

Background

Over the course of more than a decade, researchers at the Center have developed the Stroke Policy Model (SPM), a simulation model that describes the epidemiology of stroke and supports decision and cost-effectiveness analyses of stroke prevention, stroke treatment and stroke rehabilitation. Its primary inputs are the Framingham Study (event rates), Medicare claims (costs) and a large survey of patients at risk for major stroke (utilities).

Initial development took place within the Patient Outcomes Research Team for the Secondary and Tertiary Prevention of Stroke (Stroke PORT). At that time, software (no longer maintained) was created in S and C++. The model has now been reprogrammed in SAS, and is available on this website to be used for non-commercial purposes.

References using the SPM are listed below. Readers should note that many of these references use previous versions of the SPM code and/or different inputs. Accordingly, the results may differ from those obtained using the current code and inputs.

Background references

Matchar DB, Duncan P, Samsa G, Whisnant J, DeFries G, Ballard D, Paul J, Witter D, Mitchell J. The stroke prevention patient outcomes research team: goals and methods. **Stroke** 1993;24:2135-2142.

Parmigiani G, Samsa GP, Ancukiewicz M, Lipscomb J, Hasselblad V, Matchar DB. Assessing uncertainty in cost-effectiveness analyses: application to a complex decision model. **Medical Decision Making** 1997;17:390-401.

Matchar DB, Samsa GP, Matthews JR, Ancukiewicz M, Parmigiani G, Hasselblad V, Lipscomb JP. The stroke prevention policy model (SPPM): linking evidence and clinical decisions. **Annals of Internal Medicine** 1997;127(8):704-711.

Samsa GP, Reutter RA, Parmigiani G, Ancukiewicz M, Abrahamse P, Lipscomb J, Matchar DB. Performing cost-effectiveness analysis by integrating randomized trial data with a comprehensive decision model: Application to treatment of acute ischemic stroke. **Journal of Clinical Epidemiology** 1999;52:259-271.

Matchar DB, Samsa G. Using outcomes data to identify best medical practice: the role of policy models. **Hepatology** 1999;29(Suppl 1):36S-39S.

Samsa GP, Matchar DB. Have randomized trials of neuroprotective drugs been underpowered? An illustration of three statistical principles. **Stroke** 2001;32:669-674.

Samsa GP, Matchar DB, Williams GR, Levy DE for the STAT Participants. Cost-effectiveness of ancreod treatment of acute ischemic stroke: results from the Stroke Treatment with Ancrod Trial (STAT). **Journal of Evaluation in Clinical Practice** 2002;8(1):61-70.

Matchar DB, Samsa GP, Liu S. In press. Cost-effectiveness of anti-platelet agents in secondary stroke prevention: the limits of certainty. **Value in Health**

Matchar DB, Samsa GP. Linking health policy modeling with health policy formation and implementation. In Decision Making in Health Care. Chapman G, F Sonnenberg F (eds). 2000. Cambridge University Press. Cambridge UK.

Matchar DB, Samsa GP. How can modeling best contribute to the assessment of secondary stroke prevention strategies? **American Journal of Medicine** 2005;118:198-199 (letter).

Samsa G, Matchar D, Liu S. Implementing standards for documentation and validation of decision, cost-effectiveness and epidemiologic simulation models: a case study using the Stroke Policy Model (draft)

Matchar DB, Abrahamse PE, Lipscomb J, Samsa G. Policy perspective on acute stroke treatment. In Management of Acute Stroke. A Shaub, L Goldstein (ed). 1999. Marcel Dekker Inc. New York NY.

Matchar DB, Samsa GP. Secondary and Tertiary Prevention of Stroke. Patient Outcomes Research Team (PORT) Final Report – Phase 1. (Prepared by Duke University Medical Center under Contract NO. 290-91-0028). AHRQ Pub. NO. 00-N001. Rockville, MD: Agency for Healthcare Research and Quality. June 2000.

SPM Structure

The Stroke Policy Model is a micro-simulation model that operates by simulating the histories of patients, having various demographic and clinical characteristics, at risk for first or recurrent stroke. Simulated patients pass through the following states associated with model-relevant health events: asymptomatic (ASY), transient ischemic attack (TIA), myocardial infarction (MI), hemorrhagic stroke (HS), ischemic stroke (IS), and death (DTH). Patients cannot leave the DTH state, nor can those with symptoms enter the ASY state.

Otherwise, all transitions between states are possible. In particular, patients may experience a less-severe event such as a TIA after a more-severe event such as a stroke. Repeated events are also possible; for example, a patient with a HS is eligible to have a second HS.

Within the stroke state, severity is recorded using the Rankin score, with categories 0-1 (no disability), 2 (some disability), 3 (noticeable disability), 4 (significant disability) and 5 (most severe disability). Although patients often recover a significant amount of function during the first three months after stroke, the SPM associates with each stroke a single Rankin score, this reflecting the steady-state level of disability after functional recovery is completed. In effect, the SPM posits the level of recovery that will eventually be achieved, assuming that the simulated patient survives long enough to reach this steady state, but does not attempt to model the trajectory by which this steady state is reached.

When a stroke occurs, its severity is randomly chosen from a distribution of Rankin scores. No patient can ever enter a state with less disability, lower cost or higher utility. Thus, for example, if a patient has a previous stroke with a Rankin level 4 and a new stroke with a Rankin level 2, the SPM will credit that patient with the cost of a new stroke-related hospitalization, but will otherwise continue to assign the patient a Rankin 4 level disability.

The occurrence of model-relevant events is determined using waiting time distributions. In particular, we first form covariate-adjusted transition functions for time until transition to TIA, to MI, to HS, to IS, and to DTH. The covariates include history of model-relevant events such as a previous stroke, as well as other patient characteristics such as age, gender, and diabetes. The history of previous model-relevant events is updated at the time of each new event; otherwise, the covariates are not time-dependent.

In determining the next event, we sample independently from the distribution of each of the above waiting times, thus generating a time until the next TIA, a time until the next MI, and so forth. The actual event is the one with the minimum waiting time. If there are limitations on the scope of the simulation – for example, if the simulation is to end at 24 months -- then reaching this limiting point is another potential “next event” as well. The simulation can proceed for a maximum of 600 months.

Costs and utilities (both discounted) are attached to the above patient histories in monthly (30 day) intervals. While the primary influence on costs and utilities is the current health state, both can potentially depend on other patient characteristics as well. Costs are subdivided into categories (e.g., physician costs, hospital costs, durable medical equipment costs).

