IMPACT: A Generic Tool for Modelling and Simulating Public Health Policy

J. D. Ainsworth1; E. Carruthers1; P. Couch1; N. Green1; M. O’Flaherty2; M. Sperrin3; R. Williams1; Z. Asghar1; S. Capewell1; I. E. Buchan1
1North-west Institute for Bio-Health Informatics, School of Community Based Medicine, University of Manchester, UK; 2Division of Public Health, University of Liverpool, UK; 3Department of Mathematics and Statistics, Lancaster University, UK

Keywords
Discrete event simulation, decision support, public health, policy modelling, computer simulation

Summary
Background: Populations are under-served by local health policies and management of resources. This partly reflects a lack of realistically complex models to enable appraisal of a wide range of potential options. Rising computing power coupled with advances in machine learning and healthcare information now enables such models to be constructed and executed. However, such models are not generally accessible to public health practitioners who often lack the requisite technical knowledge or skills.

Objectives: To design and develop a system for creating, executing and analysing the results of simulated public health and health-care policy interventions, in ways that are accessible and usable by modellers and policymakers.

Methods: The system requirements were captured and analysed in parallel with the statistical method development for the simulation engine. From the resulting software requirement specification the system architecture was designed, implemented and tested. A model for Coronary Heart Disease (CHD) was created and validated against empirical data.

Results: The system was successfully used to create and validate the CHD model. The initial validation results show concordance between the simulation results and the empirical data.

Conclusions: We have demonstrated the ability to connect health policy-modellers and policy-makers in a unified system, thereby making population health models easier to share, maintain, reuse and deploy.

1. Introduction

Long-term conditions, such as Coronary Heart Disease (CHD), consume the largest proportion of healthcare budgets and are a major target for public health initiatives. Moving interventions “up-stream” to earlier stages of the natural histories of diseases would delay or prevent subsequent events, thereby reducing the amount of suffering over the average lifetime, and potentially saving substantial costs. Health policy-makers and those planning and managing local health services are, however, poorly served by over-simple estimates of the potential public health impacts of taking preventive public health measures or making changes to the pathways of care. These estimates are often unreliable [1] because the models do not adequately represent the complexity of the population, the disease, or care over time.

Population health impact estimation is usually done by a small group of analysts synthesising evidence and producing a report for a decision-making team. For example, to quantify the potential impact of reducing CHD in a defined population over five years, local policy-makers might ask, “how should the balance be struck between investments in statin tablets vs. smoking cessation vs. physical activity promotion?”

There are several problems with this approach:

a) The available data and literature to consider is vast, complex and increasing;
b) A static report is relatively inflexible and does not enable “what if” scenario planning; thus, relatively few options are appraised;
c) There are not enough analysts to support current decision-making needs, yet it is unlikely that health systems could afford to employ more analysts. Furthermore they are in short supply;
d) Most healthcare commissioning groups do not have the skills or time to build realistically complex models which take all reasonable factors into consideration – decisions may therefore be biased by where a narrowly defined model focuses, which may reflect the interests of service providers more than the needs of the population served.

It is possible to construct visual representations of disease and healthcare pathways, and to use the resulting networks to simulate outcomes for populations. Such a simulation system would enable the user to compare different intervention scenarios, with the ability to modify both clinical and public health interventions, and measure the effectiveness based on both clinical outcomes and costs. The system could bring together public health professionals, clini-
cians and service commissioners in interactive scenario planning activities to inform policy decisions. The ideal system would enable users to construct and share models around what if scenarios easily; to execute individual simulations quickly; and to interpret simulation results collectively. Larger simulations, in terms of the population size, provide greater accuracy but consume more computational resources. The construction of the best models requires collaboration between epidemiologists, biostatisticians, health economists and typical decision-makers/leaders (public health professionals, healthcare managers, and clinicians).

In this paper we report on the IMPACT system that has been designed to enable this approach, by bringing together model builders, model users and computational resources to participate in shared decision-making.

2. Background

There are many examples of computer modelling and simulation of a disease using a range of methods reported in the literature [2–4]. Coronary heart disease (CHD) is one of the most extensively modelled diseases, so it was chosen the focus for designing a generic system for modelling health impacts in defined populations.

