Assessing and Forecasting Population Health: Integrating Knowledge and Beliefs in a Comprehensive Framework

JEROEN VAN MEIJGAARD, MA^a
JONATHAN E. FIELDING, MD,
MPH, MBA, MA^a
GERALD F. KOMINSKI, PhD^a

SYNOPSIS

A comprehensive population health-forecasting model has the potential to interject new and valuable information about the future health status of the population based on current conditions, socioeconomic and demographic trends, and potential changes in policies and programs. Our Health Forecasting Model uses a continuous-time microsimulation framework to simulate individuals' lifetime histories by using birth, risk exposures, disease incidence, and death rates to mark changes in the state of the individual. The model generates a reference forecast of future health in California, including details on physical activity, obesity, coronary heart disease, all-cause mortality, and medical expenditures. We use the model to answer specific research questions, inform debate on important policy issues in public health, support community advocacy, and provide analysis on the long-term impact of proposed changes in policies and programs, thus informing stakeholders at all levels and supporting decisions that can improve the health of populations.

^aUniversity of California Los Angeles School of Public Health, Department of Health Services, Los Angeles, CA

Address correspondence to: Jeroen van Meijgaard, MA, University of California School of Public Health, Department of Health Services, CHS 61-253, 650 Charles E. Young Dr. South, Los Angeles, CA 90024; tel. 310-206-6236; fax 310-206-7096; e-mail <jeroenvm@ucla.edu>.

©2009 Association of Schools of Public Health

As a society, we devote considerable resources to forecasting various phenomena—economic conditions, weather and climate, agricultural yields, and effects of various technologies. All such forecasting efforts are beset by inherent complexity, technical difficulty, and reliance on multiple assumptions. In the health-care sector, forecasting has been used to project the impact of approved and potential federal policies on health care, such as their effect on insurance coverage, federal Medicare and Medicaid utilization, costs, as well as the number, type, and distribution of health-care providers.¹ With few notable exceptions, such as the development of the Future Elderly Model (FEM),² relatively little effort has been invested in using forecasting to address questions about the health status of the population—the ultimate goal of health care, disease prevention, and health promotion. How much longer will they live? What will be the burden of morbidity? Will disparities in the distribution of morbidity and mortality widen or narrow?

The development of a population health-forecasting model has the potential to interject new and valuable information about the future health status of geopolitically or otherwise defined populations, based on current conditions, socioeconomic and demographic trends, and potential changes in policies and programs. More comprehensive modeling is now possible due to dramatic improvements in forecasting methodologies, including new techniques to assess and reduce forecasting error, support for sensitivity analyses, and better estimates of uncertainty enabled by improved computing power.³ A health-forecasting model can be a critical analytic and translation tool to infuse timely and relevant information into policy debates at all levels of government, as well as in the private sector.

Forecasting is not entirely new to public health. Current and previous efforts to forecast disease- and risk-specific aspects of human health aim to better understand the impact of interventions and provide actionable information to practitioners. Each of these efforts has tended to be limited, focusing on a specific aspect of health. For example, infectious disease modeling has examined disease transmission to project disease incidence and prevalence based on patterns of contact, mode of transmission, incubation period, and vectors;⁴ the results complement disease surveillance to support epidemic preparedness.⁵ Forecasting has also been used to examine trajectories of functional capacity in populations, using demographic trends and information on disability status and patterns in chronic conditions.^{6,7} Among forecasting models analyzing the effects of changes in behavioral risk factors and environmental conditions, tobacco use is by far the most

extensively modeled behavioral risk factor, but physical activity, nutrition, and obesity are increasingly being modeled due to their linkage to diabetes and cardiovascular diseases.⁸ Some models include multiple risk factors and the estimated associated burden of disease for regional populations,⁹ while others focus on a narrow set of risk factors and trace disease trajectories for alternative intervention scenarios.^{10–17}

Population health forecasting requires rich data, an understanding of the determinants of health and their interactions, and technically innovative modeling. The evidence base for such modeling is supported by systematic reviews of the environmental and policy determinants of health and meta-analyses of specific health risk factors and related interventions.^{18–20} Both give a sense of the extent of missed opportunities for health improvement and the high cost of not undertaking policies and programs of proven effectiveness. However, there continues to be a translation gap between research and practice.^{21,22} Policy researchers have identified many reasons for this gap, including a lack of understanding or awareness, uncertainty about the relevance of research for particular situations, and the lack of confidence in information sources.^{23,24}

Most models that evaluate multiple health determinants and related outcomes are based on a microsimulation framework that allows modeling of individual units, usually individuals. Compared with approaches based on aggregate trends, microsimulation models are particularly suitable to evaluating different interventions and policy scenarios, by allowing the incorporation of data from disparate sources and inclusion of distributional information on variables of interest. Two major types are widely used. Static models use cross-sectional databases that provide a snapshot of the population at a point in time. In contrast, dynamic models build longitudinal databases of individual histories and allow behaviors and exposures to change over the time modeled.^{25,26}

