Commentary: Making risk factors more cost-effective predictors of disease

William G Rothstein

Accepted 7 January 2009

This commentary will apply Michael Marmot’s call for theories that make more accurate predictions to risk factors, which provide the etiological model for most chronic diseases. The use of risk factor research in health care will be compared with that in the life insurance industry, which developed the concept. Two proposals will be made to improve the predictive accuracy and cost-effectiveness of risk factors in health care.

Marmot uses theories of the causes of coronary heart disease to examine models of causation in epidemiology and public health. He shows that science does not proceed in a straightforward inductive chain from facts to causal hypotheses and theories. Nor can hypotheses be refuted by simply producing facts that contradict them. Consequently, he suggests that a useful theory is one that produces relationships which accurately predict health outcomes.

Accurate predictions, however, should not be the only objective of health care models, especially since predictions tend to be less accurate in chronic diseases than in most infectious diseases. A problem of equal urgency is cost-effectiveness. Too much health research fails to incorporate cost-effectiveness in its design or analyses and consequently too much health care produces only a small amount of benefit for the societal expenditures involved.

Cost-effectiveness is a key issue because of the steep increases in health care expenditures that have afflicted all advanced countries. In the USA, health care expenditures have reached one-sixth of the gross domestic product and are climbing steadily. Individual states spend up to one-quarter of their budgets on medical care for the poor. Increased public and private expenditures for health care have reduced the standard of living in many ways, including less spending for education, social welfare, income security, the environment, infrastructure and public safety.

More concern with cost-effectiveness is also warranted because of the failure of many health care expenditures to significantly improve the health of the population. Critics point out that millions of patients have undergone diagnostic procedures, ingested medications and endured surgical operations for chronic diseases with little or no benefit to their health or lifespans. Many of these patients have suffered adverse reactions, ranging from unpleasant to fatal. Major population groups in all advanced countries are being urged to take medications or change their lifestyles in ways that are unlikely to produce any improvement in their health.

Much of the blame for this situation has been directed at the use of risk factors because of the low probabilities of disease outcomes associated with them. In operational terms, risk factor research measures and compares the probabilities of the future occurrence of disease or death in subsamples with the same characteristics. The subsamples may differ from each other with regard to personal features, such as body weight, or behaviors, such as dietary fat intake. Some subsamples in a sample may receive drugs or be asked to change their lifestyles. After a period of time, the study measures the proportion of each subsample in the sample that experiences risk-factor related mortality or morbidity per year. It then compares the morbidity or mortality rates for subsamples with different personal features, behaviors, or characteristics. The objective of the research is to determine which characteristics of the subsamples produce better health and to generalize those findings to populations of interest.

Because the risk factor was developed by the life insurance industry, it is useful to compare the methodologies of risk factor research in that industry, on the one hand, and in health care, on the other. Life insurance companies have always based their selection criteria on the statistical probabilities of death of large numbers of persons with certain characteristics. The risk factor was developed as a selection tool primarily to enable the companies to insure certain classes of previously uninsurable persons by adjusting their premiums to their higher levels of risk. Risk factors, like all life insurance selection criteria, are characteristics of large groups and are not designed to predict mortality in individuals.
The objective of life insurance risk factors is to predict the probability of premature death in specific groups, not to determine the causes of death. Moreover, the life insurance industry seeks risk factors that have good predictive value and are accurately and inexpensively measured. For this reason, many life insurance risk factors, such as age, sex and occupation, are not in themselves causes of mortality.

Life insurance companies use rigorous statistical methods in their risk factor research that produce highly accurate findings. They found through long experience that future policyholders will be very similar to past and present policyholders. They use formal sampling methods to draw samples of multiple thousands of current and deceased policyholders and sometimes rejected applicants. This insures that the statistical methods used to analyse the sample data meet the assumptions of the methods. The risk factor findings obtained from these studies can be generalized with great confidence to future applicants for life insurance policies.

The risk factor research used by health researchers differs in basic ways from life insurance research. First, health researchers almost never use formal sampling methods to draw a sample from the population to which the findings are to be generalized. This violates a basic requirement of the inferential statistical tests that the researchers use to analyse the data.

Secondly, most health research studies of risk factors use selection methods that produce samples which are not representative of any specific population of interest. Clinical trials and similar studies use volunteers who are atypical in their health status and behaviours and other characteristics (pp. 233–35). Telephone and email surveys and personal interviews produce biased samples because of the high non-response rates. Studies of patient records over time produce biased samples because of the high non-response rates. Studies of patient records over time exclude persons for whom longitudinal records or records of any kind are unavailable.

Thirdly, the sample sizes used in health research are almost always a small fraction of those used by life insurance companies. This produces an emphasis on inferential statistical tests, which are strongly affected by sample size. In life insurance studies, the sample sizes are so large that the focus is on the fundamental issue—the absolute level of risk produced by the risk factor.

