Risk factor modelling in PHE

Julian Flowers, Jurgen Schmidt, Brian Ferguson et al

Table of Contents

# Introduction

Models are simplifications of reality which allow us to describe, explain, simulate and predict how systems work. Population and public health systems are inherhently complex interplays of individual, environmental and social factors so if we want to try and understand how for example changes in health behaviour may inlfuence changes in health outcomes we inevitably need to construct some kind of model to help us understand what may happen - there is often not a simple linear relationship (for example, changes in smoking behaviour do no translate into comcomitant changes in mortality outcomes because there is a more complex shift in the distribution and temporality of risk).

For PHE modelling is increasingly important. PHE is committed to make the case for prevention and supporting investment by local authorities and others to improve health and tackle health inequalities. In order to do this we need to understand trends in risk factors and morbidity; we need to make forecasts about future levels of risk, disease, burden and outcome; we need evidence of effective interventions and their uptake to estimate their impact, and we need to evaluate the effect on cost and return on investment. To do this we need public health and heath economic modelling and we need qauntification.

PHE has prioritised tackling the risk factors with the largest contribution to disease burden or mortality. Some of these are clinical, some behaviourl and some environmental. \* Smoking \* Alcoohl consupmption \* Overwwieght and obeity \* Diet (especially sugar) \* UNderactivity \* Hypertension \* Atrial fibrillation \* Diabetes \* AIr pollution \* Early years \* Mental ill health

Often they are related - for example, alcohol consumption increases the prevalence of hypertension, smoking and alochol (independently and interactively) cause heart disease which underlies atrial fibtillation, overweight and undeactivity contribute to diabetes and so on, so models which accounts for these relationships and interactions may be necessary but are inevitably more complex and data hugry.

### A typology of models

A proposed.

1. *Models developed by others* - this is where the bulk of our effort has gone (see below)
2. *Statstical models* - this includes regression models, time series models and so on. Again we have done some work in this area
3. *Machine learning models* - this is at the core of the development of data science but as yet has had little impact on PHEs work
4. *Other mathematical models* - this includes the kinds of modelling done by the EDR team on environmental hazards for example of plumes caused by fires or toxic releases

Type 1. Too complex for us to build but need skills to run and interpret. Data hungry and need data linking risk with outcome so tend ot use HSE as care data. can be built in spreadsheets, with statistical tools like R or as bespoke software developments like UKHF models.

