Is an Ounce of Prevention Worth an Ounce of Cure? Explaining the Decline in Cardiovascular Mortality, 1964-2010

Abstract

Mortality from coronary heart disease in the United States has fallen 60% from its peak. Cardiologists and epidemiologists have debated whether this decline reflects risk factor control or the power of medical therapeutics. Attempts to resolve this debate and guide health policy have generated sophisticated datasets and techniques for modeling cardiovascular mortality. Neither effort, however, has provided specific guidance for health policy. Historical analysis of the decline debate and the development of cardiovascular modeling offers valuable lessons for policymakers about tensions between medical and public health strategies, the changing meanings of disease prevention, and ability of evidence-based research and models to guide health policy. Policymakers must learn to open up the black box of epidemiological models -- and of their own decision making processes -- to produce the best evidence-informed policy.
At some unnoticed moment in the mid-1960s, mortality from coronary heart disease (CHD) in the United States peaked and began to decline. Even though CHD remains the leading cause of death worldwide, CHD mortality has fallen 60% from its zenith.¹ Although this may represent the greatest public health achievement of the twentieth century, it is not clear who or what deserves credit. Does the decline demonstrate the therapeutic power of modern medicine or the impact of lifestyle change and risk factor management? Cardiologists and epidemiologists have struggled for nearly forty years to resolve this question. They have developed sophisticated datasets and quantitative models of the factors that cause or mitigate cardiovascular mortality. They hoped that their analyses would help policy makers decide whether to invest in treatment or prevention. The analyses, instead, have consistently demonstrated the value of both.

Much can be learned by studying the history of the intersection of cardiovascular epidemiology and health policy in this decline debate. As researchers’ methods evolved from anecdote to back-of-the-envelope calculations to complex models, increasing precision came at the cost of increasing opacity. Few readers will understand the subtle mechanisms and
assumptions on which such models rely. The models, however, have become more versatile and ubiquitous. It is essential for policymakers to understand the promise and limitations of the models, the changing meanings of prevention, and the likelihood that empirical research might shape policy. Just as the decline itself provided a “natural experiment” for cardiovascular epidemiologists, the decline debate provides an opportunity to analyze the dynamic interplay between knowledge production and health care policy.

The Rise and Fall of Coronary Heart Disease

Heart disease devastated postwar America. By 1960 it killed one-third of all Americans. Led by the American Heart Association and the National Heart Institute, the country mobilized unprecedented resources against the scourge. The first tidings of changing fortune came in 1964 when state health officials reported a decline in CHD mortality in California, but this news received little attention. Well into the 1970s, cardiologists and the national media sounded the alarm about the inexorable rise of CHD. In March 1974, however, a “sign of spring” emerged. Cardiologist Weldon Walker reported that age-adjusted CHD mortality rates had actually been declining in the United States since 1963. His announcement was met with guarded enthusiasm: not everyone was certain whether the decline was
real or not. Part of the problem was the delicacy of epidemiology as a historical science, dependent upon cause of death reporting with disease taxonomies that are constantly in flux. Depending on how epidemiologists parsed mortality data, CHD reached its apogee in either 1963 or 1968.

Eager to reach consensus about the reality and causes of decline, director Robert Levy called leading researchers to the National Heart, Lung, and Blood Institute in October 1978 for what became known as the “Decline Conference.” Epidemiologists and clinicians concluded that the decline -- a 20% drop between 1968 and 1978 -- was “real.” But debate continued on the second question: what had caused the decline? The timing of the decline had coincided with too many relevant changes: with vigorous efforts to educate Americans about smoking, diet, and other CHD risk factors; with changes in medical care, including aggressive control of hypertension, specialized coronary care units, β-blockers, and bypass surgery; and with the passage of Medicare and Medicaid.

