Public Health England

Protecting and improving the nation's health

Appendix 10. Comparisons between the Air Quality tool and UK Health Forum microsimulation

About Public Health England

Public Health England exists to protect and improve the nation's health and wellbeing, and reduce health inequalities. We do this through world-leading science, knowledge and intelligence, advocacy, partnerships and the delivery of specialist public health services. We are an executive agency of the Department of Health and Social Care, and a distinct delivery organisation with operational autonomy. We provide government, local government, the NHS, Parliament, industry and the public with evidence-based professional, scientific and delivery expertise and support.

Public Health England
Wellington House
133-155 Waterloo Road
London SE1 8UG
Tel: 020 7654 8000
www.gov.uk/phe
Twitter: @PHE_uk
Facebook: www.facebook.com/PublicHealthEngland

Prepared by: UK Health Forum

Published May 2018
PHE publications
gateway number: 2017858

PHE supports the UN
Sustainable Development Goals

Contents

About Public Health England 2

Tool versus Microsimulation: differences 4

A comparison between the AQ Tool and UKHF Microsimulation Model 4

Comparisons in the literature between stochastic microsimulation models and deterministic tools 9

References 10

Tool versus Microsimulation: differences

A comparison between the Air Quality Tool and UK Health Forum Microsimulation Model

The air quality tool and the microsimulation model both generate future prevalence rates, the prevalence cases avoided with a reduction in the exposure levels and the cases attributable to air pollution. This allows for a comparison of the results generated by tool with the microsimulation. These simulations were generated alongside the development of the tool and use a 40% reduction of the published NO_2 dose response estimates for each disease. The results in this appendix are presented to illustrate the differences betweent the microsimulation model and the weighted cohort tool. The most recent version of the tool now uses a 60% reduction of the published NO_2 dose response based on updated COMEAP recommendations.

The tool simulates a weighted cohort through time whereas, the microsimulation model offers greater flexibility and enables the user to evaluate either a weighted cohort through time or a dynamic population through time. The microsimulation results presented in the main report correspond to simulating a dynamic population over time. In order to compare outputs from the tool and microsimulation, the microsimulation has been set up with a weighted cohort which is closed (unreplenished and no births) throughout the simulation. Although there are still some differences between each of the methods and assumptions. This comparison has focused on the relative changes between the baseline (no change) scenario and a scenario. In this example this is the cases attributable to NO_2. The absolute prevalence outputs estimated by the tool and microsimulation are different which is due to differenecs in the methods used to initialise both models. The initialisation methods are discussed in more detail in the later section.

A comparison has been drawn for the cases attributable to NO_2 between the tool and microsimulation. **Table 1** illustrates the prevalence cases per 100,00 attributable to NO_2 in England.

Table 1 Prevalence cases attributable per 100,000 to NO_2 in England by year for the microsimulation model and the air quality tool.

Year	Microsimulation model			Air Quality Tool		
	Asthma	Type 2 diabetes	Lung cancer	Asthma	Type 2 diabetes	Lung cancer
2015	24 [+-14]	67 [+-13]	5 [+-1]	0	0	0
2016	39 [+-14]	133 [+-13]	6 [+-1]	0	51	2
2017	56 [+-14]	197 [+-13]	8 [+-1]	0	92	3
2018	73 [+-14]	265 [+-13]	8 [+-1]	2	136	3
2019	89 [+-14]	329 [+-14]	9 [+-1]	0	178	3

2020	105 [+-14]	397 [+-14]	9 [+-1]	107	224	3
2021	123 [+-14]	463 [+-14]	11 [+-1]	109	271	3
2022	141 [+-14]	529 [+-14]	10 [+-1]	110	318	3
2023	157 [+-14]	597 [+-14]	12 [+-1]	1	356	5
2024	173 [+-16]	660 [+-14]	11 [+-1]	129	417	4
2025	190 [+-16]	726 [+-15]	12 [+-2]	131	457	3
2026	208 [+-16]	793 [+-16]	12 [+-3]	133	499	4
2027	223 [+-16]	858 [+-16]	13 [+-3]	117	530	5
2028	238 [+-16]	921 [+-16]	14 [+-3]	120	623	4
2029	255 [+-16]	987 [+-16]	15 [+-3]	141	636	4
2030	271 [+-16]	1051 [+-16]	15 [+-3]	124	631	4
2031	288 [+-16]	1117 [+-16]	15 [+-3]	146	647	4
2032	305 [+-16]	1183 [+-17]	16 [+-3]	129	770	3
2033	320 [+-17]	1247 [+-17]	16 [+-3]	283	804	4
2034	337 [+-17]	1309 [+-17]	17 [+-3]	267	934	6
2035	353 [+-17]	1374 [+-18]	18 [+-3]	272	952	5

The results from both the microsimulation and tool show the prevalence cases attributable to NO_2 increase overtime as the cohort ages during the simulation. In general, the predicted prevalence cases avoided in the tool are much lower than the microsimulation. For example in 2020 there were 105 cases of asthma compared with 107 in the tool per 100,000. For type 2 diabetes there was a much larger difference between the tool and microsimulation. In 2020 there were 224 attributable cases per 100,000 in the tool compared with 397 in the microsimulation.