A recent systematic review [5] of cardiovascular disease policy models concluded that models vary widely in their depth, breadth, quality, utility and versatility; with few models being adequately validated or replicated in different settings. Moreover, few were either available for inspection or transparent enough to enable full understanding of the underpinning methods and assumptions. As such, the strengths and limitations of most models were poorly defined; therefore, few were acceptable for use in policy making. For example, a recent model published by the English Department of Health to support cardiovascular screening appears both over-simple and opaque [6]. Out of 70 modelling attempts identified in this area, fewer than 10% published more than one paper, and very few have functioned for a decade or more.

The first IMPACT model [7] used an attributable risk fraction approach and was implemented in a spreadsheet with over 44,000 cells. However, it required extensive training of users and was difficult to deconstruct for validation. Here we report a new approach to the IMPACT model, separating the generic modelling challenge from its application to CHD. Furthermore, we separate the computation of the model from interaction with users, and address the generic problem of simulating public health impact.

3. Objectives

The mathematical methods and computing technologies required to unify model building and use are available [8, 9]. The aim of this work was to harness these methods for health policy making. The objectives were to: 1) develop a versatile, flexible, valid and credible quantitative system for executing population disease models; 2) provide a single framework for domain experts to collaborate on model design and validation; and 3) to provide a decision support capability that enables health professionals to interact with the models.

4. Method

4.1 System Requirements and Analysis

Taylor-Robinson et al. conducted an extensive consultation exercise with policymakers on their attitudes to modelling and simulation [10]. The findings of that research were used to inform our requirements for the system.

4.1.1 Versatile and Flexible

Our principal objective is to provide a generic system for simulating public health interventions, enabling users to find, ask, and reuse ‘what-if’ questions about options for preventive and clinical interventions in a population’s health. This can be contrasted with the prevailing use of bespoke models often implemented with spreadsheet applications. Consequently, the system must contain a generic execution engine, that can instantiate a given model and perform the simulation. To create models, a model design tool is required that guides the end user through model creation and ensures valid models are created. What constitutes a valid model is intrinsically linked to the design and implementation of the model execution engine. The model alone cannot be executed; it must be configured with additional parameters that define a simulation. Thus a simulation is the combination of the model and the data that characterises the population, the environment, and the interventions being considered. Therefore the system must provide a tool that enables users to define simulations for a given model. We must also consider what the system will be used for. The IMPACT system is intended for answering five types of question:

- How will the burden of disease change over time?
- What will be the impact of specific treatment interventions/technologies?
- What will be the impact of population level/public health interventions?
- In terms of life expectancy is prevention more effective than treatment?
- Are interventions targeted at high-risk groups more effective than whole population level interventions?

The system must provide a tool that enables the results of a simulation to be analysed and visualised, and for comparisons to be made between simulations.

4.1.2 Transparency

Transparency was identified as a key requirement for users to be able to trust and subsequently act on the results of simulations. By transparency we mean that the system must be open to inspection at all levels. Consequently:

- The system software must be open source, so that it can be inspected and critically appraised. The source code must have companion documentation that describes its architecture, algorithms and implementation that is accessible from the system.
- The statistical theory and algorithms underpinning the models and their
execution must be formally documented and accessible.
- For each model, the model builders are required to supply descriptive metadata that describes: the risk factors and disease groups; data sources and main assumptions; the relative risk reductions of interventions; the uptake (availability and adherence) of interventions; the nodes of the graphical model; the edges of the graphical model, defining transition probabilities between health states; the observable outputs of the model and terminology.
- For each simulation, the system must enable users to inspect the configuration that defines the population, environment, and interventions.

4.1.3 Accessible

To achieve widespread adoption, access to the system must be as easy as possible for the end user. Thus we are delivering the IMPACT model as a web application that requires no end user installation, configuration or maintenance.

The user interface must be simple and intuitive to use. In order to achieve this different classes of user are defined in terms of their intended use of the system, such that the functions and features available in each user class provides a different view of the system. This enables the complexity of the system to be hidden from the user interface if it is not required. Basic users can create and execute simulations, perform simulation comparisons, and share their results. Advanced users have access to a suite of model building tools enabling them to create new models for wider consumption.

4.1.4 Usable for Collaborative Model Creation and Decision Making

The development and validation of models requires collaboration between statisticians/modellers, epidemiologists and health economists. Health policy-making is also a multi-disciplinary process. Web-based social computing technologies are widely deployed and used across many different disciplines [11] for collaborative working. This again favours a web application such that a shared workspace can be created and technologies for storage, retrieval and search of work products can be leveraged. In essence the system must bring people, data and methods together if it is to meet our objectives.