The Coronary Heart Disease Policy Model was one of the earliest microsimulation applications to evaluate policy and behavioral changes and their impact on population health. This model assesses the impact of policy and technological advances on the incidence, prevalence, and mortality from coronary heart disease, and related changes in health-care costs.^{15,16} Several other models have been developed to estimate the impact of policy changes on smoking patterns and outcomes.¹⁴ In Canada, a dynamic, continuous-time Population Health Model (POHEM) has been developed to assess the impact of different policy interventions and technologies on the health of the Canadian population.²⁷

Using continuous-time modeling, such as used in Canada's POHEM, simplifies the modeling of multiple processes with many events that in a discrete-time model would result in an explosion of the number of possible state transitions. Continuous-time modeling reduces the complexity of modeling covarying behaviors, comorbidity, and competing risks, and it can incorporate joint distributions of variables that are determinants of health.

These existing models still have many limitations. Static models are limited in their application to different populations and time horizons. Other models have been limited in the number of variables, the scope of the model, or the exclusion of demographic and socioeconomic trends, providing only a partial picture of future health gains and changes in health-care costs. To obtain a more comprehensive understanding of the impact on health and associated outcomes, a more comprehensive model is required in which comorbid states and covarying behaviors are explicitly modeled, and unrelated interventions can be compared against a standardized baseline.

Probably the most concerted effort to forecast the health status of people in the United States is the FEM.^{2,28} This model is used to forecast the consequences of health trends and medical innovations for the Medicare population. It combines information on trends in health conditions, functional state, and risk factors (e.g., weight and smoking) with information on the availability of new medical technologies that may impact these trends, and the likely medical expenditures associated with the observed health conditions. This model exemplifies the value of building a more comprehensive model, as it can be used to more realistically anticipate future health-care expenditures.

Our approach is similar to the FEM in using simulation methods to incorporate information from different data sources and allowing for the incorporation of comorbid states and covarying health variables. We've expanded on this by (1) extending the age range that is modeled to birth to provide a full life-course model, (2) incorporating additional aspects of the dynamics of population demographics (e.g., changes due to migration), and (3) incorporating time-varying health risk factors (e.g., physical activity and obesity) to better evaluate the impact of public health programs and policies.

In this article, we focus primarily on the development of the model, its capabilities, and its application in public health practice. Although an outline of the technical specifications is provided, a detailed technical working document and updates to the model are available at www.health-forecasting.org.

METHODS

To organize data in a way that reflects the best scientific understanding of underlying causal processes,⁶ we derived our conceptual model from the Evans-Stoddart model of the determinants of health, including both physical and social environmental determinants, and allowing incorporation of government policies and systems.^{29,30} We also incorporated recent research on differential effects in high-risk, underserved, and historically disadvantaged populations³¹ into this conceptual model.

Model outline

To produce sound, valuable estimates, a population health-forecasting model needs to address several key considerations, each of which expands the data requirements for the model. For one, long-term forecasts are needed to model associations between risk factors and final outcomes because of the long induction periods for many chronic diseases. Secondly, multivariate techniques are required to incorporate the complexity of the factors that are known to influence health outcomes. Third, health impacts need to be estimated for the entire population and subpopulations, identifying the underlying cause of changes in outcomes and highlighting distributional effects within subgroups according to race/ethnicity, gender, age, and location.

Using a continuous-time microsimulation framework that simulates individuals' lifetime histories allowed the model to accommodate these methodological considerations. Statistics Canada A used a similar approach in its LifePath and POHEM models.^{21,32} The model consists of three core modules: (1) the core population module, (2) the risk factor/disease module, and (3) the forecasting module, each with its own set of variables, embedded assumptions, and techniques for data incorporation (Figure 1).

The core population module provides estimates of demographic and socioeconomic variables, as well as population movements in and out of the state and within the state. It includes demographic information such as natality, mortality, and migration data, and socioeconomic information such as education and marital status. Each variable is stratified by gender and race/ethnicity, and is modeled explicitly as joint distributions, facilitating their use in developing risk factor estimates by sociodemographic characteristic. To create accurate lifetime histories for current cohorts, we used historical data starting with the early 1900s and interpolated data where gaps existed.

The risk factor/disease module estimates exposures based on environmental conditions and health

Figure 1. Building blocks for a comprehensive health-forecasting framework

behaviors across the population and examines the impact of changes in risk factors, including changes resulting from interventions. Distributions of risk factors across the population (e.g., health behaviors) are defined, as are the relative risks for different health outcomes. These relative risks are used to simulate disease outcomes, conditional on the risk factor and disease profiles of the simulated individuals. Multiple disease processes can be modeled this way, including comorbid and competing outcomes.

