

# Linked Sensitivity Analysis, Calibration, and Uncertainty Analysis Using a System Dynamics Model for Stroke Comparative Effectiveness Research

Yuan Tian, MSc, Kristen Hassmiller Lich, PhD, Nathaniel D. Osgood, PhD,  
Kirsten Eom, MPH, David B. Matchar, MD

**Background.** As health services researchers and decision makers tackle more difficult problems using simulation models, the number of parameters and the corresponding degree of uncertainty have increased. This often results in reduced confidence in such complex models to guide decision making. **Objective.** To demonstrate a systematic approach of linked sensitivity analysis, calibration, and uncertainty analysis to improve confidence in complex models. **Methods.** Four techniques were integrated and applied to a System Dynamics stroke model of US veterans, which was developed to inform systemwide intervention and research planning: Morris method (sensitivity analysis), multistart Powell hill-climbing algorithm and generalized likelihood uncertainty estimation (calibration), and Monte Carlo simulation (uncertainty analysis). **Results.** Of 60 uncertain parameters, sensitivity analysis identified 29 needing calibration, 7 that did not need calibration but significantly influenced key stroke outcomes, and 24 not influential to calibration or stroke outcomes that were fixed at their best

guess values. One thousand alternative well-calibrated baselines were obtained to reflect calibration uncertainty and brought into uncertainty analysis. The initial stroke incidence rate among veterans was identified as the most influential uncertain parameter, for which further data should be collected. That said, accounting for current uncertainty, the analysis of 15 distinct prevention and treatment interventions provided a robust conclusion that hypertension control for all veterans would yield the largest gain in quality-adjusted life years. **Conclusions.** For complex health care models, a mixed approach was applied to examine the uncertainty surrounding key stroke outcomes and the robustness of conclusions. We demonstrate that this rigorous approach can be practical and advocate for such analysis to promote understanding of the limits of certainty in applying models to current decisions and to guide future data collection. **Key words:** sensitivity analysis; calibration; uncertainty analysis; stroke; System Dynamics; simulation model. (*Med Decis Making* 2016;36:1043–1057)

Simulation models can serve as valuable health research tools. Existing simulations have offered important insights into intervention tradeoffs and have aided in identifying drivers of trends. They have also helped in prioritizing data collection and communicating health care system structure to research teams as well as decision makers.<sup>1–3</sup> With more complex health problems being investigated, simulation modeling presents both opportunities and challenges.<sup>4</sup> The opportunities lie in the role it can play in guiding investment, but a principal

challenge derives from the fact that broader system models often come with increased uncertainty.

There are 3 major sources of uncertainty in simulation models. *Structural uncertainty* results from the fact that models are, by necessity, simplifications of reality. Structural uncertainty refers to uncertainty in the causal assumptions made in the process of conceptualizing a model. *Parameter uncertainty* stems from imperfect or unavailable data leading to imprecise estimates of model parameters. In some cases, uncertain model parameters can be informed by a *calibration process*. This process involves maximizing the consistency of the behavior of the model to empirical data and observed trends by adjusting uncertain model parameters. *Stochastic uncertainty* can also contribute to variation in outcomes due to the influence of random variables.<sup>5</sup> For example,

heterogeneity in the characteristics of individuals or the outcomes of an event might differ both within and across runs of a simulation model if draws from a predefined distribution are introduced within each run of the simulation model.

Not surprisingly, this variety of uncertainty can make analysts and decision makers distrust results and insights, particularly if they are presented without sufficient indication of the degree of confidence that accompanies them.<sup>6</sup> In this context, it is important to learn how uncertainty affects a model's behavior and whether it undermines the research team's ability to make conclusions without seeking more data. It is also critical to build confidence among decision makers and other stakeholders if the model is going to affect real-world programs, practices, and policies. To this end, the best modeling practices recommended by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) - The Society for Medical Decision Making (SMDM) Modeling Good Research Practices Task Force are useful.<sup>5,7</sup> But when working with a complex model, knowing how to integrate sensitivity analysis, calibration, and uncertainty analyses, as recommended, in an efficient process that can be explained and used to support decision making is challenging.

We have already explained calibration but would like to clarify the subtle difference between sensitivity and uncertainty analyses, as the distinction is often unclear. *Sensitivity analysis* seeks to quantify

the effects of varying structural assumptions and parameter values, or of introducing stochasticity on simulation outcomes of interest (e.g., quality-adjusted life years). In contrast, *uncertainty analysis* seeks to estimate the plausible range around the simulation outcomes of interest, given structural, parameter, and/or stochastic uncertainty in the model. Research teams building and analyzing dynamic models have long applied a basic set of techniques for sensitivity and uncertainty analyses. In particular, 1-way and multivariate sensitivity analyses have been widely employed in many domains using a variety of methods such as scenario analyses presented in the form of tables or tornado charts,<sup>5</sup> Monte Carlo-based analysis,<sup>8-10</sup> parameter screening,<sup>10,11</sup> and informal and formal Bayesian methods.<sup>2,12,13</sup> Given that calibration results are typically stochastic (e.g., reflecting a stochastic global optimization), modelers have addressed their uncertainties using bootstrap techniques.<sup>14</sup> However, calibration, sensitivity analyses, and uncertainty analyses are usually conducted as separate stages of the modeling process; a common approach is to calibrate a model and then to perform sensitivity and uncertainty analyses, separately, around the point estimates of the parameters. Unfortunately, this approach might vary previously calibrated parameters in the subsequent analyses, which might undermine the credibility of the model in replicating empirical data patterns. Furthermore, uncertainty in model calibration is too often not considered in subsequent analyses and recommendations.

Although the individual roles of sensitivity analysis, calibration, and uncertainty analysis are clear, their integration has been unclear. In this article, we showcase a systematic way to integrate sensitivity analysis, calibration, and uncertainty analyses using a System Dynamics stroke model for the US Veterans Affairs (VA) population.<sup>15</sup> The illustrative decision problem is described in more detail below, to offer a concrete example of proposed methods. In this framework, sensitivity analysis supports understanding of which uncertain parameters lead to the most significant uncertainty in key outputs being studied. Sensitivity analysis also directly informs model calibration, and results from these steps together shape the way uncertainty is operationalized in the final multivariate uncertainty analysis. Furthermore, we conclude by demonstrating additional approaches to display results from uncertainty analysis. In this project, we compared 15 distinct intervention approaches to improve stroke care in the VA health care system. In the face of broad and complex decision challenges such as the illustrative decision

---

Received 8 September 2014 from Program in Health Services & Systems Research, Duke-NUS Graduate Medical School Singapore, Singapore (YT, KE, DBM); Department of Health Policy & Management, University of North Carolina at Chapel Hill, NC, USA (KHL); Department of Computer Science, University of Saskatchewan, SK, Canada (NDO); and Department of Internal Medicine, Duke University Medical Center, Durham, NC, USA (DBM). This research was supported by the Singapore Ministry of Health's National Medical Research Council under its STaR Award Grant (NMRC/STaR 0005/2009), the Department of Veterans Affairs, Veterans Health Administration (VHA), Health Services Research and Development Service Quality Enhancement Research Initiative (RRP 07-295) and the VA Center for Applied Systems Engineering (VA-CASE), and award number KL2RR025746 from the National Center for Research Resources. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. Revision accepted for publication 15 March 2016.

Supplementary material for this article is available on the *Medical Decision Making* Web site at <http://mdm.sagepub.com/supplemental>.

Address correspondence to David B. Matchar, MD, Program in Health Services & Systems Research, Duke-NUS Graduate Medical School Singapore, 8 College Road, 169857, Singapore; e-mail: david.matchar@duke-nus.edu.sg.

problem presented in this article, this integrated analysis offers a new approach for integrating currently disconnected data to provide robust policy recommendations and prioritize future data collection efforts, when needed, to further clarify decisions.

## DECISION PROBLEM

In the United States, stroke is a leading cause of death and long-term disability,<sup>16</sup> and the same is true for the US Department of Veterans Affairs population.<sup>17</sup> To address this problem, the VA Stroke Quality Enhancement Research Initiative (QUERI) was launched with the mission of identifying and promoting practices to reduce the burden of stroke through enhanced prevention and treatment.<sup>17</sup> Given the extensive set of evidence-based options aimed at different aspects of the spectrum of stroke care and prevention, the Executive Committee of the Stroke QUERI, consisting of VA stakeholders and stroke experts, sought a strategy to prioritize potential prevention and treatment options and prioritize future research around stroke in the VA. The spectrum of care options for veterans considered by the Stroke QUERI included broad primary prevention, targeted primary and secondary prevention, stroke awareness, acute care, and rehabilitation (Table 1). Across these 15 alternative intervention approaches, their challenge was to compare the relative impact of each on key population-level stroke outcomes, including quality-adjusted life years (QALYs), strokes, and stroke fatalities. This challenge is especially relevant in the presence of uncertainty about interventions' effects on the VA population. For instance, although previous studies have documented that both acute stroke care (e.g., improving use of thrombolytic therapy) and risk factor management (e.g., lowering blood pressure among patients with a history of stroke) have a favorable impact on stroke outcomes, it was unclear which is more beneficial in an evolving VA health system serving a changing veteran population.

