Public Health England

Protecting and improving the nation's health

Appendix 1. Technical appendix and data inputs

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

Data inputs

1 Population data

Demographic data was collected for England and Wales combined (no separate data for England was available), and for each UK local authority separately. Information was collected on the age and sex distribution of the population, the distribution of births by mother's age, the total fertility rate and the distribution of deaths by age and sex.

The data were processed as text files, in a format suitable for inclusion in the microsimulation programme. The data sources were as follows

Table 1. Population data sources by geography

Demography	Geography	Source
Total population by age and sex	England and Wales	ONS. Population Estimates for UK, England and Wales, Scotland and Northern Ireland: mid-2015. ONS; 2016.(1)
	Local authority	ONS. Population Estimates for UK, England and Wales, Scotland and Northern Ireland: mid-2015. ONS; 2016.(1)
Births by mothers age	England and Wales	ONS. Birth Summary Tables – England and Wales. 2015. ONS; 2016 (2)
	Local authority	ONS. Births by mothers' usual area of residence in the UK. ONS; 2016.(3)
Total fertility rate	England and Wales	ONS. Birth Summary Tables – England and Wales. 2015. ONS; 2016 (2)
	Local authority	ONS. Births by mothers' usual area of residence in the UK. ONS; 2016.(3)
Deaths by age and sex	England and Wales	ONS. Deaths registed in Engalnd and Wales 2015. ONS; 2016 (4)
	Local authority	ONS. Mortality statistics - underlying cause, sex and age. ONS; 2016.(5)

2 Disease data

A number of air pollution-related diseases were modelled (see Table 2). The list of diseases modelled for each pollutant was determined after a review of the literature available on the dose-response relationship between exposure to air pollutants and risk of incidence of disease (Table 2). Decisions were made by the advisory committee to include or exclude diseases based on existence of a dose-response relationship and availablility of epidemiological data. For instance, while an association exists between lung function in children and both NO_2 and $PM_{2.5}$, this condition is not a fixed disease

with identifiable prevalence, incidence and mortality data, so this condition was excluded from the disease list. Pre-term birth was also identified, but low-birth weight was considered a better proxy of the process of intrauterine growth reswtriction, and excluding pre-term births may avoid double counting.

Table 2. Characteristics of diseases modelled for each pollutant

	Duration	Terminal	Age category	Pollutant	
				NO_2	$PM_{2.5}$
Respiratory outcomes					
Asthma (children)	Chronic	Yes	Child	X	X
Asthma (adults)	Chronic	Yes	Adult	X	
COPD	Chronic	Yes	Adult		X
Cardiovascular outcomes					
CHD	Chronic	Yes	Adult		X
Stroke	Chronic	Yes	Adult		X
Diabetes	Chronic	No	Adult	X	X
Cancer and other outcomes					
Dementia	Chronic	Yes	Adult	X	
Low birth weight	Acute	No	Adult	X	X
Lung cancer	Chronic	Yes	Adult	X	X

All diseases except for low birth weight were lifelong, chronic diseases, so once acquired, were prevalent for the duration of an individual's life (see section *Module 2: Microsimulation model* on modelling birth weight for further details). Individuals could develop more than one diseases, but these were considered independent of one another. All diseases apart from diabetes and low birth weight were terminal.

Epidemiological data on each disease's incidence, prevalence, mortality and survival and dose-response was collected (see Table 3). When a parameter, eg Survival was not available from the literature or national statistics, this was computed – see *Module 2: Microsimulation model* section *Approximating missing disease statistics* for methods.

2.1 Summary of data sources

Table 3. Summary of disease data sources

Diseases	Incidence	Prevalence	Mortality	Survival	Relative Risk
Asthma	BLF Asthma Statistics (6)	BLF Asthma Statistics (6)	ONS, Deaths Registrations Summary Statistics, England and Wales, 2015 (4)	Computed from prevalence and mortality	**NO_2:** Khreis *et al.* 2016 (7) In children =<6 years : OR 1.08 (1.04; 1.12) per 4μg/m^3→*Converted to OR 1.212 (1.103; 1.328)* per 10μg/m^3→*REDUCED by 60%* → **1.08 (1.01; 1.12) per 10μg/m^3**

					In children >6 years: OR 1.03 (1.00; 1.06) per 4μg/m^3→*Converted to OR 1.08 (1.00; 1.16)* per 10μg/m^3→ *REDUCED by 60%* → **1.03 (1.00; 1.06) per 10μg/m^3** Jaquemin *et al.* 2015 (8) *In adults:* OR 1.10 (0.99;1.21) per 10μg/m^3 →*REDUCED by 60%* → **1.04 (0.996; 1.08*)* per 10μg/m^3** **$PM_{2.5}$:** Khreis et al. 2016 (7) In children >6 years: OR 1.04 (1.02; 1.07) per 1μg/m^3 →Converted OR **1.48 (1.22 ; 1.97) per 10μg/m^3**
COPD	Computed from prevalence and mortality.	PHE modelled estimates, 2008 (9)	ONS, Deaths Registrations Summary Statistics, England and Wales, 2015 (4)	Computed from prevalence and mortality	**$PM_{2.5}$:** COMEAP 2016 (10) COMEAP recommend using PM_{10} estimate based on Cai et al. 2014 estimate for chronic phlegm in never smokers in sensitivity analyses: OR 1.32 (1.02; 1.71) per 10μg/m^3 of PM_{10} → scale to $PM_{2.5}$ using the conversion factor of $PM_{2.5}$-> PM_{10}: 0.7 (or PM_{10} -> $PM_{2.5}$:1.42) recently used in the air quality index, COMEAP: Converted to **1.49 (1.03; 2.14) per 10μg/m^3 of $PM_{2.5}$**
CHD	Smolina et al 2012. Corrected data on incidence and mortality in 2013 (11)	BHF, Cardiovascular Disease Statistics 2014 (12)	ONS, Deaths Registrations Summary Statistics, England and Wales, 2015 (4)	Computed from prevalence and mortality	**$PM_{2.5}$:** Cesaroni et al. 2014 (13) Estimate used in CAPTOR tool from subgroup analysis of participants with additional information on CVD risk factors: HR 1.19 (1.01; 1.42) per 5μg/m^3→ Converted to **1.41 (1.00 - 2.01) per 10μg/m^3**
Diabetes	Personal communication with dr Craig Curry from Cardiff University	National Diabetes Audit 2015-2016(14)	Non-terminal	Non-terminal	**NO_2**: Eze *et al.* 2015 (15) RR 1.12 (1.05; 1.19) per 10μg/m^3→*REDUCED by 60%* → **1.05 (1.02; 1.07) per 10μg/m^3** **$PM_{2.5}$**: Eze et al. 2015 (15) RR **1.10 (1.02; 1.18) per 10μg/m^3**
Stroke	BHF, stroke	BHF, Cardiovasc	ONS, Deaths Registrations	Computed from	**$PM_{2.5}$:** Scheers et al. 2015 (17) HR 1.064 (1.021; 1.109) per 5μg/m^3

	statistics 2009 (16)	ular Disease Statistics 2014 (12)	Summary Statistics, England and Wales, 2015 (4)	prevalence and mortality	→Converted to **1.13 (1.04; 1.23) per 10μg/m^3**
Dementia	Computed from prevalence and mortality	Dementia UK 2014 (18)	ONS, Deaths Registrations Summary Statistics, England and Wales, 2015 (4)	Computed from prevalence and mortality	**NO_2**: Oudin et al. 2016(19) HR 1.08 (1.00; 1.16) per 10μg/m^3 NOx. Scaling factor: NOx → NO2: 0.44 which was developed by Anderson et al. based on the ratio that fell midway between the average or roadside vs urban background monitoring sites in London for 2001 (see Online Supp 2) (20) Converted from NOx to NO2:HR 1.03 (1.00; 1.07) →REDUCED by 60% → **1.01 (1.01; 1.03) per 10μg/m3 of NO2**
Low birth weight	ONS Birth Characteristics, 2015 (21)	Considered equivalent to incidence	Non-terminal	Non-terminal	**NO_2**: Pedersen et al. 2013 (22) OR 1.09 (1.00; 1.19) per 10μg/m^3→*REDUCED by 60%* → **1.04 (1.00; 1.07) per 10μg/m^3** **$PM_{2.5}$:** Pedersen et al. 2013 (22) OR 1.18 (1.06; 1.33) per 5μg/m^3→Converted OR **1.39 (1.12; 1.77) per 10μg/m^3:**
Lung cancer	CRUK, 2012-14 (23)	Not required in model as model uses incidence	CRUK, 2012-14 (23)	1, 5 year: ONS, 2010-14 (24); 10 year: ONS, 2008-12 (25)	**NO_2**: Hamra et al. 2015 (26) RR 1.04 (1.01; 1.08) per 10μg/m^3→*REDUCED by 60%* → **1.02 (1.00; 1.03) per 10μg/m^3** **$PM_{2.5}$:** Hamra et al. 2014 (27) **RR 1.09 (1.04; 1.14) per 10μg/m^3**
All NO2 relative risks reduced by 60% following COMEAP recommendations(28)					

2.2 Incidence, Prevalence, Mortality data by disease

Asthma

Table 4. Asthma epidemiological data (per 100,000 population)

Incidence		**Prevalence**		**Mortality**		
BLF Asthma Statistics(6)		BLF Asthma Statistics(6)		ONS 2015(4)		
Data sourced from the THIN Database, ICD codes unclear		Data sourced from the THIN Database, ICD codes unclear		ICD 10: J45-J46		
Age group	Both genders	Age group	Both genders	Age group	Male	Female
0-5	929.0	0-5	3114.0	<1	0.0	0.0
				1-4	0.1	0.0
6-10	561.0	6-10	10079.0	5-14	0.2	0.2
11-15	356.0	11-15	15899.0			
16-20	170.0	16-20	20180.0	15-24	0.2	0.2
21-30	150.0	21-30	17306.0			
				25-34	0.2	0.2
31-40	180.0	31-40	13286.0			
				35-44	0.4	0.3
41-50	201.0	41-50	11751.0			
				45-54	0.5	1.0
51-60	204.0	51-60	10903.0			
				55-64	1.1	1.4
61-70	231.0	61-70	10848.0			
				65-74	1.5	2.9
71-80	194.0	71-80	11526.0			
				75-84	5.2	12.3
81+	111.0	81+	10135.0	85+	32.3	56.0

Chronic obstructive pulmonary disease (COPD)

COPD incidence was estimated from prevalence and mortality data, – see *Module 2: Microsimulation model section Approximating missing disease statistics* for methods.

Table 5. COPD epidemiological data (per 100,000 population)

Incidence			**Prevalence**			**Mortality**		
UKHF Derived estimates			PHE Modelled estimates (2008) (9)			ONS 2015(4)		
Computed from Prevalence and Mortality			COPD based on FEV1 measurements in the Health Survey for England 2001 using British Thoracic Society criteria			ICD 10: J40-J44		
Age group	Male	Female	Age group	Male	Female	Age group	Male	Female
0-4	1.2	0.0	0-4	4.2	5.9	<1	0.0	0.0
						1-4	0.0	0.0
5-9	2.6	3.4	5-9	10.4	6.0	5-9	0.0	0.0
10-14	2.1	1.6	10-14	23.2	23.1			
15-19	0.0	0.0	15-19	33.7	31.3	15-24	0.0	0.0
20-24	1.8	1.3	20-24	27.6	12.9			
25-29	3.4	3.4	25-29	36.8	19.6	25-34	0.1	0.1
30-34	7.9	9.8	30-34	54.0	36.5			
35-39	25.0	28.2	35-39	93.3	85.5	35-44	1.0	0.5
40-44	52.0	60.4	40-44	218.0	226.3			
45-49	106.0	121.2	45-49	477.1	527.0	45-54	6.4	4.4
50-54	205.2	174.5	50-54	1003.6	1128.5			
55-59	298.0	235.4	55-59	2015.4	1988.1	55-64	32.0	28.2
60-64	491.9	326.9	60-64	3466.8	3136.4			
65-69	437.5	216.8	65-69	5817.7	4709.4	65-74	126.6	101.4
70-74	334.8	228.0	70-74	7860.1	5737.9			
75-79	263.4	0.0	75-79	9392.4	6807.6	75-84	367.7	283.2
80-84	0.0	0.0	80-84	10579.6	6783.3			
85+	0.0	0.0	85+	9857.29	5300.6	85+	939.45	558.4

Coronary heart disease (CHD)

Table 6. CHD epidemiological data (per 100,000 population)

Incidence			**Prevalence**			**Mortality**		
Smolina et al. 2012(11)			BHF CVD Stats 2014(12)			ONS 2015(4)		
ICD 10: I21-I22			ICD 10: I21			ICD 10: I21-I22		
Age group	Male	Female	Age group	Male	Female	Age group	Male	Female
			0-44	60.0	30.0	<1	0.0	0.3
						1-4	0.0	0.0
						5-9	0.0	0.0
						15-24	0.1	0.0
						25-34	0.8	0.2
30-54	88.1	21.2						
						35-44	4.9	1.4
			45-54	1070.0	430.0	45-54	21.2	5.2
55-64	317.0	90.3	55-64	4510.0	1240.0	55-64	52.2	14.2
65-74	533.0	237.0	65-74	8660.0	2960.0	65-74	109.4	44.6
75-84	1017.0	597.0	75+	14780.0	6960.0	75-84	281.0	146.0
85+	1987.0	1395.0				85+	692.0	454.7

Stroke

Table 7. Stroke epidemiological data (per 100,000 population)

Incidence			**Prevalence**			**Mortality**		
BHF Stroke Statistics 2009(16)			BHF CVD Stats 2014(12)			ONS 2015(4)		
Based on general practice records, ICD codes not given			ICD 10: I60-I69			ICD: I60-I64		
Age group	Male	Female	Age group	Male	Female	Age group	Male	Female
0-44	7.0	6.0	0-44	110.0	110.0	<1	0.3	0.0
						1-4	0.0	0.1
						5-9	0.1	0.2
						15-24	0.5	0.3
						25-34	0.9	1.0
						35-44	3.3	2.9
45-64	114.0	69.0	45-54	890.0	790.0	45-54	9.9	8.0
			55-64	2690.0	1960.0	55-64	26.1	19.1
65-74	393.0	275.0	65-74	6400.0	4390.0	65-74	67.7	56.7
75+	794.0	879.0	75+	14890.0	12430.0	75-84	284.6	260.5
						85+	915.9	1031.7

Diabetes Type 2

Table 8. Diabetes incidence and prevalence estimates (per 100,000 population)

Incidence			**Prevalence**			**Mortality**		
Personal communication with Dr Curry from Cardiff University (29)			National Diabetes Audit 2015-2016(14)					
ICD 10 codes unknown			ICD 10 codes unknown					
Age group	Male	Female	Age group	Male	Female	Age group	Male	Female
0-4	56	53	0-4	1.999	2.727	Non terminal		
5-9	34	42	5-9	6.681	6.372			
10-14	43	40	10-14	15	19.285			
15-19	83	107	15-19	41.744	64.613			
20-24	75	145	20-24	85.329	160.621			
25-29	101	226	25-29	202.748	352.739			
30-34	150	242	30-34	561.584	684.461			
35-39	240	263	35-39	1361.296	1249.819			
40-44	355	333	40-44	2617.251	1898.323			
45-49	561	482	45-49	4338.317	2858.298			
50-54	820	636	50-54	6451.945	4227.206			
55-59	1068	847	55-59	9371.893	6188.7			
60-64	1316	965	60-64	11825.85	7780.135			
65-69	1516	1234	65-69	13621.13	9047.041			
70-74	1763	1378	70-74	16010.86	11196.63			
75-79	1677	1483	75-79	18065.24	13559.67			
80-84	1645	1336	80-84	18464.43	14217.44			
85-89	1300	1169	85+	15210.91	11513.66			
90+	546	440						

Dementia

Dementia incidence was estimated from prevalence and mortality data, – see *Module 2: Microsimulation model section Approximating missing disease statistics* for methods.