We obtained distributions of risk factor variables for the risk factor/disease module from local surveys and augmented them with nationwide surveys to obtain estimates for longitudinal patterns. We derived parameter estimates on the relationships between variables from analyses presented in the literature. For instance, we used the results of research evaluating the dose-response relationship between physical activity and coronary heart incidence.^{33–35} We parameterized the relative risk of physical activity (PA) on coronary heart disease as $(1/[1 + \alpha \text{ PA}])$ and estimated α using the reported relative risk estimates. Where no relevant analyses were available from the literature, we estimated relative risks directly from publicly available datasets. Many relative risk estimates were for populations in different geographic locales, implicitly assuming that the relative risks were similar among populations. Moreover, unless otherwise shown in the literature, it was assumed that relative risks of socioeconomic and behavior variables for health outcomes were constant across race/ethnicity, while allowing for differences

across gender. Because risk factor levels and baseline hazards were allowed to vary by race/ethnicity, disparities in outcomes across racial/ethnic groups were preserved.

The forecasting module enables users to develop forecasts of future outcomes with or without proposed interventions, while in the absence of the forecasting module, it is only possible to predict current outcomes under counterfactual conditions. Each of the variables input into the previously developed modules was projected forward in time using a variety of forecasting techniques, including econometric models, time-series models, expert judgments, and the Delphi method. For variables with large expected changes, we preferred judgmental time-series forecasting.³⁶

The California Health Forecasting Model

In the current version of the model, we simulated births inside and outside California, migration to and from California, all-cause mortality and coronary heart disease (CHD)-specific mortality, the duration and intensity of leisure-time physical activity, and obesity (i.e., body mass index [BMI] in kilograms per meter squared). We also simulated direct personal medical expenditures based on those collected in the Medical Expenditures Panel Survey. As with the demographic and socioeconomic variables, each of these rates is gender-, age-, and race/ethnicity-specific. Obesity was also conditional on physical activity, and mortality and expenditures were conditional on physical activity and obesity.

The simulation used events—such as birth, death, marriage, or disease incidence—to mark changes in the state of the individual, which are described by state variables. At the time of an event, all the state variables (e.g., health status, exercise level, and BMI) as well as the event probabilities are updated. All state variables are updated at least once each modeled year, on the individual's birthday, but often more frequently. Mean levels of leisure-time physical activity and BMI differed by gender, age, and race/ethnicity.^{37–39} We also incorporated secular trends; thus, all levels and rates were conditional on calendar time. Where evidence from the literature was available, we modeled the modifying effects of gender, age, or race/ethnicity on relative risks. For example, we modeled the relative risk of obesity on mortality as a function of age.⁴⁰

From birth onward, life and health events occur in a competing hazard framework with multiple possible outcomes. Each event probability is conditional on the health status of the individual, including behaviors and disease characteristics. Assumptions are necessary due to incomplete information on joint distributions and longitudinal relationships of the predictor variables, the mismatch of data elements when matching different data sources, and the lack of information from the literature regarding some causal relationships. The model is calibrated to match population and all-cause mortality estimates by adjusting the baseline hazards of the event processes. Although calibration is also a valuable tool for understanding feedback within the model, it needs to be used very carefully to ensure that real effects are not eliminated from the model. We used a mixture of quantitative and qualitative forecasting techniques to develop projections for a limited number of important external variables, such as births, migration, exercise and overweight status, and baseline CHD incidence and case fatality. The process was facilitated through the collection of trend data and projections of the demographic characteristics of the population by a variety of state and federal agencies.

Figure 2 lists the data sources by module for each type of information required in the California version of the population health-forecasting model. Additional details of data sources and model implementation are described in a technical working document (www.health-forecasting.org/reports.aspx). The microsimulation is programmed in the .NET version of Microsoft® Visual C++,⁴¹ using Modgen.⁴² We performed statistical analyses using SAS® version 9.1.⁴³

RESULTS

The first step in using the model was to populate it with the most up-to-date information available, and

calibrate the model to ensure it provided results that were consistent with observed data. Next, we used the model to create a reference forecast. We compared alternative scenarios with this reference forecast. Once the model was populated and calibrated, we used it to analyze different scenarios and provided these scenarios to various stakeholders who were considering, advocating, or deciding on interventions to promote improvements in population health.

Population health forecast for California

The model combines and parses a large amount of information from disparate sources into a microsimulation framework to reflect the population in California up to several decades into the future. The model has been revised and expanded periodically to reflect new data or to incorporate additional risk factors and health outcomes.

The reference forecast of future health in California is the case in which no targeted action is taken to improve population health beyond existing policies and practices. The model is calibrated by making small adjustments to the baseline mortality hazards to yield all-cause mortality and population estimates that are consistent with estimates of the California Department of Finance. Additionally, we adjusted the CHD incidence in California to yield observed prevalence rates. Calibration ensures consistency with other estimates, while providing additional information on behaviors, disease incidence and prevalence, and disease-specific mortality rates. For example, age-adjusted mortality rates are projected to decline in the future, reflecting expected changes in disease patterns such as CHD incidence and prevalence. These trends also reflect changes in the age and racial/ethnic composition of the population.

