Building public health modelling capability and capacity in PHE

Julian Flowers, Jurgen Schmidt, Brian Ferguson

library(knitcitations)  
library(RefManageR)  
library(htmlTable)  
library(tidyverse)

## Loading tidyverse: ggplot2  
## Loading tidyverse: tibble  
## Loading tidyverse: tidyr  
## Loading tidyverse: readr  
## Loading tidyverse: purrr  
## Loading tidyverse: dplyr

## Conflicts with tidy packages ----------------------------------------------

## filter(): dplyr, stats  
## lag(): dplyr, stats

library(ztable)

## Welcome to package ztable ver 0.1.5

#Sys.setenv(https\_proxy = "https://158.119.150.18:8080")  
  
#ref <- ReadPubMed("laura webber", database = "PubMed", retmax = 1)  
  
  
#citep("10.1007/s10654-014-9978-0")

# Introduction

Modelling is increasingly important in public health practice and policy and we argue will become a larger part of PHE's work and remit over the next few years. By "model" we mean a simplification of reality and we are generally referring to quantitative - numerical models, although qualitative models are important to, for example, define the system we are trying to measure.

This paper:

1. Describes the rationale for PH modelling
2. Presents a stock-take of current modelling activity
3. Set out the case for a more systematic and resourced approach to modelling in PHE on behalf of PH and the wider PH system with a core modelling team, coordinationg activity within OGDs and across academia, and investment in capacity in capability

## Some background - what do we mean by modelling?

### A typology - some definitions

Can be classified by methods

* System dynamic modelling is a methodology and mathematical modeling technique to frame, understand, and discuss complex issues and problems. Originally developed in the 1950s to help corporate managers improve their understanding of industrial processes, SD is currently being used throughout the public and private sector for policy analysis and design
* Predictive modelling. Predictive modeling uses statistics to predict outcomes. Most often the event one wants to predict is in the future, but predictive modelling can be applied to any type of unknown event, regardless of when it occurred. It is synonymous with (supervised) machine learning
* Microsimulation (from microanalytic simulation) is a category of computerized analytical tools that perform highly detailed analysis of activities such as highway traffic flowing through an intersection, financial transactions, or pathogens spreading disease through a population. Microsimulation is often used to evaluate the effects of proposed interventions before they are implemented in the real world. For example, a traffic microsimulation model could be used to evaluate the effectiveness of lengthening a turn lane at an intersection, and thus help decide whether it is worth spending money on actually lengthening the lane.
* An agent-based model (ABM) is one of a class of computational models for simulating the actions and interactions of autonomous agents (both individual or collective entities such as organizations or groups) with a view to assessing their effects on the system as a whole. It combines elements of game theory, complex systems, emergence, computational sociology, multi-agent systems, and evolutionary programming. Monte Carlo methods are used to introduce randomness.
* Forecasting is the process of making predictions of the future based on past and present data and most commonly by analysis of trends.
* Econometric models are statistical models used in econometrics. An econometric model specifies the statistical relationship that is believed to hold between the various economic quantities pertaining to a particular economic phenomenon under study. An econometric model can be derived from a deterministic economic model by allowing for uncertainty, or from an economic model which itself is stochastic.

## Why model - the case for modelling?

PHE's remit on behalf of government is to improve and protect health and reduce health inequality. It acts through a range of levers and tools including influencing policy and practice, prioritisaon, resource allocation, service design, intervention, behaviour change and so on. PHE needs to be able to systematically monitor current health, identifiy priorities for action, project and predict future health states, test scenarios and hypotheses - all of these may need modelling.

In public health practice and policy making we use models and modelling techniques for a range of purposes [@Webber14]:

* **Filling data gaps** - for example obtaining estimates of disease frequency or burden where direct measurement maybe difficult or too costly.
* **Projecting future health states** - trying to understand current and potential future trends in risk factors, health determinants and disease outcomes
* **Testing interventions or policy options** - it can be very difficult to design controlled trials or 'gold standard' methodological design for evaluating interventions, so models can help in devising and testing scenarios, or altering levels of intervention, assessing impact and so on.
* **Understanding systems** - models (both qualitative and quantitative) can help us understand how systems *work* - the dynamics and interactions of complex systems - often present in public health problems and solutions and this may help in understanding why interventions may or may not work and devise new innovations
* **Testing and challenging assumptions** - all models are based on assumptions and some times these are not explicit and therefore not open to challenge. For example,
* **Understanding costs** of future health states and savings from intervention
* **Driving improvements in measuring health states and outcomes**. For example, if we build forecasts of future prevalence of adult obesity, we need a means of monitoring adult obesity rates which we can compare against our estimates, both to evaluate our model performance, and evaluate our efforts to tackle the problem.