4.2 Model and Execution Engine

The life courses of the population of interest are modelled statistically through a two-stage procedure. The first stage is called the population model, which simulates disease incidence. The second stage is called the clinical model, which simulates the progression of diseased individuals to death. The priorities for the population and clinical models are different, so different types of model are used. The overall modelling platform is designed to be a flexible sandbox, allowing various 'what-if' scenarios to be trialled in the population.

The population model uses an accelerated failure time (AFT) approach to model the age of onset of the disease of interest [12]. That is, the covariates measured on an individual are assumed to act multiplicatively on the time scale and so affect the rate at which an individual proceeds to become an incident case. Risk factors such as cholesterol, smoking status and blood pressure are incorporated as covariates into the regression. These risk factors are allowed to change over time. The approach can be generalised to allow downstream risk factors to be controlled by upstream risk factors such as diet and exercise. It is also possible to generalise to a multivariate approach, allowing multiple diseases to be considered simultaneously. All incident cases generated by the population model are passed to the clinical model with their associated characteristics at time of incidence.

Interventions in the population model are modelled as changes to the distributions of risk factors. For example, a population level intervention on healthy eating may reduce average salt intake, and the model will propagate this automatically to downstream risk factors such as BMI and blood pressure. Through the AFT model this will thus reduce the speed of progression towards becoming an incident case. Alternatively, a targeted or medical intervention such as a change in statins prescribing trends may reduce, for example, cholesterol levels amongst those with existing high cholesterol. The population model can be run for various potential interventions, and then the incidence distribution of disease cases compared to address questions about the impact of specific interventions and the burden of disease.

The clinical model uses discrete event simulation: for each individual, a sequence of events, from a possible set of events defined by a multi-state model, occur chronologically [13–15]. Various disease states are included as nodes in a graph, and edges represent permitted transitions between the disease states. For example, the multi-state model for CHD is given in Figure 1. A subset of states are specified as entry states, where an individual has presented with CHD symptoms, and another (possibly overlapping) subset of states are defined as sink states, corresponding to death events. A continuous time multi-state model is developed here, which allows higher fidelity simulation and more flexible, realistic intervention policies than a discrete time analogue.

Transitions between nodes in the clinical model are controlled by hazard functions, which describe the instantaneous risk for a given individual making a transition between two disease states. The topology of the graph determines the competing risks for any given disease state. For non-specialists in particular, the graph representation makes constructing and editing a new model relatively straightforward.

Interventions in the clinical model affect patients indirectly, meaning that the results of a simulated intervention do not immediately alter the state of an individual (an impulse intervention) but rather alter the hazard of entering a given state. This is achieved by implementing interventions as proportional adjustments to the transition hazard functions. It is an assumption of the model that the adjustment in the hazard should be proportional. For example, a patient suffering from chronic angina and taking statins will have a reduced hazard of experiencing myocardial infarction, compared to an otherwise identical patient with chronic angina who is not taking statins. It
is possible to specify the uptake and availabilities of various interventions for different disease states.

The clinical model can be run with numerous different interventions applied, such as adjusting the uptake or availability of a particular drug, or even adding a new drug. Since the clinical model simulates patients to death, with two separate nodes, one for death from the disease of interest, one for death from other causes, the effect of these intervention strategies can be analysed on the whole life course. For simulations run under different conditions, various powerful and easy to use tools are available to statistically compare outputs to address policy questions. For example, we may interrogate the number of deaths, counts per state, life tables, time to events, common pathways, time in state, period prevalence and transition probabilities.

A major benefit of this model is the integration of the population and clinical models. This allows policy makers to answer questions such as “should I invest my money in smoking cessation as a preventive measure, or instead spend the money by prescribing more statins to diseased individuals?”

This approach is essentially a generic approach to modelling any non-communicable disease epidemiology and its control. Most diseases and conditions can be described temporally in two phases: an initial clinically not detectable phase, where most preventative interventions aimed at disease risk factors are targeted; and a clinically evident phase where therapeutic interventions are used. For most common diseases, data on risk factors and clinical effectiveness of preventative and therapeutic interventions is available from the literature or national surveys.

A model such as this requires fitting, so that the results it produces are evidence-based, robust, and reflect the population of interest. Furthermore, the model fitting procedure used is able to synthesise evidence from a range of sources.