This was not just an academic question. Everyone present felt the “urgent need” to answer the question and allow “intelligent decisions about the allocation of scarce resources between competing programs.” Clinicians, in particular, were on the defensive. Inspired by the critiques by Thomas McKeown and Ivan Illich, debates raged in the 1970s about whether medicine
made substantial contributions to the health of society and whether its contributions justified its growing cost. These critiques cast a long shadow. Speakers at the Decline Conference invoked Illich as a warning to those who would take the value of medicine for granted. Proponents of both medical care and disease prevention knew that they had to make their case carefully.

Quantifying the Value of Prevention and Treatment

The first decade of decline research saw arguments based simply on temporal association. Walker, who noted that education campaigns in 1964 coincided with the onset of decline, favored prevention. The accompanying editorial emphasized the impact of coronary care units. Although concordance might suggest cause and effect, critics reminded that the evidence was circumstantial. As preventive cardiologist Jeremiah Stamler complained in 1978, “when such multiple socio-medical trends evolve over the years, it is virtually impossible to make a definitive scientific assessment as to the role of each of them singly, and all of them together in causing the decline in mortality rates.” Amid this uncertainty, an ecumenical solution appeared necessary. As NHLBI’s Levy concluded, “both primary prevention through lifestyle changes and improved treatment regimes have played a role in the decline.”
Frustrated by their initial inability to solve “the case of the disappearing epidemic,” the researchers sought more rigorous, quantitative analyses of the impact of specific interventions.\textsuperscript{11} The Framingham Heart Study had identified specific risk factors that correlated with CHD mortality. One team of researchers used Framingham data to argue that the observed 5 mg/dl drop in cholesterol levels since the 1960s would predict a 4.3% decline in CHD mortality.\textsuperscript{12} Epidemiologist Michael Stern used Framingham algorithms to integrate changes in several risk factors. He concluded that these changes accounted for 50% of the decline in men.\textsuperscript{9} Stern, however, admitted that it was “not possible at present to quantify definitively” the relative impact of lifestyle changes and improved medical care -- “both have played a role.”

Researchers realized that they needed to distinguish between two effects: the extent to which prevention campaigns reduced the incidence of CHD and the extent to which medical care reduced fatality rates.\textsuperscript{12} This required data on both incidence and mortality. Motivated by the Decline Conference, researchers undertook community-based surveillance projects. The World Health Organization orchestrated the largest, MONICA. Researchers at 39 centers in 26 countries collected data about risk factors, medical care, event rates, case-fatality rates, and mortality from over 100,000 people.\textsuperscript{13} These efforts did not
yield decisive answers. To obtain high resolution data, researchers had to focus on specific sites which might not be representative. They struggled to ensure consistent data collection and analysis. Their analyses of aggregate population data (e.g., correlating changes in average risk factor levels and event rates) impeded their ability to decipher causal relationships. Finally they realized that their basic assumptions -- risk factors determined event rates while health care influenced case fatality -- were too simplistic.

The Emergence of Models and Simulations

Other cardiologists and epidemiologists took up the challenge from the Decline Conference and went in a different direction. They adapted analytic techniques from systems engineering to produce more precise and integrated assessments of specific preventive and therapeutic interventions. In 1984 Lee Goldman (a cardiologist) and Francis Cook (an epidemiologist) published a model that recapitulated the passage of a CHD patient through the health care system, including emergency medical services, coronary care units, and surgical and medical treatment. Each domain was divided into specific interventions that could be quantified with data from observational studies and then reassembled using simple arithmetic to calculate the number of lives saved. For example,
to enumerate the value of coronary care units, the authors estimated that 500,000 patients were hospitalized for heart attacks each year, 4.5% of whom suffered ventricular fibrillation, 88% of whom were successfully resuscitated: $500,000 \times 0.045 \times 0.88 = 19,800$ lives saved annually. Similar calculations revealed the contributions of lifestyle interventions against dietary fat, cholesterol, smoking, obesity, and exercise.

The authors acknowledged the subjective assessments, approximations, and potential errors in their model. But they celebrated when the interventions they modeled -- four therapeutic, four preventive -- combined to account for 90% of the decline between 1968 and 1976. Lifestyle changes accounted for 54% of the total, a finding they confessed “may be as serendipitous as it is accurate.” Goldman and Cook’s model, intelligible to multiple audiences, coupled rigorous literature review with explicit assumptions and transparent calculations. It remains the most-cited reference in the decline literature.