Initial estimates of future CHD mortality were in part assessed by trending CHD incidence and case fatality rates from 2002 onward and comparing the model results to observed CHD fatality cases in 2003–2006. This extrapolation yielded an overestimate of CHD fatality by as much as 15% in 2006 due to an unanticipated acceleration in the decline in new CHD cases and case fatality rates. Therefore, we made additional adjustments to the model to accommodate observed changes in the rate of change of CHD outcomes.

We will continue to validate and calibrate the model as more data become available, and compare results with those produced by other models when they become available. Results of our reference model, using either the “best assumption” based on prior data and expert opinion or, alternatively, a variety of assumptions about future changes in key variables, provide insights into the likely changes in CHD incidence, prevalence,

Figure 2. Data sources summary for California health-forecasting model

<i>Model component</i>	<i>Source(s)</i>
<i>Core population module</i>	
Historic and current population counts for California and counties, stratified by age, gender, and race/ethnicity	California Department of Finance—Demographic Research Unit. E-4 Revised Historical City, County and State Population Estimates, 1991–2000, with 1990 and 2000 Census Counts. Sacramento (CA): California Department of Finance; 2003. U.S. Census Bureau. Decennial Census Estimates
Historic and current immigration rates for California and counties, by country of origin	California Department of Finance—Demographic Research Unit
Historic and current mortality and fertility rates for California and counties, stratified by age, gender, and race/ethnicity	California Department of Finance—Demographic Research Unit U.S. Census Bureau. Decennial Census Estimates
Historic mortality for CHD for California, stratified by age, gender, and race/ethnicity	California Department of Health Services—Center for Health Statistics American Heart Association
Historic levels of educational attainment for each county, stratified by age, gender, and race/ethnicity	California Department of Education U.S. Census Bureau. Decennial Census Estimates
<i>Risk factor/disease modules</i>	
Historic levels of physical activity and obesity for Californians and counties, stratified by age, gender, and race/ethnicity	Behavioral Risk Factor Surveillance System National Health Interview Survey (NHIS) National Health and Nutrition Examination Survey (NHANES) California Health Interview Survey
Relation between weight and physical activity levels and all-cause mortality, and CHD incidence and CHD mortality, stratified by age and gender	Studies in the published literature ^{a,b,c} Alameda County Human Population Laboratory NHIS Health Promotion and Disease Prevention supplement 1990–1991 with linked mortality files
Persistence of individual physical activity levels over time, stratified by age, gender, and race/ethnicity	Studies in the published literature NHANES Epidemiologic Follow-up Study
Relation between weight and physical activity levels in childhood and adolescence and weight and physical activity levels for adults, stratified by age, gender, and race/ethnicity	Studies in the published literature
<i>Forecasting module</i>	
Projected population count for each county, stratified by age, gender, and race/ethnicity	California Department of Finance—Demographic Research Unit. P3 Population Projections by Race/Ethnicity, Gender and Age for California and Its Counties 2000–2050. Sacramento (CA): California Department of Finance; 2007.
Projected mortality and fertility rates for each county, stratified by age, gender, and race/ethnicity	California Department of Finance—Demographic Research Unit U.S. Census Bureau
Projected immigration rate for each county, by country of origin	California Department of Finance—Demographic Research Unit
Projected levels of weight and physical activity for Californians, stratified by age, gender, and race/ethnicity	Extrapolations from historical trends from population surveys
Projected mortality rates for CHD for each Californian, stratified by age, gender, and race/ethnicity	Extrapolations from historical trends using vital statistics data

NOTE: A more detailed description of the consulted literature and methodologies used can be found in the technical working document available at <http://www.health-forecasting.org/reports.aspx>

^aLee IM, Sesso HD, Oguma Y, Paffenbarger RS Jr. Relative intensity of physical activity and risk of coronary heart disease. *Circulation* 2003;107:1110–6.

^bFolsom AR, Arnett DK, Hutchinson RG, Liao F, Clegg LX, Cooper LS. Physical activity and incidence of coronary heart disease in middle-aged women and men. *Med Sci Sports Exercise* 1997;29:901–9.

^cSesso HD, Paffenbarger RS Jr, Lee IM. Physical activity and coronary heart disease in men: the Harvard Alumni Health Study. *Circulation* 2000;102:975–80.

CHD = coronary heart disease

and case fatality and how these changes impact different groups (e.g., by race/ethnicity) (Figure 3). Thus, using the reference case, stakeholders can anticipate shifts in demands for services by race/ethnicity using a quantitative framework that is internally consistent with other data sources upon which government officials rely.

The reference forecast serves as a comparison case for evaluating the impact of different interventions. One of the main advantages of comparing interventions within this comprehensive framework is that interventions that target different behaviors or have different time horizons can be compared using the

same reference. Thus, it eliminates the need to parse discrepancies in intervention impact that are due to differences in the reference case, and enables decision makers to compare interventions more directly.

