Progress in Cardiovascular Diseases 53 (2010) 62–67
www.onlinepcd.com

Deciphering the Causes of Cardiovascular and Other Complex Diseases in Populations: Achievements, Challenges, Opportunities, and Approaches

Salim Yusuf*, Sonia Anand

Population Health Research Institute, McMaster University and Hamilton Health Sciences, Hamilton, Ontario, Canada

Abstract This paper provides an overview of key achievements of the Framingham Heart Study and identifies areas and approaches for future research in cardiovascular disease epidemiology and prevention. There is a need for a range of different studies using diverse designs (i.e. case-control, cohort, multi-community, birth cohort, family-based cohorts and randomized trials) in different settings and involving multiple ethnic groups. Incorporation of a range of new disciplines, such as genetics, behavioural sciences, social epidemiology, measures of the environment, geography, and health policy are required to understand the root determinants of cardiovascular diseases. (Prog Cardiovasc Dis 2010;53:62-67)
© 2010 Elsevier Inc. All rights reserved.

Keywords: Cardiovascular diseases; Prevention; Causes; Risk factors

Although individual cases of cardiovascular disease (CVD) such as acute myocardial infarction (MI) and stroke had been described many thousands of years earlier, it was only in the middle of the 20th century that CVD and other chronic diseases reached epidemic proportions. This epidemic first affected the developed countries (especially men) and was relatively uncommon in the less developed countries.[1] A number of population-based studies in North America, UK, and Europe were established between 1950 onward and report that a few simple risk factors could separate populations into those who had a low risk of future CVD and those with up to a 30-fold higher risk. As asserted by Kannel et al[2] in 1967, coronary heart disease (CHD) affects "mainly flabby, sedentary males (or older females) given to excess of cigarette smoking, diets rich in saturated fats, and living with the conveniences of modern, labor-saving devices." Importantly, some of the risk factors identified through the early cohort studies (such as cigarette smoking, elevated blood pressure [BP], and abnormal lipids) were potentially modifiable and challenged early beliefs that atherosclerotic disease and its outcomes were not preventable. The Framingham Heart Study (FHS) was among the earliest and longest running observational studies. The FHS was relatively large for its time (but modest by modern standards), confined largely to 1 small town, and included both randomly selected individuals and volunteers. It was among the first cohort studies with prospective detailed data collection on risk factors using standardized measures. With more than a thousand publications over 6 decades, the study has contributed substantially to our understanding of the causes of CVD including the continuous nature of the association of serum cholesterol and BP to CVD, the importance of cigarette smoking, diabetes, overweight, and the higher risk of stroke associated with atrial fibrillation. The Framingham risk score has been developed to predict future events[3] (although it requires recalibration for other European versus North American populations and requires study in non–European populations). The FHS also developed

Statement of Conflict of Interest: see page 67.

* Address reprint requests to Salim Yusuf, DPhil, Population Health Research Institute, McMaster University and Hamilton Health Sciences, 237 Barton Street East, Hamilton, Ontario, Canada L8L 2X2.
E-mail address: yusufs@mcmaster.ca (S. Yusuf).

0033-0620/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.pcad.2010.02.001

Abbreviations and Acronyms
CVD = cardiovascular disease
CHD = coronary heart disease
BP = blood pressure
FHS = Framingham Heart Study
MI = myocardial infarction

novel statistical methods such as multivariate regression methods and pioneered the use of repeated measures of risk factors in the same population.[4] More recently articles have described the "life-long" risks of various outcomes, highlighted the importance of a long-term perspective when considering the impact of risk factors on chronic diseases,[5] and has contributed to our understanding of the genetic factors associated with cardiovascular risk factors such as lipids and BP. Without question, the FHS has had a positive influence in emphasizing the preventability of CHD and strokes (later confirmed unequivocally by randomized trials of BP and lipid lowering) and the predictability of CVD, development of novel statistical and epidemiologic methods and has had an impact on the design of other epidemiologic studies. Over its 6 decades of conduct, many epidemiologists and statisticians were trained, several of whom have made major scientific contributions especially in the United States. Thus, the FHS together with other major epidemiologic studies have laid the foundations for the prevention of CVD. Consequently, in many Western countries such as the United States, Canada, UK, and Finland, CVD rates have declined markedly largely due to better risk factor control and the implementation of evidence-based secondary prevention measures.