In order to align the mortality age groups with the prevalence age groups, prevalence in age groups 85 to 89, 90 to 94 and 95+ were pooled . The pooling was weighted based on the number of cases in each age group, itself a function of population in each age group.

Table 9. Dementia epidemiological data (per 100,000 population)

Incidence			**Prevalence**			**Mortality**		
UKHF Derived estimate			Dementia UK 2014 (18)			ONS 2015(4)		
Computed from Prevalence and Mortality			ICD 10: F00-F03			ICD 10: F01,F03		
Age group	Male	Female	Age group	Male	Female	Age group	Male	Female
0-59	0	0				<1		
						1-4		
						5-14		
						15-24		
						25-34		
						35-44		
						45-54	0.1	0.2
						55-64	2.2	1.9
60-64	488.06	515.88	60-64	900	900			
65-69	304.79	492.8	65-69	1500	1800	65-74	30.4	26.9
70-74	1278.18	1031.38	70-74	3100	3000			
75-79	1422.62	2875.31	75-79	5300	6600	75-84	349.9	360.4
80-84	4042.55	2985.94	80-84	10300	11700			
85-89	3363.16	8573.13	85-89	15100	20200	85+	2025.7	2686.9
			90-94	22600	33000			
			95+	28800	44200			

Low birth weight

The outcome of low birth weight is related to several health outcomes throughout the life course, however, these downstream consequences was not be modelled in this project. Low birth weight is modelled as an outcome of the mother. Breakdown by maternal age was not available, and rates of 7% of all live births being low birth weight have been stable since 2011(21). Low birth weight was the only disease not modelled as a lifelong, chronic disease. Prevalence of low birth weight in 1 year is considered equivalent to incidence, as low birth weight is considered an acute event occurring only in a given year.

Table 10. Low birth weight epidemiological data (per 100,000 population)

Incidence			**Prevalence**			**Mortality**		
ONS 2015(21)								
ICD 10: P07.1								
Age grp	Male	Female	Age grp	Male	Female	Age grp	Male	Female
16-59	NA	7000	Prevalence of low birth weight in 1 year is considered equivalent to incidence.			Not applicable (non terminal)		

Lung cancer

Prevalence data was not available on lung cancer data, but the model does not require the input of prevalence, only of incidence, so this paramtere was not required.

Table 11. Lung cancer epidemiological data (per 100,000 population)

Incidence			**Prevalence**			**Mortality**		
CRUK 2012-14 (23)			Prevalence is not a required input into the model			CRUK 2012-14 (23)		
ICD 10: C33-C34			N/A			ICD 10: C33-C34		
Age group	Male	Female	Age group	Male	Female	Age group	Male	Female
0-4	0.1	0.0				0-5	0.0	0.0
5-9	0.0	0.0				5-9	0.0	0.0
10-14	0.0	0.0				10-14	0.0	0.0
15-19	0.1	0.1				15-19	0.0	0.0
20-24	0.3	0.3				20-24	0.1	0.0
25-29	0.5	0.6				25-29	0.1	0.1
30-34	0.9	1.3				30-34	0.4	0.5
35-39	2.3	2.4				35-39	1.4	1.1
40-44	6.7	5.6				40-44	4.3	3.0
45-49	15.9	15.1				45-49	10.9	9.6
50-54	36.0	34.0				50-54	24.9	21.6
55-59	80.0	72.7				55-59	58.1	48.0
60-64	151.5	126.2				60-64	109.8	84.1
65-69	239.5	190.5				65-69	176.2	131.1
70-74	367.9	264.4				70-74	279.9	190.9

75-79	463.2	305.5	75-79	367.8	234.1
80-84	538.1	337.6	80-84	453.6	283.6
85-89	608.7	339.1	85-89	564.4	309.4
90+	542.0	253.7	90+	521.2	245.6

2.3 Survival data

Survival statistics for CHD, COPD, Stroke and dementia were not identified in the literature. We modelled these using prevalence and mortality data, – see *Module 2: Microsimulation model* section *Approximating missing disease statistics* for methods.

Asthma

Table 12. Probability of 1, 5 and 1 year survival computed from prevalence and mortality data for Asthma.

Age	Survival probability – 1 year		Survival probability – 5 years		Survival probability – 10 years	
	M	F	M	F	M	F
0-74	1.000	1.000	1.000	1.000	1.000	1.000
75-84	1.000	0.999	1.000	0.999	1.000	0.999
85-94	0.998	0.997	0.998	0.997	0.998	0.997
95-107	0.999	0.997	0.999	0.997	0.999	0.997
108+	0.999	0.998	0.999	0.998	0.999	0.998

CHD and COPD

Table 13. Probability of 1, 5 and 1 year survival computed from prevalence and mortality data fpr Coronary Heart Disease and Chronic Obstructive Pulmonary Disease.

Age	CHD						COPD					
	Survival probability – 1 year		Survival probability – 5 year		Survival probability – 10 year		Survival probability – 1 year		Survival probability – 5 year		Survival probability – 10 year	
	M	F	M	F	M	F	M	F	M	F	M	F
1-5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	0.500	1.000	0.500	1.000	0.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	0.667	1.000	0.667	1.000	0.667	1.000	1.000	1.000	1.000	1.000	1.000	1.000
8	0.750	1.000	0.750	1.000	0.750	1.000	1.000	1.000	1.000	1.000	1.000	1.000
9	0.800	1.000	0.800	1.000	0.800	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0.833	1.000	0.833	1.000	0.833	1.000	1.000	1.000	1.000	1.000	1.000	1.000
11	0.857	1.000	0.857	1.000	0.857	1.000	1.000	1.000	1.000	1.000	1.000	1.000
12	0.875	1.000	0.875	1.000	0.875	1.000	1.000	1.000	1.000	1.000	1.000	1.000
13	0.889	1.000	0.889	1.000	0.889	1.000	1.000	1.000	1.000	1.000	1.000	1.000
14	0.900	1.000	0.900	1.000	0.900	1.000	1.000	1.000	1.000	1.000	1.000	1.000
15	1.000	0.000	1.000	0.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
16	1.000	0.500	1.000	0.500	1.000	0.500	1.000	1.000	1.000	1.000	1.000	1.000
17	1.000	0.667	1.000	0.667	1.000	0.667	1.000	1.000	1.000	1.000	1.000	1.000
18	1.000	0.750	1.000	0.750	1.000	0.750	1.000	1.000	1.000	1.000	1.000	1.000
19	1.000	0.800	1.000	0.800	1.000	0.800	1.000	1.000	1.000	1.000	1.000	1.000
20	1.000	0.833	1.000	0.833	1.000	0.833	1.000	1.000	1.000	1.000	1.000	1.000

21	1.000	0.857	1.000	0.857	1.000	0.857	1.000	1.000	1.000	1.000	1.000	1.000
22	1.000	0.875	1.000	0.875	1.000	0.875	1.000	1.000	1.000	1.000	1.000	1.000
23	1.000	0.889	1.000	0.889	1.000	0.889	1.000	1.000	1.000	1.000	1.000	1.000
24	1.000	0.900	1.000	0.900	1.000	0.900	1.000	1.000	1.000	1.000	1.000	1.000
25	0.792	0.849	0.792	0.849	0.792	0.849	1.000	1.000	1.000	1.000	1.000	1.000
26	0.828	0.868	0.828	0.868	0.828	0.868	1.000	1.000	1.000	1.000	1.000	1.000
27	0.853	0.884	0.853	0.884	0.853	0.884	1.000	1.000	1.000	1.000	1.000	1.000
28	0.872	0.896	0.872	0.896	0.872	0.896	1.000	1.000	1.000	1.000	1.000	1.000
29	0.887	0.906	0.887	0.906	0.887	0.906	1.000	1.000	1.000	1.000	1.000	1.000
30	0.898	0.914	0.898	0.914	0.898	0.914	0.991	0.992	0.991	0.992	0.991	0.992
31	0.908	0.921	0.908	0.921	0.908	0.921	0.996	0.996	0.996	0.996	0.996	0.996
32	0.915	0.927	0.915	0.927	0.915	0.927	0.997	0.997	0.997	0.997	0.997	0.997
33	0.922	0.932	0.922	0.932	0.922	0.932	0.998	0.998	0.998	0.998	0.998	0.998
34	0.928	0.936	0.928	0.936	0.928	0.936	0.998	0.998	0.998	0.998	0.998	0.998
35	0.507	0.595	0.507	0.595	0.507	0.595	0.991	0.989	0.991	0.989	0.991	0.989
36	0.670	0.712	0.670	0.712	0.670	0.712	0.992	0.991	0.992	0.991	0.992	0.991
37	0.752	0.776	0.752	0.776	0.752	0.776	0.993	0.992	0.993	0.992	0.993	0.992
38	0.801	0.817	0.801	0.817	0.801	0.817	0.994	0.993	0.994	0.993	0.994	0.993
39	0.834	0.845	0.834	0.845	0.834	0.845	0.995	0.993	0.995	0.993	0.995	0.993
40	0.858	0.866	0.858	0.866	0.858	0.866	0.995	0.994	0.995	0.994	0.995	0.994
41	0.875	0.882	0.875	0.882	0.875	0.882	0.995	0.995	0.995	0.995	0.995	0.995
42	0.889	0.894	0.889	0.894	0.889	0.894	0.996	0.995	0.996	0.995	0.996	0.995
43	0.900	0.905	0.900	0.905	0.900	0.905	0.996	0.995	0.996	0.995	0.996	0.995
44	0.909	0.913	0.909	0.913	0.909	0.913	0.996	0.996	0.996	0.996	0.996	0.996
45	0.637	0.583	0.637	0.583	0.637	0.583	0.985	0.985	0.985	0.985	0.985	0.985
46	0.734	0.706	0.734	0.706	0.734	0.706	0.986	0.986	0.986	0.986	0.986	0.986
47	0.790	0.773	0.790	0.773	0.790	0.773	0.987	0.986	0.987	0.986	0.987	0.986
48	0.826	0.815	0.826	0.815	0.826	0.815	0.987	0.987	0.987	0.987	0.987	0.987
49	0.852	0.844	0.852	0.844	0.852	0.844	0.988	0.988	0.988	0.988	0.988	0.988
50	0.871	0.865	0.871	0.865	0.871	0.865	0.989	0.988	0.989	0.988	0.989	0.988
51	0.886	0.881	0.886	0.881	0.886	0.881	0.989	0.989	0.989	0.989	0.989	0.989
52	0.898	0.894	0.898	0.894	0.898	0.894	0.990	0.989	0.990	0.989	0.990	0.989
53	0.907	0.904	0.907	0.904	0.907	0.904	0.990	0.990	0.990	0.990	0.990	0.990
54	0.915	0.912	0.915	0.912	0.915	0.912	0.990	0.990	0.990	0.990	0.990	0.990
55	0.702	0.638	0.702	0.638	0.702	0.638	0.979	0.977	0.979	0.977	0.979	0.977
56	0.771	0.734	0.771	0.734	0.771	0.734	0.981	0.980	0.981	0.980	0.981	0.980
57	0.813	0.790	0.813	0.790	0.813	0.790	0.983	0.982	0.983	0.982	0.983	0.982
58	0.843	0.827	0.843	0.827	0.843	0.827	0.985	0.984	0.985	0.984	0.985	0.984
59	0.864	0.852	0.864	0.852	0.864	0.852	0.986	0.986	0.986	0.986	0.986	0.986
60	0.880	0.871	0.880	0.871	0.880	0.871	0.987	0.987	0.987	0.987	0.987	0.987
61	0.893	0.886	0.893	0.886	0.893	0.886	0.988	0.988	0.988	0.988	0.988	0.988
62	0.903	0.898	0.903	0.898	0.903	0.898	0.989	0.989	0.989	0.989	0.989	0.989
63	0.912	0.907	0.912	0.907	0.912	0.907	0.989	0.989	0.989	0.989	0.989	0.989
64	0.919	0.915	0.919	0.915	0.919	0.915	0.990	0.990	0.990	0.990	0.990	0.990