Application of the model

We are using the model to (1) evaluate research questions about the association among sets of variables that cannot be observed directly through surveys—a common use of simulation models, (2) inform debate on important policy issues in public health through issue briefs, (3) support community advocacy to strengthen local communities and efforts to improve population

Figure 3. Estimated coronary heart disease prevalence and fatality by race/ethnicity, California, 1980–2020

CHD = coronary heart disease

health, and (4) provide analysis on the long-term impact of proposed policies and programs.

For the first type of application of the model, we have examined the effects of physical activity and BMI on mortality rates and medical expenditures. For example, we have modeled estimates of total direct personal medical expenditures for the remainder of an individual's lifetime contingent on current BMI and age. The results indicate that while lifetime medical expenditures generally increase with BMI, these expenditures may decrease at very high levels of BMI due to increases in mortality rates (Figure 4), a finding that is consistent with recent simulation results from the Netherlands.⁴⁴ Our expenditure estimates are also qualitatively similar to findings from a study of the Medicare population during a 20-year period, where expenditures were estimated conditional on BMI levels prior to age 65.⁴⁵

Second, we are contributing to policy debates in California by providing information on demographic and health trends and how these trends impact future population health and demand for services. Simply providing this information (including sensitivity analyses) unrelated to possible interventions supports the larger discussion on important policy issues. For example, in a recent issue brief,⁴⁶ we outlined what an extension of current trends in overweight and obesity would mean for the future health of the population and medical

expenditures in California. We showed that if current upward trends in BMI continue, as much as 73% of the California population will be overweight or obese by 2025 (Figure 5). Unlike simple extrapolation of current aggregate trends, this simulation model is able to show how the distribution of overweight and obesity changes over time. This and other briefs are targeted at decision-making and influencing individuals and organizations, such as state legislators, state government decision makers, county elected and appointed leaders, and county departments of public health and health services.

Third, the model is being employed as a tool for community organizations and local agencies, making evidence-based research accessible and relating this information to local populations. Our Web-based interface (www.health-forecasting.org) provides intuitive access to the results from the model and is aimed at supporting public health agencies, communities, and others to advocate for new policies and implement programs to enhance population health. Small agencies and other community organizations often do not have the resources to analyze the evidence and estimate the impact of various programs on local populations. The forecasting model can be expanded to answer these specific questions. For example, we are currently working with one county to chart the impact of air pollution on population health, and support the efforts of

Figure 4. Estimated future lifetime medical expenditures conditional on age and overweight/obesity by gender, California, 2005^a

^aOverweight is defined as having a body mass index of 25–29.9 kilograms/meter squared (kg/m²). Obese is defined as having a body mass index ≥ 30 kg/m².

Figure 5. Recent trends and projections for overweight and obesity prevalence in California adults aged 18 and older^a

^aOverweight is defined as having a body mass index of 25–29.9 kilograms/meter squared (kg/m²). Obese is defined as having a body mass index ≥ 30 kg/m².

local community groups to quantify the harm of local pollution sources and better advocate for prevention policies and programs.

Fourth, the model is used to define and explore alternative scenarios of interest to policy makers and other stakeholders. The estimated health outcomes are compared with the reference case, thus providing the long-term benefits and costs of potential new or revised policies and programs. The long time horizon incorporated into the model is especially beneficial because the full effects of many interventions are often not manifested for many years after implementation.

By incorporating standard assumptions of the impact of these policies on health behaviors, this flexible simulation model allows for a quick turnaround to answer specific policy questions. For example, at the request of an advocacy organization, the model was used to evaluate the long-term impact of a proposed menu labeling policy on projected obesity trends in California. We continue to work with advocacy organizations and the California Department of Public Health to support the evaluation of state and county policies that enhance population health and expand the evidence base for such policies.

DISCUSSION

We set out to address questions about the health status of the population, how this status changes over time, and how future populations are impacted by new public health policies and programs. This is valuable information in the hands of public health practitioners and elected and appointed state and local senior leaders. With the aim of offering sound, valuable information that can help bridge the gap between research and policy, health-forecasting models provide estimates of potential health benefits and harms under different scenarios at different points in time. Alternative scenarios for analyses may include a range of potential policy or programmatic interventions. To fully realize their utility, the results of these analyses need to be communicated in easy-to-use and relevant formats. Recent advances in public health research and computing power, coupled with an expanded evidence base from systematic reviews, have reduced many of the technical barriers to population health forecasting.

Our forecasting model combines large amounts of information on demographics, risk factors, and trends of the California population, and then applies evidence-based interventions to forecast future health and health-related outcomes. These simulations allow users to explore and better understand the likely health effects of alternative approaches to reducing the disease and injury burden. These effects can be examined in aggregate for all Californians, as well as for subpopulations defined by geography, age, gender, race/ethnicity, and risk factors.

