendrick RE, Yaffe MJ, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. Radiology. 2008; 246(2):376–83. [PubMed: 18227537]
30. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med. 2005; 353(17):1773–83. [PubMed: 16169887]
31. Morrison DM, Hoppe MJ, Gillmore MR, Kluver C, Higa D, Wells EA. Replicating an intervention: the tension between fidelity and adaptation. AIDS Educ Prev. 2009; 21(2):128–140. [PubMed: 19397435]
32. Hawley ST, Volk RJ, Krishnamurthy P, Jibaja-Weiss M, Vernon SW, Kneuper S. Preferences for colorectal cancer screening among racially/ethnically diverse primary care patients. Med Care. 2008; 46(9 Suppl 1):S10–6. [PubMed: 18725820]
33. Walsh JM, Kaplan CP, Nguyen B, Gildengorin G, McPhee SJ, Perez-Stable EJ. Barriers to colorectal cancer screening in Latino and Vietnamese Americans. Compared with non-Latino white Americans. J Gen Intern Med. 2004; 19(2):156–66. [PubMed: 15009795]
34. Gaffikin LD, Lauterbach MB, Blumenthal PD. Performance of visual inspection With acetic acid for cervical cancer screening: a qualitative summary of evidence to date. Obstet Gynecol Surv. 2003; 58(8):543–550. [PubMed: 12886165]
35. Ngelangel CA, Limson GM, Cordero CP, Abelardo AD, Avila JM, Festin MR. Acetic-acid guided visual inspection vs. cytology-based screening for cervical cancer in the Philippines. Int J Gynaecol Obstet. 2003; 83(2):141–150. [PubMed: 14550588]
36. Cronjé HS. Screening for cervical cancer in the developing world. Best Prac Res Clin Obstet Gynaecol. 2005; 19(4):517–529.
37. Mandelblatt JS, Lawrence WF, Gaffikin L, et al. Costs and benefits of different strategies to screen for cervical cancer in less-developed countries. J Natl Cancer Inst. 2002; 94(19):1469–1483. [PubMed: 12359856]
38. Goldie SJ, Gaffikin L, Goldhaber-Fiebert JD, et al. Cost-effectiveness of cervical-cancer screening in five developing countries. New Engl J Med. 2005; 353(20):2158–2168. [PubMed: 16291985]

Table 1

Examples of relative regimen and program effectiveness questions in cancer screening

Relative regimen effectiveness	Relative program effectiveness
Is cervical cancer screening every 5 years with HPV tests more effective at preventing cervical cancer mortality than cervical cytology every 3 years?	Is a program that mails women kits for home-based, self-administered HPV testing every 5 years more effective at reducing cervical cancer mortality than a program that mails Pap reminders every 3 years?
Is colonoscopy every 10 years more effective at reducing colorectal cancer mortality than annual FOBT?	Is a program that mails FOBT kits to patients more effective at reducing colorectal mortality than a program that mails postcard reminders for colonoscopy screening?

FOBT: Fecal occult blood test; HPV: Human papillomavirus

Table 2
Calculation of relative regimen effectiveness and relative program effectiveness

	Program M (Regimen A recommended)		Program N (Regimen B recommended)	
	Proportion exposed	**Outcome rate**	**Proportion exposed**	**Outcome rate**
Regimen received				
A	a_1	R_{A1}	a_2	R_{A2}
B	b_1	R_{B1}	b_2	R_{B2}
Neither	$1-a_1-b_1$	R_{01}	$1-a_2-b_2$	R_{02}
Overall		R_1		R_2
Relative effectiveness				
Regimen A vs Regimen B	$RR_{regimen,1} = \frac{R_{A1}}{R_{B1}}$		$RR_{regimen,2} = \frac{R_{A2}}{R_{B2}}$	
Program M vs Program N	$RR_{program} = \frac{R_1}{R_2}$			

Note: Subscripts 1 and 2 refer to two hypothetic populations in which Programs M and N, respectively, are adopted.

Table 3

Results from a hypothetic study comparing the effect of various screening regimens and programs on cancer mortality

	Program M (Regimen A recommended)		**Program N (Regimen B recommended)**	
Regimen received	**Proportion exposed**	**Cancer mortality (per 100,000)**	**Proportion exposed**	**Cancer mortality (per 100,000)**
A	0.5	11.0	0.1	11.0
B	0.1	13.0	0.8	13.0
Neither	0.4	20.0	0.1	20.0
Overall	1.0	14.8	1.0	13.5
Relative effectiveness				
Regimen A vs Regimen B	$\frac{11.0\ /\ 100,\ 000}{13.0\ /\ 100,\ 000} = 0.85$		$\frac{11.0\ /\ 100,\ 000}{13.0\ /\ 100,\ 000} = 0.85$	
Program M vs Program N	$\frac{14.8\ /\ 100,\ 000}{13.5\ /\ 100,\ 000} = 1.10$			

Note: The example shows the calculation of relative regimen effectiveness (A vs B) in two populations, as well as relative program effectiveness (M vs N). The example assumes no confounding. Regimen A is more effective than Regimen B at reducing cancer mortality; however, a program that invites people to receive Regimen B is more effective than a program that invites people to receive Regimen A. Because interventions seek to reduce mortality, a relative risk <1 indicates greater effectiveness.