When using the SPM to assess the impact of interventions (i.e., on stroke prevention, acute treatment or rehabilitation), all of these impacts are quantified via modification of the input parameters. For example, carotid endarterectomy has various risks and costs during the 30 days after the surgery, but improves outcomes thereafter. Its impact would be represented by a modification of the transition probabilities (i.e., to reflect increased hazards during the first 30 days but decreased hazards subsequently), utilities (i.e., to reflect the disutility of surgery during the initial 30-day period) and costs (i.e., to reflect the costs of surgery). No change to the basic structure of the natural history model is required.

Critical comment on SPM structure

The initial review of the SPM’s structure was performed by a group of experts in stroke epidemiology and clinical medicine. Their overall assessment was that the SPM’s structure is sound, and that it should be a useful tool for providing insight into a number of clinical and policy issues pertaining to stroke.

Critical comment focused on various issues. The first involves the SPM's lack of detail at various points. Among these are the following:

- The model contains no information about the trajectory of recovery of functional status from the time of the stroke until the eventual attainment of the steady-state level of disability.
- The covariates (with the exception of those pertaining to history of previous model-relevant events) are fixed rather than time-dependent. For example, simulated patients are assumed to not develop new diagnoses such as hypertension or diabetes.
- By implementing the impact of interventions via changes in the SPM's input parameters rather than a more detailed decision tree, the model loses flexibility in considering complex interventions.

Regarding the decision not to model the trajectory of recovery after stroke, the primary impact is an over-estimation of quality-adjusted life expectancy, because functional status and thus utilities will tend to be over-estimated during the period immediately following the stroke. So long as the follow-up period is relatively lengthy, the impact of this over-estimation should be slight. The issue of fixed versus time-dependent covariates is subsumed within the question of how closely a model should simulate the experience of individual patients (discussed below). The considerations involved with the treatment of interventions are discussed later.

More generally, one of the most fundamental decisions in model development involves the appropriate level of detail. Here, the trade-off is that while more complex models achieve increased realism, they do so at the cost of having input parameters that are more difficult to estimate and justify. At one extreme in detail is the back of the envelope calculation so popular among pragmatic decision makers. At the other extreme is the attempt to create a simulation that is realistic at the level of the experience of the individual patient;

for example, as intended in models that study the propagation of sexually transmitted diseases and other epidemics. (Even more detailed are models that, for example, attempt to describe the development of cancer within the organs or cells of simulated individuals.) As a rule, highly detailed models are best reserved for dynamic, inter-related and non-linear systems (e.g., as illustrated above), where unexpected insights are most likely, rather than the more straightforward type decision and cost-effectiveness analysis considered here, where it usually suffices to be accurate at the level of the population rather than the individual.

The SPM has an intermediate level of detail. It accommodates a large number of covariates; otherwise, the model structure is quite straightforward. This level of detail is consistent with the views expressed by its current and potential users, whose most typical request is whether the SPM can assist them in determining where to most profitably invest their limited resources (e.g., a hypertension management that modestly improves blood pressure control for large numbers of patients at risk for stroke, versus a program that markedly improves anticoagulation management for the smaller number of patients with atrial fibrillation, versus an advertising campaign to increase the number of acute stroke patients that present to the hospital in time to receive t-PA, versus the roll-out of acute stroke services, versus an attempt to improve post-stroke rehabilitation, and so forth). The large number of covariates allows analyses to be targeted to the specific population in question, under the assumption that it is sufficient for the subsequent comparisons to be approximate accurate at the level of the population rather than precisely accurate at the level of the individual.

Critical comment about the SPM's structure has included various technical issues as well. One such question pertains to the assumed independence of the draws from the waiting time distributions. In reality, a patient that is likely to have an IS sooner than average is probably more likely to have a TIA or a MI sooner

than average as well. This lack of independence should not affect group-level results, which would remain correct in aggregate.

A more fundamental technical consideration pertains to the design decision to use waiting times rather than the more traditional format of a cycle tree. The rationale was to increase processing speed. For example, a patient living 350 months might have 3 events during that period. Using the waiting time formulation, this only requires 3 sets of random draws – one per event. Using the alternative cycle tree format, 350 random draws would be required – one per monthly cycle -- even though the result of the great majority of these draws is that no event occurs.

Here, the trade-off is that while waiting time distributions are a very efficient way to implement a natural history model, the most natural way to conceptualize a complex intervention (e.g., involving multiple triggering conditions, multiple steps, and/or multiple conditional branches) is by drawing a decision tree. In turn, such a decision tree is most straightforwardly implemented on a cycle-by-cycle basis (i.e., where the program only needs to keep track of the values of the various triggering conditions one cycle at a time). While it should be technically feasible to define even complex conditional-logic-based interventions using waiting times, we were not able to find any way to do so that avoided a proliferation of programming statements. This did not induce any major design problems, since limiting the model to simple interventions was consistent with user requirements (as discussed above).

Another technical issue pertains to the level of detail used to characterize stroke related disability. The initial version of the SPM was limited to the categories of major/minor stroke, whereas the current version recognizes that disability after stroke is significantly more nuanced, and thus uses a finer classification based on Rankin score. This decision was made with the understanding that the literature at that time did not contain information about

Rankin scores in the level of detail necessary to create the input parameters required by the SPM. (For example, while some information existed on the distribution of Rankin scores among stroke survivors some time after the event, less data could be found on the precise impact of disability on subsequent transition probabilities and costs). Nevertheless, we followed the principle that model structure should reflect the clinical characteristics of the problem under study, and not necessarily be limited by the information that is currently available. In practice, this serves to highlight gaps in current knowledge, and thus can help provide the rationale for future research that fills in those gaps.

The current version of the model includes the major/minor stroke dichotomy as a special case; that is, by setting all the input parameters pertaining to Rankin 0-2 to identical values, and doing the same for all the input parameters pertaining to Rankin 3-5. The same principle applies to patient-level covariates whose impact might be unknown or estimated imprecisely; such covariates can always be set to the group mean or another other reference value that eliminates their impact from the analysis.

As a point of software development, it is usually far easier to drop functionality than to add it later, since adding functionality has the potential to induce multiple changes to the computer code, perhaps even necessitating a fundamental restructuring. Indeed, when it is anticipated that a model will be used for multiple applications, there is much to be said for taking the extra time and other resources to develop a general version of the software having sufficient detail to address as many of the topics for which it might eventually be used as possible. Once such a model is developed and, ideally, peer-reviewed, then creating applications is not only relatively straightforward, but those applications that are created can borrow scientific credibility from the already well-documented parent model. The next best solution is to make a description of the structure of the ultimate model one of the first steps in the software

development process, and to design the early versions of the model in a fashion that is consistent with this goal of eventually migrating toward this structure.