To take this analysis further, Goldman teamed up with Milton Weinstein to form the CHD Policy Model Research Group. In 1987 they developed the first computer model to forecast CHD mortality. Their “state-transition” model simulated patient trajectories over time, considering the impact of primary prevention, the transition from health to CHD, and the impact of
treatment and secondary prevention. The model could be “run” to follow a simulated population as it aged and did or did not develop CHD. These simulations could be compared against historical data to determine how much of the actual observed decline had been captured by the model. With no adjustment for improving risk factors or treatments, the model over-estimated mortality between 1980 and 1990 by 34%. When it took these interventions into consideration, the model came within 2.8% of the actual data. The team concluded that a substantial portion of the decline must have come from these interventions.

Modeling has now become a popular tool in cardiovascular epidemiology, applied both to explain past declines and to predict future possibilities. A 2006 review found 75 articles that used 42 different models to inform CHD policy. But models, by their nature, are imperfect representations of reality. They make simplifying assumptions to facilitate methodical analysis. Goldman and Cook, for instance, cautioned that their model “must be considered approximate at best ... a perspective rather than a definitive explanation.” This is not a problem so long as readers understand a model’s limits.

The challenge for readers is two-fold. First, many different types of models exist. Some are static, calculating effects based on the prevalence and impact of specific interventions in a population. Others are simulations that
analyze a computer-generated cohort of “individuals” as they age over time. Second, the quality of models varies considerably. Are the assumptions explicit? Are the mechanisms transparent? Have sensitivity analyses been done (e.g., to test the effect of different assumptions)? Has the model been validated (e.g., tested against existing data sets)? The 2006 review found that few of the 42 models met these quality criteria.\textsuperscript{17}

Consider several prominent examples. Researchers working with the World Health Organization’s Disease Control Priorities Project developed a model of the global burden of disease. They compiled data on morbidity and mortality of over 130 diseases and calculated what share of this burden could be attributed to specific risk factors. Using assumptions about socioeconomic development and risk factor trends, they forecast the burden of disease in 2030. CHD, increased by tobacco use, hypertension, and inactivity, and decreased by alcohol use, will remain the leading cause of death worldwide.\textsuperscript{18} Such projections might help countries determine where to invest health resources.

A different model, IMPACT, focused on CHD decline. Developed by Scottish cardiologist and epidemiologist Simon Capewell, IMPACT, like Goldman’s initial analysis, quantified the utilization and impact of interventions to calculate the number of deaths prevented by each.\textsuperscript{19} Capewell demonstrated his model on data from Scotland and found that of the deaths
prevented between 1975 and 1994, 10% came from acute coronary care, 9% from treatment of hypertension, 8% from secondary risk factor management, 8% from management of heart failure, 2% from bypass surgery, 2% from aspirin, and 0.1% from angioplasty. Meanwhile, of the risk factors, smoking contributed 36%, cholesterol 6%, blood pressure 6%, and deprivation 3%. Taken together, “risk factor reductions and modern treatments contributed almost equally,” 40% treatment, 51% prevention.

Capewell shared IMPACT widely, developing a website for the model and collaborating with researchers in many countries to run analyses on New Zealand, England and Wales, Finland, Ireland, the United States, Sweden, Canada, Italy, Iceland, China, Spain, and Northern Ireland. Although results varied, they almost always shared responsibility with near equality: 42%-58% England and Wales, 47%-44% in the United States, and so forth.\textsuperscript{20,21} Only in Scandinavian countries, such as Finland, where aggressive public health campaigns reduced consumption of dairy fat consumption, did this balance shift (23%-72%).\textsuperscript{22} A microcosm of decline literature, IMPACT produced remarkably consistent results across time and place, half credit each to prevention and treatment.

IMPACT has also been adapted to predict the impact of interventions and "bring together public health professionals, clinicians and service commissioners in interactive scenario
planning activities to inform policy decisions.”