This microsimulation model has substantial advantages over other models. It supports the evaluation of multiple health conditions associated with multiple risk factors simultaneously, taking into account comorbid states and the effects that a single risk factor (e.g., smoking or physical activity) may have on many different outcomes. It also allows for examination of cascading effects of changes in disease states and their impact on more distal outcomes. By simulating the lifetime histories of individuals, it can analyze the impact of events longitudinally, including long-term effects of exposures. Additionally, it can provide cross-sectional analyses at different points in time, generating distributional information on outcomes by taking into account joint distributions of several risk factors without assuming independence.

Due to the large number of variables, extrapolations, and incomplete evidence on the causal linkages between risk factors and outcomes, any comprehensive forecasting model requires a number of assumptions and some element of expert judgment. Because these assumptions can influence results, we make them

explicit. This also facilitates sensitivity analyses, which are crucial for model validation and developing an understanding of the dynamics of this complex model. Furthermore, by making assumptions that are transparent to users, this approach facilitates comparison of results under alternative sets of assumptions and between models.

Inherent uncertainties and the need for assumptions in any comprehensive model make calibration an integral developmental component. Forecasting models can be validated in several ways, including “backcasting,” comparison of predictions with near-term realization of outcomes, and convergence with other models. We will continue to use and expand on these methods to validate our forecasting model in future iterations. Backcasting predicts present events based on previous conditions, and then compares the results of these predictions with realizations of the outcome variables. Robust, meaningful backcasting is complicated by assumptions that are embedded in the model and cannot easily be excluded, such as the choice of specific statistical processes that are used to model longitudinal patterns of physical activity and obesity and that are identified using recent data.

The second approach to model validation, near-term forecasting, may be unable to detect errors that manifest only over long time horizons. The third approach, comparison with other models, is desirable but limited, as few comparable models that cover the entire lifespan currently exist in the U.S. It is, however, possible to perform partial comparisons for those select conditions that are analyzed in other models, such as comparing the elderly results with the FEM,² or the heart disease results with those from the CHD Policy Model.¹⁶ Direct comparison between the California model and FEM are complicated by differences in variables and populations that are modeled. Nevertheless, the forecasts are similar in that there is a continuing increase in expected medical expenditures. Both models suggest that improved health status in the Medicare population may lower per capita expenditures, but does not necessarily lower total Medicare expenditures because of higher life expectancy. Yet, we find that the total savings in medical expenditures resulting from reducing obesity prevalence can be substantial in the California population, when the population younger than 65 years of age is included.⁴⁶

Limitations

Our approach had some limitations in scope and use, either by choice of modeling approach, lack of data, or computational requirements. In particular, because we do not allow for individual-to-individual interactions in

the model, it is not possible to model infectious diseases based on transmission patterns. In North America, infectious diseases account for less than 10% of the disease burden.⁴⁷ Furthermore, sophisticated models already exist to address the spread of and public health response to infectious diseases.^{4,5}

The simulation framework allows for complex joint distributions of underlying risk factors of disease, time-varying risk factors, comorbid conditions, and various disease progressions. However, information on some of these joint distributions is limited. In many instances, we do not have access to multiple variables in one dataset to generate joint distributions, longitudinal data to simulate life-course information, or relative risks in the presence of other conditions. In those instances, we make simplifying assumptions, which as much as possible are made explicit in the model interface or technical documentation.

Similarly, we developed the model around large geographic areas and populations. The scope of geography and population can be reduced, however, only insofar as there is information available on the levels of the risk factors in these populations.

Lastly, even with advances in computer power, simulating multiple complex disease processes with many underlying risk factors can require many hours of calculation on a powerful desktop. Thus, for computational purposes, it is necessary to limit details, such as by setting an individual's weekly activity level only once a year rather than letting it change every week. As with other forecasting models, keeping this forecasting model current will require periodic updates to incorporate new data on population statistics and emerging research on relations between and among risk factors and health outcomes.

CONCLUSIONS

The development of this model is only the first step in a broader effort to bridge the gap between research and practice. An intuitive interface and a broad communication, dissemination, and training program on the applicability of this model to different circumstances are necessary to support full use of this forecasting model by public health agencies, community organizations, and other stakeholders.^{22,24} The model in its current form can be most easily used by entities for which specific recent data are available (e.g., states, counties, and large cities). For smaller units such as smaller municipalities, communities, and neighborhoods, we developed a Web-based tool that interpolates results to those levels based on their key sociodemographic characteristics, using small area

estimation.⁴⁸ This tool also allows users to view and graph the model results and translate forecasts to local communities by inserting local data, thereby further supporting broad utilization.

A major advantage of a forecast that extends several decades into the future is that it provides a strong rationale for current action, even though the full benefits of policy change may not be felt for a number of years. Thus, the health-forecasting model supplies a multigenerational approach as a counterweight to short-term thinking that often dominates politics. It supplies policy makers with more relevant information to support many specific interventions, policies, and programs that can improve the health of the populations that they represent.