## SYSTEM DYNAMICS STROKE MODEL AND SOURCES OF DATA

The available VA empirical data, the current literature on stroke interventions and outcomes, and experts' opinions were synthesized into a system dynamics stroke model (Figure 1).<sup>15,18–26</sup> The Stroke QUERI Executive Committee was engaged in each phase of model construction and analysis. The stroke model was a deterministic differential equation

model containing 11 stocks (accumulating disease states) to stratify VA users into mutually exclusive health states based on their history of transient ischemic attack (TIA) and ischemic stroke. TIAs are caused by a temporary disruption in the blood supply to part of the brain. They do not result in any functional loss but are associated with increased risk for stroke. VA users were defined as the veterans who used VA primary care services in the preceding year and reflect the population of veterans most reachable with intervention. Risk factors included age (stratified into <45, 45–64, 65–75, 75+ years), hypertension measured by systolic blood pressure (SBP [mm Hg]; <140, 140–159, 160+), atrial fibrillation, diabetes mellitus type 2, smoking status, and history of cardiovascular disease (CVD). The post-TIA population was differentiated by diagnosis status (diagnosed/not) and time since the attack. The poststroke population was stratified by time since the most recent stroke and modified Rankin Scale (mRS). The modified Rankin Scale is a commonly used scale for measuring the degree of disability or dependence in the daily activities of people who have had a stroke or other causes of neurological disability.

Transitions between stocks, represented as flows, determined the events or pathways through which the health status of VA users can progress over time (Figure 1). We started with a simulated population of 4.14 million VA users in 2010. Our stroke model allowed some veteran nonusers to become new VA users, either spontaneously or in the face of a new TIA or stroke. The model was programmed in Vensim 5.11 DSS,<sup>27</sup> and more information can be found online (<http://vastrokemodel.weebly.com/>).

The stroke model required data on the utilization of VA facilities, current practices, plausible projected levels of care resulting from future interventions, and the effectiveness of the intervention approach in terms of reducing the incidence, mortality, and functional loss poststroke. Although empirical and historical stroke data were available through published and unpublished sources and scientific evidence on stroke is rich, we still faced parameter and structural uncertainty. Parameter uncertainty stemmed from the fact that data were obtained in different years or from different subpopulations (e.g., regional data) or were incomplete (e.g., not all strokes among veterans are treated within VA facilities and thus not observable in available data). Scientific evidence is often presented in ways that imply different model structure. As an example of structural uncertainty, we needed to consider alternative assumptions about whether a given therapy reduced the risk of mortality

**Table 1** Fifteen Stroke Intervention Approaches Considered by the Stroke Quality Enhancement Research Initiative (QUERI) to Improve Quality-Adjusted Life Years among Users of the Veterans Affairs System

Intervention Category	Description of Intervention Approaches
Preevent primary prevention	Improve hypertension control for all VA users with SBP >140 mm Hg Improve hypertension control for VA users with SBP >160 mm Hg Improve hypertension control for VA users with DM Improve ischemic stroke prevention for all eligible VA users with AF (antiplatelet/anti-coagulation therapy) Improve hypertension control and/or management of AF for VA users with prior cardiovascular disease
Secondary prevention	Improve accuracy/timeliness of TIA diagnosis in VA users by increasing awareness/symptom recognition Improve the immediate management of TIA through expedited evaluation and risk factor management Improve the rate of CEA for eligible individuals post-TIA Increase % of stroke patients receiving guideline concordant stroke care for secondary prevention Improve the rate of CEA for eligible individuals following acute ischemic stroke
Acute treatment/rehabilitation	Increase % of acute stroke ED arrival within 60 minutes of the symptom onset Improve the use of thrombolytic therapy with tPA within the VA health care system Increase the % of ischemic stroke patients receiving proper deep venous thromboembolism prophylaxis Increase the % of acute ischemic stroke patients receiving dysphagia screening upon admission Increase % of eligible stroke patients receiving guideline-concordant rehabilitation services

AF, atrial fibrillation; CEA, carotid endarterectomy; DM, diabetes mellitus; ED, emergency department; SBP, systolic blood pressure (mm Hg); TIA, transient ischemic attack; tPA, tissue plasminogen activator; VA, Veterans Affairs.

poststroke v. simply improving functional independence. Structural uncertainty was represented and controlled through parameters.

**METHODS**

**Methodological Overview**

To inform strategic planning by the Stroke QUERI, we sought to clarify the impact of uncertainty on the robustness of simulation-based research and practice conclusions. Figure 2 provides an overview of our integrated approach to sensitivity analysis, calibration, and uncertainty analysis. We sought a systematic and replicable approach. Before proceeding, we will clarify the symbols and terms used throughout this article. For our stroke model  $f$ , let  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  denote the set of uncertain parameters (Suppl. Table S1). For simplicity, the output measures of the stroke model  $f$  corresponding to a set of values for the parameter set  $\Theta$  are denoted by  $Y$  and expressed as  $Y = f(\Theta)$ . We further define  $Y = \{S, P\}$ , in which  $S$  and  $P$  are mutually exclusive.

$S = \{s_1, s_2, s_3\}$  represents *stroke outcomes*, including cumulative QALYs ( $s_1$ ), cumulative strokes ( $s_2$ ), and cumulative stroke fatalities ( $s_3$ ), and  $P = \{p_1, p_2, \dots, p_{13}\}$  represents *calibration objectives*. Calibration objectives  $P$  are model outputs that quantify the consistency of the model's behavior with observations we sought to match. Let  $Z = \{\zeta_1, \zeta_2, \dots, \zeta_{13}\}$  denote the specific empirical outcomes that the model seeks to match (defined in Table 2). For the  $i$ th calibration objective, let  $q_{\zeta_i, t}$  denote the simulated value of  $\zeta_i$  at time  $t$ , and let  $z_{\zeta_i, t}$  represent the empirical value of  $\zeta_i$  at time  $t$ , that is,

$$p_i = \frac{\int_{t=0}^T \sqrt{\left(\frac{q_{\zeta_i, t} - z_{\zeta_i, t}}{z_{\zeta_i, t}}\right)^2} dt}{T}$$

In some cases, the empirical observation of  $\zeta_i$  is stable (e.g., percentage of the poststroke population with mRS 2–3 is understood to have been stable in the past 3-year window). In these cases,  $z_{\zeta_i, t} = z_{\zeta_i, 0}$ .

In step 1, we used Morris method<sup>28</sup> to segment the parameter set  $\Theta$  into 3 mutually exclusive subsets:  $\alpha$ ,

