MEDINFO 2013
C.U. Lehmann et al. (Eds.)

doi:10.3233/978-1-61499-289-9-288

e-Labs and the Stock of Health Method for Simulating Health Policies

Philip Couch[a], Martin O'Flaherty[b], Matthew Sperrin[c], Benjamin Green[a], Panagiotis Balatsoukas[a], Stephen Lloyd[a], James McGrath[a], Claudia Soiland-Reyes[a], John Ainsworth[a], Simon Capewell[b], Iain Buchan[a]

[a] *Centre for Health Informatics, Institute of Population Health, University of Manchester and Manchester Academic Health Science Centre, UK*
[b] *Division of Public Health and Policy, University of Liverpool, UK*
[c] *Department of Mathematics and Statistics, Lancaster University, UK*

Abstract

Regional outcomes of national health policies are difficult to forecast. This is partly due to a lack of realistically complex models that can be used to appraise policy options and partly a lack of accessible and adaptable tools that can be used to simulate the consequences of policy decisions. These barriers might be overcome by exploiting the commoditization of massively parallel computing architectures, advances in machine learning, and the increased availability of large-scale linked healthcare data. This paper presents a novel modelling methodology, The Stock of Health, for harnessing emerging data and computational resources to simulate health policy, with application initially to coronary heart disease. We detail the use of multi-core graphical processing architectures to facilitate a micro-simulation approach. The simulation tools have been deployed through the IMPACT *Framework. We explore how this framework can be extended to support the sharing and reuse of policy models and simulations based on the digital publishing concept of e-Lab.*

Keywords:

Policy modelling, public health, in-silico parallel simulation, policy decision support.

Introduction

Cardiovascular disease (CVD) produces a massive burden of disability, distress and early death in society. It is therefore a policy priority for public health in many parts of the world [1]. In the UK, 50,000 premature and avoidable deaths are caused each year through CVD sub-diseases such as coronary heart disease (CHD), heart failure, stroke, chronic kidney disease, and peripheral vascular disease. Yet over 80% of premature CVD deaths are avoidable [2]. Premature CVD rates are three-fold higher in the most deprived groups, substantially contributing to health inequalities. CVD chronically affects over three million patients [1] with the total annual cost in the UK exceeding £30 billion [3]. Although UK CVD mortality rates have been falling since 1980, population aging will result in additional future cases. Furthermore, as therapies improve, more people are surviving their first CVD events. Thus the patient surviving CHD through better treatment may go on to have other CVD problems such as stroke or heart failure. This rising healthcare need, particularly in the late stages of disease, is compounded by escalating costs for hospital procedures and drugs [4]. Governments are therefore promoting CVD prevention through policy initiatives (smoke free legislation, salt reduction and increasing physical activity) and targeting medications at the higher risk individuals [4]. However, the current and future population impact of these strategies remains unclear [1]. Health planners and clinicians need a better understanding of the recent trends in the burden of CVD, especially the underlying epidemiological and therapeutic factors, in order to compare future policy options and plan appropriate services for CVD prevention.

Our initial work on CHD suggests a poor connection between decision making and potentially available evidence. In addition, software to support policy making is seldom used interactively by policy makers because it is seen as too difficult to use. Furthermore, the potential value of population wide strategies may be large, but is currently poorly quantified [5]. Growing evidence suggests that implementing preventive strategies throughout the entire life course will be an increasingly important policy option [6]. However, potential impacts are difficult to quantify. To maximise the benefits from scarce resources, policy-makers need reliable information on disease trends and interventions (costs & outcomes), and local planners want tools to estimate future changes in service costs. Among current CVD interventions prevention is dwarfed by treatment. Delays in identifying more effective strategies for CVD prevention will be costly; therefore 'do-nothing' is not an option politically, ethically or economically. Using modelling to compare the potential impacts of different strategies is increasingly recognized as key to evidence-based policy decision making [4].

The IMPACT programme is developing a framework for collaborative development and dissemination of public health and healthcare policy models. The central informatics challenge is to make such models locally relevant and easy to use, through incorporating local data and connecting decision makers in a professional social network of option appraisal.