For instance, an additional 372,000 deaths could be prevented if Americans achieved “ideal risk-factor levels.” IMPACT can also detect disquieting trends. Improvements in cholesterol, blood pressure, and smoking in the United States have been offset by worsening obesity and diabetes. Decline has slowed and even plateaued for younger adults, “potential warning signs” that hard fought gains might soon be lost.

The most ambitious model, Archimedes, has been developed by David Eddy and Leonard Schlessinger at Kaiser Permanente. Archimedes attempts a “full-scale simulation model of human physiology, diseases, behaviors, interventions, and healthcare systems.” It offers researchers, administrators, and policy makers the chance to “run clinically realistic virtual trials on any population and create compelling evidence to make decisions.” For instance, Eddy’s team simulated the impact of 11 prevention activities over 30 years in a representative population of 20- to 80-year olds. They found that the interventions could prevent 63% of all heart attacks (or 36% using more realistic assumptions about treatment uptake). Aspirin (in high risk patients), diabetes prevention, and weight loss had the biggest impact. Only one intervention, smoking cessation, was cost saving. The lowest value came from cholesterol reduction in low risk populations, a finding with “important policy and clinical
implications, as it is currently one of the most heavily promoted of all the prevention activities."

Models and Their Discontents

The evolution of the decline debate has been animated by the prospect that historical modeling would inform health care policy by enumerating the relative contributions of risk factor management and medical interventions. Researchers’ ambitions developed in parallel, from explaining the past to predicting the future. But as researchers’ models have grown more powerful they have become less intelligible.

Archimedes, for instance, has been critiqued for being “extraordinarily opaque.” This is a problem for anyone who believes that models are only useful if their inner workings can be understood. Eddy, in response, has argued that transparency is a poor criterion for judging a model. What matters is not how a model works, but how well it works. And even the critics of Archimedes acknowledge that its “results are astounding.” Eddy and Schlessinger ran simulations of 74 randomized clinical trials of diabetes interventions and then compared their simulated results to the results obtained in the actual trials. They found a correlation coefficient of 0.99. It remains to be seen whether researchers and policy makers will be more swayed by transparency or accuracy.
The inner workings of IMPACT, in contrast, are accessible to motivated readers. Close analysis reveals several important features. First, its outcome and implications are malleable: IMPACT can be run with different endpoints that yield different assessments. Capewell developed IMPACT to analyze deaths prevented. Beginning in the 2000s, his team ran parallel analyses of “life-years gained.” Although treatment had had the upper hand in the deaths prevented analysis in the United States (47% to 44%), it lost out to risk factors on life-years gained (35% to 65%).\textsuperscript{21,30} In England the shift was even starker, from 42%-58% to 21%-79%.\textsuperscript{20,31} Why the difference? IMPACT calculated life-years gained by multiplying the number of deaths prevented by the median survival after the intervention. Since prevention targets younger and healthier patients, they have longer median survival post-intervention, something that accentuates the benefit of prevention.\textsuperscript{32} Researchers can conduct either analysis, aware of the diverging policy implications.

Second, these models apply a thin veneer of specificity atop a messy foundation. IMPACT requires researchers to reduce the efficacy of each treatment to a single coefficient.\textsuperscript{33} When conflicting data exist, researchers seek the most recent, least biased, and most representative estimates.\textsuperscript{21} Researchers also estimate compliance rates, with a range from 50% among asymptomatic outpatients to 100% in hospitalized patients. Both
sets of estimates are, themselves, uncertain.\textsuperscript{19} Estimating the contribution of risk factor reductions remains an even “less precise science.”\textsuperscript{33} In Finland, for instance, risk factors explained either 53\% or 71.8\% of the decline, depending on whether the researchers derived their coefficients from Finnish or international studies.\textsuperscript{22} Latitude in parameter estimates raises an important question. Since the 1970s, analyses of CHD decline have generally assigned equal credit to prevention and medical care. Does this mean that the analyses have reliably revealed a correct answer? It might also mean that researchers’ expectations have subtly influenced their methods and produced expedient results.