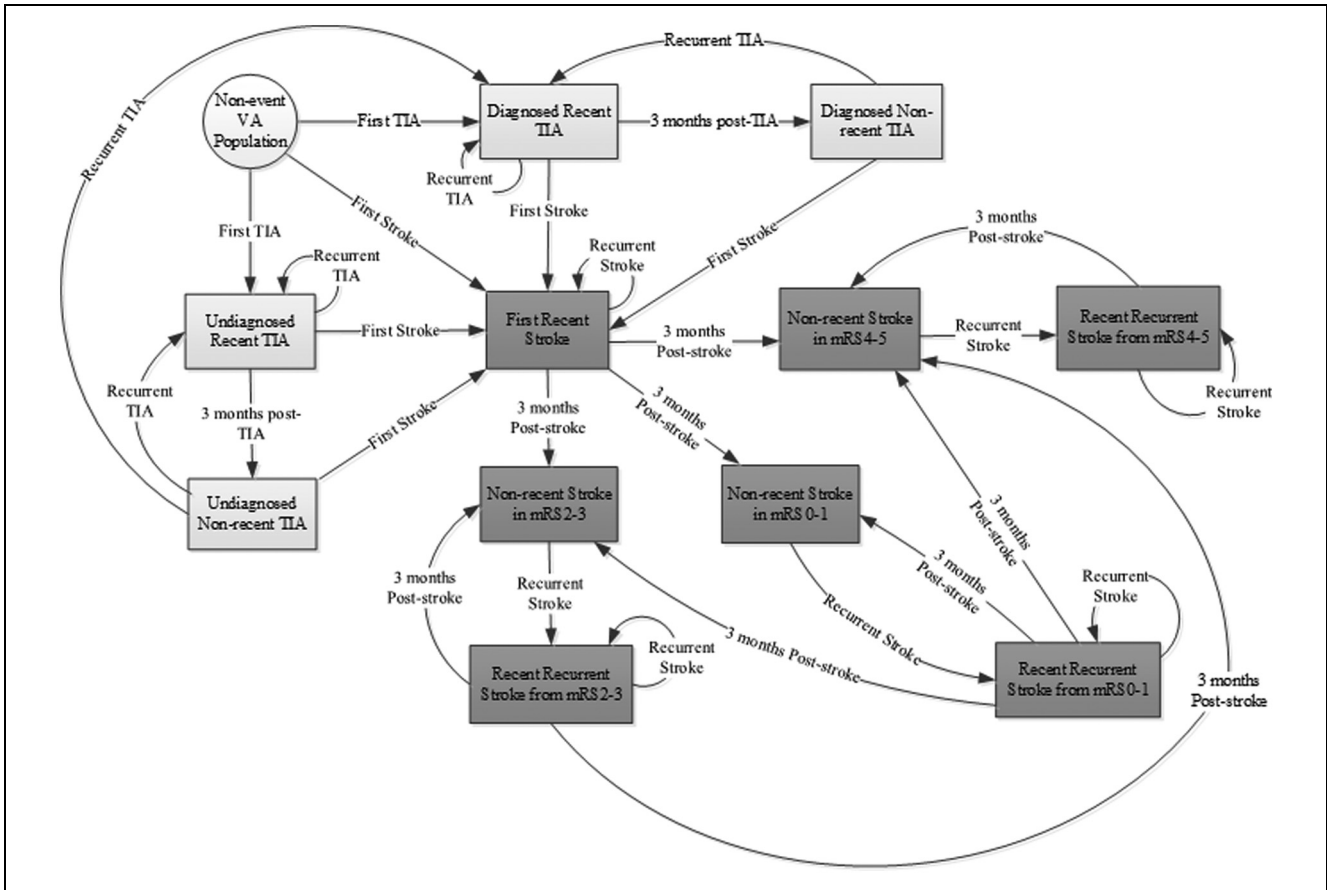


Figure 1 Diagram depicts stocks (rectangles) and flows (arrows) of the system dynamics stroke model for the Veteran Affairs (VA) population. Stocks and flows govern the states and changes in health status of veteran enrollee population over time. The flows capture the transitions between stocks that shift individuals from one to another over time. The Nonevent VA population, depicted as a circle, is implemented as time-series model inputs that influence the inflows. The stocks with light color indicate veterans with a history of transient ischemic attack (TIA) but not stroke, while the stocks with dark color represent veterans with a history of stroke. To simplify the diagram, each stock has a death outflow, which is not explicitly depicted. mRS, modified Rankin Scale.

$\beta$ , and  $\gamma$ , where  $\Theta = \{\alpha, \beta, \gamma\}$ .  $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_i\}$  includes parameters to which calibration objectives  $P$  were sensitive, regardless of their impact on stroke outcomes  $S$ .  $\beta = \{\beta_1, \beta_2, \dots, \beta_j\}$  includes parameters to which only stroke outcomes  $S$  were sensitive.  $\gamma = \{\gamma_1, \gamma_2, \dots, \gamma_k\}$  includes parameters to which neither calibration objectives  $P$  nor stroke outcomes  $S$  were sensitive. The Morris method is a global sensitivity analysis approach described in more detail below. In short, the goal is to achieve dimension reduction for calibration in step 2 as well as to understand the most substantial contributors to both parameter and structural uncertainty.

In step 2, parameter values in  $\alpha$  were estimated using generalized likelihood uncertainty estimation (GLUE) and the multistart Powell hill-climbing

optimization algorithm. Together, these methods were used to generate many well-calibrated parameter sets  $\hat{\alpha}^{(m)}$ , with  $m = 1, 2, \dots, M$ . Each  $\hat{\alpha}^{(m)}$  represents one plausible hypothesis, consistent with available historical data and observed trends. Parameter values in  $\beta$  were fixed at their best-guess values  $\bar{\beta}$  during step 2, as they did not substantively affect calibration. Parameter values in  $\gamma$  were fixed at their best-guess values  $\bar{\gamma}$  throughout the remainder of analysis.

In step 3, we conducted uncertainty analysis to estimate the relative impact of each of the 15 intervention approaches on stroke outcomes, compared with care as usual. We also sought to estimate the robustness of intervention rankings across iterations of the simulation. In this step, we introduced a new

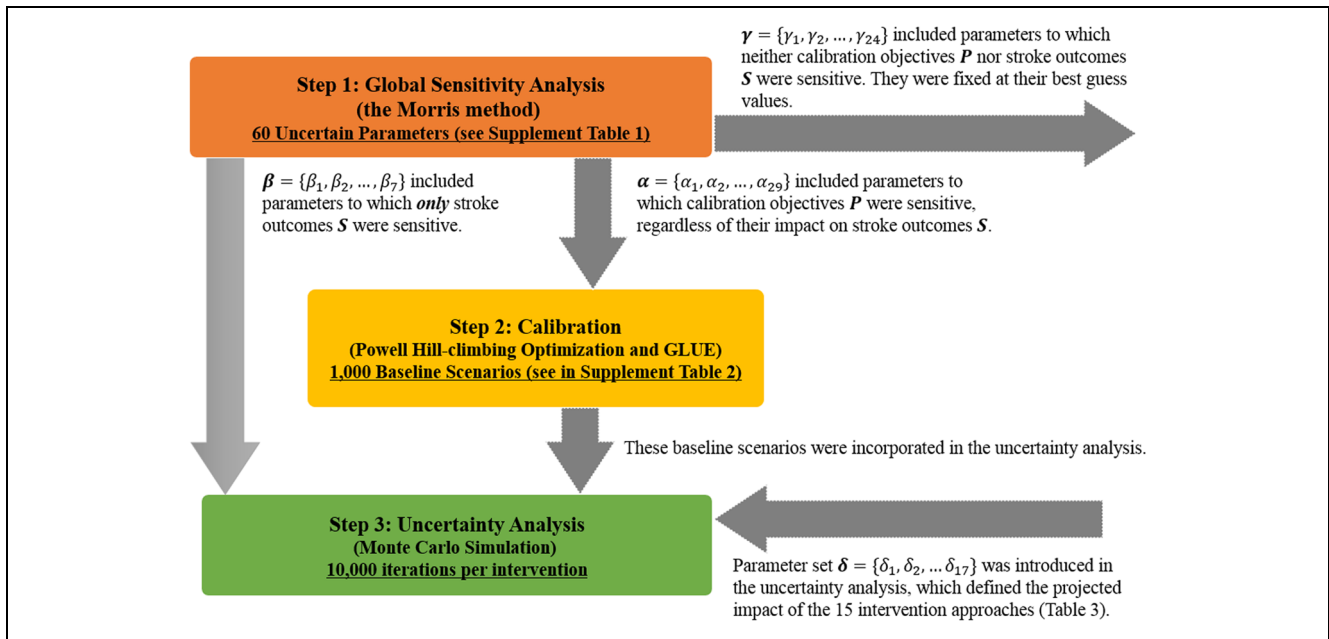


Figure 2 Flowchart depicting the integrated approach to sensitivity analysis, calibration, and uncertainty analysis. GLUE, generalized likelihood uncertainty estimation.

parameter set  $\delta = \{\delta_1, \delta_2, \dots, \delta_{17}\}$ , which defined the projected impact of the 15 intervention approaches (Table 3). Parameters in  $\delta$  do not affect calibration because these parameters are not relevant in simulating historical data. Uncertainty analysis was conducted through Monte Carlo simulation of the stroke model  $f(\hat{\alpha}^{(m)}, \tilde{\beta}, \tilde{\gamma}, \tilde{\delta})$ , in which  $\tilde{\beta}$  and  $\tilde{\delta}$  represent draws from predefined distributions for each  $\beta_i$  and  $\delta_j$  respectively.

### Global Sensitivity Analysis

The Morris method<sup>28</sup> is a global sensitivity analysis approach employing a factorial sampling strategy and consisting of randomized 1-factor-at-a-time iterations. Local 1-parameter changes are made at different points throughout the uncertain parameter space. As noted above, 60 parameters were identified,  $\Theta = \{\theta_1, \theta_2, \dots, \theta_{60}\}$ , and plausible ranges were specified for each in Supplemental Table S1.