To fit the parameters of the CHD population model, effect sizes of the risk factors on time to disease onset are estimated from various US [16] cohort studies through AFT regression. The model is also tuned against estimates of the incidence distribution of CHD in the population of interest, where this is available.

The clinical model is able to combine information from cohort studies and elicited expert opinion. For simplicity and tractability, the information obtained is converted into a collection of transition probabilities, or, more generally, constraints that the model attempts to replicate. An example of a transition probability constraint is:

\[ \Pr[55 \text{ year old male in state B at time } t | \text{ was in state A at time } t-1] \]

This model fitting construction is designed to be very flexible, to allow evidence from disparate sources to be synthesised. Constraints can be weighted depending on the value assigned to each source of evidence. These can be formal weights, depending on the statistical variability of data, or subjective evaluations of belief in different sources.

Both models are fitted using methods similar to simulated annealing [17]. For the clinical model, for example, we attempt to maximise the fit of the model to the supplied constraints by minimising the Jensen-Shannon divergence [18], which reflects how well the fitted model reproduces the constraints.

4.3 System Architecture

The system was designed around a number of architectural principles. In the interests of transparency, open source technologies were used and the IMPACT Simulator has Service Oriented Architecture to provide a clean separation between components with a view to minimizing the impact of future development and to enable scaling through flexible deployment across a range of hardware platforms. The system is composed of four components: Presentation, Data Management, Broker Service and Simulation Service (Fig. 2).

The Presentation component is a web application that provides the interface for users to interact with the system and is developed to be conformant to the Model-View-Controller design pattern. Web pages offer users functionality in the form of editors and management and reporting tools. Simulations can be configured using editors that perform general create, read, update and delete operations on simulation data. The Model Editor allows users to define models by specifying graphical structures that represent the disease pathways of interest. Graph nodes represent the disease states and arcs represent the allowed transi-
tions between those states. This editor also allows the assignment of Quality Adjusted Life Years (QALY) weights to each disease state for use in post-simulation analysis. Users can only partially specify models using this editor and model fitting has to be performed before a model can be used with a simulation. The Model Fit Editor allows users to specify initial hazard function parameter values for transitions between disease states and to optimise these by fitting to a set of inputted constraints.

Each model can be fitted many times with users controlling the fitting process by inputting simulated annealing fitting parameters. One criticism of simulated annealing is the difficulty with which fit parameters are chosen. The system provides default values that the developers have found to be sensible for most scenarios. The editor presents diagnostic information following the completion of a fit to inform the user about the suitability of these parameters, including information about the convergence and closeness of fit. The fitting process is completed separately for males and females and the most suitable fit for each gender can be marked for ‘publication’. Published fits provide the hazard functions used for a model during a simulation. The Intervention Editor allows users to define new interventions and to characterise their effect on models. This effect is represented as a reduction in one or more transition risks, inputted by the user as a set of relative risks. Regional variation in the availability of an intervention and the concordance of patients to their use (uptake) is modelled using Regions. The Region Editor allows users to define new regions characterised by uptakes and availabilities for interventions and a population size. Disease incidence in the simulated population is determined using either the Population Model or a cohort specified using the Cohort Editor. A cohort is defined by the age, gender and incident disease state of each member and can be inputted by tabular entry or, more conveniently for large cohorts, by file upload. Finally, the Simulation Editor is used to associate all of the information required to fully specify a simulation for execution. Users will need to select a fitted model, region, cohort and duration. The editor allows intervention uptakes and availabilities, selected through the region, to be altered and new interventions to be defined for the duration of the simulation. This is designed to facilitate straightforward exploration of ‘what if’ intervention scenarios. Any modifications are registered as a text description that can be retrieved later as simulation metadata. The simulation editor can then be used to execute the simulation. Management tools monitor simulations and model fits, providing the current status and allowing users to stop execution. These web pages make use of asynchronous methods and automatically update to provide user notifications such as the completion of a process. Reporting tools offer visualisation and analysis functionality, supporting the effective interpretation of simulation results. Both tabular and graphical representations are employed and full provision is made for large data sets by allowing results to be exported as Excel spread sheets. Existing reports include:

i. The number of disease and non-disease related deaths for the duration of the simulation;
ii. The number of individuals in specific disease states at a specified time;
iii. Life table;
iv. QALY;
v. The time taken for each individual to reach specific disease states;
vi. The disease state of each individual at a particular time;
vii. The complete disease history for each individual (suitable for off-line analysis);
viii. Common disease pathways;
ix. The percentage of individuals in each disease state at each age;
x. Time spent in each disease state (per visit) box plot;
xi. The prevalence of each disease state for specific years;
xii. The number of individuals that make each transition and the probability of transitions from each disease state for the duration of the simulation;
xiii. The probability of being in a particular disease state one year after being in another disease state (by direct and indirect transitions);
xiv. Comparison of simulated time to event data with time to event data uploaded by the user (comparisons include Cox regression, Kaplan–Meier curves, age at event distributions and QQ plots).