The Stock of Health

Method

The concepts of this approach are illustrated in Figure 1. We assume that each individual is born with a stock of health (SoH), and on an annual basis this stock depreciates. The rate of depreciation depends on underlying fixed and variable risk

factors and demographic factors, as well as random chance. When the health stock of an individual reaches a critical point (the dotted line in Figure 1) an event of interest occurs. The SoH approach can be used to model any event, the only constraint for reliability being the availability of data for model calibration. In the context of the application to CHD that we present here, there are two events of interest: the new presentation of a patient with symptoms of CHD (incidence), and death from CHD (mortality).

More formally, let $y_{i,t}$ denote the SoH held by individual i at age t. Then let $\eta_{i,t} = y_{i,t} - y_{i,t-1}$, the SoH lost in year t, which is modelled as

$$\log(\eta_{i,t}) = \beta_0 + \gamma_i + \beta_X X_i + \beta_Z Z_{i,t} \quad (1)$$

β_0 is the baseline, which is allowed to depend on the calendar year in order to capture secular trends in disease onset or disease specific death; $\gamma_i \sim N(0,\sigma^2)$ is an individual level random effect; X_i and $Z_{i,t}$ denote individuals' age invariant and age variant risk factors respectively (which are assumed to have population-level means subtracted) and the βs are parameters controlling how risk factors affect the SoH loss. The log-transform in (1) means that SoH annual loss is strictly positive. The model is mathematically equivalent to an accelerated failure time (AFT) model [7] with a log-normal distribution; the SoH formulation is useful for communication with policy makers, and also provides a mechanism for incorporating risk factor changes.

Risk factor changes in individuals are caused by population level or targeted interventions. Their effect is divided into two components: first, the long-term effect caused by the risk factor being different (leading to a change in the future rate of decline of SoH); second, an instantaneous effect, represented as an instant change in SoH. The instant change caused by an intervention, which shifts risk factor j by $\Delta_{j,t}$, leads to a new SoH $y^*_{i,t} = y_{i,t} + \delta_{i,t}$, where

$$\delta_{i,t} = \alpha_j \Delta_{j,t} y_{i,t}^{p_{1,j}} (y_{i,0} - y_{i,t})^{p_{2,j}} \quad (2)$$

α_j and the exponents $p_{1,j}$ and $p_{2,j}$ are estimated separately for each risk factor through optimisation (*Risk Factor Shifts* section). This is a skewed function of SoH, constrained to be 0 at the minimum (0) and maximum ($y_{i,0}$) SoH values. The instant change can, in principle, be positive or negative.

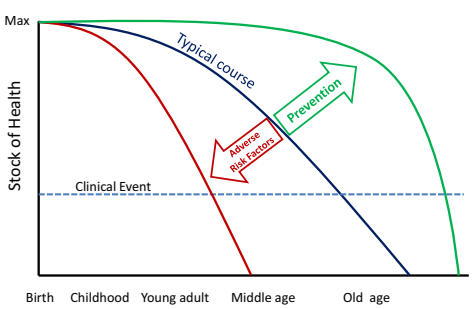

Figure 1 - Stock of Health trajectories

Coronary Heart Disease Models

We have applied the SoH approach to the IMPACT CHD modelling programme to predict the outcomes of policy options in terms of disease incidence and mortality. Polices are considered through their effect on five risk factors: systolic blood pressure, body mass index, total cholesterol, smoking and diabetes. Evidence has been gathered from the literature and a variety of databases in order to calibrate these models. The following sections describe the data sources, outline the statistical methods employed to optimize the models, and compare some SoH-simulated with real-world outcomes.

Model Calibration

Model calibration involves determining the values of the unknown parameters presented in the *Method* section. These include parameters that control the effect of risk factors on the SoH, the baseline change in the SoH and the instant effect of risk factor shifts. Separate models are developed for mortality and incidence. In the mortality model, an individual's SoH represents a point on a trajectory toward death from CHD. This is in contrast to the incidence model, where the trajectory is toward presentation of the disease. Both models are calibrated separately for males and females.

Risk Factor Effects

In order to obtain the risk factor effect parameters (β_X, β_Z) we analysed the Cardiovascular Lifetime Risk Pooling Project dataset [8]. If individual studies met the following criteria, they were included in the dataset: 1) used community- or population-based sampling or a large volunteer cohort; 2) availability of at least one baseline examination at which participants provided demographic, personal and medical history information and underwent direct measurement of physiologic and/or anthropometric variables (e.g., blood pressure and weight); 3) longitudinal follow-up of at least 10 years with complete or near-complete ascertainment of vital status; and 4) availability of cause-specific or cardiovascular mortality data with or without ascertainment of non-fatal cardiovascular events. The risk factor effect parameters are maximum likelihood estimates from an accelerated failure time model fitted to the cohort data. For the incidence model, the endpoint was taken as the onset of non-fatal myocardial infarction and for the mortality model the endpoint was death due to CHD.