Third, the models can sweep uncertainty under the carpet. Because of the potential variation in parameter estimates, IMPACT analyses include an analysis of extremes, using “maximum and minimum feasible values” to produce a range of estimates of deaths prevented.\textsuperscript{19} For the United States, sensitivity analysis showed that the modeled parameters could explain anywhere from 51\% to 160\% of the decline, not just the 91\% advertised in the abstract.\textsuperscript{21} The model could explain nearly all of the decline, or half of the decline, or substantial decline that had not actually happened.

Fourth, ambiguity about what IMPACT does or does not leave unexplained focuses attention on another issue: IMPACT and the
other models can only analyze factors that have been quantified and measured. It is no accident that they focus on the usual suspects, including smoking, blood pressure, body mass index, and the impact of specific medical interventions. Less easily measured variables, such as stress or social context, get left out. Social epidemiologist Michael Marmot conceded this point in his own foray into the debate in 1984: "'Stress' is excluded from discussions of trends in mortality because of conceptual, definitional and measurement difficulties."  

Since social factors remain un-modeled, IMPACT researchers can attribute shortfalls to these "other, unmeasured risk factors." The magnitude of this "other" could be as high as 24% (e.g., in Finland). Life-years gained analyses erased "the other" altogether. When one epidemiologist pointed out that the analysis of the United States had ignored the important role played by reduced air pollution, Capewell offered a complex response. While air pollution and other risk factors might account for the 9% unexplained, it was also possible that "imprecision in the measurement and modeling of the major risk factors (cholesterol, smoking, and blood pressure) might also account for much of the gap." How likely was this? That depends on the model’s robustness. With the sensitivity analysis revealing that IMPACT accounted for anywhere between
51% and 160% of the actual decline, the “other” might be responsible for as much as 49%.

The role of non-traditional risk factors remains controversial. As cardiovascular epidemiologists developed their models, social epidemiologists sought different causal explanations. Marmot’s Whitehall Study demonstrated that CHD mortality correlated powerfully with occupational grade within the British civil service. Subsequent work linked mortality to relative position in any status hierarchy, whether of education, income, or control. The potential importance of social factors can be seen in the United States, where decline followed different trajectories in different parts of the country. Decline began in California, and then in other regions in the West and Northeast, before spreading from the coasts to the interior and from cities to rural areas. Speakers at the Decline Conference recognized that these disparities held clues to causes of the decline, including socioeconomic status and lifestyle.

Other researchers have downplayed the significance of social variables. A 2001 review concluded that 75% of all CHD deaths could be attributed to the three major risk factors -- cholesterol, blood pressure, and cigarettes. One team re-analyzed the Whitehall results and argued that Marmot’s mortality gradients were substantially explained by risk factor
gradients along the occupational hierarchy. Capewell’s team, aware of the potential role of diet, stress, or poverty, has modeled them when adequate data exist. They found, for instance, that decreased deprivation (i.e., economic development) accounted for 3.4% of the CHD decline in England and Wales. Such modeling remains a work in progress. In the meantime, researchers must balance the appeal of the quantitative models against awareness of the potentially important factors that they exclude.

One last point deserves mention. Researchers have validated their models by testing how well they match the observed historical changes in risk factors, health care, and outcomes. However, the ability of a model to explain the past is not a perfect marker of its ability to predict the future. This will be especially true if the models are applied to understand the emerging epidemic of CHD in developing countries.

Conclusions

Protagonists in the decline debate have long sought answers that would guide policy choices between prevention campaigns (e.g., education targeting populations to bring about lifestyle change and reduce risk factors) and medical care (e.g., medical technology provided to individual patients). Sometimes researchers have been ecumenical. The final report of the
Decline Conference assumed that changes in risk factors and improved medical care had both contributed. Sometimes they have been oppositional. Ford and Capewell subtitled their 2011 review “Public Health Versus Clinical Care.” The debate, as a result, has perpetuated long-standing tensions between medicine and public health, even as it offered an olive branch by crediting both with substantial contributions to past decline and by offering each a substantial role in future policy.