In retrospect, we were not able to fully apply this principle during the development of the SPM. For example, some applications will require a new state for bleeding or other complications of stroke prevention or treatment, and adding such a new state is likely to be somewhat tedious given the structure of the current code. (That is, the number of states is fixed, whereas in retrospect it would have been preferable to write the code in sufficient generality to be able to change a single input indicating the number of states). Also, as previously discussed, the code structure that can easily create a natural history model does not comfortably extend to those decision and cost-effectiveness applications that require the inclusion of complex interventions. Consideration of these issues is ongoing.

Input description

The SPM's primary data based inputs are:

- baseline transition functions;
- regression coefficients quantifying the effect of the covariates in modifying these baseline transition functions;
- baseline cost functions;
- regression coefficients quantifying the effect of the covariates in modifying these baseline cost functions; and
- utilities.

Other inputs, such as the discount rate, are not discussed here. A comprehensive list of inputs is presented at the conclusion of this section. The Stroke PORT final report provides a detailed description of the derivation of the inputs, which is briefly summarized below.

There are 25 baseline transition functions describing transitions from each of the ASY, TIA, MI, HS and IS states to each of TIA, MI, HS, IS and DTH. Using the Cox proportional hazards model, the covariates modify these baseline transition functions using the formula $\log h(t) = \log h_0(t) + \beta X$, where $h_0(t)$ is the non-parametric estimate of the hazard rate associated with the baseline transition function, β is the vector of regression coefficients summarizing the effect of the covariates, and X is the vector of covariates. These covariates are age, gender, ever smoked, systolic blood pressure, treated hypertension, diabetes, coronary artery disease, atrial fibrillation, heart failure, intermittent claudication, bruit, history of transient ischemic attack, history of myocardial infarction, history of hemorrhagic stroke, history of ischemic stroke, Rankin category (stroke only). Many of the transition functions use a subset of these covariates, in some cases because the covariate is not predictive of outcomes and in other cases because of limitations of the input data sets.

With the exception of the impact of stroke severity, both the baseline transition functions and their associated regression coefficients were estimated using data from the Framingham Study. (For patients with stroke, Rankin score is considered to be a covariate, and thus generates its own set of regression coefficients.) The baseline hazard rates for transitions from ASY to DTH were subsequently adjusted upward, in order to be more consistent with the survival observed in other populations.

Specifically, we submitted a data request to the Framingham investigators, who re-analyzed their data in a fashion consistent with the SPM by (a) forming ASY, TIA, MI, HS and IS cohorts; and (b) censoring the follow-up of patients in these cohorts at the time of the next model-relevant event. Having thus created the appropriate data array, parameter estimation then proceeded using the usual techniques of proportional hazards modeling.

Stroke severity primarily affects quality of life and cost, but might also affect the probabilities of transition as well. For example, patients with a Rankin 5 level disability are bed-bound, and thus are at increased risk for pulmonary embolism, pressure ulcers and other potentially fatal complications. Little information for estimating the effect of stroke severity on transitions was available from the literature; accordingly, these transitions were estimated using expert judgment through a modified Delphi process. Because of the lack of a strong evidence base, in many applications the conservative assumption is made that severity of stroke is not predictive of the probability of subsequent model-relevant events. The distribution of Rankin scores after stroke is based on the literature.

The cost inputs can potentially include, for each of the ASY, TIA, MI, HS and IS states, a 600x13 matrix, with 600 months (30 day intervals) and 13 categories of cost. The cost categories are as follows: direct medical – hospital facility, hospital physician, hospital outpatient, home health, skilled nursing, other nursing, durable medical equipment, rehabilitation unit, rehabilitation hospital, outpatient drug; direct non-medical – caregiver time, environmental modifications; and indirect – lost wages and non-market productivity. For each of the states except ASY, we assume that this cost history is preceded by the cost of an initial hospitalization.

The estimates for most cost categories (e.g., excepting indirect costs, prescription drugs, durable medical equipment and some nursing home usage) were generated from an analysis of Medicare claims for the years 1991-1993. Cohorts of patients with the event in question during 1991 were followed until December 31, 1993; thus, the follow-up time for surviving patients extended from 24-36 months. Economic costs were estimated by multiplying utilization by unit prices. The difference in *per diem* costs associated with different levels of stroke severity was estimated by an expert panel.

Once the core set of Medicare-based analyses were performed, various additional analyses were performed in order to fill out the 600x13 cost grid. Costs beyond 36 months were generated by linear extrapolation, based on the slope of a line fit to the data from months 7-36. Costs of nursing home usage (i.e., beyond Medicare's 100 day payment limit) were based on an algorithm that imputed nursing home placement based on information in part A and part B claims. Prescription drug costs were obtained from the administrative records of a managed care organization. The costs of durable medical equipment were based on a survey. Indirect costs are usually excluded.

We are in the process of updating the above Medicare costs with a new analysis of Medicare costs during 2000-2001. Initial results suggest that, while total costs remain within a similar order of magnitude, the relative allocation of costs into categories has changed, with fewer costs associated with inpatient hospitalizations and rehabilitation and more costs associated with skilled nursing facilities.

Most analyses to date have focused on total costs rather than the disaggregation of cost into categories. Indeed, the core SAS code assumes that cost is a scalar rather than a vector. One, albeit tedious, way to produce an analysis of cost by category is to repeat the analysis, using one cost category at a time, using the same set of random seeds (i.e., and thus the same event history).

Utilities are quantitative representations of health-related quality of life, with a single utility value being assigned to each state in the model. (For this purpose, stroke patients in different Rankin categories are considered to be in separate states, since perceived quality of life varies with severity of stroke.) Utility values, on a 0-1 scale, for the HS, IS and TIA states were based on a large survey of patients at risk for major stroke. Utility values for the ASY and MI

states were estimated from the literature. By definition, the utility value for the DTH state is 0.