Given our deterministic stroke model  $f$ , model outputs  $Y = \{y_1, y_2, \dots, y_N\}$  can be estimated precisely as  $f(Y)$ . Let  $\theta_i^{min}$  and  $\theta_i^{max}$  denote the minimum and maximum possible values for parameter  $\theta_i$ , respectively. We discretize  $\theta_i$  by defining  $\theta_i = \theta_i^{min} + l_i(\theta_i^{max} - \theta_i^{min})$ , where  $l_i$  is drawn with equal probability from the set  $\{0, \frac{1}{h-1}, \frac{2}{h-1}, \dots, 1\}$ ;  $h$  denotes the number of levels. The region for sampling

values for  $\Theta$  becomes a 60-dimensional  $h$ -level grid. Let  $y_n(\Theta)$  denote the model output  $y_n$  for a given set of  $\Theta$  values. Thus, we define the elementary effect of  $\theta_i$  on  $y_n$  as an incremental ratio:

$$d_{i,n}(\theta) = \frac{y_n(\theta_1, \theta_2, \dots, \theta_{i-1}, \theta_i + \Delta, \theta_{i+1}, \dots, \theta_{60}) - y_n(\theta_1, \theta_2, \dots, \theta_{60})}{\Delta}$$

where  $\Delta$  is defined as a multiple of  $1/(h-1)$ . The distribution of parameter  $\theta_i$ 's elementary effect on  $y_n$  is denoted  $F_{i,n}$ . We draw independent random samples from  $F_{i,n}$  to obtain  $r$  elementary effects.<sup>10</sup> Sensitivity measurements of the Morris method include the sample mean  $\mu$  and standard deviation  $\sigma$  of each distribution  $F_{i,n}$ , which can be estimated as

$$\mu = \frac{\sum_{i=1}^r d_{i,n}}{r}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^r (d_{i,n} - \mu)^2}{r}}$$

The measures  $\mu$  and  $\sigma$  are the mean and standard deviation of the incremental ratios of randomized sampled points of the parameter space, respectively. A higher mean  $\mu$  of the distribution  $F_{i,n}$  indicates

**Table 2** Empirical Observations and Uncertain Ranges, When Appropriate, Used to Calibrate the Stroke Model (**Z**)

Variable	Value	Data Source or Reference
Prevalence of diagnosed TIA in the VA user population	Stable	36
% post-TIA who have a recurrent TIA within 90 days	10% (9.5%–10.5%)	37
% post-TIA who have a stroke within 90 days	15% (13%–17%)	38, 39
% post-TIA who ever have a TIA after 90 days (prior to stroke or death)	14% (12%–16%)	40
% post-TIA who have a stroke after 90 days (prior to death)	7%/year (6.3%–7.7%)	41
Prevalence of stroke in the VA user population in 2010	Stable	36
% of first strokes preceded by a diagnosed TIA	11% (8%–15%)	37, 42
% poststroke who have a recurrent stroke within 90 days	5% (4%–6%)	37
% poststroke who ever have a recurrent stroke after 90 days (prior to death)	20% (18%–22%)	37
% of all strokes that are recurrent	23% (17.3%–28.8%)	37, 42
Average life expectancy poststroke, y	6.5 (6–7)	43
Average life expectancy post-TIA, y	Stable (7–10.54)	40, 43
% of the poststroke population with mRS 0–1	Stable	23, 44, 45
% of the poststroke population with mRS 2–3	Stable	23, 44, 45
% of the poststroke population with mRS 4–5	Stable	23, 44, 45

Ranges in parentheses specify a permitted deviation for each estimate. “Stable” is indicated if the parameter value at year 3 of the simulation fell within a range of  $\pm 10\%$  of the initial parameter value. mRS, modified Rankin scale; TIA, transient ischemic attack; VA, Veterans Affairs.

a larger elementary (or first-order) effect of parameter  $\theta_i$  on the output  $y_n$ . A larger standard deviation  $\sigma$  of  $F_{i,n}$  implies possible interaction effects in nonlinear dynamic models, in which the elementary effect of  $\theta_i$  is influenced by the value of other parameters in the parameter space.

Given our deterministic stroke model, the examination of change in output measurement  $d_{i,n}$  attributable to parameter  $\theta_i$  was unambiguous. We set  $h = 8$  and  $r = 10$ . Currently, there is no clear best practice guiding these choices. We chose 10 for  $r$  as some studies suggest that this provides the minimum sample size to gain confidence in experiment results.<sup>10</sup> Then, we chose 8 for the number of levels ( $h$ ) because it resulted in the largest number of discretized points in the space that could be analyzed in SimLab software.<sup>29</sup> SimLab is an easy-to-use model-free sensitivity and uncertainty analysis toolkit. For a  $k$ -dimensional  $h$ -level grid, estimating  $r$  elementary effects using the Morris method requires a sample size of  $r(k + 1)$ . For our decision problem, we ran 610 iterations of the stroke model in Vensim to complete the sensitivity analysis in SimLab. In contrast to the Morris method, a random sampling strategy would require a larger sample size ( $2rk$ ) to obtain  $r$  elementary effects.

The cutoff values for classifying uncertain parameter  $\theta_i$  as influential (“sensitive”) or noninfluential (“insensitive”) are subjective. For each parameter  $\theta_i$ , we classified it as influential to calibration

objectives if any  $\mu$  or  $\sigma$  value for any calibration objective  $p_i$  was greater than 10%. For each stroke outcome  $s_j$ , we sorted  $\mu$  and  $\sigma$  values separately and identified the parameters  $\theta_i$  corresponding to the 75th percentile of the values. The de-duplicated set of these parameters was classified as influential to stroke outcomes. At the end of this step, parameter set  $\Theta$  was segmented into 3 mutually exclusive subsets  $\alpha$ ,  $\beta$ , and  $\gamma$ , where  $\Theta = \{\alpha, \beta, \gamma\}$ .  $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_i\}$  represents parameters to which at least 1 calibration objective  $p_i$  was sensitive, regardless of its impact on stroke outcomes  $S$ .  $\beta = \{\beta_1, \beta_2, \dots, \beta_j\}$  represents parameters to which only stroke outcomes  $S$  were sensitive.  $\gamma$ , parameters to which neither calibration objectives  $P$  nor stroke outcomes  $S$  were sensitive, were identified and fixed at the best-guess value  $\bar{\gamma}$  throughout the analysis.

**Calibration, Step 1: Parameter Estimation via the Multistart Powell Hill-Climbing Optimization Algorithm**

Model calibration is an important step in which uncertain parameters for which data are not directly available are estimated based on additional data. The ability to reproduce historical trends with the model through calibration can help build confidence among decision makers in its credibility.<sup>30</sup> In this modeling effort, we had many uncertain parameters  $\alpha$  that were capable of affecting the model fit (i.e., as

**Table 3** Uncertain Model Parameters Affecting Future Impacts of Evaluated Stroke Intervention Approaches but Not Model Calibration ( $\delta$ )

Postintervention Parameters	Minimum	Maximum
Proportion of preintervention SBP 140 to 159 and nondiabetic patients moveable to under 140, accounting for their CVD status	0.45	0.75
Proportion of preintervention SBP 140 to 159 and diabetic patients moveable to under 140 who have diabetes, accounting for their CVD status	0.6	0.95
Proportion of preintervention SBP 160+ and nondiabetic patients moveable to under 140 who do not have diabetes, accounting for their CVD status	0.32	0.48
Proportion of preintervention SBP 160+ and diabetic patients moveable to under 140 who have diabetes, accounting for their CVD status	0.45	0.75
Postintervention proportion of AF patients with good anticoagulation management (assumed the same for those with and without CVD currently)	0.4	0.8
Postintervention diagnosed proportion of TIAs in existing VA users	0.81	0.85
Postintervention proportion of eligible population with diagnosed TIAs and strokes getting carotid endarterectomy	0.4	0.75
Overall postintervention proportion of diagnosed TIA population getting good preventive management	0.6125	0.8375
Postintervention proportion of stroke population that receives good secondary preventive management	0.5	0.7
Postintervention proportion of first strokes timely to hospital	0.275	0.5375
Postintervention proportion of recurrent strokes timely to hospital	0.2875	0.55
Postintervention proportion of eligible strokes getting tPA effectively	0.175	0.475
Postintervention proportion of eligible strokes getting tPA ineffectively	$4.25 \times 10^{-3}$	$5.75 \times 10^{-3}$
Postintervention proportion of ineligible strokes getting tPA	$2 \times 10^{-4}$	$8.75 \times 10^{-4}$
Postintervention proportion of eligible strokes getting DVT prophylaxis	0.8075	0.935
Postintervention proportion of eligible stroke patients getting dysphagia screening	0.5	0.875
Overall postintervention proportion of eligible stroke patients getting good rehabilitation	0.55	0.725

Estimates are based on workgroup consensus. AF, atrial fibrillation; CVD, cardiovascular disease; DVT, deep vein thrombosis; SBP, systolic blood pressure (mm Hg); TIA, transient ischemic stroke; tPA, tissue plasminogen activator; VA, Veterans Affairs.

measured with calibration objectives  $\mathbf{P}$ ), reinforcing the importance of accounting for uncertainty associated with calibration.

