Public Health England

Protecting and improving the nation's health

Technical document for the diabetes prevalence model for England 2016

About Public Health England

Public Health England exists to protect and improve the nation's health and wellbeing, and reduce health inequalities. It does this through world-class science, knowledge and intelligence, advocacy, partnerships and the delivery of specialist public health services. PHE is an operationally autonomous executive agency of the Department of Health.

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

Any enquiries regarding this publication should be sent to publications@phe.gov.uk.

Published September 2016
PHE publications gateway number: 2016243

Introduction

The diabetes prevalence model provides estimates of total (diagnosed and undiagnosed) diabetes prevalence for people aged 16 years and older in England.

Diabetes refers to a condition where the amount of glucose in your blood is too high. There are two main types of diabetes, type 1 diabetes and type 2 diabetes. Type 1 diabetes develops when the body is unable to produce any insulin. Type 2 diabetes develops when the body is unable to produce enough insulin or the body's cells don't react to insulin. It is estimated that approximately 90% of diabetes is type 2.[1] The diabetes prevalence model does not make any distinction in the type of diabetes.

The model was developed using data from the latest three years of Health Surveys for England (HSE), 2012, 2013 and 2014.[2] The estimates take into account the age, sex and ethnic group distribution, as well as deprivation of the area. Estimates are created using resident populations and GP registered populations. The 2014 Office for National Statistics (ONS) population projections were used for the resident population estimates.[3] The numbers of patients registered by GP practice in April 2015 were used for the registered population estimates.[4]

Diabetes prevalence estimates have been produced for local authorities, clinical commissioning groups (CCG) and for the whole of England and are provided for 2015, 2016, 2017, 2018, 2019, 2020, 2025, 2030 and 2035. As with all modelled data, there is a degree of uncertainty associated with these estimates therefore should be considered indicative only. The estimates are available to download at: www.ncvin.org.uk.

Previous diabetes estimates

The diabetes prevalence model is an update of the Association of Public Health Observatories (APHO) diabetes prevalence model produced in 2010 [5] which in turn was built on the underlying principles and structure of the Public Health Observatory-Brent Primary Care Trust-ScHARR (PBS) diabetes prevalence model originally developed in 2004.[6]

The PBS diabetes prevalence model used data from a number of the epidemiological population-based diabetes prevalence studies from the late 1980s to the early 1990s. Diabetes was defined using oral glucose tolerance testing to identify type 2 diabetes and a capture-recapture method based study for type 1 diabetes. These prevalence estimates were applied to the resident population of England by age, sex and ethnic group with adjustments applied for deprivation.

The APHO diabetes prevalence model followed a similar structure to the PBS diabetes model, but with the underlying prevalence data taken from the 2006 HSE. In addition, the 2004 HSE was used to take into account differences in prevalence by ethnic group. Diabetes was defined as self-reported doctor-diagnosed diabetes and/or a glycated haemoglobin (HbA1c) of 6.5% of more. Self-reported diagnosed-diabetes prevalence by age and sex were calculated from the 2006 HSE with age-adjusted relative risks by ethnicity applied from the 2004 HSE. Undiagnosed diabetes was accounted for by applying the proportion of undiagnosed diabetes by sex to the age, sex and ethnic specific prevalence rates of diagnosed diabetes. A deprivation adjustment index was then applied to take into account the pattern of deprivation within each area.

Methods

The HSE is an annual survey of adults aged 16 and over living in private households in England.[7] The surveys are designed to be representative of the population living in private households in England and are weighted to match ONS population estimates by age, sex and region. The HSE uses a clustered, multi-stage, stratified probability design which means that participants in the survey are selected over two stages. This involves a random sample of primary sampling units (PSUs) based on postcode sectors and then a random sample of postal addresses.

Each survey consists of a series of core questions conducted by an interviewer, followed by a visit from a nurse in those who agree. The nurse visit includes additional questions, physiological measurements and collection of blood and saliva samples. Not all interviewees agree to a nurse visit and not all who have a nurse visit agree to a blood test. In those who had a blood test, HbA1c was measured. Non-response weights are included in the HSE dataset, including weighting factors for respondents who had a blood sample. The blood weight adjusts for selection, non-response and the population profile of the sample that receives the nurse visit.

Using the HSE data, diabetes can be defined using two methods:

- self-reported doctor-diagnosed diabetes
- the results of the blood hba1c measurement data

Self-reported doctor-diagnosed diabetes was defined based on a positive response to both of the following questions at interview (with the exception of women who only had diabetes during pregnancy):

- Do you now have, or have you ever had diabetes?
- Were you told by a doctor that you had diabetes?