SPM input list

Transition functions

- 25 baseline transition functions (from ASY, TIA, MI, IS, HS) to (TIA, MI, HS, IS, DTH)
- For each transition function, a vector of regression coefficients describing the effect of the covariates on baseline transitions

Costs

- 5 baseline cost functions (1 per fundamental model state), these reflecting the cost of the initial hospitalization (if any), then costs in 30-day intervals
- For each cost function, a vector of coefficients describing the effect of the covariates on baseline costs

Utilities

- For ASY, TIA,MS and death, a single utility
- For HS and IS, one utility per Rankin score

Other

- A file containing characteristics (covariates) for each of the simulated patients
- Rate at which to discount costs
- Rate at which to discount utilities
- Various parameters to control the output (e.g., to save a “patient history file” containing each event observed during the course of the simulation)
- Various parameters internal to the simulation (e.g., maximum number of events per patient, maximum duration of the simulation)

Critical comment on SPM inputs

Regarding transition probabilities, although the Framingham Study is an internationally recognized epidemiologic resource, it does have some weaknesses in addition to its notable strengths. First, TIAs were only intermittently recorded, and those that were recorded did not always have a precise date. Accordingly, in most cases transition probabilities from the TIA state were derived from those of patients having ischemic stroke (but excluding the high initial mortality associated with stroke). This is used as a first approximation until information from other cohorts of patients with TIA becomes available.

Second, the size of the cohorts (particularly for HS) imposed various limitations on the estimation of regression parameters. Most importantly, in order to avoid over-fitting it was not possible to include all of the covariates in each of the 25 transition functions.

Moreover, because these regression parameters were estimated within the context of models having multiple and correlated predictor variables, various statistical issues must be considered. For example, the set of covariate values to which the transition functions apply might be much smaller than the range of the individual predictors observed in the various Framingham cohorts. In order to avoid the circumstance where the impact of the covariates is both large and outside the range of the Framingham data, the user might wish to set a constraint that the covariates can affect the log hazard by no more than a certain factor (when considered as a group). Alternatively, a sensitivity analysis can be performed using different values of the covariates.

Also, while the set of covariates followed in the Framingham Study is extensive, it is not exhaustive. For example, although race was recorded, the cohort lacks diversity. No information is available about biomarkers and other

currently topical predictors. Moreover, it is not a statistically trivial matter to supplement the Framingham-based coefficients with parameter values derived from other sources, since these other sources tend to be based on models with different sets of covariates, and multi-variable regression coefficients from models using different sets of covariates are not necessarily comparable. (Regarding race, we nevertheless do use regression coefficients derived from other cohorts, and note the above methodological problems as a limitation).

In total, the implication of these limitations is that although the SPM is structured in a way to potentially account for a large number of patient characteristics, in practice the number of covariates that can realistically be included in any given analysis is moderate to small. This observation applies to virtually all epidemiologic models, and not just to the SPM.

Regarding costs, the most relevant issues involve extrapolation, comparability, and timeliness. All of the Medicare-based costs beyond month 36 are extrapolated. However, total costs for months 7-36 post-event were observed to decline slowly, and in approximately linear fashion, the likely explanation being that (a) these costs are primarily disability related, the level of disability being stable this long after the event in question; and (b) over time, patients with more severe strokes die at a higher rate than those with less severe strokes, tilting the distribution of disability among survivors toward those with lower costs. Since these trends would likely continue over time, making a linear extrapolation beyond month 36 seems reasonable. In addition, in order to ensure internal consistency, various constraints on the extrapolated values were included so that, for example, monthly costs of ASY patients never exceed those of patients in the symptomatic states. Finally, the transition probabilities imply that relatively few symptomatic patients will go for long periods between events, and for those that do the impact of down-stream costs will be increasingly diluted by the impact of discounting.

Medicare claims are only pertinent to the approximately 70% of stroke patients aged 65 years and above. Costs for younger patients required extrapolation. To the degree that indirect costs are included among the categories, this extrapolation may be particularly inexact (e.g., due to different proportions of patients working for pay). Most analyses to date using the SPM have excluded indirect costs.

Regarding comparability, although the core of the cost estimates is based on a single source (i.e., Medicare), various other data sets were used as well. If these sources are differently calibrated, then their costs will not be strictly comparable. This problem is endemic to modeling, since input parameters are often derived from multiple sources, and the principle of comparability often comes into conflict with the equally important principle of using each data set for its best purpose. In practice, these are issues that can only be resolved on a case-by-case basis.

As a more technical point, it should be noted that there is a slight inconsistency between time as measured by the transition functions and time as measured by the cost functions. In particular, for counting costs the 30-day intervals of follow-up time were measured from the date of discharge from the initial hospitalization; in essence, assuming that the length of the initial hospitalization was 0 days. Second, during their derivation no patients were censored from the cost cohorts due to newly incident events. As a result, as the follow-up time increases the “uncensored” cost inputs will increasingly diverge from those “censored” inputs we would have preferred to have. These technical issues were induced by the methods by which the Medicare claims files were processed and, as argued above, their overall impact is unlikely to be serious.

A limitation of the Medicare files is that relatively little clinical information is available. Accordingly, even though the SPM is designed in to allow costs to vary according to the same clinical characteristics that affect event rates, in

practice the only covariates included to date are age, gender and Rankin score (the impact of this latter covariate being derived from an expert panel). Here, the critical assumption being made is that the primary driver of costs is the nature of the clinical event in question (e.g., TIA, MI, HS, IS) and not the other characteristics of patients experiencing that event.

Finally, perhaps the greatest problem with the cost inputs is that they are now a decade out of date. In distinction to event rates and utilities, which should change relatively slowly if at all, changes in the organization and funding of health care have likely affected not only the overall level of costs over time but also the patterns of costs by category. As noted above, we are in the process of addressing this problem by updating the cost estimates.

The question of the timeliness of the cost estimates has not only been of particular salience to users, but is also a fundamental problem for a model whose goal is to provide an ongoing repository of the best current information about the epidemiology of stroke. As a technical exercise in computer programming, it is certainly true that the SPM is designed in such a way as to facilitate the incorporation of new cost inputs (and, indeed, the updating of any of its parameters); that is, as new cost inputs become available the old input files (e.g., 600x13 cost matrices) are simply replaced with new input files, with no other changes to the model being required. Indeed, the SPM is also structured to facilitate tailoring by local users; for example, if a particular user is responsible for only a subset of costs and/or has a wildly different set of unit prices than was applied to the 1991-1993 Medicare files, then she can simply replace the Medicare-based cost inputs alternative inputs better suited to her needs. In practice, though, such inputs are strongly desired by the user yet seldom available. On the other hand, it is prohibitively expensive to generate new cost inputs by analyzing claims data every few years, especially in the absence of a research infrastructure that supports the ongoing maintenance of the SPM. A fully ideal solution to this problem is not immediately apparent.