65	0.757	0.766	0.757	0.766	0.757	0.766	0.981	0.973	0.981	0.973	0.981	0.973
66	0.804	0.810	0.804	0.810	0.804	0.810	0.983	0.976	0.983	0.976	0.983	0.976
67	0.836	0.840	0.836	0.840	0.836	0.840	0.984	0.979	0.984	0.979	0.984	0.979
68	0.859	0.862	0.859	0.862	0.859	0.862	0.985	0.981	0.985	0.981	0.985	0.981
69	0.876	0.879	0.876	0.879	0.876	0.879	0.986	0.983	0.986	0.983	0.986	0.983
70	0.890	0.892	0.890	0.892	0.890	0.892	0.987	0.984	0.987	0.984	0.987	0.984
71	0.901	0.902	0.901	0.902	0.901	0.902	0.988	0.985	0.988	0.985	0.988	0.985
72	0.910	0.911	0.910	0.911	0.910	0.911	0.988	0.986	0.988	0.986	0.988	0.986
73	0.917	0.918	0.917	0.918	0.917	0.918	0.989	0.987	0.989	0.987	0.989	0.987
74	0.923	0.924	0.923	0.924	0.923	0.924	0.989	0.988	0.989	0.988	0.989	0.988
75	0.817	0.825	0.817	0.825	0.817	0.825	0.975	0.966	0.975	0.966	0.975	0.966
76	0.845	0.851	0.845	0.851	0.845	0.851	0.977	0.970	0.977	0.970	0.977	0.970
77	0.865	0.870	0.865	0.870	0.865	0.870	0.978	0.973	0.978	0.973	0.978	0.973
78	0.881	0.884	0.881	0.884	0.881	0.884	0.980	0.976	0.980	0.976	0.980	0.976
79	0.893	0.896	0.893	0.896	0.893	0.896	0.981	0.978	0.981	0.978	0.981	0.978
80	0.903	0.906	0.903	0.906	0.903	0.906	0.982	0.980	0.982	0.980	0.982	0.980
81	0.911	0.914	0.911	0.914	0.911	0.914	0.983	0.981	0.983	0.981	0.983	0.981
82	0.918	0.920	0.918	0.920	0.918	0.920	0.984	0.982	0.984	0.982	0.984	0.982
83	0.924	0.926	0.924	0.926	0.924	0.926	0.984	0.983	0.984	0.983	0.984	0.983
84	0.929	0.931	0.929	0.931	0.929	0.931	0.985	0.984	0.985	0.984	0.985	0.984
85	0.846	0.879	0.846	0.879	0.846	0.879	0.966	0.957	0.966	0.957	0.966	0.957
86	0.865	0.892	0.865	0.892	0.865	0.892	0.969	0.962	0.969	0.962	0.969	0.962
87	0.880	0.902	0.880	0.902	0.880	0.902	0.971	0.965	0.971	0.965	0.971	0.965
88	0.892	0.910	0.892	0.910	0.892	0.910	0.973	0.968	0.973	0.968	0.973	0.968
89	0.902	0.917	0.902	0.917	0.902	0.917	0.974	0.971	0.974	0.971	0.974	0.971
90	0.910	0.923	0.910	0.923	0.910	0.923	0.975	0.973	0.975	0.973	0.975	0.973
91	0.917	0.928	0.917	0.928	0.917	0.928	0.977	0.975	0.977	0.975	0.977	0.975
92	0.922	0.933	0.922	0.933	0.922	0.933	0.978	0.976	0.978	0.976	0.978	0.976
93	0.927	0.937	0.927	0.937	0.927	0.937	0.979	0.977	0.979	0.977	0.979	0.977
94	0.932	0.940	0.932	0.940	0.932	0.940	0.979	0.979	0.979	0.979	0.979	0.979
95	0.936	0.943	0.936	0.943	0.936	0.943	0.980	0.980	0.980	0.980	0.980	0.980
96	0.939	0.946	0.939	0.946	0.939	0.946	0.981	0.981	0.981	0.981	0.981	0.981
97	0.942	0.948	0.942	0.948	0.942	0.948	0.982	0.981	0.982	0.981	0.982	0.981
98	0.945	0.951	0.945	0.951	0.945	0.951	0.982	0.982	0.982	0.982	0.982	0.982
99	0.947	0.953	0.947	0.953	0.947	0.953	0.983	0.983	0.983	0.983	0.983	0.983
100	0.949	0.955	0.949	0.955	0.949	0.955	0.983	0.984	0.983	0.984	0.983	0.984
101	0.951	0.956	0.951	0.956	0.951	0.956	0.984	0.984	0.984	0.984	0.984	0.984
102	0.953	0.958	0.953	0.958	0.953	0.958	0.984	0.985	0.984	0.985	0.984	0.985
103	0.955	0.960	0.955	0.960	0.955	0.960	0.985	0.985	0.985	0.985	0.985	0.985
104	0.956	0.961	0.956	0.961	0.956	0.961	0.985	0.986	0.985	0.986	0.985	0.986
105	0.958	0.962	0.958	0.962	0.958	0.962	0.985	0.986	0.985	0.986	0.985	0.986
106	0.959	0.963	0.959	0.963	0.959	0.963	0.986	0.986	0.986	0.986	0.986	0.986
107	0.960	0.964	0.960	0.964	0.960	0.964	0.986	0.987	0.986	0.987	0.986	0.987
108	0.962	0.965	0.962	0.965	0.962	0.965	0.986	0.987	0.986	0.987	0.986	0.987

109	0.963	0.966	0.963	0.966	0.963	0.966	0.986	0.987	0.986	0.987	0.986	0.987
109+	0.963	0.966	0.963	0.966	0.963	0.966	0.986	0.987	0.986	0.987	0.986	0.987

Stroke and Dementia

Table 14. Probability of 1, 5 and 1 year survival computed from prevalence and mortality data fpr Stroke and Dementia.

Age	Stroke						Dementia					
	Survival probability – 1 year		Survival probability – 5 year		Survival probability – 10 year		Survival probability – 1 year		Survival probability – 5 year		Survival probability – 10 year	
	M	F	M	F	M	F	M	F	M	F	M	F
1	1.000	0.988	1.000	0.988	1.000	0.988	1.000	1.000	1.000	1.000	1.000	1.000
2	1.000	0.994	1.000	0.994	1.000	0.994	1.000	1.000	1.000	1.000	1.000	1.000
3	1.000	0.996	1.000	0.996	1.000	0.996	1.000	1.000	1.000	1.000	1.000	1.000
4	1.000	0.997	1.000	0.997	1.000	0.997	1.000	1.000	1.000	1.000	1.000	1.000
5	0.998	0.995	0.998	0.995	0.998	0.995	1.000	1.000	1.000	1.000	1.000	1.000
6	0.998	0.996	0.998	0.996	0.998	0.996	1.000	1.000	1.000	1.000	1.000	1.000
7	0.998	0.996	0.998	0.996	0.998	0.996	1.000	1.000	1.000	1.000	1.000	1.000
8	0.998	0.997	0.998	0.997	0.998	0.997	1.000	1.000	1.000	1.000	1.000	1.000
9	0.999	0.997	0.999	0.997	0.999	0.997	1.000	1.000	1.000	1.000	1.000	1.000
10	0.999	0.997	0.999	0.997	0.999	0.997	1.000	1.000	1.000	1.000	1.000	1.000
11	0.999	0.998	0.999	0.998	0.999	0.998	1.000	1.000	1.000	1.000	1.000	1.000
12	0.999	0.998	0.999	0.998	0.999	0.998	1.000	1.000	1.000	1.000	1.000	1.000
13	0.999	0.998	0.999	0.998	0.999	0.998	1.000	1.000	1.000	1.000	1.000	1.000
14	0.999	0.998	0.999	0.998	0.999	0.998	1.000	1.000	1.000	1.000	1.000	1.000
15	0.995	0.997	0.995	0.997	0.995	0.997	1.000	1.000	1.000	1.000	1.000	1.000
16	0.995	0.997	0.995	0.997	0.995	0.997	1.000	1.000	1.000	1.000	1.000	1.000
17	0.996	0.997	0.996	0.997	0.996	0.997	1.000	1.000	1.000	1.000	1.000	1.000
18	0.996	0.997	0.996	0.997	0.996	0.997	1.000	1.000	1.000	1.000	1.000	1.000
19	0.996	0.997	0.996	0.997	0.996	0.997	1.000	1.000	1.000	1.000	1.000	1.000
20	0.996	0.997	0.996	0.997	0.996	0.997	1.000	1.000	1.000	1.000	1.000	1.000
21	0.997	0.998	0.997	0.998	0.997	0.998	1.000	1.000	1.000	1.000	1.000	1.000
22	0.997	0.998	0.997	0.998	0.997	0.998	1.000	1.000	1.000	1.000	1.000	1.000
23	0.997	0.998	0.997	0.998	0.997	0.998	1.000	1.000	1.000	1.000	1.000	1.000
24	0.997	0.998	0.997	0.998	0.997	0.998	1.000	1.000	1.000	1.000	1.000	1.000
25	0.948	0.994	0.948	0.994	0.948	0.994	1.000	1.000	1.000	1.000	1.000	1.000
26	0.950	0.994	0.950	0.994	0.950	0.994	1.000	1.000	1.000	1.000	1.000	1.000
27	0.952	0.994	0.952	0.994	0.952	0.994	1.000	1.000	1.000	1.000	1.000	1.000
28	0.953	0.994	0.953	0.994	0.953	0.994	1.000	1.000	1.000	1.000	1.000	1.000
29	0.955	0.994	0.955	0.994	0.955	0.994	1.000	1.000	1.000	1.000	1.000	1.000
30	0.956	0.995	0.956	0.995	0.956	0.995	1.000	1.000	1.000	1.000	1.000	1.000
31	0.958	0.995	0.958	0.995	0.958	0.995	1.000	1.000	1.000	1.000	1.000	1.000
32	0.959	0.995	0.959	0.995	0.959	0.995	1.000	1.000	1.000	1.000	1.000	1.000
33	0.960	0.995	0.960	0.995	0.960	0.995	1.000	1.000	1.000	1.000	1.000	1.000
34	0.962	0.995	0.962	0.995	0.962	0.995	1.000	1.000	1.000	1.000	1.000	1.000
35	0.987	0.986	0.987	0.986	0.987	0.986	1.000	1.000	1.000	1.000	1.000	1.000
36	0.987	0.987	0.987	0.987	0.987	0.987	1.000	1.000	1.000	1.000	1.000	1.000

37	0.987	0.987	0.987	0.987	0.987	0.987	1.000	1.000	1.000	1.000	1.000	1.000
38	0.988	0.987	0.988	0.987	0.988	0.987	1.000	1.000	1.000	1.000	1.000	1.000
39	0.988	0.988	0.988	0.988	0.988	0.988	1.000	1.000	1.000	1.000	1.000	1.000
40	0.988	0.988	0.988	0.988	0.988	0.988	1.000	1.000	1.000	1.000	1.000	1.000
41	0.989	0.988	0.989	0.988	0.989	0.988	1.000	1.000	1.000	1.000	1.000	1.000
42	0.989	0.989	0.989	0.989	0.989	0.989	1.000	1.000	1.000	1.000	1.000	1.000
43	0.989	0.989	0.989	0.989	0.989	0.989	1.000	1.000	1.000	1.000	1.000	1.000
44	0.989	0.989	0.989	0.989	0.989	0.989	1.000	1.000	1.000	1.000	1.000	1.000
45	0.977	0.976	0.977	0.976	0.977	0.976	1.000	1.000	1.000	1.000	1.000	1.000
46	0.982	0.980	0.982	0.980	0.982	0.980	1.000	1.000	1.000	1.000	1.000	1.000
47	0.985	0.983	0.985	0.983	0.985	0.983	1.000	1.000	1.000	1.000	1.000	1.000
48	0.987	0.985	0.987	0.985	0.987	0.985	1.000	1.000	1.000	1.000	1.000	1.000
49	0.989	0.987	0.989	0.987	0.989	0.987	1.000	1.000	1.000	1.000	1.000	1.000
50	0.990	0.988	0.990	0.988	0.990	0.988	1.000	1.000	1.000	1.000	1.000	1.000
51	0.991	0.989	0.991	0.989	0.991	0.989	1.000	1.000	1.000	1.000	1.000	1.000
52	0.992	0.990	0.992	0.990	0.992	0.990	1.000	1.000	1.000	1.000	1.000	1.000
53	0.993	0.991	0.993	0.991	0.993	0.991	1.000	1.000	1.000	1.000	1.000	1.000
54	0.993	0.992	0.993	0.992	0.993	0.992	1.000	1.000	1.000	1.000	1.000	1.000
55	0.983	0.981	0.983	0.981	0.983	0.981	1.000	1.000	1.000	1.000	1.000	1.000
56	0.984	0.982	0.984	0.982	0.984	0.982	1.000	1.000	1.000	1.000	1.000	1.000
57	0.985	0.984	0.985	0.984	0.985	0.984	1.000	1.000	1.000	1.000	1.000	1.000
58	0.986	0.984	0.986	0.984	0.986	0.984	1.000	1.000	1.000	1.000	1.000	1.000
59	0.987	0.985	0.987	0.985	0.987	0.985	1.000	1.000	1.000	1.000	1.000	1.000
60	0.988	0.986	0.988	0.986	0.988	0.986	0.845	0.858	0.845	0.858	0.845	0.858
61	0.988	0.987	0.988	0.987	0.988	0.987	0.866	0.876	0.866	0.876	0.866	0.876
62	0.989	0.987	0.989	0.987	0.989	0.987	0.882	0.890	0.882	0.890	0.882	0.890
63	0.989	0.988	0.989	0.988	0.989	0.988	0.894	0.901	0.894	0.901	0.894	0.901
64	0.990	0.988	0.990	0.988	0.990	0.988	0.904	0.910	0.904	0.910	0.904	0.910
65	0.977	0.970	0.977	0.970	0.977	0.970	0.431	0.438	0.431	0.438	0.431	0.438
66	0.980	0.974	0.980	0.974	0.980	0.974	0.637	0.640	0.637	0.640	0.637	0.640
67	0.982	0.977	0.982	0.977	0.982	0.977	0.734	0.735	0.734	0.735	0.734	0.735
68	0.983	0.979	0.983	0.979	0.983	0.979	0.790	0.791	0.790	0.791	0.790	0.791
69	0.985	0.981	0.985	0.981	0.985	0.981	0.826	0.827	0.826	0.827	0.826	0.827
70	0.986	0.983	0.986	0.983	0.986	0.983	0.852	0.852	0.852	0.852	0.852	0.852
71	0.987	0.984	0.987	0.984	0.987	0.984	0.871	0.871	0.871	0.871	0.871	0.871
72	0.988	0.985	0.988	0.985	0.988	0.985	0.886	0.886	0.886	0.886	0.886	0.886
73	0.989	0.986	0.989	0.986	0.989	0.986	0.897	0.898	0.897	0.898	0.897	0.898
74	0.989	0.987	0.989	0.987	0.989	0.987	0.907	0.907	0.907	0.907	0.907	0.907
75	0.960	0.949	0.960	0.949	0.960	0.949	0.482	0.445	0.482	0.445	0.482	0.445
76	0.964	0.956	0.964	0.956	0.964	0.956	0.658	0.642	0.658	0.642	0.658	0.642
77	0.967	0.962	0.967	0.962	0.967	0.962	0.744	0.736	0.744	0.736	0.744	0.736
78	0.969	0.966	0.969	0.966	0.969	0.966	0.796	0.790	0.796	0.790	0.796	0.790
79	0.972	0.969	0.972	0.969	0.972	0.969	0.830	0.826	0.830	0.826	0.830	0.826
80	0.973	0.972	0.973	0.972	0.973	0.972	0.854	0.852	0.854	0.852	0.854	0.852