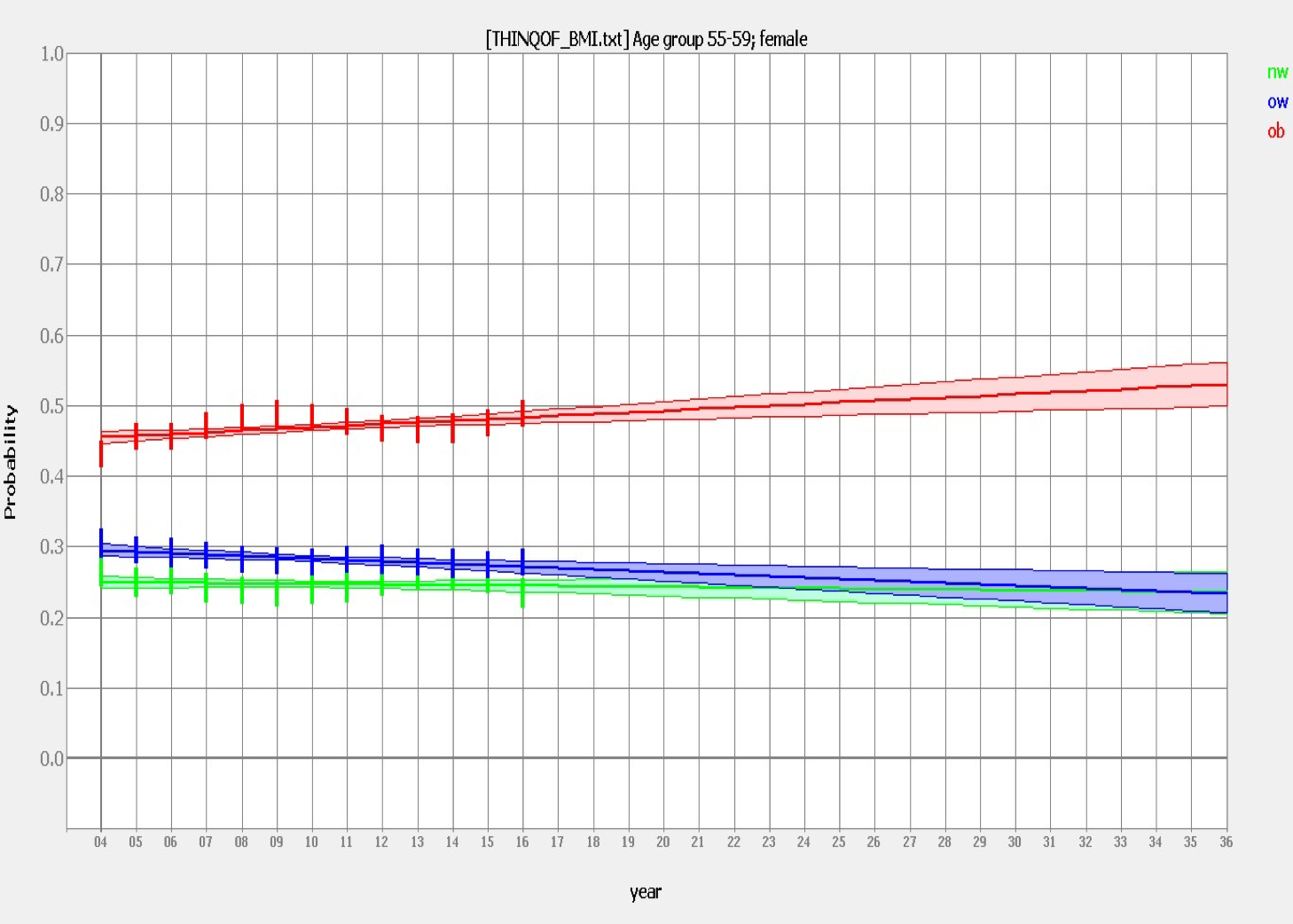
Current examples:

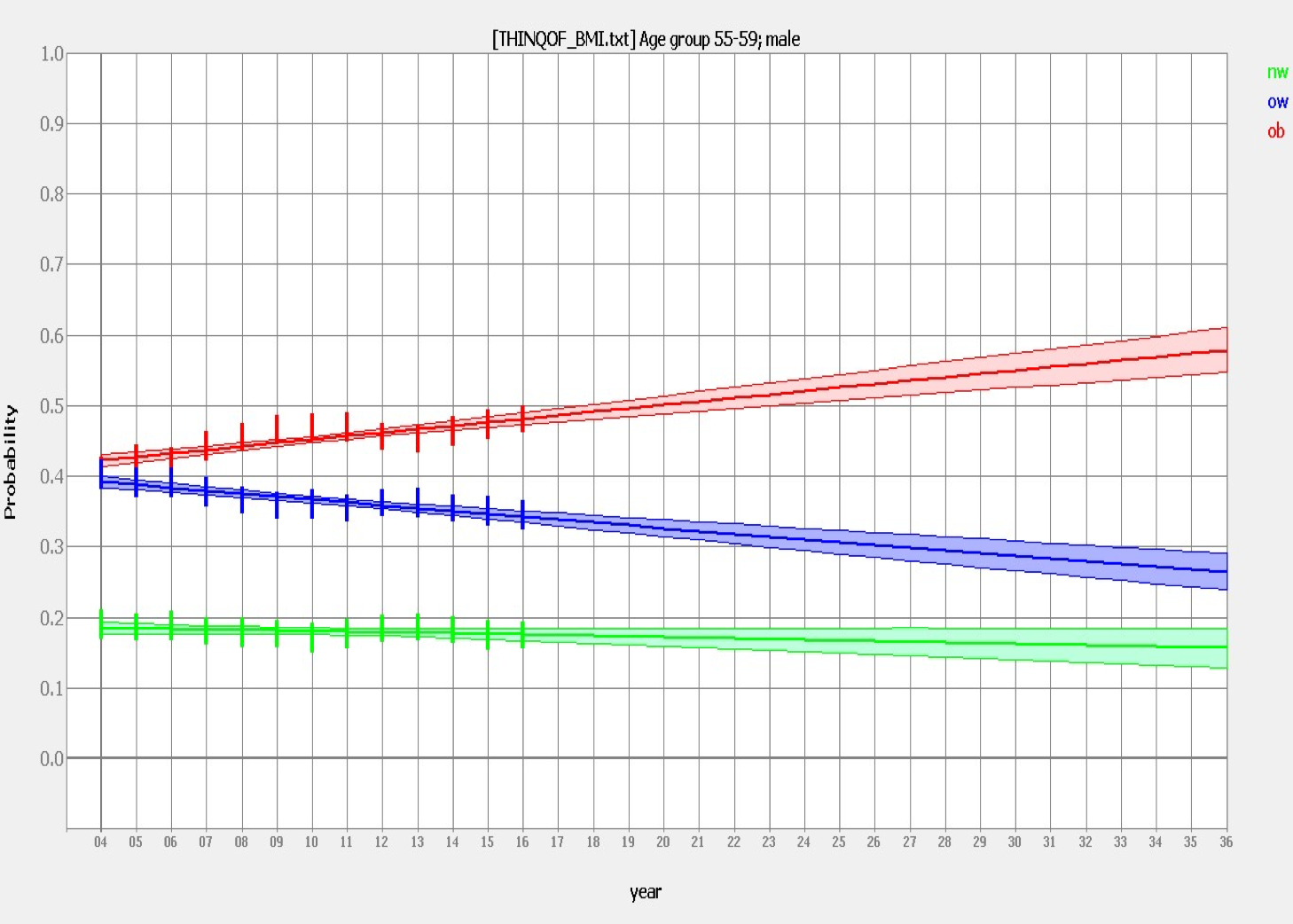
* UKHF Midrif model. Designed to predict a range of outcomes - disease, mortality adn cost - for single risk factors. A microsimulation model - that is it creates a distribtiun of populations with a set of characteristics who are aged over a lifetime and experience a set of events - birth, deaths, acquisition of ddisease and so on. It can account forthe templral relationship between risks and disease. It generally simulates large populations (say 20 million) and gives precise results. This model was used to estimate the impact of various scenarios tackling obesity, smoking and salt consumption, and a cut down version for air pollution.

The model is potentially very powerful and PHE owns the software licence, but the model is enorumously complex and our experience is that it requires at least one FULL TIME analyst to lear and run the model in its current form for us to benefit. The licence require software development on the part of UKHF to simplify thte user interface so that anlaysts in PHE can more easily run the model. This work is only part complete. If the model were considerablyt more user firendly then we cuold run a wide range of scenarios on national and simulated local populations.

THe model has 2 parts - forcasting risk factors and simulating health outcomes and costs. The forecasting risk factors is relatively straightforward compared to the full simulation and there has been some rcent owr on obesity forecasts using the UKHF tool fed with GP data from teh THIN database (a laorge database of GP records containing ~ 36 miliion BMI measurements) and comparision with the Health Survey for Enfalnd.

The figures belw show obesity projections based no model fitting to 12 years of THIN frmo 2004 (whe QOF began) to 2016 data for men and women aged 55-59. If ttrends contineu as they are by 2035 60% of men and 55% of women in this age group will be obese.





Our activity is focussed on 1, and to a lesser degree 2, but we need to develop capability to do 2 and 3. There will always be a need to commission complex simulation and systematic dynamic models - but we need them to be developed so they can be run in house and we have the capability to understand how they work and interpret the results.