Many of the reports can be obtained for an inputted list of age ranges and for either or both genders. Comparisons between simulations can be made simply by incorporating information about multiple simulations in the same report. Tools make use of R packages using a COM-based interoperability layer to perform some of the statistical analysis.

User accounts and role-based access control are also managed through the Presentation component. Users are required to log on with a username and password before using the simulator. An initial registration process is required to obtain a user account with default ‘Guest’ privileges. Users with Guest privileges are able to run simulations and generate reports but are not able to perform more advanced tasks such as creating and editing models, interventions and regions. Elevated privileges may be obtained by contacting the site administrator. To provide consistency of functionality and appearance across a range of web browsers and therefore ensuring availability to a wide community of users, graphical components are expressed in both the Vector Mark-up Language (VML) and Scalable Vector Graphics (SVG).

An important feature of the user interface is its ability to allow simulation data to be associated with metadata. Users are able to make general comments about data and provide references for data sources. This metadata could provide, for example, references for intervention hazard adjustments or a description that clarifies the meaning of a disease state. As discussed earlier, this information is seen as critical for the correct interpretation and re-use of the simulation data. The system automatically records the users that create metadata and the dates that this information is provided. The metadata can be updated at any time and this update history is recorded. The IMPACT system supports a form of peer review by allowing users to verify and approve data provided by others. This is achieved by users marking data for review and adding to the data comments. The review date and reviewer user name are automatically captured by the system.

The Data Management component provides data storage, retrieval and validation services for other system components. It exposes an interface that abstracts away details relating to the physical storage of data,
allowing data to be managed more simply in terms of an object model. The creation of this shared component allows the localisation of the business logic to a single piece of software, reducing the likelihood of inconsistency across the system. This component makes use of the NHibernate Framework to map the domain object model to a relational model, persisting data in a SQL Server database. A full recovery model is employed to backup the database, permitting recovery of all data to any point in time following a failure. Further, the databases are regularly and frequently copied to a remote site to allow disaster recovery should any physical damage occur to the system.

The Presentation component interacts with the Broker Service to execute simulations and model fits. This service uses the Data Management component to retrieve information required to configure simulations and model fits and to persist the results. The Broker Service is a web service and requires each client to identify itself using an X.509 certificate, which is validated by comparing it to existing client public certificates held on the host. The use of certificates enables the service to guarantee the integrity of the transferred data, by including a digital signature in the request message, and to authenticate the client. The Simulation Engine is the sub-component that performs the simulation – it was developed in Python and uses some mathematical functions from the SciPy library. The other system components were developed to use the Microsoft .NET Framework. In order to support the execution of simulations in a wide range of operating environments, a channel adapter was developed. The adapter exposes a web service interface to the Simulation Engine, allowing simulations to be configured and results to be returned using SOAP messages over HTTP. This architecture allows multiple Simulation Services to be deployed to execute simulations, ensuring the future scalability of the system to many concurrent simulations and improving the tolerance of the software to problems such as hardware failures and network faults.

5. Validation

The system has been tested by using it to implement and validate the IMPACT model of CHD. The validation process is an integral part of model development. It helps in identifying issues with model implementation, data and assumptions. More important, it is a key element in increasing the model value to policy makers. However, this aspect of model development has been frequently overlooked in cardiovascular disease modelling [5].

We validated the model by using it to simulate the SLIDE cohort, a cohort of survivors of acute coronary syndromes in Scotland [20] (n = 80,241). Our aim was to compare the actual CHD mortality experience of the acute myocardial infarction sub-cohort, with that predicted by the model. For this, we simulated a population with the same age and gender structure as the real sub-cohort, and constructed the simulation taking into account historically plausible treatment effects, using data from systematic reviews and randomized clinical trials [21]. Treatment uptakes were obtained from surveys and audits from a contemporary source to the SLIDE cohort [22]. We generated censoring times in the simulated data with the same distribution as the censoring times in the actual data.