81	0.975	0.974	0.975	0.974	0.975	0.974	0.872	0.870	0.872	0.870	0.872	0.870
82	0.977	0.976	0.977	0.976	0.977	0.976	0.887	0.885	0.887	0.885	0.887	0.885
83	0.978	0.978	0.978	0.978	0.978	0.978	0.898	0.896	0.898	0.896	0.898	0.896
84	0.979	0.979	0.979	0.979	0.979	0.979	0.907	0.906	0.907	0.906	0.907	0.906
85	0.935	0.922	0.935	0.922	0.935	0.922	0.645	0.581	0.645	0.581	0.645	0.581
86	0.938	0.926	0.938	0.926	0.938	0.926	0.734	0.699	0.734	0.699	0.734	0.699
87	0.941	0.930	0.941	0.930	0.941	0.930	0.787	0.764	0.787	0.764	0.787	0.764
88	0.944	0.933	0.944	0.933	0.944	0.933	0.821	0.805	0.821	0.805	0.821	0.805
89	0.946	0.936	0.946	0.936	0.946	0.936	0.846	0.833	0.846	0.833	0.846	0.833
90	0.948	0.939	0.948	0.939	0.948	0.939	0.864	0.853	0.864	0.853	0.864	0.853
91	0.950	0.942	0.950	0.942	0.950	0.942	0.878	0.869	0.878	0.869	0.878	0.869
92	0.951	0.944	0.951	0.944	0.951	0.944	0.889	0.881	0.889	0.881	0.889	0.881
93	0.953	0.946	0.953	0.946	0.953	0.946	0.899	0.891	0.899	0.891	0.899	0.891
94	0.955	0.948	0.955	0.948	0.955	0.948	0.906	0.900	0.906	0.900	0.906	0.900
95	0.956	0.950	0.956	0.950	0.956	0.950	0.913	0.907	0.913	0.907	0.913	0.907
96	0.957	0.951	0.957	0.951	0.957	0.951	0.918	0.912	0.918	0.912	0.918	0.912
97	0.958	0.953	0.958	0.953	0.958	0.953	0.923	0.917	0.923	0.917	0.923	0.917
98	0.960	0.954	0.960	0.954	0.960	0.954	0.927	0.922	0.927	0.922	0.927	0.922
99	0.961	0.956	0.961	0.956	0.961	0.956	0.931	0.926	0.931	0.926	0.931	0.926
100	0.962	0.957	0.962	0.957	0.962	0.957	0.934	0.929	0.934	0.929	0.934	0.929
101	0.963	0.958	0.963	0.958	0.963	0.958	0.937	0.932	0.937	0.932	0.937	0.932
102	0.964	0.959	0.964	0.959	0.964	0.959	0.939	0.935	0.939	0.935	0.939	0.935
103	0.964	0.960	0.964	0.960	0.964	0.960	0.942	0.937	0.942	0.937	0.942	0.937
104	0.965	0.961	0.965	0.961	0.965	0.961	0.944	0.939	0.944	0.939	0.944	0.939
105	0.966	0.962	0.966	0.962	0.966	0.962	0.946	0.941	0.946	0.941	0.946	0.941
106	0.967	0.963	0.967	0.963	0.967	0.963	0.948	0.943	0.948	0.943	0.948	0.943
107	0.967	0.964	0.967	0.964	0.967	0.964	0.949	0.945	0.949	0.945	0.949	0.945
108	0.968	0.965	0.968	0.965	0.968	0.965	0.951	0.946	0.951	0.946	0.951	0.946
109	0.969	0.965	0.969	0.965	0.969	0.965	0.952	0.948	0.952	0.948	0.952	0.948
109 +	0.969	0.965	0.969	0.965	0.969	0.965	0.952	0.948	0.952	0.948	0.952	0.948

Lung cancer

Table 15. Probability of 1, 5 and 1 year survival computed from prevalence and mortality data for Lung Cancer

Age	Survival probability – 1 year		Survival probability – 5 year		Survival probability – 10 year	
	M	F	M	F	M	F
15-39	0.615	0.663	0.615	0.663	0.615	0.663
40-49	0.751	0.774	0.751	0.774	0.751	0.774
50-59	0.784	0.802	0.784	0.802	0.784	0.802
60-69	0.790	0.818	0.790	0.818	0.790	0.818
70-79	0.825	0.859	0.825	0.859	0.825	0.859
>79	0.945	0.988	0.945	0.988	0.945	0.988

2.4 Relative risks

We searched all relevant and latest reports from Committee on the Medical Effects of Air Pollutants (COMEAP), Environmental Protection Agency (EPA) International Science Assessments (ISA), World Health Organization (WHO) for long-term effect estimates of exposure to the 2 pollutants on chronic diseases, focussing on respiratory and cardiovascular disease, and lung cancer. We also conducted PubMed searches for association with long-term effects by disease group, using a validated and published search strategy (30), see example Pubmed searches:

COPD:
((((("air pollution"[Title/Abstract] OR ozone[Title/Abstract] OR "particulate matter"[Title/Abstract] OR PM[Title/Abstract] OR "nitrogen dioxide"[Title/Abstract] OR "NO2"[Title/Abstract])) AND (COPD OR "chronic obstructive pulmonary disease" OR "chronic bronchitis" OR emphysema)) AND systematic[sb] AND "last 5 years"[PDat])

Cardiovascular outcomes:
(("air pollution"[Title/Abstract] OR ozone[Title/Abstract] OR "particulate matter"[Title/Abstract] OR PM[Title/Abstract] OR "nitrogen dioxide"[Title/Abstract] OR "NO2"[Title/Abstract])) AND (CVD[Title/Abstract] OR cardiovascular[Title/Abstract] OR stroke[Title/Abstract] OR cerebrovascular[Title/Abstract] OR "blood pressure"[Title/Abstract] OR hypertension[Title/Abstract] OR diabetes[Title/Abstract])

Effect estimates were included using the following inclusion exclusion and preference criteria:

Exposure: Measured or modelled annual exposure to NO_2 or $PM_{2.5}$. Short term studies including 24hour, 8 hour concentration data were not extracted.

Outcome: Incidence or prevalence of respiratory diseases, including asthma, COPD, bronchitis; cardiovascular diseases, including CHD, stroke, diabetes and lung cancer and emerging diseases related to air pollution, including dementia and low birthweight.

Source: COMEAP publications were prioritised, after which systematic reviews and meta-analyses preferred to RCT and single cohort studies, but these secondary latter sources were considered if no systematic reviews and meta-analyses for a given pollutant-outcome pair were identified. The most recent review and meta-analysis was preferred, but the inclusion/exclusion criteria and list of included studies was checked against other reviews to ensure the analysis was as complete and relevant as possible. When cohort and case-control/cross-sectional studies were included in the systematic review, these were included, but estimates from subgroup analyses of longitudinal cohort studies were preferred if available. Estimates from random effects meta-analyses

were extracted as between-study heterogeneity is expected in studies of air pollution and health outcomes. When studies presented models adjusting for confounding, the model adjusting for the most variables, was selected.

Effect estimates: Effects presented as relative risk ratios (RR), hazard ratios (HR) and odds ratios (OR) were extracted. ORs are considered to approximate RR if the effect is rare (which is not the case for the majority of outcomes included) or when the effect size is small ie above 20% increased odds (31). For estimates above OR =1.2, the baseline risk was assessed and was found to range between 2.4 and 13.1%. This would result in an overestimation of the RR by less than 10%, so all ORs were considered to approximate RRs. HRs were also considered to approximate RRs.

Effect estimates were presented for a range of exposure units and were standardised to 10μg/m^3 for all pollutants. When presented in parts per billion (ppb), conversion factors from Defra (32) were used to obtain estimates in μg/m^3. Where effect estimates are based on PM_{10}, a conversion factor of 1.43 was applied to convert the effect estimate to $PM_{2.5}$, as recommended and performed in the a review of the quality of air index(33). When using the. dose-response estimate for effect of NO_2 on dementia, the effect estimate was for NO_x which was converted to NO_2 using a scaling factor of 0.44. This factor was developed by Anderson et al. based on the ratio that fell midway between the average or roadside vs urban background monitoring sites in London for 2001 (see online supplement number 2) (20).

Communication with COMEAP and steering group members led to the decision to reduce all NO_2 effect estimates to 40% of the original RR (a reduction by 60%) to adjusts for the effect of $PM_{2.5}$ and other pollutants. This represents the mid point of range of 25-55% reduction recommended to be applied to unadjusted coefficients to account for the effect of PM 2.5 and other pollutants (28). Table 16 presents features from the studies from which dose-response estiamtes were obtained.

Table 16. Summary of studies identified potential associations between long-term exposure to NO_2 and $PM_{2.5}$ of air pollution on chronic disease

Outcome	Reference / Source	Population	Exposure measurement	Outcome measurement	Effect size
		NO_2 dose-response estimates			
Asthma (children)	Khreis et al. 2016 (7) meta-analysis of 7 studies	Children aged ≤6 years	Monitoring data, LUR or dispersion modelling of annual average (per 4μg/m3)	Varied: parental/self-report of doctor diagnosis; on treatment; episodes of wheeze; hospital discharge;	OR 1.08 (1.04; 1.12)
	Khreis et al. 2016 (7)	Children			OR 1.03

	meta-analysis of 14 studies	aged >6 years			(1.00; 1.06)
Asthma (adults) [59]	Jaquemin et al. 2015 (8) ESCAPE project (6 European Cohorts) including the ECHRS – mentioned in EPA ISA 2016	23704 participants with 1257 incident cases of asthma	LUR models of annual average at home address (per 10μg/m3)	Questionnaire (excludes asthmatics & symptomatics within 1 year of baseline)	OR 1.10 (0.99;1.21)
Diabetes	Eze et al. 2015 (15) meta-analysis of 5 longitudinal studies (other results include CC)	Males & females	LUR models & satellite-derived estimates of annual average at home address (per 10μg/m3)	Doctor diagnosed or antidiabetic medication use	RR 1.12 (1.05; 1.19)
Lung cancer	Hamra et al. 2015(26) meta-analysis of 15 studies	Europe, North American & rest of world	Fixed site monitoring data; LUR, dispersion, spatiotemporal, inverse-distance modelling of annual average exposure (per 10μg/m3)	Not reported (lung cancer incidence & mortality)	RR 1.04 (1.01; 1.08)
Low birth weight	Pedersen et al. 2013 (22) ESCAPE project	61452 women with singleton term births in Europe. 4.2% LBW	LUR models of annual average at home address (per 10μg/m3)	Questionnaire & interview	OR 1.09 (1.00; 1.19)
Dementia	Oudin et al. 2016(19) Prospective cohort study (Sweden)	1806 participants from the Betula study in Sweden with no AD at baseline	LUR models of annual average of NOx at home address (per 10μg/m3).	Medical record linkage for dementia diagnosis codes (DSM-IV criteria)	HR 1.08 (1.00; 1.16) * note, for NO_x not NO_2. Converted from NO_x to NO_2: HR 1.03 (1.00; 1.07)
		$PM_{2.5}$ dose-response estimates			
Asthma (children)	Khreis et al. 2016 7) (Meta-analysis) 8 studies (including all relevant studies previously identified from ISA EPA 2016	Children aged >6 years	Monitoring data, LUR or dispersion modelling of annual average (per 1μg/m3)	Varied: parental/self-report of doctor diagnosis; on treatment; episodes of wheeze; hospital discharge; combinations of above criteria	OR 1.04 (1.02; 1.07)
COPD/ Chronic Bronchitis	COMEAP 2016 (10) Most recent review– decided to focus on	No association with PM2.5	Annual average (per 10μg/m3 of PM10)		OR 1.32 (1.02; 1.71)

	Chronic Bronchitis instead of COPD	but recommend using PM 10 estimate based on Cai et al. 2014 35 estimate for chronic phlegm in never smokers in sensitivity analyses:			Can scale to $PM_{2.5}$ using the conversion factor of $PM_{2.5}$-> PM_{10}: 0.7 (or PM_{10} -> $PM_{2.5}$:1.42) recently used in the air quality index, COMEAP: Converted to per 10μg/m3 of $PM_{2.5}$: OR 1.49 (1.03; 2.14)
CHD	Cesaroni et al. 2014 (13) ESCAPE project: 11 European Cohorts: Estimate used in CAPTOR tool from subgroup analysis of participants with additional information on CVD risk factors	10,166 participants with 5157 incident events of acute MI or other acute IHD	LUR models of annual average at home address (per 5μg/m3)	Linkage to hospital discharge & mortality registries, excluding coronary hospitalisations 28d before event	HR 1.19 (1.01; 1.42)
Stroke	Scheers et al. 2015 (17) (Meta-analysis) 10 studies from	>10 million people Europe & America	Fixe monitoring data; LUR models of annual average at home address (per 5μg/m3)	Not reported (but outcome stroke incidence or mortality, so death or hospitalisation registry linkage)	HR 1.064 (1.021; 1.109)
Diabetes	Eze et al. 2015 (15) (meta-analysis) 5 longitudinal studies (other results include case-control design)	Males & females	LUR models & satellite-derived estimates of annual average at home address (per 10μg/m3)	LUR models & satellite-derived estimates of annual average at home address (per 10μg/m3)	RR 1.10 (1.02; 1.18)
Lung cancer	Hamra et al. 2014 (27) (meta-analysis) 13 studies	Europe, North American & rest of world	Fixed site monitoring data; LUR, dispersion, spatiotemporal, inverse-distance modelling of annual average exposure (per 10μg/m3)	Not reported (lung cancer incidence & mortality)	RR 1.09 (1.04; 1.14)
Low birth weight	Pedersen et al. 2013 (22) ESCAPE project	50151 women with singleton term births in Europe. 1.3% LBW	LUR models of annual average at home address (per 5μg/m3)	Questionnaire & interview	OR 1.18 (1.06; 1.33)

3 Health economic data

3.1 Sources of cost data

Inpatient costs came from the Hospital Episodes Statistics (HES) dataset. We identified all episodes in 2010 with a primary diagnosis matching the list of ICD-10 codes in **Table 17**. The episode's HRGs version 4.0 were matched to the national tariffs and adjusted for the Market Factor Forces (MFF).