Lost in the debate is recognition of how much the categories of medical care and prevention have changed. Prevention once meant ensuring the healthiness of lived environments: clean air, clean water, and clean food. In the closing decades of the twentieth century, prevention has transmogrified and been integrated into biomedical regimes of surveillance and control. Future health is increasingly ensured through compliance with pharmaceutical regimens, whether for diabetes, hypertension, or high cholesterol. The persistent debate between treatment and prevention polices a boundary which becomes less meaningful each year. The role of social forces in producing and ameliorating disease, in contrast, has largely been left without voice in the debates.

Is it likely, in the end, that the decline debate and its models will provide useful guidance to policy makers? When the debate began, no one doubted whether particular preventive or
therapeutic interventions had potential value: nearly all of them do. The question was whether researchers could demonstrate a large enough differential value to justify difficult decisions about resource investment. This has not happened. Instead of favoring prevention or treatment, researchers’ findings have motivated calls for increased investment in both. Better utilization of evidence-based therapies could save even more lives. More aggressive campaigns against risk factors could prevent even more deaths. The United States needed a “comprehensive strategy,” as did Finland and England and Wales.

Given the powerful interests at stake, it is no surprise that expedient results -- half credit each to medicine and public health -- emerged time and time again. As Lewis Carroll’s Dodo bird observed, “everyone has won, and all must have prizes.” Such conclusions provide little guidance to policy makers. But even if a definitive answer did emerge for or against a particular strategy, should health policy necessarily follow suit? The Archimedes analysis of CHD, for instance, did make a strong critique of statins and cholesterol lowering medications. Confronted with such findings, policy makers face a difficult challenge. First, they must decide if they can trust the result. Models, like randomized clinical trials, need to be read critically. Whether policy makers
scrutinize a model’s inner workings (e.g., IMPACT) or check its validation studies (e.g., Archimedes), they must make the effort to understand the quality of the result. If the model passes muster, then policymakers must weigh it against other factors that inevitably influence decisions. When the IMPACT team examined this process in 2011, they were dismayed by what they found. Policymakers felt that existing research was too uncertain, had poor local applicability, paid too little attention to social determinants, and was poorly communicated. They often gave more weight to their own intuitions, expert consensus, public opinion, stakeholder pressure, financial sustainability, and political viability.

Researchers and policymakers face a delicate situation. As the Patient Centered Outcomes Research Institute takes shape amid increasing pressure to improve the efficiency and quality of health services, stakeholders will demand that research findings actually guide health policy. But it is unlikely that any single model or research study will produce findings that are clear and reliable enough to justify transformative policy. Capewell’s team has emphasized that researchers and policymakers must both be aware of how “the concept of evidence is negotiated and socially constructed by and between individuals.” Policymakers must work to understand what kinds of knowledge are made and obscured in researchers’ analyses, they must learn what
lessons can be drawn from a particular study despite its limitations, and they must be consciously aware of what else they consider (and the many limits of these considerations) when they formulate policy. Just as they must open the black box of medical research and modeling, they must open the black box of their own decision making processes. This will not guarantee perfect policy, but it will at least make clear how and why the policy was made.

Acknowledgments:

The authors gratefully acknowledge feedback from many people, including Jerry Avorn, Henry Blackburn, Simon Capewell, Niteesh Choudhry, Lee Goldman, Joel Howell, Zubair Kabir, Aaron Kesselheim, Todd Olszewski, Gerry Oppenheimer, Will Shrank, Sarah Tracy, and six anonymous reviewers. The research was supported by a RWJF Investigator Award in Health Policy Research from the Robert Wood Johnson Foundation®, Princeton, New Jersey.
REFERENCES


24. Ford ES, Capewell S. Proportion of the decline in cardiovascular mortality disease due to prevention versus


38. Magnus P, Beaglehole R. The real contribution of the major risk factors to the coronary epidemics: time to end the "only-50%" myth. Archives of Internal Medicine 2001;161:2657-60.