Regarding utilities, their estimation involved no unusual problems. The remaining issues (e.g., whether to use utilities estimated *ex post* or *ex ante*, whether the ASY health state should receive a utility value that is less than 1 since it is the most preferred state in the model yet represents less than perfect health, whether a major stroke is perceived by patients to be worse than death), although by no means trivial, are encountered on a regular basis by the users of utilities. As with costs, although the SPM can potentially accommodate the impact of covariates on utilities, at present our input data are not sufficiently detailed to allow us to do so.

To comment on the above critique, the SPM is the current state-of-the-science in modeling for stroke, and producing its inputs required extensive efforts to re-analyze one of the world's premier epidemiologic studies, to process very large numbers of nationally-representative administrative claims, and to design and implement a large survey of quality of life for patients at risk for stroke. The critical review of the data inputs is not intended to suggest that these inputs are unsound; indeed, they are the best that are currently available and have usually been quite adequate for their intended purpose. This commentary does, however, illustrate that even excellent data sets have their limitations, and that it is the responsibility of the modeler to provide a thorough reporting of those limitations. The degree to which these potential limitations have a tangible impact on a data analysis is context-dependent; that is, data limitations are not absolute but in the last resort depend upon the use to which a model is placed.

One consistent set of questions from users pertains to the degree to which the SPM's results can be generalized beyond those data sets on which its input parameters were derived. For example, recognizing that the participants in the Framingham Study are disproportionately Caucasian, to what degree does the SPM generalize to other, more racially and ethnically diverse, populations? How serious is the limitation caused by the use of outdated cost estimates?

Ultimately, the best answers to questions such as these lie in external validation; that is, in a comparison between the actual experience of populations other than those used to generate the SPM's inputs and the experience that is predicted by the SPM.

Finally, it might be noted that there has been no formal discussion of the validation of the SPM's input parameters in the sense of, for example, the statistical validation of the proportional hazards models fit to the Framingham data in order to derive the transition probabilities. The fact that such parameters will ultimately be used within the context of an epidemiologic simulation model should have no effect on these issues of initial statistical validation. From the perspective of this subsequent modeling, it is sufficient that the model developer:

- have confidence that the analyses used to estimate the input parameters were consistent with the state of the science in statistics;
- have generated confidence intervals or similarly derived estimates of precision that can be used as ranges across which the model's inputs parameters can be allowed to vary; and
- report the derivation of the input parameters in detail, as illustrated by the documentation on our website.

Input files

The SPM's inputs consist of a set of flat SAS files, which are typically housed within a single SAS library. It is recommended that users print these files, in order to better visualize the requirements and characteristics of the input data. As noted above, users can replace these default inputs with their own files, so long as these files are identically structured, by replacing the pointers in the initial part of the SAS program used to read in the data.

SAS code description

The SPM is written in SAS-IML, a structured matrix language. Briefly, the program first collects all the data inputs, whether as defaults or as tailored values provided by the user. Inputs can potentially be based on a random selection (for example, from a prior distribution) performed outside the main body of the code, thus facilitating probabilistic sensitivity analysis.

Next, the program collects various functions, these functions being self-contained entities that perform the main operations of the SPM. These functions/operations are listed below.

Next, the program initializes various population-level counters, and begins a “do loop” with one iteration per patient. Within this do loop, the program initializes various patient-level counters, then begins a nested do loop with one iteration per event. (The number of events per patient is variable.) For each event, the sequence of steps, accomplished by ordering the above functions and passing parameters (i.e., function inputs and outputs) through these functions as appropriate, is as follows:

1. Modify baseline transition function to account for covariates;
2. Randomly select the waiting times until each possible event;
3. Select the next event as the one with the minimum waiting time;
4. Update the summary of survival;
5. Update the summary of quality-adjusted survival;
6. Modify the baseline cost function to account for covariates;
7. Calculate cost based on time since the previous event;
8. Update the summary of costs;
9. Check whether the event in question is the final one for that patient;
10. Update the counters; and
11. Based on whether the event is the final one for that patient, transfer control to the appropriate location in the program.

Critical comment on main software algorithms

During the initial development of the SPM, we requested comment from various software designers, including a faculty member in the computer science department at our university. Components of this review included the level of detail included in our specification of the programming task, the structure of the basic algorithms, and the choice of programming language. Comment on the choice of programming language is postponed until the discussion section.

The basic algorithms were found to be straightforward, and the use of functions was believed to be likely to achieve an adequate degree of modularization in the code. As a general principle, such modularization is critical to subsequent development and maintenance of the code, as it allows large tasks to be disaggregated into smaller sub-tasks that can be accomplished independently of the others.

Our reviewers found that the level of specification of the programming task was far less than is the standard current among professional software developers. Indeed, this pointed to an ongoing difficulty we (and many others that are primarily content experts rather than software designers) have faced; namely, that the task of creating formal specifications, in the sense as understood by software designers, is extremely tedious and time consuming, as well as being difficult to accomplish. On the other hand, without such detailed specifications, the risk is significant that what a programmer creates will not be what is intended. This risk is increased by the high level of clinical and scientific content that is typically contained within the sort of model required to provide information on questions of clinical and policy importance. In addition, the risk is further increased by the fact that a non-trivial portion of this information is housed as background (i.e., tacit) knowledge of the content experts, and thus might not be explicitly communicated to the software designer. In summary, we found that when attempting to create a model with a high level of epidemiological and

clinical content, the task of precisely communicating expectations between clinical experts and the software designers is very difficult; indeed, for us it was sufficiently difficult as to induce more than one false start the process of software development.

Ultimately, our solution to this problem was to have one of the substance-matter experts program a prototype of the software. This ensured that the basic structure and algorithms were communicated as intended, thus allowing the designer to focus on issues of code development. This prototype served as a partial statement of the programming specifications (albeit not in the traditional format), and also facilitated a much more fruitful dialogue between the software engineers and the clinical experts. In essence, the prototype provided our software designer with sufficient background in order to be able to ask informed questions, and then translate the answers of these questions into the types of internal specifications that are typical of that discipline.

We have found prototyping by content experts to be a very useful approach, and now use it whenever possible.