This research was supported by a grant from The California Endowment.

REFERENCES

1. Martini EM, Garrett N, Lindquist T, Isham GJ. The boomers are coming: a total cost of care model of the impact of population aging on health care costs in the United States by major practice category. *Health Serv Res* 2007;42(1 Pt 1):201-18.
2. Goldman DP, Shang B, Bhattacharya J, Garber AM, Hurd M, Joyce GF, et al. Consequences of health trends and medical innovation for the future elderly. *Health Aff (Millwood)* 2005;W5:R5-17.
3. Armstrong JS. Principles of forecasting: a handbook for researchers and practitioners. Norwell (MA): Kluwer Academic Publishers; 2001.
4. Koopman JS, Chick SE, Riolo CP, Simon CP, Jacquez G. Stochastic effects of disseminating versus local infection transmission. *Math Biosci* 2002;180:49-71.
5. Myers MF, Rogers DJ, Cox J, Flahault A, Hay SI. Forecasting disease risk for increased epidemic preparedness in public health. *Adv Parasitol* 2000;47:309-30.
6. Manton KG, Stallard E, Singer B. Projecting the future size and health status of the US elderly population. *Int J Forecast* 1992; 8:433-58.
7. Manton KG, Land KC. Multidimensional disability/mortality trajectories at ages 65 and over. The impact of state dependence. *Soc Indicators Res* 2000;51:193-221.
8. Bibbins-Domingo K, Coxson P, Pletcher MJ, Lightwood J, Goldman L. Adolescent overweight and future adult coronary heart disease. *N Engl J Med* 2007;357:2371-9.
9. World Health Organization. World health report 2002: reducing risk, promoting healthy life. Geneva: WHO; 2002.
10. Gunning-Schepers LJ. Models: instruments for evidence based policy. *J Epidemiol Community Health* 1999;53:263.
11. Warner KE, Smith RJ, Smith DG, Fries BE. Health and economic implications of a work-site smoking-cessation program: a simulation analysis. *J Occup Environ Med* 1996;38:981-92.
12. Tsevat J. Impact and cost-effectiveness of smoking interventions. *Am J Med* 1992;93(Suppl):S43-7.
13. Gavin NI, Wiesen C, Layton C. Review and meta-analysis of the evidence of the impact of smoking on perinatal conditions built into SAMMEC II. Atlanta: Centers for Disease Control and Prevention (US); 2001.
14. Tengs TO, Osgood ND, Lin TH. Public health impact of changes in smoking behavior: results from the Tobacco Policy Model. *Med Care* 2001;39:1131-41.
15. Weinstein MC, Coxson PG, Williams LW, Pass TM, Stason WB, Goldman L. Forecasting coronary heart disease incidence, mortality, and cost: the Coronary Heart Disease Policy Model. *Am J Public Health* 1987;77:1417-26.
16. Goldman L, Phillips KA, Coxson P, Goldman PA, Williams L, Hunink