Although the FHS remains an important foundation of our knowledge of CVD risk factors, much remains to be learned and implemented. Studies such as the FHS (and other studies) point the way to newer, larger, and more diverse studies with expanded goals and using some of the new technologic advances in biology and computation. Below, we will discuss some pertinent issues.

Need for studies in diverse regions, multiple ethnic groups, and in urban and rural settings

Over the last 50 years, CVD rates have been increasing in low- and middle-income countries, whereas it has been declining in most high-income countries. Despite these trends, little was known regarding whether CVD risk factors were similar or different in various regions and ethnic groups. The INTERHEART case-control study involving 27,000 subjects from 52 countries representing all inhabited continents of the world has demonstrated the universal impact of 9 common and potentially modifiable risk factors for MI, which account for more than 90% of the population attributable risk of MI globally.[6] Ideally, this study should be complemented by cohort studies to ensure that the results are replicated using different methodologies and to obtain information on absolute age-adjusted event rates in multiple regions of the world. Given the marked variations in lifestyles in different geographic regions, information is needed from both urban and rural settings in low-, middle-, and high-income countries. By studying diverse populations with marked variations in risk factor levels, a more complete picture of the relationships from very low to very high levels of risk factors versus disease rates can be obtained. Although relating biologic risk factors (eg, BP, lipids, or glucose) to outcomes across diverse populations would be expected to be relatively straightforward, measuring health-related behaviors is more challenging. For example, dietary assessments (not available in several of the early studies such as Framingham) are complicated by the immense variations in dietary patterns, portion sizes, food types, and cultural factors (such as an entire family eating from common dishes) that make comparable quantitative assessments of energy or nutrient content of different diets across different regions and ethnic groups particularly challenging. Furthermore, methods of cooking may alter the nutritional content of some foods (eg, folate may be destroyed by prolonged cooking),[7] or the impact of salted foods may differ from that of the same foods, if they are unsalted. Similarly, the types of physical activities vary across societies where energy expenditure during utilitarian activity (either at work or during household chores) is much greater than that during leisure time activity. Even a relatively straightforward habit such as tobacco use shows marked variations (eg, smoking of cigarettes, indigenous smoking like the water pipe or beedies, reverse smoking, or use of smokeless tobacco of various types).[8] Although such diversity creates challenges in measurement and comparability, it also creates new opportunities; for example, very low levels of regular smoking of 2 to 3 cigarettes a day can be best studied in societies where low intensity of smoking is prevalent, thereby characterizing a "dose-response" relationship at the low-end of exposure (which is not readily possible in Western countries). Demonstrating the adverse effects of both smoked and smokeless tobacco indicates that the harm from tobacco is inherent to the leaf and may not be solely related to toxins produced when tobacco is smoked.[8]

Very large cohorts needed to study the impact of health behaviors and various exposures on multiple diseases and outcomes

It is now apparent that the same populations and individuals are often at risk of multiple conditions and

diseases (eg, CHD, stroke, diabetes, common cancers, and chronic obstructive airways disease), which are often caused by the same risk factors (poor diet, lack of activity, tobacco use, especially smoking, etc). It is also possible that a significant proportion of cognitive impairment at older ages may be associated with the same risk factors suggesting that cognitive decline may be a disease manifestation of the small vessels of the brain.[9-11] Therefore, future studies should aim to study the impact of multiple exposures and risk factors on several conditions simultaneously. Such an approach, although initially more expensive, will provide substantially richer data on the impact of the same risk factor on multiple diseases (which provides better quantification of the overall health impact of specific behaviors) as well as studying the interrelationship between different diseases. For example, one might gain insights into biologic mechanisms by understanding why a poor diet or smoking is associated with quite different diseases (eg, CVD and some cancers).