Baseline

The parameters that control the baseline rate of change of SoH (σ, β_0) for the England and Wales mortality model were estimated by minimizing the distance between observed and simulated mortality statistics. A long time-series of ischemic heart disease mortality is required for this purpose. The UK Office of National Statistics (ONS) 20th and 21st Century mortality datasets were used. These files are a record of mortality in England and Wales from 1901 to 2011. They consist of an aggregated database of deaths by age group, sex, year and underlying cause, and include populations for England and Wales. International Classification of Diseases 10 (ICD 10) codes I20-15 and ICD 9-8 codes 410-414 were used to estimate coronary heart disease mortality rates. The ONS mortality dataset was also used to construct a life table, determining the probability of individuals dying each year from any cause other than CHD.

The parameters were optimized using a Nelder-Mead simplex approach. England and Wales birth cohorts from 1901 – 2010 were simulated until all individuals died, either from coronary heart disease or any other cause. The cohort demographics were based on birth tables from the ONS. An individual's risk factors levels were determined by sampling distributions obtained from the UK Health Survey for England data. Data were obtained for total cholesterol, systolic blood pressure,

smoking, body mass index and diabetes from the 2004 wave. Calendar time changes in risk factor distributions were captured by trends in β_0 (Equation 1). The optimization criterion was minimizing the distance between the observed and simulated total and age group specific number of CHD deaths (1985 – 2010), and the mean and variance of the age at CHD death (1993 – 2004). Immigration effects were considered by adjusting simulated mortality based on the ratio of simulated to observed population sizes. Population trends and projections were obtained from the ONS; the main variant was used for projections. Figure 2 illustrates the best match between simulated and observed mortality for males and females, including a simulated projection to 2030. Baseline parameters for the England and Wales incidence model have yet to be optimized for the England and Wales population. This model currently uses parameters determined from the US cohorts during the AFT regression.

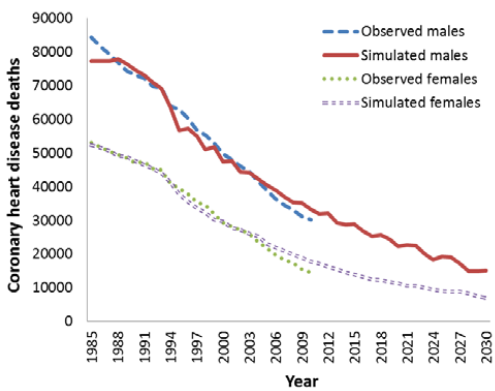

Figure 2 - A comparison between simulated and observed ischemic heart disease mortality for England and Wales, with simulated projection to 2030

Risk Factor Shifts

The outcomes of public health policy are predicted through the expected shifts in risk factor levels. These shifts affect the future rate of change of the SoH, but also lead to an instant change (Equation 2). The parameters that control the instant change (α_j, $p_{1,j}$ and $p_{2,j}$) are optimized by matching simulated hazard or odds ratios for risk factor shifts with evidence presented by Ford et al. [9]. Each risk factor was optimized separately for males and females across a range of risk factor shifts and age groups. Excellent agreement was found between simulated and observed hazard ratios.

Model Uncertainty

There are a number of sources of uncertainty in the model, including uncertainty in parameter estimates, Monte Carlo error, model assumptions and external assumptions. The complexity of the model and heterogeneity of data sources preclude calculations of strict confidence limits for most of the parameters. For example, risk factor effect parameters are estimated using an accelerated failure time regression to US cohorts and are applied to the England and Wales population. The behaviour of the SoH due to risk factor shifts is an important model assumption. At the time of an intervention, an individual's SoH is immediately shifted towards a counterfactual trajectory with the post-intervention risk factor levels from birth. External assumptions include different population projections, different future risk factor trends and novel future treatments.