Table 17. List of ICD-10 codes

Disease	ICD code
Asthma	J45
COPD	J40-J47
CHD	I20-I25
Stroke	I60-I63
Diabetes	E10, E11 and O24.4
Dementia	F00-F03
Lung Cancer	C34
Respiratory	J00-J99
Cardiovascular	I00-I99

Outpatient costs, extracted from the literature (see Appendix 5 for more details), were combined with the inpatient costs to estimate the hospitalisation costs.

All the other costs came from the literature, with the exception of outpatient costs for asthma which we proxied with the NHS programme budget.

The microsimulation mdoel uses cost per case to calculate the total healthcare costs incurred due to the prevalence of disease in a scenario.

Costs were available from the Imperial Business School in 3 different formats, see appendix 5: they were provided as total costs (in £ million) for England or the UK depending on the study; as costs per case, or as cost per death:
All of the costs were inflated from the original cost year to 2015. In order to obtain a cost per case for each health care category, and for each disease:

1. Total costs were divided by the total number of cases in the cost year for the country. This method assumed that the prevalence of diseases obeained for the model for each disease was constant over time, as it was not possible to obtain prevalence estimates for each of the diseases, for each of the years of cost. However, the prevalence used in the microsiimulation model was scaled up to the population in the year of the cost, in order to

more accurately reflect the total number of cases of diseases which contributed to the total cost in that year.

2. Cost per case inflated to 2015 (eg for asthma outpatient costs for instance) were used directly.
3. When costs were given per death, they were inflated to 2015 and then adjusted for the ratio of deaths to cases for that disease.

Table 18. Summary of the cost per case identified in the literature per type of care by chronic disease (annual average £ per case)

Cost type (£ per case)	Primary Care	Social Care	Medication	Hospitalisation
Asthma	21.28	0.50	87.57	27.02
Chronic Obstructive Pulmonary Disease (COPD)	400.43	85.30	126.79	587.48
Coronary Heart Disease (CHD)	71.57	109.70	818.60	1460.46
Stroke	36.45	76.05	504.10	722.84
Diabetes	375.00	601.56	276.88	536.75
Lung cancer	51.73	89.38	35.10	466.63
Dementia	430.62	12348.93	310.24	197.24

UKHF microsimulation methodology

1 Microsimulation framework

Our simulation consists of 2 modules. The first module calculates the predictions of risk factor trends over time based on data from rolling cross-sectional studies. The second module performs the microsimulation of a virtual population, generated with demographic characteristics matching those of the observed data. The health trajectory of each individual from the population is simulated over time allowing them to contract, survive or die from a set of diseases or injuries related to the analysed risk factors. The detailed description of the 2 modules is presented below.

2 Microsimulation Module one: Predictions of NO_2 and $PM_{2.5}$ over time

NO_2 and $PM_{2.5}$ are analysed within the model as risk factors (RF), as described in Table 19.

Table 19 Description of the categories used for the risk factors NO_2 and $PM_{2.5}$

Risk factor (RF)	Number of categories (N)	Categories
Nitrogen dioxide (NO_2)	3	NO_2 < 20.5 µg m-3 NO_2 from 20.5 to 28.5 µg m-3 NO_2 ≥ 28.5 µg m-3
Particulate matter ($PM_{2.5}$)	3	$PM_{2.5}$ < 12.3 µg m-3 $PM_{2.5}$ from 12.3 to 13.5 µg m-3 $PM_{2.5}$ ≥ 13.5 µg m-3

For the RF, let *N* be the number of categories for a given risk factor, eg *N* = 3 for NO_2. Let k = 1, 2, …, *N* number these categories and $p_k(t)$ denote the prevalence of individuals with RF values that correspond to the category k at time *t*. We estimate $p_k(t)$ using multinomial logistic regression model with prevalence of RF category k as the outcome, and time *t* as a single explanatory variable. For $k < N$, we have

$$\ln\left(\frac{p_k(t)}{p_1(t)}\right) = \beta_0^k + \beta_1^k t \qquad (0.1)$$

The prevalence of the first category is obtained by using the normalisation constraint $\sum_{k=1}^{N} p_k(t) = 1$. Solving equation (0.1) for $p_k(t)$, we obtain

$$p_k(t) = \frac{\exp\left(\beta_0^k + \beta_1^k t\right)}{1 + \sum_{k'=1}^{N} \exp\left(\beta_0^{k'} + \beta_1^{k'} t\right)}, \qquad (0.2)$$

which respects all constraints on the prevalence values, ie normalisation and [0, 1] bounds.

2.1 Multinomial logistic regression for each risk factor

Measured data consist of sets of probabilities, with their variances, at specific time values (typically the year of the survey). For any particular time the sum of these probabilities is unity. Typically such data might be the probabilities of low, medium and high pollution exposure, as they are extracted from the survey data set. Each data point is treated as a normally distributed[1] random variable; together they are a set of N groups (number of years) of K probabilities $\{\{t_i, \mu_{ki}, \sigma_{ki} | k \in [0,K\text{-}1]\} | i \in [0,N\text{-}1]\}$. For each year the set of K probabilities form a distribution – their sum is equal to unity.

The regression consists of fitting a set of logistic functions $\{p_k(\mathbf{a}, \mathbf{b}, t) | k \in [0,K\text{-}1]\}$ to these data – one function for each k-value. At each time value the sum of these functions is unity. Thus, for example, when measuring NO_2 in the 3 states already mentioned, the $k = 0$ regression function represents the probability of low pollution exposure over time, $k = 1$ the probability of medium pollution exposure and $k = 2$ the probability of high pollution exposure.

The regression equations are most easily derived from a familiar least square minimization. In the following equation set the weighted difference between the measured and predicted probabilities is written as S; the logistic regression functions $p_k(\mathbf{a},\mathbf{b};t)$ are chosen to be ratios of sums of exponentials (This is equivalent to modelling the log probability ratios, p_k/p_0, as linear functions of time).

$$S(\mathbf{a},\mathbf{b}) = \tfrac{1}{2} \sum_{k=0}^{k=K-1} \sum_{i=0}^{i=N-1} \frac{\left(p_k(\mathbf{a},\mathbf{b};t_i) - \mu_{ki}\right)^2}{\sigma_{ki}^2} \qquad (0.2)$$

[1] Depending on the circumstances this assumption will be more or less accurate and more or less necessary. In general, it is both extremely useful and accurate. For simple surveys the individual Bayesian prior and posterior probabilities are Beta distributions – the likelihood being binomial. For reasonably large samples, the approximation of the beta distributions by normal distributions is both legitimate and a practical necessity. For complex, multi-PSU, stratified surveys, it is again assumed that these base probabilities are approximately normally distributed and, again, it is an assumption that makes the analysis tractable.

Depending on the nature of the raw data set it may be possible to use non-parametric statistical methods for this analysis. This is possible for the HSE and GHS data sets of this study but when this has been done the authors can report no discernible difference in the results.

$$p_k(\mathbf{a},\mathbf{b},t) \equiv \frac{e^{A_k}}{1+e^{A_1}+..+e^{A_{K-1}}}$$
$$\mathbf{a} \equiv (a_0,a_1,..,a_{K-1}), \quad \mathbf{b} \equiv (b_0,b_1,..,b_{K-1}) \qquad (0.2)$$
$$A_0 \equiv 0, \quad A_k \equiv a_k + b_k t$$

The parameters A_0, a_0 and b_0 are all zero and are used merely to preserve the symmetry of the expressions and their manipulation. For a K-dimensional set of probabilities there will be 2(K-1) regression parameters to be determined.
For a given dimension K there are K-1 independent functions p_k – the remaining function being determined from the requirement that complete set of K form a distribution and sum to unity.

Note that the parameterization ensures that the necessary requirement that each p_k be interpretable as a probability – a real number lying between 0 and 1.
The minimum of the function S is determined from the equations

$$\frac{\partial S}{\partial a_j} = \frac{\partial S}{\partial b_j} = 0 \qquad \text{for j=1,2,....,k-1} \qquad (0.2)$$

noting the relations

$$\frac{\partial p_k}{\partial A_j} = \frac{\partial}{\partial A_j}\left(\frac{e^{A_k}}{1+e^{A_1}+..+e^{A_{K-1}}}\right) = p_k\delta_{kj} - p_k p_j$$
$$\frac{\partial}{\partial a_j} = \frac{\partial}{\partial A_j} \qquad (0.3)$$
$$\frac{\partial}{\partial b_j} = t\frac{\partial}{\partial A_j}$$

The values of the vectors **a**, **b** that satisfy these equations are denoted $\hat{\mathbf{a}},\hat{\mathbf{b}}$. They provide the trend lines $p_k(\hat{\mathbf{a}},\hat{\mathbf{b}};t)$, for the separate probabilities. The confidence intervals for the trend lines are derived most easily from the underlying Bayesian analysis of the problem.

2.2 Bayesian interpretation

The 2K-2 regression parameters {**a**,**b**} are regarded as random variables whose posterior distribution is proportional to the function exp(-S(**a**,**b**)). The maximum likelihood estimate of this probability distribution function, the minimum of the function S, is obtained at the values $\hat{\mathbf{a}},\hat{\mathbf{b}}$. Other properties of the (2K-2)-dimensional probability distribution function are obtained by first approximating it as a (2K-2)-dimensional normal distribution whose mean is the maximum likelihood estimate. This amounts to expanding the function S(**a**,**b**) in a Taylor series as far as terms quadratic in the differences $(\mathbf{a}-\hat{\mathbf{a}}),(\mathbf{b}-\hat{\mathbf{b}})$ about the maximum likelihood estimate $\hat{\mathbf{S}} \equiv S(\hat{\mathbf{a}},\hat{\mathbf{b}})$. Hence

$$
\begin{aligned}
S(\mathbf{a},\mathbf{b}) &= \tfrac{1}{2}\sum_{k=0}^{k=K-1}\sum_{i=0}^{i=N-1}\frac{\left(p_k(\mathbf{a},\mathbf{b};t_i)-\mu_{ki}\right)^2}{\sigma_{ki}^2} \\
&\equiv S(\hat{a},\hat{b}) + \tfrac{1}{2}\left(a-\hat{a},b-\hat{b}\right)P^{-1}\left(a-\hat{a},b-\hat{b}\right)+\ldots \\
&\approx S(\hat{a},\hat{b}) + \tfrac{1}{2}\sum_{i,j}(a_i-\hat{a}_i)\frac{\partial^2\hat{S}}{\partial\hat{a}_i\partial\hat{a}_j}(a_j-\hat{a}_j) + \tfrac{1}{2}\sum_{i,j}(a_i-\hat{a}_i)\frac{\partial^2\hat{S}}{\partial\hat{a}_i\partial\hat{b}_j}\left(b_j-\hat{b}_j\right) + \\
&+ \tfrac{1}{2}\sum_{i,j}\left(b_i-\hat{b}_i\right)\frac{\partial^2\hat{S}}{\partial\hat{b}_i\partial\hat{a}_j}(a_j-\hat{a}_j) + \tfrac{1}{2}\sum_{i,j}\left(b_i-\hat{b}_i\right)\frac{\partial^2\hat{S}}{\partial\hat{b}_i\partial\hat{b}_j}\left(b_j-\hat{b}_j\right)
\end{aligned}
\qquad (0.3)
$$

The (2*K*-2)-dimensional covariance matrix *P* is the inverse of the appropriate expansion coefficients. This matrix is central to the construction of the confidence limits for the trend lines.