Need for very large studies

Future studies need to be planned to be sufficiently large so that several thousands of specific events and their subtypes accrue within a relatively short period for several reasons. First, subtypes of specific events, such as stroke (ischemic due to small, large vessel disease or embolism; intracerebral hemorrhage, or subarachnoid hemorrhage) may have differing risk factors and at times clarifying the varying strengths of relationship between a risk factor (BP), and the various subtypes of the stroke (eg, strong relationship of BP with ischemic stroke, but an even stronger relationship with intracerebral hemorrhage) will provide greater insights into disease causation. Similarly, is the relationship between BP or lipids and the different types of ischemic strokes (eg, small vessel, large vessel, or embolic) the same or different? To accrue large numbers (eg, 1000-2000 cases of intracerebral hemorrhage) of a specific subtype of stroke, either large case-control studies or much larger cohort studies are needed. Second, the "strong" risk factors (odds ratio of 2.0 or greater) for common conditions have probably already been discovered by previous studies, and it is likely that any remaining undiscovered risk factors (eg, genetic) may have much more modest effects (odds ratio of 1.2 to 1.5), which necessitates studies with a large number of events to not only detect main and independent effects (after multivariate adjustments) but also to assess whether the strength of relationship varies in different ethnic groups, by age or sex. To make such studies feasible, they may have to be "streamlined" in their approach to data collection, store blood and urine samples for future analyses using a nested case-control approach, and have repeated sampling and collection of key data in only a small portion of the population (eg, 5%-10%) at periodic intervals (eg, every 3 to 5 years). The latter information can be used to assess changes in risk factors over time as well as used to mathematically calculate "usual" levels of risk factors (by overcoming "regression-dilution" biases) to obtain a more reliable estimate of the strength of relationship of a specific risk factor with specific outcomes.

Need to study the "causes" of the "causes": importance of gene-environment interactions

A large proportion of the risk for some diseases such as MI can be explained by a handful of risk factors (eg, just 3 risk factors: tobacco use, elevated BP, and abnormal lipids, probably account for >80% of the population attributable risk of acute MI). It is possible that most of the risk of many other chronic diseases (eg, stroke, common cancers, etc) may be explained by a handful of risk factors. Given that many of these risk factors are likely to be the result of environmental and behavioral factors interacting with genetic endowment throughout life, a comprehensive understanding of what causes the risk factors ("causes of causes") is usually needed to shed light on the whole chain of causation of

Table 1
Key design features of studies to understanding the causes of CVD and other complex adult conditions

1. Large size so that several thousands of relevant *events* accrue
2. Characterization of outcomes so that specific *subtypes of events* which may have different pathogenesis may be identified
3. Focused data collection to facilitate large studies with focused data collection to enable the above 2 goals
4. Enrollment of individuals from a large number of communities that are geographically diverse to obtain a broad range of environments, health behaviors, risk factors, and disease rates
5. Repeat measurement of risk factors and other exposures in a small subset of the study population to estimate "usual" levels of exposures
6. Standardized methods of data collection and outcomes ascertainment with high rates of follow-up
7. Collaborations with scientists in social epidemiology, behavioral sciences, geography, economics, urban planning and architecture, and food and tobacco policy
8. Characterization of the environment, health behaviors, physical characteristics, biologic samples, and genes to explore environment-behavior-genetic interactions on the development of risk factors and disease

chronic diseases.[12] Future studies should invest in developing measures of the environment (eg, built-environment that may influence physical activity; nutrition environment, which in turn affects diet; tobacco environment, which influences rates of initiating smoking or quitting; air quality—both indoor and outdoor; water quality; and soil quality)—that may all be related to health. Developing detailed measurements of the environment requires collaboration with social epidemiologists, health economists, geographers, and soil and water experts. Given that environmental factors often tend to affect groups of individuals living in common environments, studies with a large number (several hundred) of geographically diverse communities (with each community involving several hundred individuals) are necessary.

