



Individual Risk Prediction and Population-wide Disease Prevention

Citation

Rockhill, B., I. Kawachi, and G.A. Colditz. 2000. "Individual Risk Prediction and Population-Wide Disease Prevention." *Epidemiologic Reviews* 22 (1): 176–80.
<https://doi.org/10.1093/oxfordjournals.epirev.a018017>.

Published version

<https://doi.org/10.1093/oxfordjournals.epirev.a018017>

Link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:41275574>

Terms of use

Metadata Only

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

Individual Risk Prediction and Population-wide Disease Prevention

B. Rockhill,¹ I. Kawachi,² and G. A. Colditz^{1,3}

INTRODUCTION

Epidemiologists seem to be engaged in a ceaseless search for new risk factors for disease. An increasingly important consequence of this activity is the categorization of individuals according to personal disease risk. This search for risk factors is being conducted with the hope of more accurately isolating groups of high risk individuals; that is, the goal is to identify risk factors with high relative risks and with a correspondingly good ability to discriminate between diseased and nondiseased individuals. However, most known risk factors associated with chronic diseases (with the important exceptions of heavy smoking and certain occupational exposures, such as asbestos) have modest relative risks. At the individual level, most risk factors have poor discriminatory power, or poor positive predictive value. Some persons in the "low risk" or unexposed category will develop disease, while the majority of those who are exposed will remain healthy.

One of the goals in the burgeoning field of genetic epidemiology is to extend the search for risk factors into the human genome, to uncover high risk individuals who were hitherto hidden within exposures defined by conventional lifestyle or environmental factors. Speculating on the future of genetic epidemiology, Feigelson et al. noted that "expanded knowledge in this area should lead to more refined definitions of 'high-risk' individuals... and, hopefully, lead us to be able to better define who may benefit from specific interventions based on underlying genetic susceptibility" (1, p. 20). In a similar vein, Khoury stated that "new gene discoveries, which occur on an almost daily basis, require that the public health community take a leadership role in translating the results of these dis-

coveries into effective and appropriate strategies to prevent disease and disability in the general population by targeting environmental, behavioral, and medical interventions to each person's genetic susceptibility" (2, p. 176).

Is the current level of enthusiasm for individual risk prediction and identification of high risk individuals really warranted? Can findings from genetic epidemiology contribute greatly to this endeavor? In this paper, we revisit Geoffrey Rose's ideas comparing the high risk and population-based strategies of disease prevention and apply them to recent debates about the course of modern epidemiology.

THE PROBLEM OF IDENTIFYING HIGH RISK INDIVIDUALS

Geoffrey Rose's important observation, made nearly two decades ago, was that most cases of chronic disease arise from the mass of the population with risk factor values close to the average (3). While he made this point mostly with regard to risk factors for coronary heart disease (4), the principle turns out to hold for almost all identified risk factor-disease relations. Most cases of disease do not arise from the "high risk" tail of the risk factor distribution. There is a continuum of disease risk associated with most exposures, and the decision to label a specific exposure level as demarcating low risk versus high risk is often arbitrary. Stated another way, the population attributable fractions for many chronic diseases can be inflated only by defining risk factors in such a way that nearly the entire population is labeled "exposed."

The components of a population attributable fraction calculation, including the selected risk factors and their estimated relative risks, derive from epidemiologic risk prediction models. These same models can be used to produce individual estimated probabilities of disease. Clinicians are increasingly interested in providing patients with such estimated personal risks. The Gail et al. model of breast cancer risk (5) is one risk model that has received much attention. Developed originally for the purpose of estimating sample sizes needed for clinical trials of

Received for publication April 13, 1999, and accepted for publication November 3, 1999.

¹ Channing Laboratory, Harvard Medical School and Brigham and Women's Hospital, Boston, MA.

² Department of Health and Social Behavior, Harvard School of Public Health, Boston, MA.

³ Department of Epidemiology, Harvard School of Public Health, Boston, MA.