Validation

Our validation and reporting strategy used the following steps:

1. Describe the SPM's structure and obtain critical comment about that structure.
2. Describe the SPM's inputs and obtain critical comment about their values.
3. Describe the SPM's main algorithms and obtain critical comment about their implementation.
4. Perform unit testing (i.e., test each of the major software components on a stand-alone basis.)
5. Perform system testing (i.e., test the software when operated as a whole).

6. Assess the degree to which the SPM's results can be generalized by applying it to databases that were not involved in its original development.

The first three steps have been illustrated above.

Unit testing

Unit testing was applied to each of the functions/modules of the SPM. We first tracked the code using hand calculations. Then, we tested the module using a more extensive set of inputs and assessed the outputs. In assessing the pattern of results, we focused on both magnitude and direction. For example, consider the SPM module that replaces a baseline transition function input with its patient-specific equivalent. Here, decreasing the value of one of the regression coefficients associated with that transition should lead to a monotonic increase in the relevant transition function; decreasing this regression coefficient by a larger amount should lead to a greater change in this transition function. All function outputs should be within their anticipated ranges.

The results of the unit testing were for the most part positive. However, it was discovered that out of range values of the inputs would either cause the program to stop executing, or (perhaps even worse) to propagate errors throughout the remaining calculations. In response, we added a module that verifies that the SPM's inputs are all within their expected ranges. In the current application, it is sufficient for this module to perform its range checks one variable at a time (i.e., if each individual variable is as expected, no combination of these inputs will lead to function outputs that are out of range). However, this relationship does not hold in general and, indeed, failure to anticipate pathological combinations of seemingly harmless input values is a common failing of many validation plans.

System testing

In contrast with unit testing, system testing involves testing the SPM when operating as a whole. The tests were all constructed in a similar fashion; treating the SPM as a “black box”, inputs were varied (either individually or in groups), and the impact of these inputs on the model outputs (typically, survival, quality-adjusted survival, and costs) was examined. As before, the reasonableness of these outputs was assessed using the criteria of direction and magnitude. For example, increasing the value of a cost input should increase lifetime costs while leaving survival and quality-adjusted survival unchanged. Decreasing the utility for MI should decrease the lifetime QALYs (but only if transitions to MI are possible), yet leave survival and cost unchanged. Decreasing the values of a transition function should (in general) decrease overall survival, and thus affect lifetime cost and QALYs as well. When assessing transitions, it was often helpful to make all the other transitions impossible, this being achieved by setting all but the last value of these other transition functions to 1. We also tracked the pattern of events for individual simulated patients; for example, in order to verify that no impossible transitions occurred and that no possible transition was inadvertently excluded by the logic of the code.

Whenever possible, the simulations in question utilized the same seed value for the random number generator. For example, this implies that, for two realizations with identical inputs except for costs the lifetime QALYs should be identical, since exactly the same pattern of events would be observed in both cases. This principle does not apply to unit tests involving transition functions, however, since once the initial pattern of events is changed the same set of seed values need not imply that subsequent events will be the same.

System testing was preceded by an assessment of the magnitude of Monte Carlo variation attributable to changes in the model’s random seed. All tests used a sample size that was sufficiently large so that the effects under study were expected to be larger than this Monte Carlo variation. As a general rule, for the SPM simulated populations of size 10,000-20,000 are sufficient when

following symptomatic cohorts. Asymptomatic cohorts require larger sample sizes, for example, in the range of 100,000, because not all patients suffer vascular events before death.

It might be noted that, in practice, the precise point of delineation between unit testing and system testing is not always clear. For our purposes, we defined unit testing as being limited to that form of validation whereby portions of the model were run. For the example in the previous section, the validation begins when the module implementing the change from baseline to patient-specific transition functions is called, and ends when that function returns an output (i.e., the patient-specific transition function) back to the main program. In system testing, the SPM as a whole was run, and thus its usual outputs (e.g., life expectancy, QALYs, costs) were the focus of attention.

The system testing involved two types of analyses. First, we made all transitions other than death impossible, and verified that the estimated survival curves (i.e., the outputs) matched the Framingham-based survival curves used as inputs (data not shown).

Second, we examined the impact of changing various inputs, as discussed above. The table at the conclusion of this section presents illustrative results, focusing on the portion of the validation dealing with the IS state. Row 1 describes the life expectancy, QALYs and costs for a reference population of 70-year old men with new IS and Rankin level 3. The values of their covariates (other than age, gender and stroke severity) were set to the means of the Framingham population with IS. The life expectancy is approximately 6 years, consistent with the literature, and the quality-adjusted life expectancy approximately 2 years. (The utility for a Rankin level 3 stroke is .5, so the quality-adjusted life expectancy can be no more than $(.5)(6.250)$ years. Since patients suffer subsequent and possibly more disabling strokes, the QALYs should and do drop below this threshold. QALYs a decreased even more by the default

discounting factor of 3%.) The lifetime cost (excluding indirect costs) is estimated to be \$77,623.

Row 2 presents results for a population with identical characteristics, except that the type of stroke is HS rather than IS. Because of the higher short-term mortality associated with HS, lifetime survival is dramatically decreased. Row 3 considers a population with new IS, but with the transition function to survival intermediate between the reference case and that of HS. As anticipated, the outcomes are intermediate as well.

Rows 4-8 illustrate the impact of stroke severity on outcomes. With increasing stroke severity, survival decreases yet cost increases, this latter phenomenon occurring because of the greater daily costs associated with more severe strokes. Row 9 illustrates the impact of doubling the costs associated with the IS state. Survival and QALYs are unchanged, but costs are increased. Row 10 illustrates the impact of doubling the costs associated with the MI state. Costs increase, but the magnitude of this increase is much smaller, since only a subset of patients with IS will suffer a MI. Row 11 illustrates the impact of decreasing the utility associated with stroke. QALYs are decreased, while life expectancy and costs remain unchanged.