As an example, it can be seen that the model produces an age distribution of CHD deaths that resembles the actual cohort (Fig. 3 and 4). These results are very encouraging, although further work on model constraints is still ongoing. Treatment uptakes have been updated and most of the edges are satisfactorily parameterised. The Kaplan-Meier survival function produced using the model output is similar to the observed (Fig. 5), where the outcome of interest is CHD death. These preliminary results demonstrate that the system is usable for model creation and execution and that the predictions of such models are consistent with observed data. More validation work is needed, specifically regarding comparisons with different cohorts and populations. Because the system can generate many outcome measures that can be derived from time to event data (e.g. life years and quality adjusted life years), as well as prevalence of individuals at individual states at specified times, comparisons with different modeling approaches producing those outcome (such as Markov models for health economics analyses) will be possible to cross compare and to explore the strengths and limitations of different approaches. In addition, the validation process will offer valuable insights towards improving the model ability to produce more accurate estimates of the number of CHD deaths, data visualization and model functionality.
6. Case Study

In this section, we step through the key elements of using the policy model.

The first step is to define the research context and question motivating the use of the model. For the population model, define the necessary population input data. Birth statistics are taken from the Office of National Statistics UK National Statistics database. The risk factor distributions and the correlation between risk factors are obtained from the Health Survey for England (HSE) database. Risk factors included are gender, Body Mass Index (BMI), Systolic Blood Pressure (SBP), cholesterol, diabetes and smoking prevalence. Estimates of the effect of these risk factors, including sex, are obtained using an available US cohort data set [23]. It is assumed that the effect of these risk factors is the same across populations [24–26].

Fit the model to this data, including the effect of the risk factors, using AFT regression.

The remaining model parameters are then calibrated, using the iterative simulated annealing-type approach, so that it achieves the correct disease incidence for the UK population. That is, the parameters of the model are tuned to obtain correct mortality age distributions and total number of fatalities, taken from the General Practice Research Database (GPRD).

Specify any population level or targeted interventions, and their effects on the risk factor distributions. Interventions are assumed to adjust risk factor distributions. For example, a smoking cessation policy will reduce the probability of an individual being affected by this risk factor.

For the clinical model, set the simulation duration and the initial population either from a specified cohort or the population model output. For exposition, we will investigate the effect on a population of size 10,000 of different levels of preventative treatments regimes. Define the discrete states for the disease of interest: for example, Figure 1 shows the CHD graph structure and Table 1 gives the node descriptions from this graph. Derive the constraint values from relevant data sources and expert elicitation. Fit the model

Table 1 Description of discrete states in CHD model

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Name of Discrete State</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>Chronic angina</td>
<td>Long term chest pain</td>
</tr>
<tr>
<td>AMI</td>
<td>Acute myocardial infarction</td>
<td>Heart attack</td>
</tr>
<tr>
<td>SD</td>
<td>Sudden death</td>
<td>Heart attack, leading almost immediately to death.</td>
</tr>
<tr>
<td>MI Surv</td>
<td>Myocardial infarction surv</td>
<td>Has experienced one heart attack in the past</td>
</tr>
<tr>
<td>MI Recur</td>
<td>Myocardial infarction recur</td>
<td>Has experienced two or more heart attacks, usually associated to higher subsequent risk of mortality</td>
</tr>
<tr>
<td>UA</td>
<td>Unstable angina</td>
<td>Triggered severe chest pain that merits admission for evaluation or treatment. Could lead to a heart attack if untreated.</td>
</tr>
<tr>
<td>Early HF</td>
<td>Early heart failure</td>
<td>Patients cannot exercise because of shortness of breath and other symptoms caused by the failure of the heart to supply blood to the rest of the body organs and systems.</td>
</tr>
<tr>
<td>Severe HF</td>
<td>Severe heart failure</td>
<td>A more severe stage of heart failure, usually associated to a very high risk of mortality.</td>
</tr>
<tr>
<td>Non CHD Death</td>
<td>Non CHD death</td>
<td>Death whose primary cause is not recorded as CHD</td>
</tr>
<tr>
<td>CHD Death</td>
<td>CHD death</td>
<td>Death whose primary cause is recorded as CHD</td>
</tr>
</tbody>
</table>

Fig. 6 Comparison of CHD deaths with statins at current levels compared with statins removed

through the calibration of the hazard function parameter values on each edge against the constraints using simulated annealing. Define the interventions and the related states on which they act; define the probability of adherence of each individual to the intervention regime, which may depend on, for example, their age and gender.