Policy Scenarios

The SoH tools can be used by policy makers and NHS planners who wish to explore a variety of issues, including the potential impact of population aging and recent trends on future overall mortality rates, mortality burden and age at death. It also permits a quantitative estimation of the potential benefits of small reductions in systolic blood pressure (-1 mmHg), as might be easily achieved by reducing the salt hidden in bread or in processed food. Likewise the potential public health benefits of small reductions in the trans-fats or saturated fats hidden in snacks or fast food, which would reduce blood cholesterol levels by at least 0.1 mmol/l. These scenarios can then be compared with alternative targeted health care strategies involving the detection and lifelong treatment of individual patients with elevated blood pressure (or cholesterol).

Simulation Acceleration

The current Stock of Health model optimization typically requires 100 simulations to be executed in series before the criteria for reaching the error surface minimum are achieved. In addition, large populations need to be simulated in order to reduce the Monte Carlo error to a level that does not prevent a correct traversal of the error surface. The requirement to simulate large populations and to execute large numbers of simulations during the calibration process requires careful consideration of the software implementation. The initial software written in R (http://cran.r-project.org) would typically complete in a timescale of a week. Although this may be sufficient for research, we aim to create a system that runs quickly enough to support the collaborative and iterative development of policy models, which may need to be optimized many times. In order to improve runtime performance we wrote tools to exploit parallelism in common computing hardware. In an initial approach, the software was re-written in F#, a functional language encouraging code that can be easily or even automatically parallelized (http://research.microsoft.com/fsharp). The asynchronous programming model of F# was employed to create software that executed a simulation across multiple CPU cores, leading to a significant improvement in performance [10]. However, this proved insufficient; simulations still completed in a timescale that hindered the exploration of different models. The SoH approach affords trivial parallelization. There is no coupling between simulated individuals and each individual or group of individuals can be simulated separately and in parallel. This allows SoH simulations to be viewed as high throughput tasks that are able to make efficient use of capacity computing infrastructures.

There have been a number of claims that the use of graphical processing unit (GPU) accelerators can provide significant performance gains over the use of CPUs for certain tasks [11]. These accelerators are particularly beneficial when tasks exhibit data parallelism and many thousands of threads can execute concurrently with minimal synchronization. This is true for SoH simulations, and therefore we ported the F# code for execution on GPUs. The host application was written in C++ and OpenCL was chosen for the kernel. OpenCL was adopted because it allows execution on a wide range of devices with platform vendors that include Intel, AMD and NVIDIA. The following discussion uses terms defined in the OpenCL ab-

stract models[1]; the initial use of these terms is italicized. A meta-programming approach was used, with the host application writing the OpenCL kernel source. This allowed kernel optimizations to be performed based on the SoH model. These included loop unrolling, performing some calculations in advance, and a reduction of the number of variables to reduce register pressure. Many GPUs contain Special Function Units (SFUs) that can be used to execute transcendental functions. The SoH simulations make considerable use of such functions and they target SFUs through OpenCL 'native' functions. Each compute unit of a GPU has an area of memory that can be used to store constant, read-only data. This memory is low latency but small in size (typically 64 kB). GPUs support a broadcast mechanism that allows constant data to be accessed by many threads using a single read, therefore greatly reducing the required memory bandwidth. The SoH simulations require a significant amount of constant data; this includes the risk factor distributions, life tables and trends. To prevent the necessary use of high latency global memory for more complex models, the host application executes a simulation using multiple kernel dispatches with each dispatch using different constant data.

The choice of global and local work item index spaces is critical for optimal performance. An N-dimensional global index space determines the number of work items, with each work item being assigned a unique N-dimensional global index. Each work item is also assigned an N-dimensional local index that groups the work items into work groups. Specific device architectures need be considered when choosing an index space. The total number of work groups was chosen to be a multiple (typically 4) of the number of compute units to ensure that concurrent work groups could be used to hide latency (e.g., global memory access). The work group size was chosen to be a multiple of the SIMT (Single Instruction Multiple Threads) width of the compute unit (32 for NVIDIA or 64 for AMD). The size was chosen ensuring that the maximum number of work items for each compute unit was not exceeded and that there were sufficient resources available (e.g., shared memory). The simulated population was grouped into birth cohorts, with each kernel dispatch simulating a fixed number of cohorts. Each processing element executes a work item that simulates the life course of a number of individuals across all birth cohorts. Further optimizations included the use of single precision and MAD instructions to perform simultaneous multiply and add operations.