Given the complex nonlinear relationship between the parameters and model outcomes in the stroke model, it was too complicated to derive the analytical solution. Instead, we turned to numerical solution by employing the multistart Powell hill-climbing optimization method<sup>31</sup> embedded in Vensim software to search through the parameter space. Independent samples were drawn from predefined ranges for each uncertain and sensitive parameter  $\alpha_i$ . The multistart option was used to increase the chances that a global optimum was identified. The aggregated calibration objective in this step was a payoff function, which measured the weighted sum of relative discrepancies of the simulated value of empirical outcome  $\zeta_i$  and the observed (or target) value of  $\zeta_i$  (Table 2). Then the aggregated payoff function to be minimized was calculated as

$$v = \sum_{i=1}^N \int_{t=0}^T C_{i,t} \left( \frac{q_{\zeta_i,t} - z_{\zeta_i,t}}{(q_{\zeta_i,t} + z_{\zeta_i,t})/2} \right)^2 dt.$$

Penalty  $C_{i,t}$  was assigned a very large value in the payoff function if the simulated outcome fell outside of the assigned target range in the calibration; otherwise,  $C_{i,t}$  was set to 1. The aggregated payoff function was dimensionless and assigned an equal weight to each calibration objective (Table 2) in this project, although other weights could be set if calibration objectives are not equally important. As a result of the calibration, the optimal estimates of parameter set  $\alpha$  (denoted as  $\hat{\alpha}$ ) that minimized the payoff function were identified.

### Calibration, Step 2: Multiple Calibrations via GLUE

The single optimal estimates  $\hat{\alpha}$  would only represent one reasonable baseline scenario. Although

this baseline scenario is consistent with available data, it is likely that  $\hat{\alpha}$  might behave poorly with additional data. Therefore, we sought to identify alternative estimates of parameter set  $\alpha$  (denoted as  $\hat{\alpha}^{(m)}$  with  $m = 1, 2, \dots, \bar{M}$ ) that are nearly as consistent with available data. Doing so allowed calibrated parameter sets to enter uncertainty analysis, since varying parameter values that stroke outcomes were sensitive to outside of the calibration process could (and, in our experience, did) result in model scenarios no longer consistent with calibration objectives.

We identified GLUE as a method for securing alternate baseline scenarios ( $\hat{\alpha}^{(m)}$ ) that were also quite consistent with observed stroke patterns in the VA (i.e., well-calibrated parameter sets) and for estimating the relative likelihood of each. GLUE<sup>32,33</sup> has been used extensively in the evaluation of prediction uncertainty in hydrological and environmental models,<sup>33,34,35</sup> which exhibit a similar degree of complexity as health care policies. GLUE is a Monte Carlo (MC)-based approach coupled with informal Bayesian analysis. It is based on the concept of equi-finality, which argues that we should not expect a unique global optimum parameter set within a particular model structure, and thus alternative parameter sets that fit model predictions reasonably well should be recognized and evaluated. Within a pseudo-Bayesian MC framework, GLUE performs a large number of iterations of the model. It incorporates numerous sampled parameter values drawn from their prior probability distributions and uses a likelihood function to estimate the likelihood of each parameter set.

Many sampling strategies can be coupled with GLUE, such as Latin hypercube or random sampling. We initially applied Latin hypercube sampling with GLUE; however, we were unable to obtain even one plausible parameter set  $\hat{\alpha}^{(m)}$  after millions of iterations given the high-dimensional parameter space. Integrating the multistart linear Powell hill-climbing optimizer with GLUE dramatically increased efficiency by reducing the time and number of model iterations needed to obtain enough plausible baseline estimates. Each run of the optimizer may produce millions of iterations, with the number varying dramatically from run to run. We ran the optimizer 20 times with different random number seeds to increase the chances of identifying heterogeneous well-calibrated parameter sets  $\hat{\alpha}^{(m)}$ . All iterations conducted across all runs of the optimizer (approximately 10,000,000 in total) were retained for GLUE.

Let  $\tilde{\alpha}$  denote the sample values of  $\alpha$  for each iteration. The likelihood of each  $\tilde{\alpha}$  was operationalized as the reciprocal of the payoff value  $v$ . The higher the

likelihood, the smaller the payoff value. We chose to define the 1000 most likely parameter sets  $\tilde{\alpha}$  as the plausible ones ( $\hat{\alpha}^{(m)}$  with  $m = 1, 2, \dots, 1000$ ), which we used as alternate baseline scenarios in the uncertainty analysis (next step). The choice of 1000 alternate baseline scenarios is an arbitrary decision made by the modelers, as GLUE requires subjective decisions on the cutoff value separating plausible from implausible estimates of the parameter set. The choice was made looking at the resulting range of likelihood values and based on our desire to include many (i.e., far more than one) estimated parameter sets in subsequent uncertainty analysis in light of the extent of uncertainty in calibration. Alternate acceptable approaches to distinguish plausible parameter sets include using a likelihood threshold value or a predetermined percentage of total iterations. The implausible parameter set estimates were removed from further analysis.

### Uncertainty Analysis

An important goal of analysis was to better understand the relative gains in stroke outcomes  $S$  under each of the 15 stroke intervention approaches considered by the Stroke QUERI, one at a time, compared with care as usual. The QUERI executive committee also sought to understand the effort that would be required to implement each intervention, in terms of the number of veterans who would need to be “touched” with the intervention to save 1 QALY. Referred to as the number needed to treat (NNT) per QALY, this measure served as a surrogate for cost given uncertainty at this stage of specific interventions (only intervention targets or approaches) to further inform strategic planning.<sup>15</sup> Understanding the effect of underlying uncertainty within the model on key outcomes was recognized to be critical for Stroke QUERI decision makers, both because they wanted to know if the ranking of interventions was robust and because they wanted to know which uncertain parameters should be prioritized for further data collection.

To complete uncertainty analysis, information from earlier steps needed to be integrated to construct a set of parameter values that would reflect the composite effect of the recognized sources of uncertainty on stroke outcomes. Again, this uncertainty relates to 4 key sets of model parameters:  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$ . Parameter set  $\delta$  defined the projected impact of the 15 intervention approaches (Table 3). We integrated sources of uncertainty by conducting more conventional Monte Carlo simulation on the stroke model

$\mathbf{S} = f(\hat{\boldsymbol{\alpha}}^{(m)}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\delta}})$  with  $m = 1, 2, \dots, 1000$ , in which  $\tilde{\boldsymbol{\beta}}$  and  $\tilde{\boldsymbol{\delta}}$  represent draws from predefined distributions for each  $\beta_i$  and  $\delta_j$ , respectively. The objectives of uncertainty analysis were to 1) estimate the distribution around each projected change in stroke outcome  $s_i$  (with  $i = 1, 2, 3$ ) under each intervention approach, compared with care as usual, and 2) assess the robustness of intervention approach rankings compared with each other within each iteration of the simulation. Uniform distributions were assumed for parameters in  $\boldsymbol{\beta}$  and  $\boldsymbol{\delta}$ , in the absence of more information. Latin hypercube sampling was applied to draw 10 random samples of  $\boldsymbol{\beta}$  and  $\boldsymbol{\delta}$  for each of the 1000 baseline scenarios, resulting in 10,000 iterations per intervention and the care-as-usual scenario. Absolute and relative intervention impacts were documented within and ultimately across iterations for each intervention approach.

## RESULTS

Sensitivity analysis identified 29 (of the 60) uncertain parameters ( $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \dots, \alpha_{29}\}$ ) that were influential to at least 1 calibration criterion (Figure 2); some of these also influenced stroke outcomes (Suppl. Figure S1). As a result, we reduced the number of parameters that needed to be calibrated from 60 to 29. Seven additional parameters ( $\boldsymbol{\beta} = \{\beta_1, \beta_2, \dots, \beta_7\}$ ) were identified as having a significant impact on 1 or more stroke outcomes  $\mathbf{S}$  (Figure 2) but did not need to be calibrated. Sensitivity measures for each individual calibration objective, cumulative QALYs, cumulative strokes, and cumulative stroke fatalities can be found in Supplemental Figures S1 and S2. Notably, the stroke incidence rate in the pre-event VA population was found to be the most influential uncertain parameter across all outputs studied. As the largest source of outcome uncertainty, further empirical research should be directed toward precise estimation of it to reduce uncertainty in stroke outcome projections.

The optimal calibrated parameter estimates and the descriptive statistics for alternate baseline scenarios obtained via GLUE are presented in Supplemental Table S2. These well-calibrated parameter sets  $\hat{\boldsymbol{\alpha}}^{(m)}$  were heterogeneous. Not all 1000 plausible parameter sets/alternate baselines are shown but are available upon request.