We now illustrate a workflow in the clinical model by considering a toy example. Suppose that, due to budget cuts, a policy maker wishes to consider the removal of statins prescriptions for 50-year-old males who present with heart attacks (note 50-year-old males are considered in isolation here for simplicity of exposition. In reality, one would consider a representative population). In order to do this, a policy maker runs simulations for the policy model under two scenarios: 1) statins set at current uptake levels; 2) statins uptake set to zero. We simulate 10,000 males who have an AMI at age 50 under each scenario, and compare their outcomes. A simple outcome we may wish to compare is how many of the males ultimately die of CHD under each scenario. A graph of this output is given in Figure 6. The overall number of deaths is, under current statin levels, 4144 (95% confidence interval [4018.8, 4272.1]); and under statins removed, 4723 (95% confidence interval [4589.3, 4859.7]). So we conclude that removal of statin prescriptions for 50-year-old males suffering AMI would lead to an increased cumulative CHD death rate of 579 per 10,000. It would then be up to the policy maker to trade off the increased death rate against the financial saving, and possibly consider other ways to make the financial saving that may have less impact.

7. Discussion

We have shown that policy modelling can be made accessible and transparent via a web-based system. Unified modelling frameworks such as the one described here may encourage epidemiologists, biostatisticians, modellers, health economists and public health practitioners to contribute to open, accessible policy models rather than creating a blizzard of niche models.

A limitation of the described approach is that the clinical model only considers the public health burden of the specified disease(s) in isolation, and not the overall public health burden. This means that caution is needed in interpreting the results of interventions since demonstrated benefit in the context of a single disease may has simply transferred the burden elsewhere.

Furthermore, the clinical model and the population model in particular are highly dependent on data. The calibration approach adopted for the clinical model does provide flexibility about the type of data suitable for this but there is still a need for a relatively large amount of data, or at least reliable expert opinion, to obtain models with good face and predictive validity.

In on-going work we are returning to the community of planners and policy makers [10] to assess the usability, accessibility and utility of the system and the IMPACT CHD model. The uptake and usage of the system will be monitored, as these will be the key measures of success.

Future work is planned to parallelise the simulation engine to take advantage of multi-core and cluster computing. This will dramatically reduce the simulation run-time making the system more usable for complex models and large populations. The modular nature of the architecture enables the use of cloud computing infrastructure in the future.

The IMPACT simulator will be integrated into the nascent e-Lab population health information system [27]. This will lever electronic health record data to refine, extend and localise models. The e-Lab platform provides the Work Object [27] mechanism as a way of exchanging knowledge between federated e-Labs in different localities. The IMPACT Simulator already has the capability to export IMPACT Simulation Work Object Archives for a specific simulation, to demonstrate proof of concept. These are ZIP archives containing files that represent all of the simulation data and metadata in a semantically explicit way. The metadata includes that inputted by the user, such as the references to data sources, descriptions and general comments. It also includes metadata automatically captured by the system, such as information about users and dates. In addition, the system also provides information about the statistical methods used and their implementation details, such as software versions. Semantic Web standards are adopted to ensure that the meaning of the data is unambiguous. RDF is used to represent the data, relating it to concepts that are specified in an ontology. This ontology is being developed for use with the simulator and builds on other conceptual models including that proposed for Research Objects [28].

In conclusion, rising computing power coupled with advances in machine learning and healthcare information now enables more user-friendly models to be constructed and executed. The IMPACT Simulator represents a system for creating, executing and analysing the results of simulated public health and healthcare policy interventions, which is accessible and usable by modellers and policy-makers alike.

The IMPACT simulator is deployed and available on the Internet at http://www.impactsimulator.org.uk.

Acknowledgments

This work was supported by funding from the UK NIHR Greater Manchester CLAHRC programme; UK MRC IMPACT project, the UK NPR13 scheme and the EU MedCHAMPS project.

References

6. Response from the Faculty of Public Health to the Department of Health Consultation on Economic Modeling of a Policy of Vascular Checks through the calibration of the hazard function parameter values on each edge against the constraints using simulated annealing.