2.2.1 Estimation of the confidence intervals

The logistic regression functions $p_k(t)$ can be approximated as a normally distributed time-varying random variable $N\left(\hat{p}_k(t),\sigma_k^2(t)\right)$ by expanding p_k about its maximum likelihood estimate (the trend line) $\hat{p}_k(t)=p\left(\hat{\mathbf{a}},\hat{\mathbf{b}},t\right)$

$$
\begin{aligned}
p_k(\mathbf{a},\mathbf{b},t) &= p_k\left(\hat{\mathbf{a}}+\mathbf{a}-\hat{\mathbf{a}},\hat{\mathbf{b}}+\mathbf{b}-\hat{\mathbf{b}},t\right) \\
&= \hat{p}_k(t) + \left(\nabla_{\hat{a}},\nabla_{\hat{b}}\right)\hat{p}_k(t)\begin{pmatrix}\mathbf{a}-\hat{\mathbf{a}}\\ \mathbf{b}-\hat{\mathbf{b}}\end{pmatrix}+\ldots
\end{aligned}
\qquad (0.3)
$$

Denoting mean values by angled brackets, the variance of p_k is thereby approximated as

$$
\begin{aligned}
&\sigma_k^2(t) \equiv \left\langle\left(p_k(\mathbf{a},\mathbf{b},t)-\hat{p}_k(t)\right)^2\right\rangle = \left(\nabla_{\hat{a}}\hat{p}_k(t),\nabla_{\hat{b}}\hat{p}_k(t)\right)\left\langle\begin{pmatrix}\mathbf{a}-\hat{\mathbf{a}}\\ \mathbf{b}-\hat{\mathbf{b}}\end{pmatrix}\begin{pmatrix}\mathbf{a}-\hat{\mathbf{a}}\\ \mathbf{b}-\hat{\mathbf{b}}\end{pmatrix}^T\right\rangle\times \\
&\left(\nabla_{\hat{a}}\hat{p}_k(t),\nabla_{\hat{b}}\hat{p}_k(t)\right)^T = \left(\nabla_{\hat{a}}\hat{p}_k(t),\nabla_{\hat{b}}\hat{p}_k(t)\right)\ P\ \left(\nabla_{\hat{a}}\hat{p}_k(t),\nabla_{\hat{b}}\hat{p}_k(t)\right)^T
\end{aligned}
\qquad (0.3)
$$

When *K*=3 this equation can be written as the 4-dimensional inner product

$$
\sigma_k^2(t) = \begin{pmatrix}\frac{\partial\hat{p}_k(t)}{\partial\hat{a}_1} & \frac{\partial\hat{p}_k(t)}{\partial\hat{a}_2} & \frac{\partial\hat{p}_k(t)}{\partial\hat{b}_1} & \frac{\partial\hat{p}_k(t)}{\partial\hat{b}_2}\end{pmatrix}\begin{bmatrix}P_{aa11} & P_{aa12} & P_{ab11} & P_{ab12}\\ P_{aa21} & P_{aa22} & P_{ab21} & P_{ab22}\\ P_{ba11} & P_{ba12} & P_{bb11} & P_{bb12}\\ P_{ba21} & P_{ba22} & P_{bb21} & P_{bb22}\end{bmatrix}\begin{pmatrix}\frac{\partial\hat{p}_k(t)}{\partial\hat{a}_1}\\ \frac{\partial\hat{p}_k(t)}{\partial\hat{a}_2}\\ \frac{\partial\hat{p}_k(t)}{\partial\hat{b}_1}\\ \frac{\partial\hat{p}_k(t)}{\partial\hat{b}_2}\end{pmatrix} \qquad (0.3)
$$

where $P_{cdij} \equiv \left\langle (c_i - \hat{c}_i)(d_j - \hat{d}_j) \right\rangle$. The 95% confidence interval for $p_k(t)$ is centred given as $[\hat{p}_k(t) - 1.96\sigma_k(t), \mathrm{p_k}(t) + 1.96\sigma_k(t)]$.

3 Module 2: Microsimulation model

3.1 Microsimulation initialisation: birth, disease and death models

Simulated people are generated with the correct demographic statistics in the simulation's start-year. In this year women are stochastically allocated the number and years of birth of their children – these are generated from known fertility and mother's age at birth statistics (valid in the start-year). If a woman has children then those children are generated as members of the simulation in the appropriate birth year.
The microsimulation is provided with a list of air pollution-related diseases. These diseases used the best available incidence, mortality, survival, relative risk and prevalence statistics (by age and gender). Individuals in the model are simulated from their year of birth (which may be before the start year of the simulation). In the course of their lives, simulated people can die from one of the diseases caused by an air pollutant that they might have acquired or from some other cause(s). The probability that a person of a given age and gender dies from a cause other than the disease are calculated in terms of known death and disease statistics valid in the start-year. It is constant over the course of the simulation.

The microsimulation incorporates a sophisticated economic module. The module employs a Markov-type simulation of long-term health benefits and health care costs. It synthesises and estimates evidence on cost-utility analysis. The model is used to project the differences in quality-adjusted life years (QALYs), and direct lifetime health-care costs over a specified time scale. The direct healthcare costs are presented separately in terms of hospital admissions, general practitioner costs, medication costs and social care costs. Outputs can be discounted for any specific discount rate.
This following section provides an overview of the main assumptions of the model.

3.2 Population models

Populations are implemented as instances of the TPopulation C++ class. The TPopulation class is created from a population (*.ppl) file. Usually a simulation will use only one population but it can simultaneously process multiple populations (for example, different ethnicities within a national population).

3.2.1 Population Editor

The Population Editor Allows editing and testing of TPopulation objects. The population is created in the start-year and propagated forwards in time. An example population

pyramid which can be used when initialising the model is shown in Figure 1 shows the population distribution for England in 2015 used in the initialisation of the model.

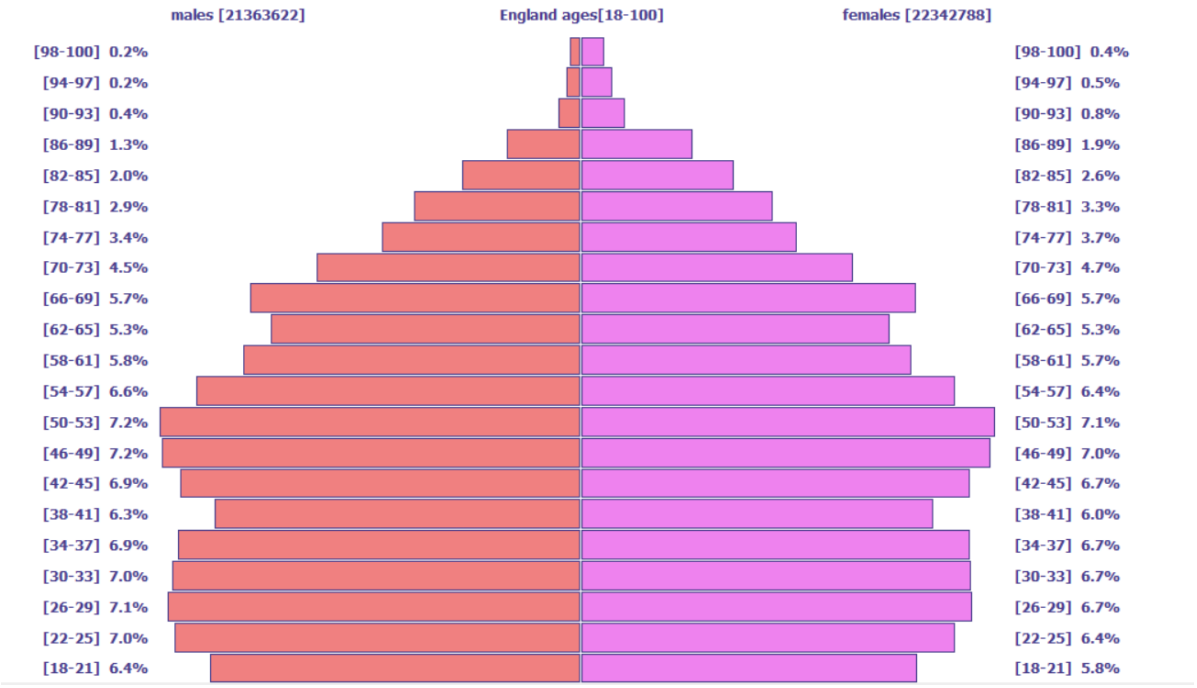


Figure 1 Population pyramid for England in 2015

People within the model can die from specific diseases or from other causes. A disease file is created within the program to represent deaths from other causes. The following distributions are required by the population editor (Table 20).

Table 20 Summary of the parameters representing the distribution component

Distribution name	symbol	note
MalesByAgeByYear	$p_m(a)$	Input in year0 – probability of a male having age a
FemalesByAgeByYear	$p_f(a)$	Input in year0 – probability of a female having age a
BirthsByAgeofMother	$p_b(a)$	Input in year0 – conditional probability of a birth at age a| the mother gives birth.
NumberOfBirths	$p_\lambda(n)$	λ≡TFR, Poisson distribution, probability of giving birth to n children

3.2.1.1 Birth model

Any female in the child bearing years {*AgeAtChild.lo, AgeAtChild.hi*} is deemed capable of giving birth. The number of children, n, that she has in her life is dictated by the Poisson distribution $p_\lambda(n)$ where the mean of the Poisson distribution is the Total Fertility Rate (TFR) parameter[2].

The probability that a mother (who does give birth) gives birth to a child at age a is determined from the BirthsByAgeOfMother distribution as $p_b(a)$. For any particular mother the births of multiple children are treated as independent events, so that the probability that a mother who produces N children produces n of them at age a is given as the Binomially distributed variable,

$$p_b\left(\text{n at a} \mid N\right) = \frac{N!}{n!(N-n)!}\left(p_b(a)\right)^n\left(1-p_{bm}(a)\right)^{N-n} \qquad (0.4)$$

The probability that the mother gives birth to n children at age a is

$$p_b(n\,at\,a) = e^{-\lambda}\sum_{N=n}^{\infty}\frac{\lambda^N}{N!}\,p_b\left(\text{n at a} \mid N\right) = e^{-\lambda}\sum_{N=n}^{\infty}\frac{\lambda^N}{n!(N-n)!}\left(p_b(a)\right)^n\left(1-p_b(a)\right)^{N-n} \qquad (0.5)$$

Performing the summation in this equation gives the simplifying result that the probability p_b(*n at a*) is itself Poisson distributed with mean parameter $\lambda p_b(a)$,

$$p_b(n\,at\,a) = e^{-\lambda p_b(a)}\frac{\left(\lambda p_b(a)\right)^n}{n!} = p_{\lambda p_b(a)}(n) \qquad (0.6)$$

Thus, on average, a mother at age a will produce $\lambda p_b(a)$ children in that year.
The gender of the children[3] is determined by the probability p_{male}=1-p_{female}. In the baseline model this is taken to be the probability $N_m/(N_m+N_f)$.

The Population editor' menu item Population Editor\Tools\Births\show random birthList creates an instance of the TPopulation class and uses it to generate and list a (selectable) sample of mothers and the years in which they give birth.

3.2.2 Deaths from modelled diseases

The simulation models any number of specified diseases some of which may be fatal. In the start year the simulation's death model uses the diseases' own mortality statistics to adjust the probabilities of death by age and gender. In the start year the net effect is to maintain the same probability of death by age and gender as before; in subsequent

[2] This could be made to be time dependent; in the baseline model it is constant.

[3] The probability of child gender can be made time dependent.

years, however, the rates at which people die from modelled diseases will change as modelled risk factors change.

3.3 The risk factor model

The distribution of risk factors (RF) in the population is estimated using regression analysis stratified by both sex *S* = {male, female} and age group *A* = {0-4, 5-9, ..., 70-74, 75+}. The fitted trends are extrapolated to forecast the distribution of each RF category in the future. For each sex-and-age-group stratum, the set of cross-sectional, time-dependent, discrete distributions $D = \{p_k(t)|k = 1, \dots N;\ t > 0\}$, is used to manufacture RF trends for individual members of the population. Each air pollutant (eg NO_2, PM2.5) is modelled as a continuous risk factor.

3.3.1 Continuous risk factors

In the case of a continuous RF, for each discrete distribution D there is a continuous counterpart. Let β denote the RF value in the continuous scale and let $f(\beta|A,S,t)$ be the probability density function of β for age group A and sex S at time t. Then

$$p_k(t|A,S) = \int_{\beta \in k} f(\beta|A,S,t)\, d\beta. \qquad (0.7)$$

Equations (0.2) and (0.7) both refer to the same quantity. However, equation (0.7) uses the definition of the probability density function to express the age-and-sex-specific percentage of individuals in RF category *k* at time *t*. Equation (0.2) gives an estimate of this quantity using equation (0.1) for all *k* = 0, …, N. The cumulative distribution function of β is

$$F(\beta|A,S,t) = \int_0^{\beta} f(\beta|A,S,t)\, d\beta. \qquad (0.8)$$

At time *t*, a person with sex S belonging to the age group A is said to be on the p–th percentile of this distribution if $F(\beta|A,S,t) = p/100$. Given the cross-sectional information from the set of distributions D, it is possible to simulate longitudinal trajectories by forming pseudo-cohorts within the population. A key requirement for these sets of longitudinal trajectories is that they reproduce the cross-sectional distribution of RF categories for any year with available data. The method adopted here and in our earlier work (1) is based on the assumption that person's RF value changes throughout their lives in such a way that they always have the same associated percentile rank. As they age, individuals move from one age group to another and their RF value changes so that they have the same percentile rank but of a different RF distribution. Crucially it meets the important condition that the cross-sectional RF distributions obtained by simulation match the RF distributions of the observed data. The above procedure can be explained using the example of the NO_2 distribution. The NO_2 distributions are known for the population stratified by sex and age for all years of the simulation (by extrapolation of fitted model, see equation (0.1)). A person who is in

age group A and who grows ten years older will at some time move into the next age group A' and will have a BMI that was described first by the distribution $f(\beta|A,S,t)$ and then at the later time t' by the distribution $f(\beta|A',S,t')$. If the NO_2 exposure level of that individual is on the p-th percentile of the NO_2 distribution, their NO_2 exposure level will change from β to β' so that

$$\beta = F^{-1}\left(\frac{p}{100}|A,S,t\right) \tag{0.9}$$

$$\beta' = F^{-1}\left(\frac{p}{100}|A',S,t'\right) \Rightarrow \beta' = F^{-1}\left(F(\beta \mid A,S,t)|A',S,t'\right) \tag{0.10}$$

Where F^{-1} is the inverse of the cumulative distribution function of β, which we model with a continuous uniform distribution within the RF categories (see **Table 19**). Equation (0.10) guarantees that the transformation taking the random variable β to β' ensures the correct cross-sectional distribution at time t'.