The main reason for the differences observed in both of these predictions is related to the different methodological approaches used and the assumptions made in both the tool and microsimulation. These are described below.

Difference 1: Disease class

The key method of the disease class is to calculate an individual's risk (transition probability) of getting a disease based on their age, sex, current disease state, medical history and risk factor level. For stochastic transitions (microsimulation) this probability is compared to an application-generated random number to determine if the transition takes place. This allows for random effects which are a closer fit to reality.

In the deterministic tool this probability is included in the relevant life-disease table that both computes and lists the probabilities of being alive with no disease, within possible exclusive discrete disease states and dead. A maximum of 4 disease states are assumed to be possible for a given individual in the tool. Whereas, in the microsimulation there is no limit on the number of diseases that a person may have at any one time. The microsimulation model is more accurate because of the assumptions

used to model new cases of disease and that restrictions are not applied to the number of diseases that an individual may have in a year.

Difference 2: Risk factor trajectories

Categorical risk factor trends for NO_2 and $PM_{2.5}$ were generated from 2015 exposure data. The microsimulation model and tool both use these exposure data sets. The microsimulation uses a representative distribution of NO_2 and $PM_{2.5}$ trajectories over the whole population estimated from these categorical trends. For NO_2 and $PM_{2.5}$ a value will be sampled from this distribution and allocated to an individual in the simulation. Whereas, the tool uses only a small set of risk factor trajectories by age and sex group. The tool initialises three individuals from each age and sex group with an exposure level equal to the midpoint of each exposure group. Therefore, the tool will not model individuals at the boundaries of the distribution and risk factor groups. This method reduces the number of individuals that need to be simulated which keeps the computational time low, while ensuring that the whole distribution is sampled for each age and sex group within the population. The tool simulations take around 1 to 2 minutes to complete compared to the microsimulation model which takes around 8 hours to run a single scenario. The sampling method used in the tool is less accurate when compared to the method used in the microsimulation. This will have an impact on the output results because the relative risk data for some diseases is more granular than the risk factor groups sampled in the tool.

Difference 3: Population class

The microsimulation model is more flexible than the tool and is able to process any specified population or cohort; the deterministic tool processes only cohorts. A population is a specified number of males and females whose age distributions and risk factor distributions are input as appropriate tab delimited text files; for the tool, a *cohort* made up of weighted individuals is used where the weight is calculated as shown in equation (0.1).

$$\begin{aligned} &\text{cohort member weight}[i, j, k, l] = p_{sex}(i) \times p_{age}(j \mid i) \times p_{rf}(k \mid i, j) \\ &\text{where } i \in [0,1],\ j \in [0, n], k \in [0, 2] \end{aligned} \qquad (0.1)$$

Where,

$p_{sex}(i)$ is the probability of being male or female
$p_{age}(j|i)$ is the probability of having a certain age given sex
$p_{rf}(k|i,j)$ is the probability of being in a certain category (i.e. high NO_2 exposure, low $PM_{2.5}$ exposure, etc…) given sex and age.

The tool can be used to provide policy makers with the future impact of an intervention on a particular cohort in time. The microsimulation enables the impact of an intervention to be studied within a population which is dynamically changing through births and deaths.

Difference 4: Scenarios

Two scenarios were developed in the microsimulation to assess the impact of different scenarios on health and cost outcomes now and in the future (to 2035):

- an annual decrease by 1 µg/m^3 in $PM_{2.5}$ and NO_2 exposure for each individual.
- a European standard scenario whereby all the highly exposed (>40 µg/m^3) individuals in the population of interest decrease their NO_2 exposure to the exact European threshold of 40 µg/m^3.

It was not possible to model these interventions in the AQ tool because there are only 3 trajectories simulated which are based on the England tertile exposure cuts (for NO_2, the exposures in the first year are 10.5, 24.5 and 52.7 µg/m^3 and for $PM_{2.5}$, the exposures in the first year are 7.67, 12.9, 17.1 µg/m^3).

Consequently, decreasing the annual exposure by a 1 µg/m^3 or applying a European standard scenario in $PM_{2.5}$ and NO_2 would only affect 3 trajectories and might lead to greater uncertainties if further assumptions are not being made (compared to the microsimulation model where each individual has their own trajectory). Future work could involve the evaluation of such assumptions. However, taking into consideration the structure of the tool, we have modelled an intervention whereby a percentage of individuals in a ‘high risk’ exposure group can be moved to a ‘low risk’ exposure group.

The attributable prevalence cases have been compared in the tool and microsimulation as shown in **Table 1**. There are differences in the methods used to calculate the attributable prevalence cases. In the microsimulation each individuals NO_2 exposure level is set to 0.4 µg/m^3 and compared against a baseline simulation where an individuals exposure is sampled from the NO_2 distribution. Conversely, in the tool the attributable cases are calculated by setting the medium and high exposure group weightings to zero and redistributing these weightings into the low exposure group. The NO_2 level used to represent the low risk group is approximately 10.05 µg/m^3. These assumptions will lower the attributable prevalence cases predicted by the tool compared to the microsimulation.