Given that the impact of most common genetic markers (eg, frequency >5%) for common complex conditions such as CVD, diabetes, and cancers tend to have modest effects (odds ratios between 1.1 and 1.4), large studies are needed to discover these reliably. Alternatively, a few rare variants may have large effects. Identifying both these types of genetic markers requires very large studies. Demonstrating the replicability (or lack of replicability if variations in linkage disequilibrium of various single nucleotide polymorphism (SNP)s exist) in different ethnic groups will provide better characterization of the genetic locus and may assist in understanding the functional relevance of a given SNP more readily. The detection of gene-environment interactions (eg, the observation that the risk associated with genetic variant 9p21 on MI is attenuated by high consumption of fruits and vegetables)[13] provides insights into the biologic mechanisms by which genes and health behaviors interact, as well as provides approaches by which those at high genetic susceptibility can modify their risks. Genetic markers can also provide insights into whether any observed association of a clinical marker (eg, C-reactive protein) with CVD is casual or due

Table 2
Potential strengths and weakness of different approaches to understanding disease causation of CVD in populations

1. Case-control study
   Strengths:
   1. Ability to enroll a large number of well-characterized cases [eg, stroke or its subtypes (eg, intracerebral hemorrhage, large vessel cerebral infarction, or small vessel infarcts)], especially at young ages, across different countries, ethnic groups, and both sexes
   2. Best design to study acute or recent precipitators of an event
   3. Cost and time efficient
   Weaknesses:
   1. Potential for reverse causality and confounding, but this can be minimized by careful and standardized selection of cases and controls and explored through careful statistical analyses
   2. Information generally limited to disease under study
2. Cohort study
   Strengths:
   1. Ability to measure risk factors before an event has occurred and less likely to be subject to biases in comparing cases versus controls
   2. Multiple disease outcomes can be studied.
   Weaknesses:
   1. Inefficient and expensive and requires very large populations followed for at least 5-10 y
   2. Biases can arise due to differential rates of follow-up by risk factor levels. Reverse causality may still be a concern if subclinical disease influences health behaviors or risk factors.
   3. Not suitable for detection of acute/recent precipitators of disease
   4. Difficult to accrue a large number of events in those at lower risk, eg, young women for MI
3. Comparisons of individuals across communities
   Strengths:
   1. Marked variations in exposures and disease rates and the ability to assess environmental exposures that affect entire communities
   Weaknesses:
   1. Potential for additional confounding (eg, “ecologic fallacy”), which can be overcome by obtaining data at both the group and individual levels
   2. Need for a large number of diverse and distinctive communities
4. Life-course studies
   Strengths:
   1. Potential to characterize exposures and intermediate (eg, risk factors) outcomes during a prolonged period of time
   Weaknesses:
   2. Feasibility of large studies needed to assess clinical outcomes doubtful; high rates of attrition during prolonged follow-up
5. Family-based cohort studies
   Strengths:
   1. Potential to characterize entire families, thereby providing information on intergenerational tracking of risk factors
   Weaknesses:
   1. Unsuitable for studying clinical outcomes because of low event rates
   2. Potential for high attrition rates during prolonged follow-up, as young families are more likely to move

to confounding, by using the approach of Mendelian randomization.[14] Thus, future large studies will benefit from inclusion of a large number of individuals of diverse ethnic groups, living in a range of settings, among whom the environment, health behaviors, and anthropometric and biologic measures are obtained and are related to specific disease outcomes, which may at times require careful characterization by using advanced imaging modalities (eg, computed tomography or magnetic resonance imaging scans to distinguish different subtypes of strokes) or tissue characterization (eg, histologic diagnosis of the type of cancer, eg, adenocarcinoma or squamous cell carcinoma of the lung). Such a comprehensive approach is required to best understand the causes of disease and requires very large case-control studies (eg, 10,000-20,000 cases of stroke are needed to obtain information on various stroke subtypes and a similar number of controls) or unusually large cohort studies (eg, involving more than 100,000-200,000 middle-aged or older individuals followed for 5 to 10 years) so that several thousands of specific events of interest can be identified (Tables 1 and 2).