The microsimulation first generates individuals from the RF distributions of the set D and, once generated, grows the individual's RF in a way that is also determined by the set D. It is possible to implement equation (0.10) as a suitably fast algorithm.

3.4 Relative risks

Suppose that α is a risk factor state of some risk factor A and denote by p_A (d|α,a,s) the incidence probability for the disease d given the risk state, α, the person's age, a, and gender, s. The relative risk ρ_A is defined by equation (0.10).

$$\begin{aligned} p_{\mathrm{A}}\left(d|\alpha,a,s\right) &= \rho_{\mathrm{A}|d}\left(\alpha|a,s\right) p_{\mathrm{A}}\left(d|\alpha_0,a,s\right) \\ \rho_{\mathrm{A}|d}\left(\alpha_0|a,s\right) &\equiv 1 \end{aligned} \tag{0.10}$$

Where α_0 is the zero risk state.
The incidence probabilities, as reported, can be expressed in terms of the equation,

$$\begin{aligned} p\left(d|a,s\right) &= \sum_{\alpha} p_{\mathrm{A}}\left(d|\alpha,a,s\right)\pi_A\left(\alpha|a,s\right) \\ &= p_A\left(d|\alpha_0,a,s\right)\sum_{\alpha}\rho_{\mathrm{A}|d}\left(\alpha|a,s\right)\pi_A\left(\alpha|a,s\right) \end{aligned} \tag{0.10}$$

Combining these equations allows the conditional incidence probabilities to be written in terms of known quantities

$$p\left(d|\alpha,a,s\right) = \rho_{A|d}\left(\alpha|a,s\right)\frac{p\left(d|a,s\right)}{\sum_{\beta}\rho_{A|d}\left(\beta \mid a,s\right)\pi_A\left(\alpha|a,s\right)} \tag{0.10}$$

Previous to any series of Monte Carlo trials the microsimulation program pre-processes the set of diseases and stores the *calibrated* incidence statistics p_A ($d|\alpha_0$, a, s). These incidence statistics are calibrated to national level data sets for both national level and local authority model simulations. In this project the risk factor distributions and incidence risks for England are used to calculate the calibrated risks.

3.5 Modelling diseases

Disease modelling relies heavily on the sets of incidence, mortality, survival, relative risk and prevalence statistics. In some cases where a data set is unavailable or not available is the specified form for the model, data has been approximated from the known sets of the data.

The microsimulation uses risk dependent incidence statistics and these are inferred from the relative risk statistics and the distribution of the risk factor within the population. In the simulation, individuals are assigned a risk factor trajectory giving their personal risk factor history for each year of their lives. Their probability of getting a particular risk factor related disease in a particular year will depend on their risk factor state in that year.

Once a person has a fatal disease (or diseases) their probability of survival will be controlled by a combination of the disease-survival statistics and the probabilities of dying from other causes. Disease survival statistics are modelled as age and gender dependent exponential distributions.

3.5.1 Mortality statistics

In any year, in some population, in a sample of N people who have the disease a subset N_ω will die from the disease.

Mortality statistics record the cross sectional probabilities of death as a result of the disease – possibly stratifying by age

$$p_\omega = \frac{N_\omega}{N} \tag{0.11}$$

Within some such subset N_ω of people that die in that year from the disease, the distribution by year-of-disease is not usually recorded. This distribution would be most useful. Consider 2 important idealised, special cases
Suppose the true probabilities of dying in the years after some age a_0 are
$\{p_{\omega 0}, p_{\omega 1}, p_{\omega 2}, p_{\omega 3}, p_{\omega 4}\}$
The probability of being alive after N years is simply that you don't die in each year

$$p_{survive}(a_0 + N) = (1 - p_{\omega 0})(1 - p_{\omega 1})(1 - p_{\omega 2})..(1 - p_{\omega N-1}) \tag{0.12}$$

3.5.2 Survival rates

It is common practice to describe survival in terms of a survival rate R, supposing an exponential death-distribution. In this formulation the probability of surviving t years from some time t_0 is given as

$$p_{\text{survival}}(t) = 1 - R\int_0^t du e^{-Ru} = e^{-Rt} \tag{0.13}$$

For a time period of 1 year

$$\begin{aligned} p_{survival}(1) &= e^{-R} \\ &\Rightarrow \\ R &= -\ln\left(p_{survival}(1)\right) = -\ln(1-p_{\omega}) \end{aligned} \tag{0.14}$$

For a time period of, for example, 4 years,

$$p_{\text{survival}}(t=4) = 1 - R^{-1}\int_0^4 du e^{-Ru} = e^{-4R} = (1-p_{\omega})^4 \tag{0.15}$$

In short, the Rate is minus the natural log of the 1-year survival probability.

3.5.3 The survival models

For any potentially terminal disease the model can use any of the 3 survival models, numbered ((0, 1, 2)). The parameters describing these models are given below.

Survival model 0

A single probability of dying $\{p_{\omega 0}\}$, where $p_{\omega 0}$ is valid for all years. Given the 1-year survival probability $p_{survival}(1)$

The model uses 1 parameter ((R))

$$R = -\ln\left(p_{survival}(1)\right) \tag{0.16}$$

Survival model 1

Two different probabilities of dying $\{p_{\omega 0}, p_{\omega 1}\}$, where $p_{\omega 0}$ is valid for the first year; $p_{\omega 1}$ thereafter. The model uses 2 parameters ((p_1, R)). Given the 1-year survival probability $p_{survival}(1)$ and the 5-year survival probability $p_{survival}(5)$

$$\begin{aligned} p_1 &= 1 - p_{survival}(1) \\ R &= -\frac{1}{4}\ln\left(\frac{p_{survival}(5)}{p_{survival}(1)}\right) \end{aligned} \tag{0.17}$$

Survival model 2

Three different probabilities of dying $\{p_{\omega 0}, p_{\omega 1}, p_{\omega 5}\}$, where $p_{\omega 0}$ is valid for the first year; $p_{\omega 1}$ for the second to the fifth year; $p_{\omega 5}$ thereafter. The model uses 3 parameters ((p_1, R, $R_{>5}$))

Given the 1-year survival probability $p_{survival}(1)$ and the 5-year survival probability $p_{survival}(5)$

$$p_1 = 1 - p_{survival}(1)$$
$$R = -\frac{1}{4}\ln\left(\frac{p_{survival}(5)}{p_{survival}(1)}\right) \quad (0.18)$$
$$R_{>5} = -\frac{1}{5}\ln\left(\frac{p_{survival}(10)}{p_{survival}(5)}\right)$$

Remember that different probabilities will apply to different age and gender groups. Typically the data might be divided into 10 year age groups.

3.5.4 Modelling low birth weight

The modelling method assumes that low birth weight (LBW) is a disease associated with a women who gives birth. The method also assumes that LBW is an acute disease; an incidence case in any year affects the prevalence rate in that year only. In the start year of the simulation the total of number of births associated to a woman and the year of each birth is computed. The probability of a newborn being LBW is calculated using the risk factor level (ie, air pollution level) in the year of birth and the associated relative risk. This approach is used when modelling other diseases in the simulation.

There are 2 differences between modelling LBW and other diseases. Firstly, a mother can have multiple births in a given year which can result in multiple incident cases of LBW. In comparison other diseases can be contracted only once in a year. Secondly, it is possible that in some years of a mother's life she does not give birth. The probability of contracting a LBW in these years is therefore zero.

Limitations

The modelling method assumes that LBW is a disease per se. A limitation extending from this would be that we do not take account of subsequent diseases brought about by LBW, eg, diabetes or CHD. The model therefore underestimates the long-term economic costs of LBW associated with air pollution. Another limitation is that we allow multiple births in the simulation (eg twins), but we do take account of the possible impact of multiple births on LBW. Multiple births are simulated as a list of independent births having the same probability of causing LBW.

3.6 Approximating missing disease statistics

A number of tools have been developed in the model in order to compute missing disease statistics data such as incidence or prevalence.

3.6.1 Approximating survival data from mortality and prevalence

An example is provided here with a standard life-table analysis for a disease *d*.
Consider the 4 following states:

state	Description
0	alive without disease *d*
1	alive with disease *d*
2	dead from disease *d*
3	dead from another disease

p_{ik} is the probability of disease *d* incidence, aged *k*
$p_{\omega k}$ is the probability of dying from the disease *d*, aged *k*
$p_{\bar{\omega} k}$ is the probability of dying other than from disease d, aged *k*
The state transition matrix is constructed as follows

$$\begin{bmatrix} p_0(k+1) \\ p_1(k+1) \\ p_2(k+1) \\ p_3(k+1) \end{bmatrix} = \begin{bmatrix} (1-p_{\bar{\omega}k})(1-p_{ik}) & (1-p_{\bar{\omega}k}-p_{\omega k})\,p_{\alpha k} & 0 & 0 \\ (1-p_{\bar{\omega}k})\,p_{ik} & (1-p_{\bar{\omega}k}-p_{\omega k})(1-p_{\alpha k}) & 0 & 0 \\ 0 & p_{\omega k} & 1 & 0 \\ p_{\bar{\omega}k} & p_{\bar{\omega}k} & 0 & 1 \end{bmatrix} \begin{bmatrix} p_0(k) \\ p_1(k) \\ p_2(k) \\ p_3(k) \end{bmatrix} \quad (0.19)$$

It is worth noting that the separate columns correctly sum to unity.
The disease mortality equation is that for state-2,

$$p_2(k+1) = p_{\omega k}\,p_1(k) + p_2(k) \quad (0.20)$$

The probability of dying from the disease in the age interval [*k*, *k*+1] is $p_{\omega k}p_1(k)$ - this is otherwise the (cross-sectional) disease mortality, $p_{mor}(k)$. $p_1(k)$ is otherwise known as the disease prevalence, $p_{pre}(k)$. Hence the relation

$$p_{\omega k} = \frac{p_{mor}(k)}{p_{pre}(k)} \quad (0.21)$$

For exponential survival probabilities the probability of dying from the disease in the age-interval [*k*, *k*+1] is denoted $p_{\Omega k}$ and is given by the formula

$$p_{\omega k} = 1 - e^{-R_k} \quad \Rightarrow \quad R_k = -\ln(1-p_{\omega k}) \quad (0.22)$$

When, as is the case for most cancers, these survival probabilities are known the microsimulation will use them, when they are not known or are too old to be any longer of any use, the microsimulation uses survival statistics inferred from the prevalence and mortality statistics (equation (0.21)). An alternative derivation equation (0.21) is as

follows. Let N_k be the number of people in the population aged k and let n_k be the number of people in the population aged k with the disease. Then, the number of deaths from the disease of people aged k can be given in 2 ways: as $p_{\omega k} n_k$ and, equivalently, as $p_{mor}(k) N_k$. Observing that the disease prevalence is n_k/N_k leads to the equation

$$\begin{aligned} p_{\Omega k} n_k &= p_{mor}(k) N_k \\ p_{pre}(k) &= \frac{n_k}{N_k} \\ &\Rightarrow \\ p_{\Omega k} &= \frac{p_{mor}(k)}{p_{pre}(k)} \end{aligned} \tag{0.23}$$

3.6.2 Approximating disease incidence from prevalence

The algorithm estimates the probability of contracting a disease given age and sex, $\hat{p}(d \mid a, s)$ from prevalence rates, survival rates and mortality rates.

Step 1: State transition matrix of the algorithm

$$\begin{pmatrix} p_{\bar{d}}(a+1 \mid s) \\ p_{d1}(a+1 \mid s) \\ p_{d}(a+1 \mid s) \\ p_{dead}(a+1 \mid s) \end{pmatrix} = \begin{pmatrix} (1-p_{\bar{w}}(a \mid s))(1-\hat{p}(d \mid a, s)) & 0 & 0 & 0 \\ (1-p_{\bar{w}}(a \mid s))\hat{p}(d \mid a, s) & 0 & 0 & 0 \\ 0 & 1-p_{w1+\bar{w}1}(a \mid s) & 1-p_{w+\bar{w}}(a \mid s) & 0 \\ p_{\bar{w}}(a \mid s) & p_{w1+\bar{w}1}(a \mid s) & p_{w+\bar{w}}(a \mid s) & 1 \end{pmatrix} \begin{pmatrix} p_{\bar{d}}(a \mid s) \\ p_{d1}(a \mid s) \\ p_{d}(a \mid s) \\ p_{dead}(a \mid s) \end{pmatrix} \tag{0.24}$$

The probability of being in a set of states:

S_0	$p_{\bar{d}}(a \mid s)$	**The probability of being alive without disease at age** a
S_1	$p_{d1}(a \mid s)$	The probability of being alive with new disease (contracting within a year) at age a
S_2	$p_{d}(a \mid s)$	The probability of being alive with old disease at age a
S_3	$p_{dead}(a \mid s)$	The probability of being dead for any reason (from the disease or other reasons) at age a

$\hat{p}(d \mid a, s)$ The estimated incidence probability at age of a given sex type s .
$p_{\bar{w}}(a \mid s)$ The probability of dying from other causes at age of a given sex type s.
$p_{w1+\bar{w}1}(a \mid s)$ The probability of dying from any reason within the first years of contracting the disease at the age of a given sex type s.
$p_{w+\bar{w}}(a \mid s)$ The probability of dying from any reasons after the first years of contracting the disease at the age a given sex type s.
$p_{survival1st}(a \mid s)$ The probability of surviving the first year after contracting the disease at the age of a given sex type s.

$p_{survival1}(a \mid s)$ The probability of surviving the year at the age of a given sex type s.

Step 2: The prevalence for a particular age group
Estimated prevalence rate can be expressed by,

$$\hat{P}_{pre\_mean}(agegroup \mid s) = \frac{\sum_{\min\_a}^{\max\_a} \hat{P}_{pre}(a \mid s) \cdot \pi(a \mid s)}{\sum_{\min\_a}^{\max\_a} \pi(a \mid s)} \quad (0.25)$$

where

$$\hat{P}_{pre}(a \mid s) = \frac{p_d(a \mid s) + p_{d1}(a \mid s)}{p_d(a \mid s) + p_{d1}(a \mid s) + p_{\bar{d}}(a \mid s)} \quad (0.26)$$

where $\min\_a$ is the youngest age in that age group and $\max\_a$ the oldest. $\pi(a \mid s)$ is the population distribution stratified by age given sex.