## Current activity

There a currently 3 models of increasing complexity designed to help us assess the impact of tackling these

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 PHE and the wider PH system with a core modelling team, coordinating activity within OGDs and across academia, and investment in capacity and capability

## Summary

PHE has a vital role to play in public health modelling to help central and local governemnt invest approprately in prevention and population health improvement.

PHE lacks a coherent modelling strategy for what is rapidly becoming core business. PHE does have access to a range of models of variable utility and quality, and although it has some capacity and capability to run and interpret models developed by others it has no in-house development capacity. As a result it has relied on re-using models produced by others, or commissioning work from academics or private providers. There is inevitable duplication and inconsistency, for example different models coming to widely differening consclusions, models using different input data and so on.

PHE needs to take advantage of the guidance on Quality Analysis provided by the [Aqua Book](https://www.gov.uk/government/publications/the-aqua-book-guidance-on-producing-quality-analysis-for-government) to assure the quality of models PHE uses for its business rqeuirements.

A strategic approach to modelling would read across to wider strategies on, for example:

* Data - quality assured and open data, building a more comprehensive and systematic population health information system to ensure we have access to all the data we need to deliver our remit
* Data science - developing the platforms, tools, capacity and skills to make the best use of data we collect or reuse

It will ensure we have the input data to build models and that outputs are made open for sharing and reuse.

## Modelling risk factors

Tobacco, alcohol, obesity

### Prevalence

Surveys GP data New tools Granular and contemporaneous

GBD

### Forecasts and projections

* UKHF - single and multiple

### 

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

### A typology - some definitions

A typology proposed by Matthew Barclay (see annex) is:

1. *Models developed by others* - this is where the bulk of our effort has gone (see below)
2. *Statstical models* - this includes regression models, time series models and so on. Again we have done some work in this area
3. *Machine learning models* - this is at the core of the development of data science but as yet has had little impact on PHEs work
4. *Other mathematical models* - this includes the kinds of modelling done by the EDR team on environmental hazards

Our activity is focussed on 1, and to a lesser degree 2, but we need to develop capability to do 2 and 3. There will always be a need to commission complex simulation and systematic dynamic models - but we need them to be developed so they can by *run* in house and we have the capability to understand how they work and interpret the results.

### Some specifics

There are many types of quantitative modelling which are relevant to the work of Public Health England and public health practice including:

* *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 or may need to 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. For example, to monitor adult obesity as part of the Public Health Outceoms Frameowork (PHOF) we currently rely on an (expensive) survey. This does not provide data which is sufficiently granular or timely for our requirements. To fill the gap work on small area estimation of prevalence has been commissioned. There has been a limited amount of work on using exsiting sources and novel datasets to develop more timely estimates. We have developed national obesity estimates from a large sample GP dataset and cross checked these with the Health Survey for England.
* **Projecting future health states** - trying to understand current and potential future trends in risk factors, health determinants and disease outcomes. Some work has been done using the projection tool in the UKHF model to predict future rates of obesity by age. These have then been used to try adn simulate future health states related to obesity such as diabetes, heart disease and overall life expectancy.
* **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. Previous work was done using the UKHF model (see below) to test scenarios of obesity, salt consumption and smoking. The Future of Ill Health Model also provides tools to help model scenarios.
* **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.

Much of the modelling work done in PHE involves using or contributing to models created by others - there is very little in-house model development for non-communicable disease or public health risks and determinants.

|  |  |
| --- | --- |
| Modelling.activity | PHE.input |
| Global burden of disease | Data provision to IHME, commissioning, interpetation, networks, explanation, publication |
| MidRif (UKHF) | Data input, software development (by UKHF) to make model user friendly for PHE to run, limited projection and simulation |
| Economic models | Range of externally developed models for ROI and economic evaluation |
| Predictive models and forecasts | Limited activity - early experimentation with machine learning, Commissioned prevalence estimates from Imperial and Southampton |

## Current activities in the wider PH system

## Nexts steps

### Resources

### Standards

## References

### Annex