When do risk factors develop? Need for a life course approach

It is likely that risk factors for complex diseases start early in life, and there are some claims that even intrauterine influences may have an important effect on risk factors as well as some adult diseases (eg, type 2 diabetes). Although there is some evidence that use of hormones (eg, stilboestrol) during pregnancy predisposes the baby to developing certain cancers (eg, vaginal cancer) in adult life,[15] there are very few other examples where intrauterine exposures affect adult diseases. It is highly likely that the environment and health behaviors in early life (eg, 5-10 years) will have an important impact on elevations of risk factors in early adulthood (20-30 years), which, in turn, influences disease development in middle and old age (40-70 years). Although large and prolonged studies (eg, for 5-6 decades), which are required to characterize the influence of early-life exposures on disease in middle age, are impractical, an alternative is to study the impact of gene-behavior-environmental effects on the development of major risk factors for CVD (eg, lipids, BP, glucose, and body fat) in early childhood (first 5 years), later childhood (5-10 years), adolescence, teens, and among young and middle aged adults. Such studies could be conducted as separate cohorts enrolling individuals during the intrauterine phase, at birth, in childhood, adolescence, and adult life using standardized measures or alternatively studying entire families, with the "inception" being when a woman is pregnant. Integrated information from such groups of cohorts or families (involving individuals at different age groups) followed for 10 to 20 years, using meta-regression methods, may allow characterization of gene-environment interactions throughout life and provide projections early in life as to who will develop CVD in middle and old age.

Using population-based cohort studies to track the use of prevention strategies for risk factor control and for secondary prevention

Hospital-based registries tend to suggest relatively high rates of use of secondary prevention strategies in those discharged after acute coronary syndrome or stroke, whereas community-based studies have suggested much lower rates of use of such strategies. Hospital-based registries tend to be based on centers who volunteer to participate in such studies. These centers may pay greater attention to prescribing key drugs and tend to reflect the use of drugs in the early months or years after an event. By contrast, community-based data obtained by cross-sectional or cohort studies represent the practice of a broader range of physicians in a community, reflect longer term use and incorporate nonadherence of subjects, and are a better reflection of the actual implementation of preventive practices in the community. Such studies can also assess the levels of BP control, tobacco control (ratio of current versus former smoking by age and sex), lipid, and diabetes management in communities; such studies can identify the "prevention gap"; and by investigating the reasons for the underuse of proven strategies in different settings, it will assist in developing new strategies to reduce and bridge the prevention gap.

From observations to interventions and policy

During the last few decades, the modifiability of risk factors and consequent diseases have been demonstrated for BP lowering (at least for levels approximately 160 mm Hg systolic) and for those with elevated low-density lipoprotein.[16] Few studies, if any, have either tested these approaches together, especially in those with "average levels" of these risk factors. Wald and Law[17] have claimed that combined BP lowering (using low doses of multiple drugs) and lipid lowering (with a statin) could potentially reduce CVD safely in the entire adult population by more than 75%. These concepts have led to the development of the polypill or the polycap,[18] with large trials underway to evaluate the effects of combined lipid lowering and BP lowering on clinical outcomes in "average" risk individuals. Demonstration of the efficacy, cost-effectiveness, and tolerability of the polypill/polycap has the potential to revolutionize the approach to primary prevention, by using

simple clinical markers to assess risk (eg, age) and demonstrating benefits at average levels of risk factors. Results from large trials are needed to use the polypill/polycap in primary prevention. By contrast, given that each component of the polypill/polycap (other than folate proposed by Law and Wald) has been shown to reduce mortality and morbidity after a vascular event, such therapy should be readily used in secondary prevention, once safety of administration has been demonstrated.