An intervention has been implemented in the microsimulation model to reflect the intervention in the tool, where the attributable prevalence cases are calculated from setting the individuals risk to 10.05 µg/m^3. The results are shown in **Table 2**.

Table 2 Prevalence cases attributable per 100,000 to NO_2 in England by year for the microsimulation model and the air quality tool. The tool calculation has been implemented in the microsimulation model.

Year	Microsimulation model			Air Quality Tool		
	Asthma	Type 2 diabetes	Lung cancer	Asthma	Type 2 diabetes	Lung cancer
2015	11 [+-14]	45 [+-13]	3 [+-1]	0	0	0
2016	22 [+-14]	95 [+-13]	4 [+-1]	0	51	2
2017	33 [+-14]	146 [+-13]	6 [+-1]	0	92	3
2018	44 [+-14]	198 [+-13]	6 [+-1]	2	136	3
2019	55 [+-14]	251 [+-14]	7 [+-1]	0	178	3
2020	67 [+-14]	306 [+-14]	7 [+-1]	107	224	3
2021	80 [+-14]	362 [+-14]	8 [+-1]	109	271	3
2022	93 [+-14]	416 [+-14]	9 [+-1]	110	318	3
2023	105 [+-14]	471 [+-14]	11 [+-1]	1	356	5
2024	118 [+-16]	526 [+-14]	11 [+-2]	129	417	4
2025	131 [+-16]	586 [+-15]	12 [+-3]	131	457	3
2026	144 [+-16]	645 [+-16]	12 [+-3]	133	499	4
2027	158 [+-16]	705 [+-16]	13 [+-3]	117	530	5
2028	170 [+-16]	763 [+-16]	13 [+-3]	120	623	4
2029	183 [+-16]	824 [+-16]	14 [+-3]	141	636	4
2030	195 [+-16]	884 [+-16]	14 [+-3]	124	631	4
2031	209 [+-16]	947 [+-16]	15 [+-3]	146	647	4
2032	224 [+-16]	1008 [+-17]	16 [+-3]	129	770	3
2033	238 [+-17]	1072 [+-17]	17 [+-3]	283	804	4
2034	253 [+-17]	1134 [+-17]	18 [+-3]	267	934	6
2035	268 [+-17]	1199 [+-18]	18 [+-3]	272	952	5

The difference between the air quality tool and the microsimulation model predictions are much lower compared with the results in **Table 1**. For example in 2035 the tool predicts that 272 prevalence cases of asthma are attributable to NO_2 compared with 268 [±17] in the microsimulation model.

Based on these methodological differences with the interventions the attributable NO_2 prevalence cases predicted in the tool are likely to be lower than those predicted in the microsimulation model as referenced in the main report.

Difference 5: Initalisation

The microsimulation initialises individuals with diseases by simulating them from birth, before the start year of the simulation. The disease incidence is used as opposed to the prevalence data for the initialisation process. This method is used to allow the prevalence of disease to be initialised based on the risk factor exposure level as opposed to randomly within the population in the start year of the simulation. In the tool

the prevalence is initialised using the prevalence distribution for each disease. The prevalence rates are distributed randomly throughout each age and sex group as the prevalence rates by exposure group are unknown.

Comparisons in the literature between stochastic microsimulation models and deterministic tools

A main cause of the difference between results by the microsimulation programme and the Tool is from the discretization error. Both these simulations involve individuals being sampled from a distribution which will cause discretization error in both models. This level of error can be decreased by sampling more individuals from the distribution but it cannot be eliminated completely. This error has been identified in simulations in various fields (1-3).

The discretization error of a simulation is usually affected by 2 factors. The first one is the sampling rate or the sampling interval. In the case, where a function cannot be solved analytically, it is approximated with a list of sampling values. How well the sampling values approximate the function is decided by its sampling rate or sampling interval. The bigger the sampling rate (i.e., the smaller the sampling interval is), the better the approximation is. In other words, if the sample rate is not large enough, the sampling values cannot reflect the original function and this leads to discretization errors. Therefore, the microsimulation will produce more accurate results compared to the tool as it has much bigger sampling size (i.e., 50 million). The second factor is the regularity of the function. Given the same sampling rate or sampling interval, a function with a higher degree of regularity leads to less discretization errors than a function with a low level of regularity. The reason is that a higher degree of regularity means a lower level of complexity, therefore the function can be better approximated with the same sampling rate. However, as the microsimulation and the tool are using the same set of functions, the second factor does not affect our conclusion.

References

1. Hisham A. Al-Mharmah JMC. Comparison Of Monte Carlo And Deterministic Methods For Non-adaptive Optimization. Proceedings of the 1997 Winter Simulation Conference. 1997:348 -51.
2. J. Xu SBP. Assessment of Numerical Accuracy of PDF/Monte Carlo Methods for Turbulent Reacting Flows. Journal of Computational Physics. 1999:192 230.
3. Soren Asmussen PG, and Jim Pitman. Discretization Error in Simulation of One-Dimensional Reflecting Brownian Motion. The Annals of Applied Probability. 1995:875 - 96.