Reprint requests to Dr. Beverly Rockhill, Channing Laboratory, 181 Longwood Avenue, Boston, MA 02115.

breast cancer prevention with tamoxifen, the Gail et al. model is now being used to provide women with individual risk estimates. These estimates are intended to be used by the women and their clinicians in making decisions about chemoprevention, lifestyle behaviors (such as postmenopausal hormone use), and screening mammography (6). In reality, such equations lack the ability to single out the truly "high risk" individuals; individual risk estimates from epidemiologic models tend to cluster around 0. This clustering at the low end of the spectrum is not surprising, given that "individual" risk is really the average risk of a group of "similar" persons. In almost all such groups, even those defined by exposure to several risk factors, only a small minority of individuals will develop disease over a specified time period, and average risk will thus be low. Put differently, for most diseases there is an enormous overlap in estimated probabilities between the group that eventually develops disease and the group that does not (e.g., see Spiegelman et al. (7), Heller et al. (8), Katz and Foxman (9), and Rose (4)). This is true even for lung cancer, although the smoking-lung cancer relative risks can be as high as 10–15, depending on the specific definition of smoking exposure.

Can knowledge of genetic susceptibility result in more accurate risk estimation at the individual level? That is, can such knowledge separate the distributions of estimated disease risk for cases and noncases, and serve as a screening tool with which to pinpoint the minority of individuals who will go on to develop disease? Such seems to be the intent behind much of the current research in genetic epidemiology.

Genetic factors that contribute to disease susceptibility include infrequent, highly penetrant, dominant mutations, such as the well publicized *BRCA1* and *BRCA2* mutations, as well as more prevalent genetic polymorphisms that influence hormone metabolism and susceptibility to effects of environmental exposures (1). On the basis of current evidence, while such mutations are often associated with high lifetime relative risks, only a very small proportion of disease cases can be attributed to them, because of their rarity. For instance, it is believed that fewer than 5 percent of all cancers arise from the effects of single, dominant mutations (10). As a result, much attention in genetic epidemiology is devoted to the study of the more prevalent, lower-penetrance polymorphisms, and the phrase "gene-environment interaction" has attained recent prominence.

Khoury and Wagener (11) have noted that knowledge of gene-environment interaction can increase the positive predictive value of a risk factor, if the gene-environment interaction is strong—i.e., if the effect of

the risk factor is strong and is concentrated among persons with the susceptible genotype and is practically null among those without the genotype. However, it is not commonly recognized that a risk factor must have a *very* strong relative risk (i.e., >50) associated with it before it can serve as a useful screening tool at the individual level—i.e., before its measures of sensitivity and specificity with regard to future disease status approach levels regarded as "adequate" by those who evaluate disease screening technologies. On the basis of current knowledge, it appears unlikely that such conditions are common with respect to gene-environment interactions. Despite much research, evidence for important interactions remains scant. Studies of genetic polymorphisms have so far failed to uncover groups of individuals at markedly higher risk from the effects of exposures.

This lack of findings is not surprising, given the knowledge that is emerging regarding the complexity of gene-environment interactions. In most cases, it appears to be not one gene but multiple genes that are involved in determining the degree of interaction with disease-related exposures and hence the degree of an individual's susceptibility to disease (12). The effects of these multiple genes are in turn modified by ethnicity, age, gender, nutritional status, duration and dose of exposure, and other known and unknown factors (13). In short, one could just as easily change the focus from genetic susceptibility to ask: "What (non-genetic) factors make individuals susceptible to having harmful gene-environment interactions expressed?" In other words, the same genetic polymorphism can be associated with different levels of risk, and with risks of different types of chronic disease, in different individuals (12), just as is the case with nongenetic risk factors; and the absence of a genetic "risk factor" rarely (if ever) signifies that an individual's probability of developing disease is much below "average."

The disadvantages of the high risk approach to prevention cited by Rose (4) are particularly relevant when genetic susceptibility is used to determine risk status. First, labeling a person genetically susceptible to a major disease would probably have important psychological, social, and economic costs. Health researchers are only now beginning to explore such consequences. Second, the genetic approach to risk identification might detract attention from the exposures themselves. High risk strategies in general, and those based on genetics in particular, do not seek to alter the underlying factors which determine exposure. They simply attempt to truncate the tail of the distribution, while leaving unchanged the hazardous exposures for the rest of the population.