During development and testing, Stock of Health kernels were executing on GPU nodes of the Computational Shared Facility at the University of Manchester. These made use of NVIDIA Tesla C2050 devices operating under Scientific Linux 5.5 and controlled using the NVIDIA CUDA toolkit 4.0.17. An impressive level of performance was achieved, with 100 simulations of the life course of all individuals born in England and Wales, UK between 1901 and 2010 (8×10^7 individuals per simulation) executing in 61s. The NVIDIA Visual Compute Profiler was used to tune the performance. This shows the current kernels to achieve 50% occupancy with 1.6 instructions being issued per cycle (maximum two). This gives a throughput of 80% of the theoretical maximum of the GPU device. Using this setup, the Nelder-Mead optimizer can calibrate the baseline England and Wales CHD mortality model in a few minutes.

It was unclear whether the significant performance increase was due to the adoption of OpenCL or use of a GPU. To facilitate a comparison with execution on a CPU, the host application was further developed to include some optimizations to improve CPU performance. These included an organization of the memory structures to improve cache performance. SSE extensions were used to target SIMD (Single Instruction Multiple Data) units through the use of OpenCL vector types. The index space was chosen based on the number of processor cores and the availability of hyper-threading to hide latency. The kernel was executed on a dual Intel Xeon E5506 system running Windows 7 Enterprise using Intel OpenCL SDK 1.5. Simulations were found to run 24x slower (using 8 CPU cores and single precision) than on the Tesla C2050, which is significantly higher than the theoretical ratio of 7.6 for throughput. An initial investigation into the possible reasons suggests this is due to the heavy use of transcendental functions. The development of this software for GPU execution has delivered increases in runtime performance large enough to make wider uses of SoH tractable. This has been an important development, facilitating a micro-simulation approach to health policy simulation.

Simulation Sharing and Reuse

Information Design

The SoH tools are being deployed in a staged manner via the IMPACT Framework. Such deployments pose a number of challenges that require careful consideration. When setting up models and simulations and interpreting results, users may need to work with large and complex data. Such datasets can become overwhelming when viewed in their entirety. Here we draw from the field of information design, using a combination of visual, statistical and psychological principles to make data more ergonomic. Recent progress in the HTML standard makes it possible to create native animations using scalable vector graphics (SVG) and JavaScript. No additional client or development software is needed and development tools are free. This has led to a range of third party libraries that leverage this new technology. An example is D3, a JavaScript data visualization library released under a permissive BSD-style license (http://d3js.org). It is entirely standards compliant, making it future proof and stable, and includes capabilities and features that meet the requirements of the IMPACT interface.

Information Management

Policy model development is a complex process that often requires the collaboration of researchers and practitioners across different disciplines, including statisticians, epidemiologists, informaticians, health economists and public health researchers. The e-Lab is an emerging eScience and Health Informatics paradigm concerned with the social aspects of scientific collaboration and the provision of generic services that can be used to manage shared information [12]. A discussion of the benefits of adopting the e-Lab paradigm and the subsequent implications for the management of data are presented by Couch et al. [10]. Here we argue for the use of a domain independent data model for the exchange of information between the IMPACT tools and e-Lab services, and we propose the Resource Description Framework (RDF). There are benefits to using this model for all information managed by the system, not just for exchange.

We are developing a new data management library for IMPACT that allows software developers to map a domain object model to a conceptual model by annotating computer

[1] http://www.khronos.org/registry/cl/specs/opencl-1.2.pdf

code using attributes; classes can be related to concepts and member declarations to properties. The library exposes an API that can be used to manage the persistence of domain objects in a relational database. The relational model is domain independent, reducing data migration issues associated with using the same database with different versions of an application. The RDF data model is a model for making statements and the domain independence of the model means that statements can be made about anything. The software uses the concept of RDF graphs to group statements, which can then be described. This supports the capture of semantically rich metadata including data provenance. The software supports the creation of different data views based on RDF graphs. For example, different application versions may have different business logic and some data should only be accessible by specific versions; views could be created and shared between versions. Social networks of simulation can be developed in this way. Views will allow fine grained authorization at the statement level. Users will be able to specify a network of trust and only view statements that form part of that network.