Figure 3 presents comprehensive uncertainty analysis results for the effect of each intervention compared with care as usual. Interventions are presented in descending order of mean impact, while

histograms represent variation in estimates across the 10,000 Monte Carlo iterations.

Mean and 95% uncertainty bounds of cumulative, undiscounted stroke outcome measures and NNT for each intervention over 5 years are presented in Table 4. The intervention approach involving improving hypertension control for all VA users yielded the largest population-level gains in QALYs (mean = 5162), strokes prevented (mean = 4639), and stroke fatalities prevented (mean = 537). Use of tissue plasminogen activator (tPA), a protein involved in the breakdown of protein, only generated 153 QALYs gained, although it ranked first across all assessed intervention approaches in terms of NNT per QALY gained (5.9). This is largely due to a small number of eligible patients, given current context. The scatter plot of the incremental QALYs gained v. the incremental NNT per stroke interventions is available at <http://vastrokemodel.weebly.com/>.

Figure 4 presents a scatter plot reporting the ranking of each intervention approach in terms of cumulative 5-year QALYs gained within each of the 10,000 Monte Carlo iterations. Interventions are presented from most to least impactful in terms of mean QALYs gained along the x-axis. For each intervention approach, a dot is placed along the y-axis to indicate the ranking of that intervention in each of the 10,000 iterations. Greater density in a given ranking on the y-axis indicates greater robustness of the approach's ranking, while greater dispersion indicates heterogeneity (uncertainty) in rankings.

## DISCUSSION

System Dynamics models are frequently used to support broad strategic planning in the face of complex systems problems, and they often involve extensive uncertainty. As such, careful sensitivity analysis, calibration, and uncertainty analysis are critical, and they can be challenging to design. Moreover, they can also be computationally expensive to implement and evaluate given a high-dimensional uncertain parameter space. Without these analyses, decision makers are often left feeling unsure of the robustness of deterministic results. In response to this challenge, we present a rigorous and analytically grounded process for integrating sensitivity analysis, calibration, and uncertainty analysis. This process enabled us to understand the impact of parameter and structural uncertainty on calibration and outcome prediction, as well as assess the robustness of results. Not only do we illustrate how Monte Carlo

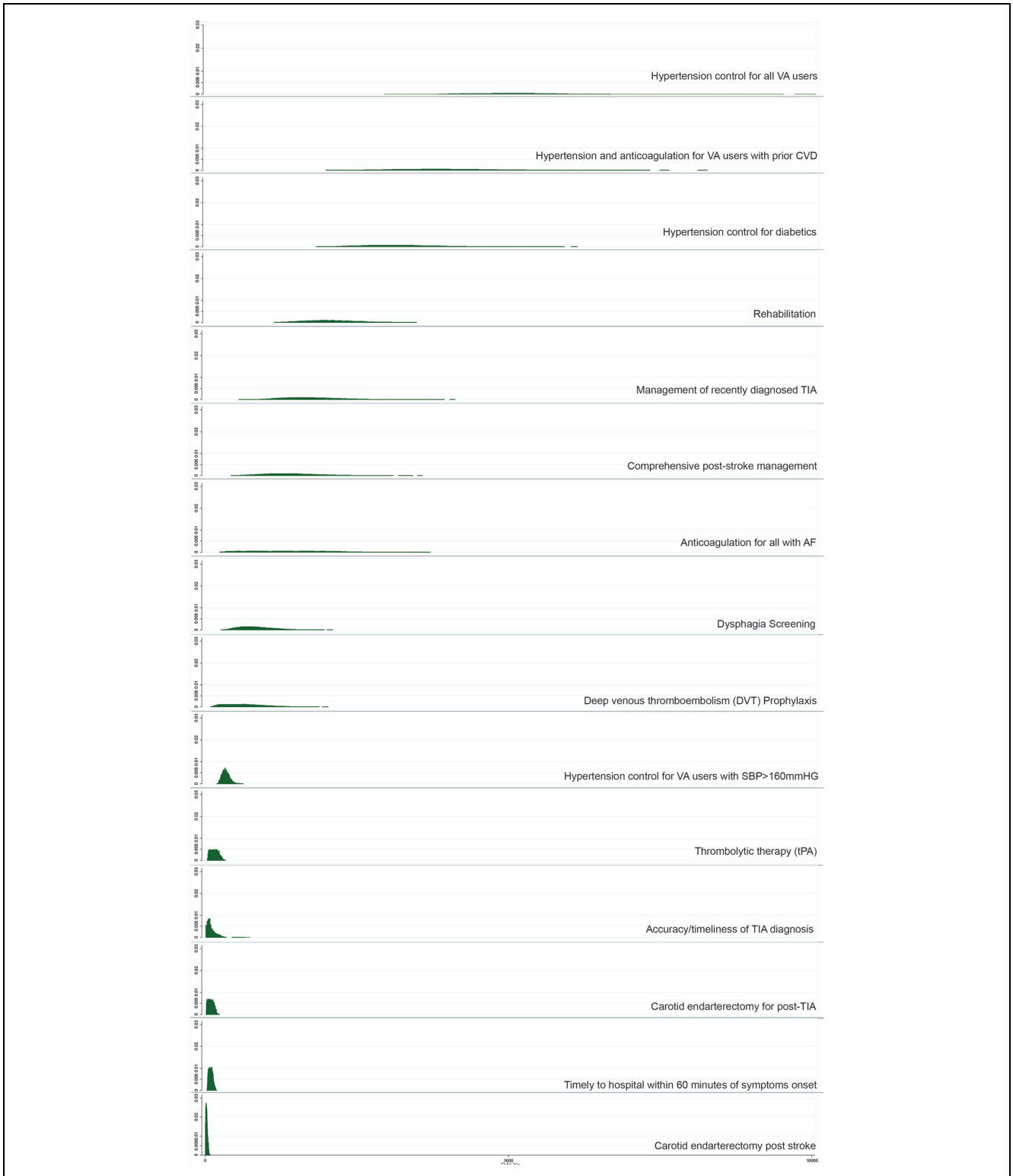


Figure 3 Histograms depicting the distribution around quality-adjusted life years gained per intervention approach simulated. AF, atrial fibrillation; CVD, cardiovascular disease; SBP, systolic blood pressure (mm Hg); TIA, transient ischemic stroke; tPA, tissue plasminogen activator; VA, Veterans Affairs.

**Table 4** Mean (95% Uncertainty Intervals) for Key Stroke Outcomes for Each Stroke Intervention Approach Compared with the Current Levels of Care over 5 Years

Intervention Name	QALYs Gained	Strokes Prevented	Stroke Fatalities Prevented	NNT per QALY Gained
Hypertension control for all VA users <sup>PP</sup>	5162 (3438 to 6555)	4639 (3438 to 6555)	537 (333 to 867)	82 (56.7 to 103.5)
Hypertension control and anticoagulation for those with prior CVD <sup>PP</sup>	3896 (2581 to 5916)	3507 (2366 to 5224)	406 (232 to 681)	34.8 (22.3 to 48.8)
Hypertension control for diabetes patients <sup>PP</sup>	3163 (2212 to 4721)	2850 (2009 to 4144)	330 (200 to 542)	62.1 (43.4 to 78.1)
Rehabilitation <sup>TR/R</sup>	2049 (1379 to 3005)	29 (2 to 76)	9 (2 to 18)	9.2 (6.6 to 11.3)
Management of recently diagnosed TIA <sup>SP</sup>	1734 (989 to 2741)	1541 (977 to 2312)	179 (96 to 295)	9.6 (7.2 to 12.5)
Comprehensive poststroke management <sup>SP</sup>	1405 (698 to 2445)	2630 (1560 to 4040)	347 (188 to 575)	20.5 (14.1 to 29.3)
Anticoagulation for all with AF <sup>PP</sup>	1390 (362 to 2754)	1249 (336 to 2482)	145 (36 to 314)	56.2 (39.3 to 70.6)
Dysphagia screening <sup>TR/R</sup>	830 (409 to 1292)	—	169 (87 to 293)	56.3 (38.1 to 82.4)
DVT prophylaxis <sup>TR/R</sup>	652 (182 to 1402)	—	132 (38 to 274)	13.5 (9.2 to 19.8)
Hypertension control for VA users with SBP >160 <sup>PP</sup>	338 (241 to 493)	304 (218 to 432)	35 (21 to 56)	39.4 (27.6 to 49.7)
Thrombolytic therapy <sup>TR/R</sup>	153 (52 to 279)	1 (−1 to 3)	8 (2 to 16)	5.9 (2 to 8.4)
Accuracy/timeliness of TIA diagnosis <sup>SP</sup>	99 (22 to 303)	90 (21 to 264)	10 (2 to 35)	21.6 (16.7 to 27.9)
CEA for post-TIA <sup>SP</sup>	97 (25 to 185)	88 (23 to 159)	10 (2 to 21)	18.6 (14.5 to 23.5)
Timely to hospital within 60 minutes of symptom onset <sup>TR/R</sup>	95 (45 to 160)	0 (0 to 2)	5 (2 to 9)	238 (167.5 to 307.6)
CEA for poststroke <sup>SP</sup>	26 (6 to 59)	65 (17 to 123)	8 (2 to 16)	133.4 (87.8 to 210.7)

AF, atrial fibrillation; CEA, carotid endarterectomy; CVD, cardiovascular disease; DVT, deep vein thrombosis; NNT, numbers needed to treat; PP, primary prevention; QALY, quality-adjusted life years; SBP, systolic blood pressure (mm Hg); SP, secondary prevention; TIA, transient ischemic attack; TR/R, treatment/rehabilitation; VA, Veterans Affairs; —, Nil.

simulations can be used to quantify uncertainty in intervention effects (Table 4 and Figure 3), but we also present an approach to study the relative effects of interventions compared with each other (Figure 4). Estimating intervention effects and uncertainty intervals is important for informing intervention prioritization and quantifying potential variation in outcomes. It is critical to note that in situations of substantial uncertainty, intervals often overlap as was the case in this research. Visualizing outcome distributions (Figure 3) can help guide decision making by clarifying the probability of various effect levels.