THE PROBLEM WITH THE LOGIC OF INDIVIDUAL RISK

Is the only alternative to the "high risk" approach to prevention, then, necessarily the simplistic approach of trying to convince *all* individuals to modify their unhealthy exposures, as some genetic epidemiologists have implied (14)? This is an oversimplification of Rose's argument. It is precisely our inability to predict, with accuracy, the diverse futures of individuals that lends strength to population-level approaches to shifting risk factor distributions. In Rose's words, "The population-wide approach seeks to move the whole distribution of a risk factor, including its low tail, in a favorable direction. Some individuals will stand to benefit much more than others, although ideally everyone would hope to gain something (for example, by control of hazardous environmental pollution)" (4, p. 106). He notes, however, that "this ideal would not be achieved if the curve relating exposure to risk is not linear, and especially if it is U- or J-shaped" (4, p. 106). Rose acknowledged that certain preventive interventions are clearly inappropriate from a population standpoint, because they carry potential risks that would not be outweighed by the potential benefits for persons at the low risk end of the distribution. Rose cited interventions with pharmacologic agents, such as cholesterol-lowering drugs, as examples of inappropriate population-level interventions.

In his writings, Rose compared and contrasted the advantages and disadvantages of high risk and population-based approaches to prevention, recognizing that the strategies are not mutually exclusive. When considered from a public health perspective, though, high risk strategies face important limitations. The aggregate-level laws of probability dictate that the majority of disease cases will not arise among persons who have unusually high estimated risk; a very large "N" or population size multiplied by a relatively low average probability will generate more cases than a very small population size multiplied by a high probability.

A historical understanding of disease patterns and public health initiatives, as well as examination of current trends, demonstrates that significant changes in population disease burdens are most often the result of transformations in determinants of risk factor distributions. This was one of Rose's central points. The mass production and marketing of cigarettes was a 20th century phenomenon in the United States and Great Britain; lung cancer consequently went from being a relatively rare disease to being one of the major causes of death by the middle of the century (16, 17). More recently, in post-Soviet Russia, social disintegration and economic frustration appear to have produced a large increase in alcohol consumption, which in turn

has led to a dramatic surge in premature male mortality (18). On a positive note, in the United States, a population-based strategy for increasing seat belt use (in the form of a legal requirement) has been followed by a significant drop in traffic fatalities (19).

The examination of such historical trends in disease highlights the importance of broad forces in determining "average risk" in a population, and illustrates problems with the logic of individual risk. The words "cause" and "prevent," as they pertain to probabilistic risk factor logic, are concepts most clearly applied to aggregates of individuals, not a specific individual. Risk factor findings, which are by necessity couched in probabilistic language, call for aggregate-level policies: If exposure can be eliminated for 1,000 individuals, five cases (for example) will be averted over a 10-year time period. Knowledge that a factor is associated with increased risk of disease obviously does not translate into the premise that a case of disease will be prevented if a specific individual eliminates exposure (or takes a chemopreventive agent); disease pathogenesis at the individual level is a very complex process. The misleading message that an individual can prevent a particular disease by altering a particular behavior or exposure (and its converse, that an individual will develop a particular disease if such behavior is not changed) has unfortunately been widely conveyed. Rose coined the term "prevention paradox" to convey the irony that "many people must take precautions in order to prevent illness in only a few... *a preventive measure that brings large benefits to the community offers little to each participating individual*" (3, p. 12; Rose's emphasis). The prevention paradox is illustrated clearly in the above example of seat belts and traffic fatalities. Even assuming that seat belt-wearing halves the risk of death in an automobile accident, the probability that a given individual will benefit is small, since only a small minority of the population are killed in automobile accidents. Yet a small change in many small "individual risks" can have important effects on population disease burden.