Conclusion

We have presented a novel informatics approach for developing and disseminating health policy models that are useful at both the national and local population level. This approach combines three key elements: 1) a flexible modelling methodology, SoH, enabling preventive and treatment healthcare options to be appraised in the context of local patterns of health and care services; 2) a software engineering approach that exploits commonplace GPU hardware to accelerate the simulation of population health in a naturally parallel way; and 3) an information architecture that exposes the digital artefacts of population health simulation so that they are easier to develop, share and reuse across social networks of health professionals. The IMPACT programme is deploying this new methodology into real-world CHD policy-making with partners such as the Collaboration for Applied Health Research and Care in the UK National Health Service. The same framework could be employed for developing and disseminating a wider range of models.

There remain significant public health informatics challenges here, namely: i) to increase the computational tractability of more complex simulations to make them even more relevant to local decision making; ii) to find the most engaging forms of information for interactive and interdisciplinary health policy option appraisal; and iii) to connect population health modellers and decision makers in social networks of simulation that lead to better models and decisions. We hope that such developments will help to move health systems from the current position of a blizzard of isolated and little-used models to an ecosystem of simulation that supports more evidence-based health policies at all levels.

Acknowledgements

Couch, Green, Lloyd, Soiland-Reyes, McGrath, Balatsoukas, Ainsworth and Buchan were funded by the UK National Institute for Health Research (NIHR) as part of the Greater Manchester Collaboration for Leadership in Applied Health Research and Care (CLAHRC). O'Flaherty is partly funded by the UK Medical Research Council and the European Union and EC FP7 grant no. 223705. Capewell is supported by the UK Higher Education Funding Council.

References

[1] British Heart Foundation. CHD statistics [internet]. 2012 [cited 2012 Nov 5]. Available from: http://www.heartstats.org.

[2] Stamler J, Stamler R, Neaton JD, Wentworth D, Daviglus ML, Garside D, Dyer AR, Liu K, Greenland P. Low risk-factor profile and long-term cardiovascular and noncardiovascular mortality and life expectancy: findings for 5 large cohorts of young adult and middle-aged men and women. JAMA 1999; 282(21):2012-8.

[3] Scarborough P, Bhatnagar P, Wickramasinghe KK, Allender S, Foster C, Rayner M. The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006–07 NHS costs. J Public Health 2011 Dec; 33(4):527-35.

[4] Bajekal M, Scholes S, Love H, Hawkins N, O'Flaherty M, Raine R, Capewell S. Analysing recent socioeconomic trends in coronary heart disease mortality in England, 2000-2007: a population modelling study. PLoS Medicine 2012; 9(6):e1001237.

[5] Barton P, Andronis L, Briggs A, McPherson K, Capewell S. Effectiveness and cost-effectiveness of cardiovascular disease prevention in whole populations: modelling study. BMJ 2011; 343:d4044.

[6] Labarthe DR. Preventing the risk to heart health, from womb to tomb. CVD Prevention 1998; 1:259-265.

[7] Wei LJ. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. StatMed 1992; 11:1871-1879.

[8] Berry JD, Dyer A, Cai X, Garside DB, Ning H, Thomas A, Greenland P, Van Horn L, Tracy RP, Lloyd-Jones DM. Lifetime risks of cardiovascular disease. N Engl J Med 2012 Jan 26; 366(4):321-9.

[9] Ford ES, Ajani UA, Croft JB, Critchley JA, Labarthe DR, Kottke TE, Giles WH, Capewell S. Explaining the decrease in US deaths from Coronary Disease, 1980 - 2000. N Engl J Med 2007; 356:2388-98.

[10] Couch PA, Ainsworth J, Buchan I. Sharable simulations of public health for evidence based policy making. Computer-Based Medical Systems (CBMS), 2011 24th International Symposium on; 2011 June 27-30; IEEE; 2011. p. 1-6.

[11] Lee A, Yau C, Giles MB, Doucet A, Holmes CC. On the Utility of Graphics Cards to Perform Massively Parallel Simulation of Advanced Monte Carlo Methods. J Comput Graph Stat 2010; 19(4):769-789.

[12] Ainsworth J, Buchan I. e-Labs and Work Objects: Towards Digital Health Economies. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2009; 16:205-216.

Address for correspondence

Philip Couch, Centre for Health Informatics, Institute of Population Health, University of Manchester, M13 9PL, UK.
Email: philip.couch@manchester.ac.uk