There may be considerable epistemic uncertainty regarding model output on account of parameters' values or uncertainty in structural assumptions. However, if altering model assumptions does not change the relative desirability of possible policies or interventions, then it may have limited impact on decision making. To this end, we compared interventions to each other under a number of distinct but feasible scenarios, where uncertain parameters take on

plausible values and the model is well calibrated to data. We can learn much about the relative impact of each intervention despite substantial uncertainty. We can also learn if/what more data are needed to improve model-based insights. For example, referring to Figure 4, this analysis clearly illustrates the robust dominance of the broad hypertension intervention in terms of its impact on QALYs gained. After that, while interventions do not consistently dominate each other, they do fall into clusters. For example, 2 targeted prevention approaches (one for veterans with a history of CVD and another for diabetes) are most impactful. This level of insight may be sufficient to support decision making, which was the case for our illustrative decision problem. If better distinctions within a cluster are required or if stroke incidence/prevalence estimates are needed, the model can be used to identify which uncertain parameter contributes most to that specific uncertainty.

Overall, prioritizing impactful model parameters 1) simplifies the model structure by fixing (or

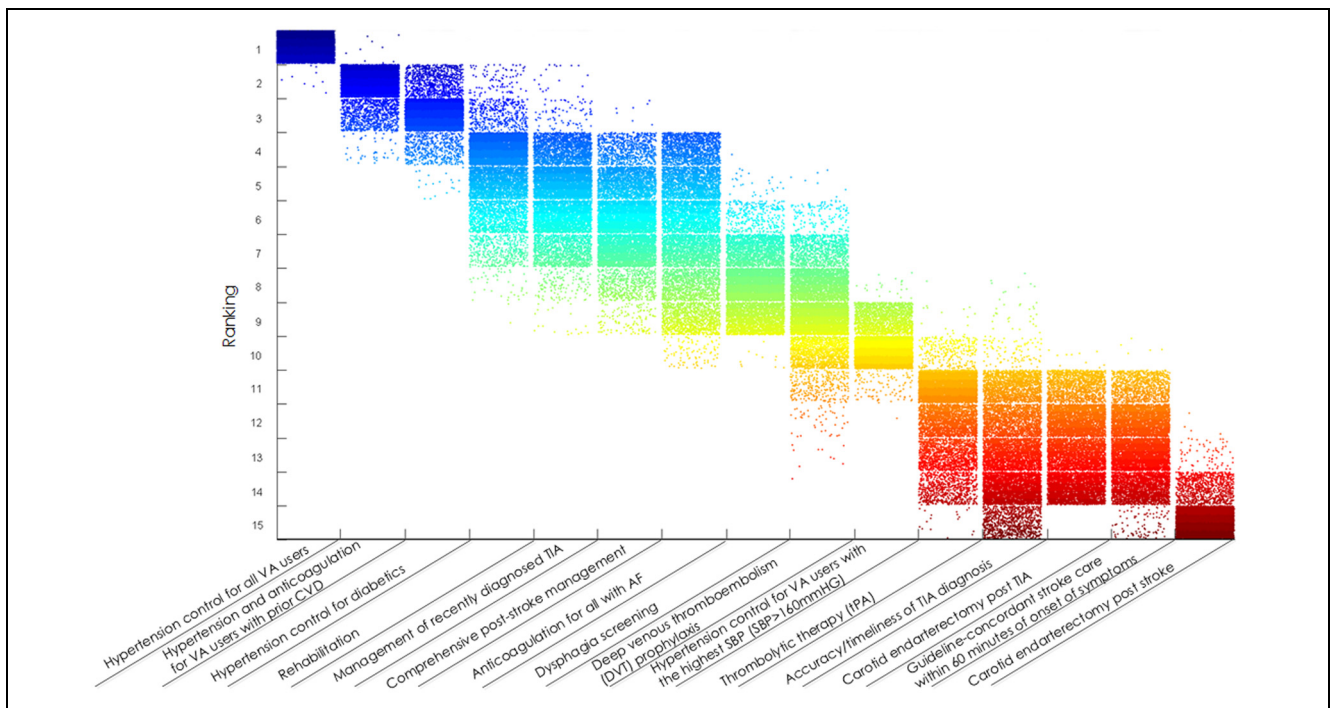


Figure 4 Ranking of policy options across iterations of the model, based on impact of quality-adjusted life years (QALYs). The smaller the ranking value, the larger the QALYs gained. AF, atrial fibrillation; CVD, cardiovascular disease; SBP, systolic blood pressure (mm Hg); TIA, transient ischemic stroke; tPA, tissue plasminogen activator; VA, Veterans Affairs.

eliminating) the less relevant sources of uncertainty, 2) allows a more judicious calibration procedure, 3) identifies the parameters that could be important leverage points for the interventions (or for future vulnerabilities), and 4) informs and prioritizes future data collection. In addition, improving the estimates of the influential factors can further increase the confidence in and credibility of the model predictions.

A number of techniques are capable of conducting sensitivity analyses to achieve such model parameter reduction, and some of these (e.g., univariate, Monte Carlo simulation) are commonly used in health research.<sup>5</sup> A more recent recommendation in health, it is also important to use a *global* sensitivity analysis approach, which explores the *entire* parameter space rather than assessing the impact of variation in one parameter at a time.<sup>11</sup> Such an approach will also investigate nonlinear relationships and possible interaction effects between uncertain parameters. However, when conducting global sensitivity analysis, thoughtful sampling of the parameter space can be critical to reaching accurate conclusions efficiently. We demonstrate one such efficient sampling strategy, the Morris method. With our illustrative decision problem, Morris increased efficiency somewhat with

610 samples compared with the 1200 required for random sampling. This gain in efficiency becomes much more substantial with a larger number of uncertain parameters to screen. Efficiency in identifying an appropriate number of well-calibrated parameter sets is also important. In our experience, using Latin hypercube sampling with GLUE was incredibly inefficient, while the multistart Powell hill-climbing optimization algorithm identified a set of 1000 well-calibrated but heterogeneous parameter sets within approximately 10,000,000 iterations.

As with any simulation-based research, simplified assumptions must be made. First, a number of decisions in our approach are subjective: the values chosen to segment the parameter set with Morris methods, the thresholds for distinguishing sensitive from insensitive uncertain parameters, the number of sampled parameter sets to include as “plausible” with GLUE, how calibration objectives are weighted/aggregated in a single likelihood function for GLUE, the subjective values in the sensitivity analysis, and how many random draws to include for sampled parameters in uncertainty analysis. Reasonable decisions can be made based on modeling objectives and intermediate results (in part). However, future research

is needed to inform these decisions. Second, while other options are available, we selected the Morris method as an efficient sampling strategy for sensitivity analysis since it requires a relatively small number of draws to reach sensitivity conclusions. Moreover, it is not dependent on any specific assumptions and is model free. The relative efficiency and results using other global sensitivity analysis methods should be further studied. Third, we chose to assume uniform distributions to characterize uncertainty in noncalibrated uncertain parameters varied in the Monte Carlo simulations. This decision seemed prudent given the paucity of additional data. However, we acknowledge that with better data, more specific distributions (e.g., log-normal, gamma) could be adopted to better represent uncertainty in these parameters.