Summary

The "first-generation" epidemiologic studies like the FHS were usually conducted in single locations in Western countries and have provided important information that has been the foundation of CVD prevention. Future studies are needed to address the increasing burden of CVD in developing countries and should be much larger and involve populations from urban and rural communities from low- and middle-income countries. By characterizing the environment of communities, health behaviors, and genetic makeup of these populations, such studies will provide unique insights into the biologic causation of CVD. Careful and precise characterization of potential risk factors and outcomes (eg, using advanced imaging modalities for stroke or tissue histologies for cancers) can provide more accurate assessments of the strengths of associations overall and in key subgroups. Studies incorporating different designs (case-control studies, prevalence studies, cohort studies, studies in individuals at different ages, and comparison of communities) are required because each approach has unique strengths that compliment the drawbacks of the other methods. Integrating modern epidemiologic methodologies with technologic advances in genetics and imaging, emerging insights into biology, expertise from other disciplines, and powerful computational methods into large studies involving diverse populations will lead to substantial new knowledge about the causation of CVD and several other common complex adult diseases. Ultimately, where feasible, the modifiability of risk factors should be demonstrated in large trials which include individuals from different ethnic groups and regions of the world.

Statement of Conflict of Interest

All authors declare that there are no conflicts of interest.

References

1. Yusuf S, Reddy S, Ôunpuu S, et al: Global burden of cardiovascular diseases: part I: general considerations, the epidemiologic transition, risk factors, and impact of urbanization. Circulation 2001;104: 2746-2753.
2. Kannel WB, Castelli WP, McNamara PM: The coronary profile: 12-year follow-up in the Framingham study. J Occup Med 1967;9:611-619.
3. D'Agostino Sr RB, Grundy S, Sullivan LM, et al: CHD Risk Prediction Group. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA 2001;286:180-187.
4. Truett J, Cornfield J, Kannel W: A multivariate analysis of the risk of coronary heart disease in Framingham. J Chronic Dis 1967;20:511-524.
5. Lloyd-Jones DM, Leip EP, Larson MG, et al: Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. Circulation 2006;113:791-798.
6. Yusuf S, Hawken S, Ounpuu S, et al: Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet 2004;364: 937-952.
7. Kelemen LE, Anand SS, Hegele RA, et al: Associations of plasma homocysteine and the methylenetetrahydrofolate reductase C677T polymorphism with carotid intima media thickness among South Asian, Chinese and European Canadians. Atherosclerosis 2004;176: 361-370.
8. Teo KK, Ounpuu S, Hawken S, et al: Tobacco use and risk of myocardial infarction in 52 countries in the INTERHEART study: a case-control study. Lancet 2006;368:647-658.
9. Anstey KJ, Lipnicki DM, Low EF: Cholesterol as a risk factor for dementia and cognitive decline: a systematic review of prospective studies with meta-analysis. Am J Geriatr Psychiatry 2008;16: 343-354.
10. Nagai M, Hoshide S, Kario K: Hypertension and dementia. Am J Hypertens 2010;23:116-124.
11. Peters R, Poulter R, Warner J, et al: Smoking, dementia and cognitive decline in the elderly, a systematic review. BMC Geriatrics 2008;8: 36-42.
12. Chow CK, Lock K, Teo K, et al: Environmental and societal influences acting on cardiovascular risk factors and disease at a population level: a review. Int J Epidemiol 2009;38:1580-1594.
13. Engert JC, Do R, Zhang X, et al: Diet modifies the myocardial infarction risk of chromosome 9p21 variants globally. Poster session presented at: 59th Annual Meeting of the American Society of Human Genetics; 2009 Oct 20-24; Honolulu, Hawaii; 2009.
14. CRP CHD Genetics Collaboration: Collaborative pooled analysis of data on C-reactive protein gene variants and coronary disease: judging causality by Mendelian randomisation. Eur J Epidemiol 2008;23:531-540.
15. Herbst AL, Ulfelder H, Poskanzer DC: Adenocarcinoma of the vagina: association of maternal stilbestrol therapy with tumor appearance in young women. N Engl J Med 1971;284:878-881.
16. Yusuf S: Two decades of progress in preventing vascular disease. Lancet 2002;360:2-3.
17. Wald NJ, Law MR: A strategy to reduce cardiovascular disease by more than 80%. BMJ 2003;326:1419.
18. Indian Polycap Study (TIPS), Yusuf S, Pais P, et al: Effects of a polypill (Polycap) on risk factors in middle-aged individuals without cardiovascular disease (TIPS): a phase II, double-blind, randomised trial. Lancet 2009;373:1341-1351.