Fourthly, prospective studies of common risk factors in samples of healthy adults, such as serum cholesterol or blood pressure, are of particular concern because they can have a major impact on health costs. They can lead millions of healthy persons to receive diagnostic tests, take medications indefinitely, change their lifestyles and have frequent medical examinations. The risk factor-related mortality and morbidity rates in these studies are often ~1% per year and the differences among the groups in the sample are a fraction of 1%. Such small differences based on samples of unrepresentative volunteers cannot be generalized with confidence to any groups in the population.

The deficiencies of risk factor research in health care do not mean that the risk factor concept is fundamentally flawed. The life insurance experience demonstrates conclusively that the risk factors can be both accurate predictors of health outcomes and very cost-effective. Two underutilized properties of risk factors can greatly improve their predictive value and cost-effectiveness in actual health care applications.

The first property is a clearer recognition that some risk factors interact with each other and produce greatly elevated levels of risk, when they are combined in a given individual. Starting about 1950, the Framingham Heart Study conducted biennial medical examinations of a representative sample of more than 5000 of the 28,000 residents of Framingham, MA, a suburb of Boston (pp. 279–85). The city’s annual census was used to draw the sample and permitted the researchers to evaluate the representativeness of the sample. The biennial medical examinations enabled the researchers to measure risk factors frequently and diagnose diseases accurately. This method was far superior to the typical multi-year study that draws an unrepresentative sample, measures the risk factors only at the onset of the study and uses death certificates to measure the cause of death.

Framingham researchers examined combinations of three risk factors in men aged 30–59: smoking 21 or more cigarettes daily, serum cholesterol levels of 250 mg/dl, and blood pressures of 160/95 mmHg. After 8 years, men with none of the three risk factors had a risk of developing coronary heart disease (excluding angina pectoris) that was one-half of the average of their age group; those with one of the risk factors had a risk slightly above the average of their age group; those with two of the risk factors had a risk twice that of their age group; and the very few men with all three risk factors had a risk five times that of their age group.

This finding indicates that some combinations of risk factors can produce very large increases in mortality and morbidity rates among groups. The 10-fold increase in risk of persons with all three risk factors compared with those with none is many orders of magnitude greater than the typical increase in the risk found in risk factor studies. These much larger differences can be generalized to populations with much greater confidence. Interventions will benefit practically all of the persons involved and be highly cost-effective.

A second underutilized property of health-related risk assessment is recognition of the usefulness of non-causal risk factors. Causal risk factors are of little value to health professionals when there are
large numbers of them and they vary greatly in importance among persons. In these cases, it is more useful to identify high-risk patients using combinations of non-causal risk factors that have high predictive value. Patients with those risk factors can then be treated individually.

Non-causal risk factors are most useful when they can be observed easily and inexpensively, as the life insurance industry has shown. Readily observable characteristics of all persons include age, sex, race and body weight. Characteristics that should be known to the patient’s health care provider include education, occupation, ethnicity, residence, smoking, physical activity, lifestyle, regular source of medical care, family circumstances, living arrangements and medical history.

Two examples of useful non-causal risk factors can be cited. One is the strong positive correlation between age and hypertension in advanced societies. Age is not a cause of hypertension because age is not related to hypertension in less developed societies. Yet, health care providers know that measuring blood pressure is much more important with older than with younger patients. Another example is the finding that the living arrangements of older persons are related to their health. In both cases, knowledge of a non-causal risk factor can improve the delivery of health care and lower morbidity and mortality.

A situation where the application of non-causal risk factors would be extremely cost-effective but have not been identified is medication non-adherence. Failure to adhere to a medication regimen is a major problem in chronic diseases because medications must be taken indefinitely. Many patients do not admit that they are non-adherers and require considerable time and effort to be identified. Patients have many and varied reasons for non-adherence that are not highly correlated with observable non-causal risk factors such as age, sex and race. As a result, most non-adherence research has focused on psychological and other non-observable causes that can only be identified by expensive and time-consuming discussions with the patient. Even then, patients will often make misleading statements to please the health care provider. It is highly unlikely that this type of research can reduce non-adherence cost-effectively.

Greater emphasis on identifying observable combinations of non-causal personal characteristics or behaviours that have a high correlation with non-adherence would have many advantages. Probable non-adherers could be identified and health care providers could discuss many possible reasons for non-adherence with them. The same model would be equally useful for other types of non-adherence, such as dietary restrictions in diabetes and obesity.

The benefits of non-causal correlations extend beyond individual patient characteristics to include neighbourhoods, schools, work sites, occupational groups and others. Public health departments have used these non-causal correlations for over a century to provide cost-effective public health programmes to groups that can benefit the most from them.

This commentary has shown the feasibility and benefits of more accurate and cost-effective predictions using combinations of causal and non-causal risk factors. Identifying high-risk patients using combinations of risk factors enables health care providers and health educators to focus their resources on fewer patients who are very likely to benefit from them. Patients can be graded by their level of risk and their treatments adjusted accordingly. This produces more cost-effective health care with greater benefit to patients, builds public confidence in the health care system and frees already scarce public and private resources for other social needs.

Acknowledgements
The author expresses his appreciation to James Mohr and Michael Weinrich for their helpful comments.

Conflict of interest: None declared.

References