Rows 12 and 13 illustrate the impact of changing one of the patient characteristics. There, the percentage of diabetic patients is changed from the Framingham mean to 0% (row 12) and 100% (row 16), respectively. Survival increases in row 12 and decreases in row 13, as expected. Because the default value for the percentage of diabetes is much closer to 0% than 100%, the results of row 12 are much closer to those of the reference population than those of row 13. Moving from the non-diabetic population to the diabetic population leads to a drop in survival of approximately 2 years, illustrating the potentially strong impact that the covariates can have on the results. Row 14 illustrates the impact of gender. With similar covariates, survival for females is slightly better than for

males. (In practice, survival for females might also be increased by having superior values of the covariates.) Row 15 illustrates the results of changing the age at stroke from 70 to 80 years. Overall survival decreases noticeably. The impact of the other covariates was in the anticipated direction (data not shown). Rows 16 and 17 illustrate the cumulative impact of the covariates for a 70-year old man. In row 16 all covariates are set to a minimum value (decreasing risk), while in row 17 all covariates are set to a maximum value (increasing risk). The reduction in survival in row 17 is dramatic. (Indeed, very few patients would have the maximum value of all the covariates; this point is almost certainly outside the range of the Framingham data used to estimate the regression parameters associated with these covariates.) Finally, rows 18 and 19 illustrate the results of changing the discount rate associated with costs from the default of 5% to 0% and 8%, respectively. (Discounted) costs are affected, but not survival or QALYs. Similarly, changing the discount rate for the utilities affects QALYs but not survival or costs (rows 20 and 21).

Combining the results of the unit and system tests, it appeared that, when problems occurred, they tended not to be attributable to the actions of the individual modules. Instead, they were concentrated in the code that linked these modules together. This “glue” serves two main functions: (a) providing a set of counters (e.g., internal counters such as the number of events per patient, external counters such as the total QALYs for the cohort as a whole) to facilitate the exchange of information across different portions of the program; and (b) keeping track of the implications of less serious events that follow more serious events. Fixing the counters is a standard problem, which was adequately solved by the usual techniques of diagramming the code, printing intermediate results, verifying that the counters remained in their expected ranges, and so forth.

Regarding the implications of less serious events that follow more serious events, the program logic requires (a) identifying the situation using various indicator variables and counters; and (b) acting on the situation by altering the

treatment of transitions, costs, and utilities. The design problem was that implementing this logic seemed to require either (a) rewriting the functions in order to keep track of these issues, which would add to their complexity; or (b) having the “glue” between the modules use conditional logic in order to keep track of each special case (e.g., less disabling stroke following a more disabling stroke, TIA following a stroke). We opted for the latter, noting that if the blocks of code containing this conditional logic became too large or grossly violated the goal of modular organization, then the program as a whole would become difficult to debug, maintain, and extend. For now this solution works, although we do harbor some residual doubts about its adequacy as the functionality of the model is extended over time. In any event, the results reported in table 3 reflect the impact of all of the changes to the SPM code suggested by the unit and system tests. Any inconsistencies would call into question the validity of the code as currently written.

Illustrative System Testing Results

Row	Survival	QALY	Cost (\$)	Description
1	6.250	2.060	77,623	IS, Rankin 3, 70-year old male, baseline covariates
2	3.758	1.347	49,894	HS, Rankin 3, baseline IS covariates
3	4.991	1.647	61,929	IS, Rankin 3, baseline covariates, transition probabilities between usual IS and HS
4	7.652	3.379	49,114	IS, Rankin 1
5	7.031	2.754	57,346	IS, Rankin 2
6	6.250	2.060	77,623	IS, Rankin 3
7	5.413	1.362	138,159	IS, Rankin 4
8	3.334	0.537	141,388	IS, Rankin 5
9	6.250	2.060	149,107	IS, Rankin 3, IS costs doubled
10	6.250	2.060	82,924	IS, Rankin 3, MI costs doubled
11	6.250	1.762	77,623	IS, Rankin 3, utilities of IS and HS decreased
12	6.645	2.190	80,505	IS, Rankin 3, 0% diabetic
13	4.778	1.595	66,597	IS, Rankin 3, 100% diabetic
14	6.643	2.238	78,842	IS, Rankin 3, 70-year old female
15	2.708	1.060	42,005	IS, Rankin 3, 80-year old male
16	7.414	2.426	84,733	IS, Rankin 3, all covariates at minimum
17	0.384	0.164	10,562	IS, Rankin 3, all covariates at maximum
18	6.250	2.060	98,570	IS, Rankin 3, cost discount rate from 5% to 0%
19	6.250	2.060	68,880	IS, Rankin 3, cost discount rate from 5% to 8%
20	6.250	2.739	77,623	IS, Rankin 3, utility discount rate from 5% to 0%
21	6.250	1.779	77,623	IS, Rankin 3, utility discount rate from 5% to 8%

External validation

The final step in the validation process involves comparison of the model against an external standard or standards. Although often considered to be a form of validation, such an assessment might equally well be understood as a test of generalizability. In other words, a model might work exactly as intended (as demonstrated by the unit and system tests), and thus be considered “internally valid”, yet still yield unexpected results when applied to an external cohort.

In pursuit of the goal of external validation, numerous potential comparisons can be considered. Indeed, no single analysis would suffice to declare a model to be “externally valid”, but the more evidence that can be accumulated the better.

In order to illustrate the process of external validation, we focus on the portion of the SPM pertaining to patients in the ASY state. Two analyses were performed. First, taking a cohort of white males aged 70 years and with covariates set to the average values of the ASY cohort from Framingham, the SPM was run for 100,000 patients, all of whom were followed until death. From the resulting “patient history file” (i.e., from the list of times until death), an actuarial life table was created, this life table including the proportion surviving each successive year, and the overall life expectancy at each point in time. This life table was compared with a national life table compiled by the Centers for Disease Control.

Second, for the same simulated cohort we estimated, for ages 72, 77, 82, 82, 92 and 97, the number of patients per 100,000 having either a stroke (either IS or HS) during the next year. We then multiplied this incidence rate by an estimate of the United States population aged 70-74, 75-79, 80-84, 85-89, 90-94 and 95-99, compiled from the 2000 census, in order to generate a national estimate of the number of strokes among persons aged 70 and above. This

figure was compared with an estimate of this quantity obtained by the Stroke Prevention PORT, which in turn was based on an analysis of Medicare claims during 1991-1993.

The figure below plots survival curves based on life tables derived from United States population data and the SPM. The survival projected by the SPM is modestly better than for the overall United States population, but the results are encouragingly within the same order of magnitude. For example, life expectancy at age 70 is 14.4 years for the SPM versus 13.2 years for the overall population. The annual mortality rates upon which these survival curves were based were quite similar at younger ages (e.g., at age 72, these mortality rates were 3.35% for the SPM and 3.45% for the United States; these figures were 5.36% and 5.38% at age 77 and 8.16% and 8.25% at age 82, respectively). For older ages, the SPM's figures were a bit lower (e.g., 11.86% versus 13.50% at age 87, 17.64% versus 20.08% at age 92, and 24.77% versus 28.14% at age 97, respectively.)