## Some modelling principles

There are some important principles:

1. Models are only as good as the data feeding them
2. Models are only as good as the underlying assumptions (implicit or explicit) that are built in to the model
3. Producing quantitative outputs is important - numerical scale matters (how big is the problem? What practical scope is there for change?)
4. Models should be as simple as they need to be but no simpler - if models are too complex they may be unstable or give unpredictable results or be too difficult to explain - if they are too simple they may miss important interactions which affect the results or the model may lack crediblity
5. Models are representations not actualities - the results will be inherently uncertain even if presented as a point estimate - as far as possible results should be give with uncertainty estimates or sensitviity analysis
6. The best models should be a trade off between
   * parsimony (so we can understand the model),
   * accuracy (so we can trust the results),
   * interpretability (so we can understand the results)
   * and be implementable (so we can act on the results)
7. Models should behave - that is if we change the data the results should change in a plausible way. For example reducing smoking rates will create more ex-smokers who carry residual risk of a range of disease so we would expect prevalences of these disesae attributalbel to smoking to continue to rise for a period until, if we found that in 20 years time the model predicted increases in the populatoin admission rate, this would suggest some issue with the model. Similarly if we found that the outcome was very sensitive to one or more inputs, this may suggest a problem with the way the model captures the relationship between the input variables and the output.

\*\* Examples

* Agent based (< 80 Pumed articles pa with recent peak)
  + Examples are mainly focussed on the spread of communicable iidsease inlcuding flu and HIV, but there are a small number of examples of its application to physical activity, especially promoting walking, tackling obesity, smoking. Agent-based modelling approaches use microsimulation.
* Microsimulation
* Predictive modelling

|  |  |  |  |
| --- | --- | --- | --- |
| Model.type | Uses | Examples | In.house.Capacity.and.Capability |
| Agent-based | Complex systems and dynamic modelling | Interventions to promote physical activity or tackle obesity, Complex community interventions, Built environment impacts on health | No - no-in house capacity for application in non-communicable disease |
| Predictive modelling and machine learning | Prevalence estimates, trends, predictions, explanatory analysis | Estimating and predicting disease and risk factor prevalence, Identifying important variables to explain variation in activity, Estimating impacts of changes in risk factors | Some - need data science platform and wider use of R/ Python |
| Forecasting | Projections, time series, future states | Future health states, Activity trends | Limited - see predictive modelling |

## Currrent activities in PHE

There are a number of modelling teams or functions in PHE:

1. Health Protection Economic Modelling Team (check) led by Peter White and Andre Charlett. They create several models related to pandemic flu, vaccine control
2. The EDR team of mathematical modellers who model environmental and other hazards likely to affect crowds...
3. CRCE?
4. Economics team
5. Data science team in K&I
6. Other

## Activities in PHE.

[look up ]

## Current activities in the wider PH system

## Nexts steps

### Resources

### Standards

## References

### Annex

My understanding is that the modelling work in PHE falls broadly into four categories. CKO modelling work is primarily in categories 1. and 2. 1. Using models produced by others . Dispersion modelling o Dstl and MO are cross-gov leads on large scale dispersion modelling for deliberate and natural events respectively, for complex terrain and variable met. This means we take outputs from NAME/STE or HPAC on faith . Stuff that is simply too complex to produce in-house o Global burden of disease o Climate projections o Archimedes model . Predictive models - QRISK2, Framingham etc o These are typically based on logistic regression . Interpreting local estimates from prevalence and projection models I assume that most analytical and epidemiological staff can effectively use and interpret models and model results produced by others. 2. Statistical models Stuff underlined and in red are things that, as far as I know, PHE has not done - though that is not to say that staff do not have those skills (eg. my MSc dissertation uses both mixed models and GEE) . Standard statistical models o Linear regression o Generalised linear models ??? Logistic regression ??? Multinomial models ??? Poisson regression o Survival models ??? Time-to-event models . Cox regression . Competing risks regression . Accelerated failure time models ??? Relative/net survival models o More complex or unusual statistical models ??? Generalised additive models ??? Mixed models ??? Population-average models (GEE) . Time series methods o ARIMA o Complex time series models (State-space etc) o Forecasting methods (eg. age-period-cohort models, Holt-Winters method etc) . Monte Carlo simulation o Health economic models (though these may well be Bayesian) . Structural equation modelling o The Archimedes model appears to be a very complex example of SEM . Bayesian statistical models o Bayesian approaches to all other model types ??? Markov chain Monte Carlo models This category includes most of the modelling work done by staff in CKO. It includes projections modelling, prevalence modelling etc. It looks like some of the Health Protection modellers have more experience in the more complex things on this list than CKO staff, particularly Bayesian models. 3. Machine learning models . Classification models o Trees and random forests o Cluster analysis o Machine learning techniques o Support Vector Machines o Neural networks o Image analysis These can also be seen as statistical models, but I think are more viewed as machine learning; you can argue that all of machine learning is statistics if you want to - it annoys machine learning specialists and is quite fun. While I call these classification models, they can be used for prediction. We don't seem to use these very often in CKO, but I've used classification trees and cluster analysis myself and Neo has used random forests. It looks like these may be used more by the Health Protection modelling teams. 4. Other Mathematical models .  
. Mathematical modelling ??? . Special functions, probability theory o Dose response o Random variable based analysis (convolutions) o. ODEs and PDEs o Dispersion models ??? Gaussian puff/plume models underpin the reverse epidemiology tools and are more simple version of dispersion models used by Dstl/MO mentioned above. o Compartmental models ??? Basic building blocks for epidemiological analysis, uses ODE basis o Meta-population modelling o Network modelling o Pseudo-individual models o Stochastic ODEs and PDEs ??? All of the above again o Numerical models  
??? Numerical simulation . Agent based simulation . Discrete event simulation . Operational research models I don't think any CKO staff specialise in these models. It looks like these may be used more by the Health Protection modelling teams.