The effort to prevent breast cancer provides a current example of some of the methodological problems inherent in the concept of individual risk, which by necessity underlies the high risk approach to prevention. Attempts to shift social and reproductive norms to reduce the breast cancer burden (for example, through the promotion of early and frequent child-bearing) would be considered unethical or undesirable in most societies. Such attempts to alleviate the public health problem of breast cancer could bring with them a net loss of public health. For this reason, there has been much discussion of strategies targeted toward women at high individual risk of the disease. However, the

recently proposed strategy of preventing breast cancer with administration of tamoxifen (or any other chemopreventive agent) demonstrates the prevention paradox well. Although it is being touted as a strategy that will be targeted only toward women with "high" estimated individual risk, many women must still engage in this preventive action in order for disease to be prevented in only a few. Currently, women with an estimated 5-year risk of breast cancer of 0.0166 (the 5-year risk for an "average" 60-year-old American woman) are advised to consider taking tamoxifen (20). Consider 100 women with an estimated 5-year risk of 0.04. Of this group, approximately four women will develop breast cancer over a 5-year period. If tamoxifen use reduces risk of breast cancer by approximately 50 percent (20), two of these four women will have their breast cancer prevented, while two of the women will still develop breast cancer. The remaining 96 women, who would have remained free of breast cancer without tamoxifen, will be exposed to increased risk of the adverse outcomes associated with this agent.

Another way to describe the above situation is in terms of individual risk: Each of the 100 women has her "high," "individual" 5-year risk of breast cancer reduced from 0.04 to 0.02. However, this statement is practically meaningless. It disguises the reality that "risk" refers to an aggregate-level state of health located outside of any one particular individual.

CONCLUSION

Recently, there have been calls for epidemiologists to return to a population perspective (21–28). Much of modern epidemiology continues to make use of the population perspective in terms of study design and analysis. However, results from these population-level analyses have been inappropriately limited to discussion of individual risk, with the consequent focus on identifying high risk individuals. This preoccupation has spawned a seemingly ceaseless search for disease risk factors that has extended into the human genome.

The vast majority of chronic disease risk factors uncovered during the past half century are associated with very modest increases in disease risk. Nearly all of them are unnecessary and insufficient to cause disease in individuals. This continues to hold true for many of the genetic "risk factors" being discovered now. These risk factors serve as extremely poor screening tools at the individual level. The equating of such factors with the causes of individual cases helps to foster an indifference to the social determinants of risk factor distributions, and may contribute to ineffectual disease prevention policies at the population level. Even for a disease such as breast cancer, where, unusually, there appear to be few or no primary prevention

strategies that can be ethically advocated at the population level, the logic of individual risk illustrates the prevention paradox. Many women will need to engage in preventive action (e.g., tamoxifen use) in order to prevent disease in a few, because knowledge of individuals' risk factor profiles does not allow accurate identification of the minority who will go on to develop the disease.

The increasingly sharp focus on the individual, and now on the genes and molecules of individuals, ironically comes at a time of growing inequity within our society and around the world and of growing possibilities for major, adverse, systemic changes in the global physical environment (30). Indeed, according to McMichael (31), over the next few decades humankind will probably need to achieve more rapid social, economic, and political change than ever before to sustain public health gains that have been made in this century. Rose's thoughtful work on the comparison between the individual and population-level perspectives on disease prevention is perhaps more relevant today than when he began discussing these ideas nearly 20 years ago. If epidemiologists wish to remain relevant to society, we should not shy away from the task of envisioning strategies by which to reduce risk at the population level when ethically possible—that is, to change the causes of causes.