Step 3: Regression
We have 2 algorithms to find the optimum value of $\hat{p}(d \mid a, s)$: simplex algorithm and cauchy algorithm. Simplex algorithm finds an optimum set of incidence rates of all age groups by minimising the distance between the estimated global prevalence rate and the actual global prevalence rate, shown in (0.27). We use simplex algorithm for most diseases as it is faster.

$$\underset{set(\hat{p}(d|a,s))}{\arg\min}\, S = \underset{set(\hat{p}(d|a,s))}{\arg\min}\, S\left(\sum_{age\_group} (P_{pre\_mean}(agegroup \mid s) - \hat{P}_{pre\_mean}(agegroup \mid s)) \right) \quad (0.27)$$

Cauchy algorithm finds an optimum incidence rate for each individual age group by minimising the distance between the estimated prevalence rate and the actual prevalence rate of the age group, shown in (0.28). We use Cauchy algorithm for diseases which are associated to certain age groups, eg, dementia which is only associated to people older than 60.

$$\underset{\hat{p}(d|a,s)}{\arg\min}\, S = \underset{\hat{p}(d|a,s)}{\arg\min}\, S\left(P_{pre\_mean}(agegroup \mid s) - \hat{P}_{pre\_mean}(agegroup \mid s) \right) \quad (0.29)$$

3.7 Model scenarios

A baseline case and 3 additional scenarios were modelled. The baseline related to the current exposure data, which included background levels of air pollution. The second scenario modelled the impact of the background air pollutions levels. This scenario was used to calculate the diseases and associated costs related to air pollution. The final 2 scenarios were used to assess the impact of a 1% and 5% drop in annual exposure levels each year.

The final scenarios are currently being finalised through discussions with Public Health England.

3.8 Microsimulation model outputs

The microsimulation model outputs will be as follows:

- Prevalence cases avoided per 100,000 by disease by year
- Incidence avoided per 100,000 by disease by year
- Cumulative incidence avoided per 100,000 by disease by year
- GP, medication, hospital and social costs by disease by year (£million per 100,000)
- Costs avoided by year (£million per 100,000)

Air Quality tool methodology

1 Summary

The tool simulates a closed weighted cohort through time.

2 Initialising the weighted cohort

The weighted cohort is initialised based on the distribution of sex, age given sex and the risk factor group given both age and sex. These weights remain constant throughout the simulation for baseline.

$$\begin{aligned} &p(sex) = \text{probability of a given sex in the population} \\ &p(age \mid sex) = \text{probability of a specific age in the population given the sex} \\ &p(rf \mid age, sex) = \text{probability of a specific risk factor group given the age and sex} \end{aligned} \quad (0.30)$$

$$weight(age, sex, rf) = p(sex) * p(age \mid sex) * p(rf \mid age, sex) \quad (0.31)$$

With the condition that all the weights sum to 1.

3 Initialising individual attributes

Individuals are assigned a RF value based on the midpoint value from each RF group NO_2 and $PM_{2.5}$ will be the RFs simulated in this tool (see Table 19). The current risk factor trends are used to determine the corresponding percentile. The individuals' percentile is fixed throughout the simulation. The RF values each year for an individual are calculated from the RF trends.

4 Projecting prevalence into the future

In a given year the model calculates the probability of an individual entering a new state. These states include disease states and a death state. The transition probabilities for each potential new state are calculated based on the calibrated incidence (*BaseRisk*) and RR's. Calibrated incidence is calculated in the microsimulation program and stored within each disease file. For LA's it was assumed that the calibrated incidence calculated from national level risk factor trends was the same for each individual.

The transition probability for an individual in a given year moving to a new disease state is shown in equation (0.32).

$$T(j \mid i, age, sex, rf) = BaseRisk(age, sex) * RR(j \mid age, sex, rf) \quad (0.32)$$

For each individual the probability of being in a disease state (*j*) is based on their probability of being in the disease state in the previous year and the transition probability (T_{ij}) (see equation (0.33)).

$$p(j \mid age, sex, rf, year) = \sum_{i=1}^{N_{states}} T(j \mid i, age, sex, rf) * p(j \mid age, sex, rf, year-1) \quad (0.33)$$

The probability of being in a disease state is calculated for each individual (*n*) within the cohort.

5 Calculating the population level prevalence each year

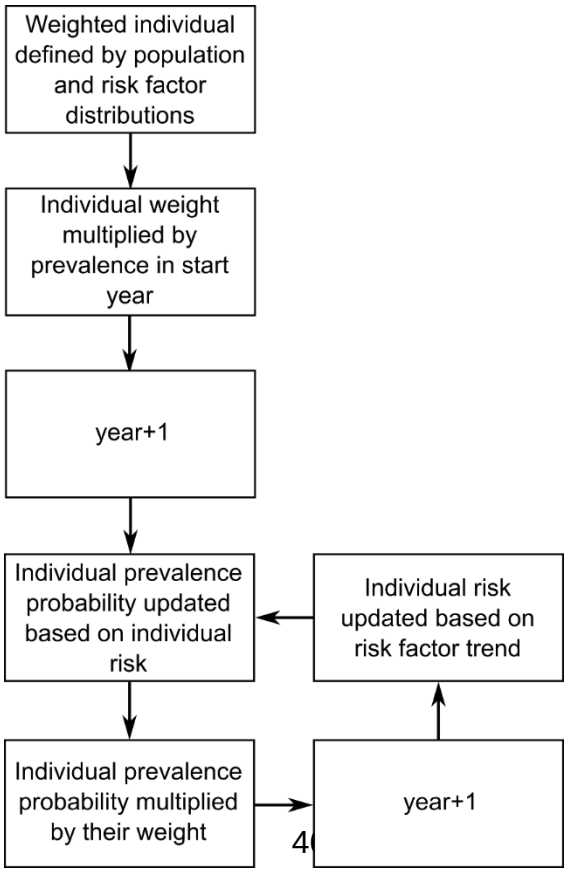
In each year of the model the prevalence is deterministically computed from the weight of each individual within the cohort multiplied by the prevalence of the group.

$$prevalence(j \mid age, sex, rf, year) = \sum_{n=1}^{N} weight_n(age, sex, rf, HI) * p_n(j \mid age, sex, rf, year) \quad (0.34)$$

6 Scenarios

The simulated scenarios impact on the risk factor trends. These trends are altered by adjusting the risk factor distribution for each age and sex group.

7 Model system diagram



8 Piloting survey questionnaire

Question	Response option
Which organisation do you work for?	• Free text
What is your role?	• Director of Public Health • Public health service commissioner or programme manager • Public health researcher • Other local government employee (please give job title below) • Employee of another organisation (please give job title below) • Free text
Is your organisation currently implementing, or planning to implement, interventions(s) to reduce air pollution in your area?	• Yes (please describe) • No • Don't know • Free text
How high a priority is air pollution for your organisation?	• Very high • High • Medium • Low • Very low • Don't know • Comments box
Would a tool that quantifies the health benefits (such as prevalence cases avoided) of changes in air pollution exposure be of value to your organisation?	• Yes • No • Comments box
And would a tool that quantifies the economic benefits (direct, indirect, GP, hospital and medication costs) of changes in air pollution exposure be of value to your organisation?	• Yes • No • Comments box
Are you currently using any tools for this purpose?	• Yes • No • *Comments box – which tools?*

The tool can be more or less flexible in terms of data inputs. We will preload the tool with national-level data from England. Please select from the list below data inputs you would like to be able to change:	• Population data • Air pollution exposure data • Disease prevalence data for your population • Disease incidence data for your population • Disease mortality data for your population • Cost data associated with disease in your population • Cost data for social care
And please consider which of the following outputs would be of use for you:	• Disease prevalence • Premature mortality avoided • Healthcare costs (GP, hospital, medications) • Social-care costs • Non-health care costs (eg lost productivity) • Cost-benefit ratio • Life expectancy • Quality-adjusted life years • Disability-adjusted life years • Years of potential life lost • *Comments box*
Which platform will you use to run the tool?	• PC • Mac • *Comments box - other*
Do you have access to MS Excel?	• Yes • No

9 Outputs generated by the tool

The ouputs generated by the tool are as follows:

- GP, medication, hospital and social care costs by year
- Prevalence by disease by year per 100,000

References

1. Office for National Statistics. Mid-2015 Population estimates: Pivot table analysis tool for the UK. 2016.
2. Office for National Statistics. Birth Summary Tables - England and Wales. 2016.
3. Office for National Statistics. Births by mothers' usual area of residence in the UK. 2016.
4. Office for National Statistics. Deaths registered in England and Wales: 2015. 2016.
5. Office for National Statistics. Mortality statistics - underlying cause, sex and age 2015 2016. Available from: https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&dataset=161&version=0.
6. Foundation BL. Asthma Statistics 2015. Available from: https://statistics.blf.org.uk/asthma.
7. Khreis H, Kelly C, Tate J, Parslow R, Lucas K, Nieuwenhuijsen M. Exposure to traffic-related air pollution and risk of development of childhood asthma: A systematic review and meta-analysis. Environment international. 2016.
8. Jacquemin B, Siroux V, Sanchez M, Carsin AE, Schikowski T, Adam M, et al. Ambient air pollution and adult asthma incidence in six European cohorts (ESCAPE). Environ Health Perspect. 2015;123(6):613-21.
9. Public Health England. COPD Prevalence Estimates London: Public Health England; 2008.
10. Committee On The Medical Effects Of Air Pollutants. Long-term Exposure term to Air Pollution and Chronic Bronchitis 2016.
11. Smolina K, Wright FL, Rayner M, Goldacre MJ. Determinants of the decline in mortality from acute myocardial infarction in England between 2002 and 2010: linked national database study. Corrected data on incidence and mortality in 2013 at http://www.bmj.com/content/347/bmj.f7379.abstract. BMJ. 2012;344:d8059.
12. British Heart Foundation. Cardiovascular Disease Statistics 2014. 2015.
13. Cesaroni G, Forastiere F, Stafoggia M, Andersen ZJ, Badaloni C, Beelen R, et al. Long term exposure to ambient air pollution and incidence of acute coronary events: prospective cohort study and meta-analysis in 11 European cohorts from the ESCAPE Project. Bmj. 2014;348:f7412.
14. NHS Digital. National Diabetes Audit 2015/2016 NHS Digital2017. Available from: http://www.content.digital.nhs.uk/catalogue/PUB23241.
15. Eze IC, Hemkens LG, Bucher HC, Hoffmann B, Schindler C, Kunzli N, et al. Association between ambient air pollution and diabetes mellitus in Europe and North America: systematic review and meta-analysis. Environ Health Perspect. 2015;123(5):381-9.
16. British Heart Foundation. Stroke Statistics 2009. 2009.
17. Scheers H, Jacobs L, Casas L, Nemery B, Nawrot TS. Long-Term Exposure to Particulate Matter Air Pollution Is a Risk Factor for Stroke: Meta-Analytical Evidence. Stroke. 2015;46(11):3058-66.
18. Alzheimer's Society. Dementia UK - second edition. Alzheimer's UK; 2014.
19. Oudin A, Forsberg B, Adolfsson AN, Lind N, Modig L, Nordin M, et al. Traffic-Related Air Pollution and Dementia Incidence in Northern Sweden: A Longitudinal Study. Environ Health Perspect. 2016;124(3):306-12.
20. Anderson HR, Favarato G, Atkinson RW. Long-term exposure to air pollution and the incidence of asthma: meta-analysis of cohort studies. Air Quality, Atmosphere & Health. 2013;6(1):47-56.
21. Office for National Statistics. Birth characteristics 2015 2016. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/birthcharacteristicsinenglandandwales.

22. Pedersen M, Giorgis-Allemand L, Bernard C, Aguilera I, Andersen AM, Ballester F, et al. Ambient air pollution and low birthweight: a European cohort study (ESCAPE). The Lancet Respiratory medicine. 2013;1(9):695-704.
23. UK CR. Statistics by cancer type - Average Number of New Cases Per Year and Age-Specific Incidence Rates per 100,000 Population, UK 2012-2014 2017. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type.
24. Statistics OfN. Cancer Survival in England - Adults Diagnosed: 2010 to 2014, followed up to 2015. 2016.
25. Office for National Statistics. Cancer Survival in England: adults diagnosed 2008–2012 and followed up to 2013. 2014.
26. Hamra GB, Laden F, Cohen AJ, Raaschou-Nielsen O, Brauer M, Loomis D. Lung Cancer and Exposure to Nitrogen Dioxide and Traffic: A Systematic Review and Meta-Analysis. Environ Health Perspect. 2015;123(11):1107-12.
27. Hamra GB, Guha N, Cohen A, Laden F, Raaschou-Nielsen O, Samet JM, et al. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. Environ Health Perspect. 2014;122(9):906-11.
28. Public Health England. UK Plan for tackling roadside nitrogen dioxide concentrations: Technical report. 2017.
29. Dr. Craig Currie at Cardiff University.
30. Curti S, Gori D, Di Gregori V, Farioli A, Baldasseroni A, Fantini MP, et al. PubMed search filters for the study of putative outdoor air pollution determinants of disease. BMJ Open. 2016;6(12):e013092.
31. Davies HT, Crombie IK, Tavakoli M. When can odds ratios mislead? BMJ. 1998;316(7136):989-91.
32. Department for Environment FaRA. Conversion Factors Between ppb and µg m-3 and ppm and mgm-3 2014. Available from: https://uk-air.defra.gov.uk/assets/documents/reports/cat06/0502160851_Conversion_Factors_Between_ppb_and.pdf.
33. COMEAP. Review of the UK air quality index. 2011.
34. Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D Scores for the United Kingdom. Med Decis Making. 2011;31(6):800-4.