One possible explanation for the above discrepancy in overall survival is that the United States population life table includes not just those in the ASY state but also those with symptomatic cerebro-vascular disease, and thus should have somewhat poorer survival. Another possible explanation pertains to the characteristics of the Framingham data set. In particular, to the extent that the Framingham population is in better health than the nation as a whole, better survival would be expected. At the time of our analysis, the Framingham Study had few participants aged 90 and above, so the hazard rates for the older age group might not have been precisely estimated. Also, and perhaps not immediately apparent from the figure, for technical reasons our analysis of the Framingham data proceeded in 5-year age intervals (e.g., age 70-74). Within each interval, the slope of the relationship between age and event rates was stronger than observed in the United States population data. As a result, for example, in comparison with the United States population data the event rates

for 70-year olds were low, the event rates for 74-year olds were high, and the event rates for 72-year olds were almost perfectly calibrated. The same phenomenon is observed for ages 75-79, ages 80-84, and beyond. The net impact of this discrepancy is to slightly increase life expectancy as projected by the SPM. Smoothing the transition function from the ASY state to the death state in order to enforce monotonicity on its age-specific hazards would be a reasonable response, and would serve to push life expectancy toward that observed in the overall population. Finally, in using the SPM to generate a life table for a 70-year old male, we set the covariates equal to the characteristics of a typical 70-year old male, and then ran the model until all its simulated patients had died. We did not update the covariates to reflect the fact that, for example, 80-year olds tend to have more risk factors for stroke than do 70-year olds. The impact of this decision is to over-estimate life expectancy as estimated by the SPM.

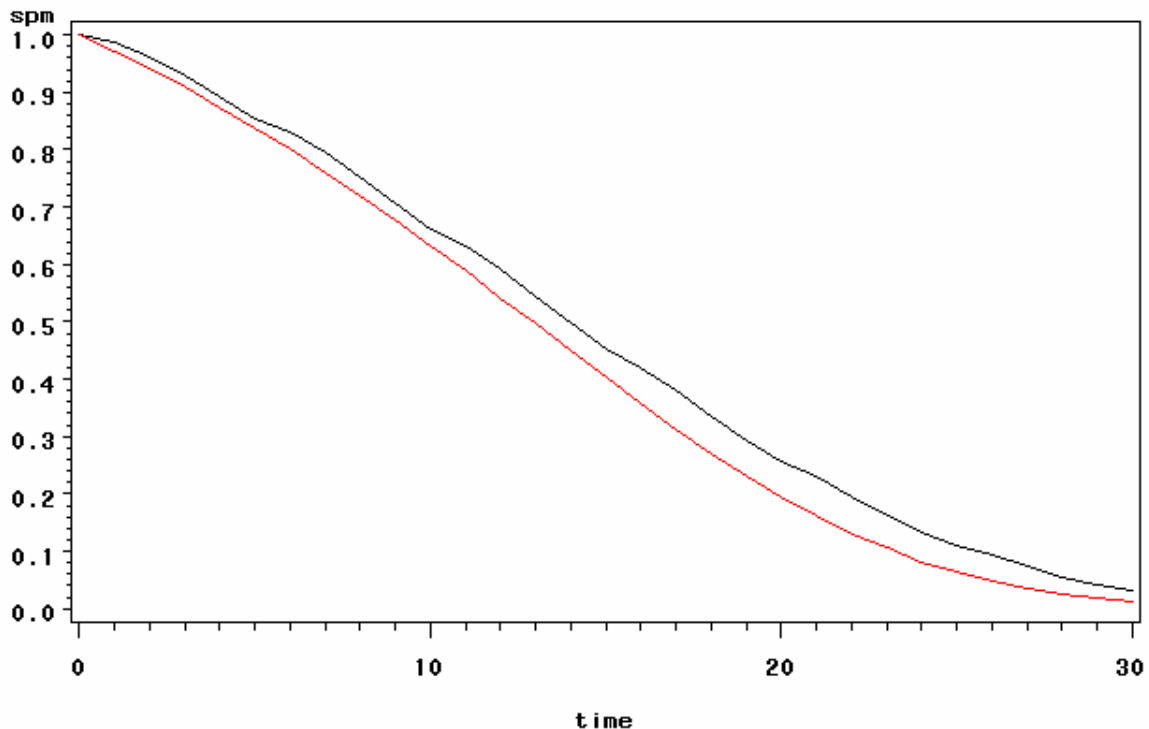
The table below illustrates the use of the SPM in order to estimate the total number of first strokes (i.e., transitions from ASY to either IS or HS) expected in the United States in any given year. For the nation as a whole, the annual number of incident strokes is typically estimated within the range of 500,000-700,000. Taking the midpoint of 600,000, and noting that approximately two thirds of strokes occur in persons aged 70 and above, yields a figure of approximately 400,000 incident strokes in patients aged 70 and above, which is in substantial agreement with the estimate provided by the SPM.

Although the above comparisons are encouraging, it should be emphasized the assessment of generalize-ability is ongoing, and that no single standard of comparison can be considered as universally applicable. Stroke incidence rates differ across populations, and also in when comparing population-based epidemiologic cohorts versus the control groups of randomized trials. A similar phenomenon applies to survival. Both incidence and survival may change over time. Our viewpoint is that the most appropriate set of inputs to

use is application-specific rather than universal, and that two of the most important things for the modeler to keep in mind are documentation and calibration. Documentation ensures that the user of the model is informed of which inputs have been selected. Calibration ensures that, if necessary, the default inputs are modified in order to produce event rates and survival that is consistent with the population under study. When calibrating, our recommendation for a typical application is to modify the baseline transition functions (e.g., by multiplying by a constant hazard ratio), while leaving the regression coefficients that quantify the impact of the covariates in modifying these hazards unchanged.

Our efforts at model development and validation are ongoing.

Illustrative External Validity Testing Results: Survival Curves for 70-Year Old White Males, United States Population (observed) Versus SPM (projected)



Legend: Top curve is projected by SPM, bottom curve is from United States population data.

Illustrative External Validity Testing Results: Expected Number of Strokes

Age	SPM: Strokes / 100,000 at risk	United States population size	SPM: Expected strokes in United States
70-74	594	8,500,000	50,490
75-79	904	7,736,000	69,333
80-84	1,523	5,576,000	84,922
85-89	2,728	3,206,000	87,460
90-94	5,095	1,431,000	72,909
95-99	10,002	412,000	41,208
Total			405,622