REFERENCES

1. Feigelson H, Ross R, Yu M, et al. Genetic susceptibility to cancer from exogenous and endogenous exposures. *J Cell Biochem* 1996;25(suppl):15–22.
2. Khoury MJ. Genetic epidemiology and the future of disease prevention and public health. *Epidemiol Rev* 1997;19:175–80.
3. Rose G. Strategy of prevention: lessons from cardiovascular disease. *Br Med J (Clin Res Ed)* 1981;282:1847–51.
4. Rose G. *The strategy of preventive medicine*. New York, NY: Oxford University Press, 1992.
5. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;81:1879–86.
6. Benichou J, Gail MH, Mulvihill JJ. Graphs to estimate an individualized risk of breast cancer. *J Clin Oncol* 1996;14:103–10.
7. Spiegelman D, Colditz GA, Hunter D, et al. Validation of the Gail et al. model for predicting individual breast cancer risk. *J Natl Cancer Inst* 1994;86:600–7.
8. Heller RF, Chinn S, Pedoe HD, et al. How well can we predict coronary heart disease? Findings in the United Kingdom Heart Disease Prevention Project. *Br Med J (Clin Res Ed)* 1984;288:1409–11.
9. Katz D, Foxman B. How well do prediction equations predict? Using receiver operating characteristic curves and accuracy curves to compare validity and generalizability. *Epidemiology* 1993;4:319–26.
10. Tomatis L, ed. *Cancer: causes, occurrence, and control*. (IARC Scientific Publication no. 100). Lyon, France: International Agency for Research on Cancer, 1990.

11. Khoury MJ, Wagener DK. Epidemiological evaluation of the use of genetics to improve the predictive value of disease risk factors. *Am J Hum Genet* 1995;56:835-44.
12. Ponder B. Genetic testing for cancer risk. *Science* 1997;278:1050-4.
13. Hubbard R, Lewontin RC. Pitfalls of genetic testing. *N Engl J Med* 1996;334:1192-4.
14. Perera FP. Environment and cancer: who are susceptible? *Science* 1997;278:1068-73.
15. Rose G. Sick individuals and sick populations. *Int J Epidemiol* 1985;14:32-8.
16. Doll R, Hill A. Smoking and carcinoma of the lung: preliminary report. *Br Med J (Clin Res Ed)* 1950;September 30:739-48.
17. Brandt AM. The cigarette, risk, and American culture. In: Burger EJ Jr, ed. *Risk*. Ann Arbor, MI: University of Michigan Press, 1990:155-76.
18. Leon DA, Chenet L, Shkolnikov VM, et al. Huge variations in Russian mortality rates 1984-94: artefact, alcohol, or what? *Lancet* 1997;350:383-8.
19. Insurance Institute for Highway Safety. Fatality Analysis Reporting System (FARS), US Department of Transportation. Fatality facts 1998. Arlington, VA: Insurance Institute for Highway Safety, 1999. (http://www.highwaysafety.org/safety_facts/fatality_facts/facts.htm).
17. Burris S. The invisibility of public health: population-level measures in a politics of market individualism. *Am J Public Health* 1997;87:1607-10.
18. Rothman KJ, Greenland S. *Modern epidemiology*. Philadelphia, PA: Lippincott-Raven Publishers, 1998.
19. Schwartz S. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *Am J Public Health* 1994;84:819-24.
20. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst* 1998;90:1371-88.
21. Susser M. *Causal thinking in the health sciences: concepts and strategies of epidemiology*. New York, NY: Oxford University Press, 1973.
22. Beaglehole R, Bonita R. *Public health at the crossroads: achievements and prospects*. New York, NY: Cambridge University Press, 1997.
23. Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998;88:216-22.
24. McKinlay JB. The promotion of health through planned sociopolitical change: challenges for research and policy. *Soc Sci Med* 1993;36:109-17.
25. Pearce N. Traditional epidemiology, modern epidemiology, and public health. *Am J Public Health* 1996;86:678-83.
26. Shy CM. The failure of academic epidemiology: witness for the prosecution. *Am J Epidemiol* 1997;145:479-84.
27. Wing S. Limits of epidemiology. *Med Global Survival* 1994;1:75-86.
28. Diez-Roux AV. On genes, individuals, society, and epidemiology. *Am J Epidemiol* 1998;148:1027-32.
29. Brandt AM. Behavior, disease, and health in the twentieth-century United States: the moral valence of individual risk. In: Brandt AM, Rozin P, eds. *Morality and health*. New York, NY: Routledge, Chapman and Hall, Inc, 1997:53-78.
30. McMichael AJ. Prisoners of the proximate: loosening the constraints on epidemiology in an age of change. *Am J Epidemiol* 1999;149:887-97.
31. McMichael AJ. From hazard to habitat: rethinking environment and health. *Epidemiology* 1999;10:460-4.