In summary, rather than focusing on the stroke intervention recommendations for the Stroke QUERI,<sup>15</sup> this article demonstrates the practical application of an integrated and efficient approach for assessing sensitivity of key stroke outcomes to uncertain model parameters, calibrating sensitive parameters to maximize model fit to available data, and conducting uncertainty analysis around stroke outcomes and relative intervention effects. We advocate systematic exploration of uncertainty surrounding the intervention outcome and its impact on the corresponding decision to convey to decision makers the confidence in the model projection and its robustness to different but plausible assumptions about uncertain parameters. For the decision problem illustrated, decision makers' questions about the relative impact of alternate intervention approaches for strategic planning could be well informed, despite uncertainty in the model.

## ACKNOWLEDGMENTS

The authors appreciate the support of Duke-NUS/SingHealth Academic Medicine Research Institute and the medical editing assistance of Taara Madhavan (Associate, Clinical Sciences, Duke-NUS Medical School).

## REFERENCES

- Morrissey JP, Lich KH, Price RA, Mandelblatt J. Computational modeling and multilevel cancer control interventions. *J Natl Cancer Inst Monogr*. 2012;44:56–66.
- Osgood N, Liu J. Bayesian parameter estimation of system dynamics models using Markov chain Monte Carlo methods: an informal introduction. Verbal and poster presentation at the 30th International conference of the System Dynamics Society; 22 July 2013; Cambridge, MA.
- Matchar DB, Samsa GP, Matthews JR, et al. The stroke prevention policy model: linking evidence and clinical decisions. *Ann Intern Med*. 1997;127(8, Pt 2):704–11.
- Homer JB, Hirsch GB. System dynamics modeling for public health: background and opportunities. *Am J Public Health*. 2006;96(3):452–8.
- Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group–6. *Med Decis Making*. 2012;32:722–32.
- Caro JJ, Möller J. Decision-analytic models: current methodological challenges. *Pharmacoeconomics*. 2014;32:943–50.
- Caro JJ, Briggs AH, Siebert U, Kuntz KM; ISPOR-SMDM Modeling Good Research Practices Task Force. Modeling good research practices—overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–1. *Med Decis Making*. 2012;32:667–77.
- Helton JC, Johnson JD, Sallaberry CJ, Storlie CB. Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliab Eng Syst Safe*. 2006;91(10–11):1175–209.
- Saltelli A, Ratto M, Tarantola S, Campolongo F. Sensitivity analysis for chemical models. *Chem Rev*. 2012;112(5):PR1–21.
- Saltelli A, Tarantola S, Campolongo F, Rat M. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. New York: John Wiley; 2004.
- Iooss B, Lemaitre P. A review on global sensitivity analysis methods. In: Meloni C, Dellino G, eds. *Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications*. New York: Springer; 2015.
- McKinley TJ, Ross JV, Deardon R, Cook AR. Simulation-based Bayesian inference for epidemic models. *Comp Stat Data Anal*. 2013;71:434–47.
- Ratto M, Tarantola S, Saltelli A. Sensitivity analysis in model calibration: GSA-GLUE approach. *Comput Phys Commun*. 2001;136(3):212–24.
- Dogan G. Bootstrapping for confidence interval estimation and hypothesis testing for parameters of system dynamics models. *Syst Dynam Rev*. 2007;23(4):415–36.
- Hassmiller-Lich K, Tian Y, Beadles C, et al. Strategic planning to reduce the burden of stroke among veterans: using simulation modeling to inform decision making. *Stroke*. 2014;45(7):2078–84.
- Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics—2013 update: a report from the American Heart Association. *Circulation*. 2013;127(1):e6–245.
- QUERI. *Stroke QUERI Fact Sheet*. Stroke Quality Enhancement Research Initiative; 2013; Indianapolis, IN.
- Shen YJ, Findley PA, Maney M, et al. Department of Veterans Affairs—Medicare dual beneficiaries with stroke: where do they get care? *J Rehab Res Dev*. 2008;45(1):43–51.
- Chumbler NR, Jia HG, Phipps MS, et al. Does inpatient quality of care differ by age among US veterans with ischemic stroke? *J Stroke Cerebrovasc Dis*. 2012;21(8):844–51.
- D'Agostino RB, Wolf PA, Belanger AJ, Kannel WB. Stroke risk profile: adjustment for antihypertensive medication: the Framingham study. *Stroke*. 1994;25(1):40–3.

21. Ekundayo OJ, Vassar SD, Williams LS, Bravata DM, Cheng EM. Using administrative databases to calculate Framingham scores within a large health care organization. *Stroke*. 2011;42(7):1982–7.
22. Reker DM, Reid K, Duncan PW, et al. Development of an integrated stroke outcomes database within Veterans Health Administration. *J Rehab Res Dev*. 2005;42(1):77–91.
23. Kwon S, Hartzema AG, Duncan PW, Lai SM. Disability measures in stroke: relationship among the Barthel index, the functional independence measure, and the modified Rankin Scale. *Stroke*. 2004;35(4):918–23.
24. Vogel WB, Rittman M, Bradshaw P, et al. Outcomes from stroke rehabilitation in Veterans Affairs rehabilitation units: detecting and correcting for selection bias. *J Reh Res Dev*. 2002;39(3):367–83.
25. US Department of Veterans Affairs. Veteran population [updated 25 June 2013; cited 1 September 2013]. Available from: [http://www.va.gov/vetdata/veteran\\_population.asp](http://www.va.gov/vetdata/veteran_population.asp).
26. Bravata D, Ordin D, Vogel B, Williams L. The Quality of VA Inpatient Ischemic Stroke Care, FY2007: Final National and Medical Center Results of the VHA Office of Quality and Performance (OQP) Special Study. Washington, DC: US Department of Veterans Affairs; 2009.
27. Ventana Systems. Vensim DSS Software 2011 [cited 1 January 2013]. Available from: <http://vensim.com/>.
28. Morris MD. Factorial sampling plans for preliminary computational experiments. *Technometrics*. 1991;33(2):161–74.
29. SimLab. Software package for uncertainty and sensitivity analysis [cited 3 July 2013]. Available from: <http://simlab.jrc.ec.europa.eu/2011>.
30. Stahl JE. Trust and recognition: coming to terms with models. *Med Decis Making*. 2015;35(2):136–8.
31. Press WH, Teukolsky SA, Vetterling WT. *Numerical Recipes in C: The Art of Scientific Computing*. 2nd ed. New York: Cambridge University Press; 1992.
32. Beven K, Binley A. The future of distributed models—model calibration and uncertainty prediction. *Hydrol Process*. 1992; 6(3):279–98.
33. Beven KJ, Freer J. Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology. *J Hydrol*. 2001; 249:11–29.
34. Muleta MK, Nicklow JW. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *J Hydrol*. 2005;306:127–45.
35. Blasone R, Vrugt JA, Madsen H, Rosbjerg D, Robinson BA, Zyvoloski GA. Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov chain Monte Carlo sampling. *Adv Water Resour*. 2008;31:630–48.
36. Cornell JE, Pugh JA, Williams JW Jr, et al. Multimorbidity clusters: clustering binary data from multimorbidity clusters: clustering binary data from a large administrative medical database. *Appl Multivar Res*. 2009;12(3):163–82.
37. Roger VL, Go AS, Lloyd-Jones DM, et al. Heart disease and stroke statistics—2012 update: a report from the American Heart Association. *Circulation*. 2012;125(1):e2–220.
38. Giles MF, Rothwell PM. Risk of stroke early after transient ischaemic attack: a systematic review and meta-analysis. *Lancet Neurol*. 2007;6(12):1063–72.
39. Wu CM, McLaughlin K, Lorenzetti DL, Hill MD, Manns BJ, Ghali WA. Early risk of stroke after transient ischemic attack: a systematic review and meta-analysis. *Arch Intern Med*. 2007;167(22): 2417–22.
40. van Wijk I, Kappelle LJ, van Gijn J, et al. Long-term survival and vascular event risk after transient ischaemic attack or minor ischaemic stroke: a cohort study. *Lancet*. 2005;365(9477): 2098–104.
41. Giles MF, Rothwell PM. Prediction and prevention of stroke after transient ischemic attack in the short and long term. *Expert Rev Neurother*. 2006;6(3):381–95.
42. Chumbler NR, Jia H, Phipps MS, et al. Does inpatient quality of care differ by age among US veterans with ischemic stroke? *J Stroke Cerebrovasc Dis*. 2012;21(8):844–51.
43. Roger VL, Go AS, Lloyd-Jones DM, et al. Heart disease and stroke statistics—2011 update: a report from the American Heart Association. *Circulation*. 2011;123(4):e18–209.
44. Reker DM, Reid K, Duncan PW, et al. Development of an integrated stroke outcomes database within Veterans Health Administration. *J Rehab Res Dev*. 2005;42(1):77–91.
45. Amarenco P, Benavente O, Goldstein LB, et al. Results of the Stroke Prevention by Aggressive Reduction in Cholesterol Levels (SPARCL) trial by stroke subtypes. *Stroke*. 2009;40(4):1405–9.