- MG, et al. The effect of risk factor reductions between 1981 and 1990 on coronary heart disease incidence, prevalence, mortality and cost. *J Am Coll Cardiol* 2001;38:1012-7.
17. Akushevich I, Kravchenko JS, Manton KG. Health-based population forecasting: effects of smoking on mortality and fertility. *Risk Analysis* 2007;27:467-82.
 18. Scholten RJ, Clarke M, Hetherington J. The Cochrane Collaboration. *Eur J Clin Nutr* 2005;59 Suppl 1:S147-9.
 19. Bjørndal A. Improving social policy and practice: knowledge matters. *Lancet* 2009;373:1829-31.
 20. Truman BI, Smith-Akin CK, Hinman AR, Gebbie KM, Brownson R, Novick LF, et al. Developing the guide to community preventive services—overview and rationale. The Task Force on Community Preventive Services. *Am J Prev Med* 2000;18(1 Suppl):18-26.
 21. Berthelot J-M, Le Petit C, Flanagan W. Use of longitudinal data in health policy simulation models. Proceedings of the American Statistical Association; 1997 Aug 10–14; Anaheim, California.
 22. Innvaer S, Vist GE, Trommald M, Oxman AD. Health policy-makers' perceptions of their use of evidence: a systematic review. *J Health Serv Res Policy* 2002;7:239-44.
 23. Sorian R, Baugh T. Power of information: closing the gap between research and policy. *Health Aff (Millwood)* 2002;21:264-73.
 24. Lavis JN, Robertson D, Woodside J, McLeod CB, Abelson J; the Knowledge Transfer Study Group. How can research organizations more effectively transfer research knowledge to decision makers? *Milbank Q* 2003;81:221-48, 171-2.
 25. Citro CF. Introduction. In: Citro CF, Hanushek B, editors. Improving information for social policy decisions—the uses of microsimulation models, volume 2: technical papers. Washington: National Academy Press; 1991. p. 1-7.
 26. Galler HP. Discrete-time and continuous-time approaches to dynamic microsimulation reconsidered. Canberra (Australia): University of Canberra National Centre of Social Economic Modelling; 1997. Technical Paper 13.
 27. Will BP, Berthelot JM, Nobrega KM, Flanagan W, Evans WK. Canada's Population Health Model (POHEM): a tool for performing economic evaluations of cancer control interventions. *Eur J Cancer* 2001;37:1797-804.
 28. Goldman DP, Shekelle PG, Bhattacharya J, Hurd M, Joyce GF, Lakdawalla DN, et al. Health status and medical treatment of the future elderly—implications for Medicare program expenditures: final report. Santa Monica (CA): RAND Corp.; 2004.
 29. Evans RG, Stoddart GL. Producing health, consuming health care. In: Evans RG, Barer ML, Marmor TR, editors. Why are some people healthy and others are not? The determinants of health of populations. New York: Aldine de Gruyter; 1994. p. 27-66.
 30. Arah OA, Westert GP, Delnoij DM, Klazinga NS. Health system outcomes and determinants amenable to public health in industrialized countries: a pooled, cross-sectional time series analysis. *BMC Public Health* 2005;5:81.
 31. Feldman PJ, Steptoe A. How neighborhoods and physical functioning are related: the roles of neighborhood socioeconomic status, perceived neighborhood strain, and individual health risk factors. *Ann Behav Med* 2004;27:91-9.
 32. Wolfson MC. POHEM—a framework for understanding and modeling the health of human populations. *World Health Stat Q* 1994;47:157-76.
 33. Lee IM, Sesso HD, Oguma Y, Paffenbarger RS Jr. Relative intensity of physical activity and risk of coronary heart disease. *Circulation* 2003;107:1110-6.
 34. Folsom AR, Arnett DK, Hutchinson RG, Liao F, Clegg LX, Cooper LS. Physical activity and incidence of coronary heart disease in middle-aged women and men. *Med Sci Sports Exercise* 1997;29:901-9.
 35. Sesso HD, Paffenbarger RS Jr, Lee IM. Physical activity and coronary heart disease in men: the Harvard Alumni Health Study. *Circulation* 2000;102:975-80.
 36. Webby R, O'Connor M, Lawrence M. Judgmental time-series forecasting using domain knowledge. In: Armstrong JS, editor. Principles of forecasting: a handbook for researchers and practitioners. New York: Springer Publishing Co.; 2001. p. 389-404.
 37. Bassett DR Jr, Fitzhugh EC, Crespo CJ, King GA, McLaughlin JE. Physical activity and ethnic differences in hypertension prevalence in the United States. *Prev Med* 2002;34:179-86.
 38. Crespo CJ, Keteyian SJ, Heath GW, Sempas CT. Leisure-time physical activity among US adults. Results from the Third National Health and Nutrition Examination Survey. *Arch Intern Med* 1996;156:93-8.
 39. Lantz PM, Lynch JW, House JS, Lepkowski JM, Mero RP, Musick MA, et al. Socioeconomic disparities in health change in a longitudinal study of US adults: the role of health-risk behaviors. *Soc Sci Med* 2001;53:29-40.
 40. Stevens J, Cai J, Pamuk ER, Williamson DF, Thun MJ, Wood JL. The effect of age on the association between body-mass index and mortality. *N Engl J Med* 1998;338:1-7.
 41. Microsoft Corp. Visual C++: 2005 SP1. Redmond (WA): Microsoft Corp.; 2005.
 42. Statistics Canada. Modgen 9. Ottawa (ON): Statistics Canada; 2007.
 43. SAS Institute, Inc. SAS®: Version 9.1 for Windows. Cary (NC): SAS Institute, Inc.; 2007.
 44. van Baal PH, Polder JJ, de Wit GA, Hoogenveen RT, Feenstra TL, Boshuizen HC, et al. Lifetime medical costs of obesity: prevention no cure for increasing health expenditure. *PLoS Med* 2008;5:e29.
 45. Daviglus ML, Liu K, Yan LL, Pirzada A, Manheim L, Manning W, et al. Relation of body mass index in young adulthood and middle age to Medicare expenditures in older age. *JAMA* 2004;292:2743-9.
 46. Fielding JE, Kominski GF, Hayes-Bautista DE, van Meijgaard J. Trends and forecast of health and economic costs of overweight and obesity in California. Los Angeles: University of California Los Angeles School of Public Health; July 2007. Also available from: URL: <http://www.health-forecasting.org/reports.aspx> [cited 2008 Jun 9].
 47. World Health Organization. Global burden of disease: 2004 update: disease and injury regional estimates for 2004. Geneva: WHO; 2008. Also available from: URL: http://www.who.int/healthinfo/global_burden_disease/estimates_regional/en/index.html [cited 2009 Feb 19].
 48. Jia H, Muennig P, Borawski E. Comparison of small-area analysis techniques for estimating county-level outcomes. *Am J Prev Med* 2004;26:453-60.