No attempt was made to validate this data and no distinction was made between type 1 and type 2 diabetes. Using data from the National Diabetes Audit (NDA), it is estimated that approximately 90% of diabetes is type 2.[1]

In 2011, the World Health Organization (WHO) recommended that HbA1c could be used as an alternative to standard glucose measures to diagnose a person with type 2 diabetes and that HbA1c levels of 6.5% (48mmol/mol) or above indicated that a person has type 2 diabetes.[8] Using the results of the blood data in the HSE, individuals who had a HbA1c value of 6.5% or more from the blood test and had not been previously diagnosed with diabetes were assumed to have undiagnosed diabetes.

Total diabetes (diagnosed and undiagnosed) was defined as self-reported doctor diagnosed diabetes or an HbA1c of 6.5% of more who had not reported that they had been diagnosed with diabetes.

The diabetes prevalence model was developed using data from the latest three years of HSE, 2012, 2013 and 2014 using a multivariate logistic regression model to predict total diabetes (diagnosed and undiagnosed). A calibration correction was applied from September 2013 onwards. [9] The data was weighted using the blood weight included in the HSE dataset. Robust standard errors were calculated to take into account the complex sampling of the survey design. The model was developed using STATA version 13.

Variables considered for inclusion in the model

The risk factors for developing type 2 diabetes are well known and include: increasing age, ethnicity, a family history of diabetes, increased body mass index (BMI) and/or waist circumference and high blood pressure, a heart attack or stroke.[10] Male gender and socio-economic deprivation are also common associations with diagnosed type 2 diabetes.[1,11] In addition, individuals with non-diabetic hyperglycaemia, also known as impaired glucose control or prediabetes, are at increased risk of type 2 diabetes [12, 13]. The risk factors for type 1 diabetes are less well known and are still being researched, however, they also include an increased risk if there is a family history of diabetes. Other risk factors include environmental and exposure to some viral infections.[10]

The HSE is a cross-sectional survey, therefore non-diabetic hyperglycaemia cannot be used as a risk factor to predict diabetes. With the exception of family history of diabetes, all other risk factors for type 2 diabetes are available from the HSE and so could be considered for inclusion in the model. However, while they are available from the HSE, only a limited number are directly available (eg age) or can be estimated at a local level, such as local authority or CCG. Risk factor variables that can be estimated at a local level (age, sex and ethnicity) were considered for inclusion in the prevalence model. This is consistent with the APHO and PBS prevalence models. Table 1 lists the variables used from the HSE and their description.

Table 1. Variables used from the HSE

Variable	HSE variable	Description
Age	Age	Age last birthday
Gender	Sex	Sex of individual
Ethnicity	Origin	Ethnic origin of individual 18 ethnic categories available

Multivariate logistic regression

A multivariate logistic model was used to produce local and national level estimates of the number of people with diabetes (diagnosed and undiagnosed). All variables were added as categorical variables with the effects estimated relative to a reference category. Age was grouped into 10-year intervals; 16-24, 25-34, 35-44, 45-54, 55-64, 65-74 and 75 plus. The 18 ethnic categories were collapsed into two groups, ‘South Asian and black’ and ‘white, mixed, other’. Any data records with missing values were excluded from the multivariate analysis, giving an unweighted sample size of 12,238. Table 2 summarises the model output.

All variables were found to be significant in the model. Interactions were considered but were not found to be significant. For age group, the adjusted odds ratio increased as the age group increased, showing a clear association between increased age and increased odds of developing diabetes. The reference category for sex was male, and for females the odds of diabetes were approximately 68% lower relative to males. For ethnic group, the reference group was the ‘white, mixed and other’ ethnic group, and for the ‘South Asian and black’ ethnic group, the odds ratio implies an increase of nearly four times relative to the reference group.

Validation was carried out by refitting the model on 70% of the data (randomly selected) and using the remaining 30% to assess model fit. Good agreement was found between the co-efficients produced using the full dataset compared with the refitted model. Using the validation data, and optimising for sensitivity and specificity, a sensitivity of 70.0% and specificity of 73.8% was found. The area under the curve was 0.78.

Table 2. Multivariate model output

Variable	Coefficient	P value	Odds ratio	95% CI lower	95% CI higher
Age group (16-24)			1.0		
Age group (25-34)	.411	.081	1.509	0.634	3.590
Age group (35-44)	1.508	<0.0001	4.516	2.080	9.801
Age group (45-54)	2.606	<0.0001	13.551	6.452	28.464
Age group (55-64)	3.039	<0.0001	20.887	10.021	43.536
Age group (65-74)	3.428	<0.0001	30.818	14.706	64.582
Age group (75+)	3.901	<0.0001	49.457	23.510	104.039
Sex (Male)			1.0		
Sex (Female)	-.386	<0.0001	0.680	0.595	0.778
Ethnic group: White, Mixed, Other			1.0		
Ethnic group: South Asian & black	1.358	<0.0001	3.889	2.928	5.166
Constant	-4.907	<0.0001			

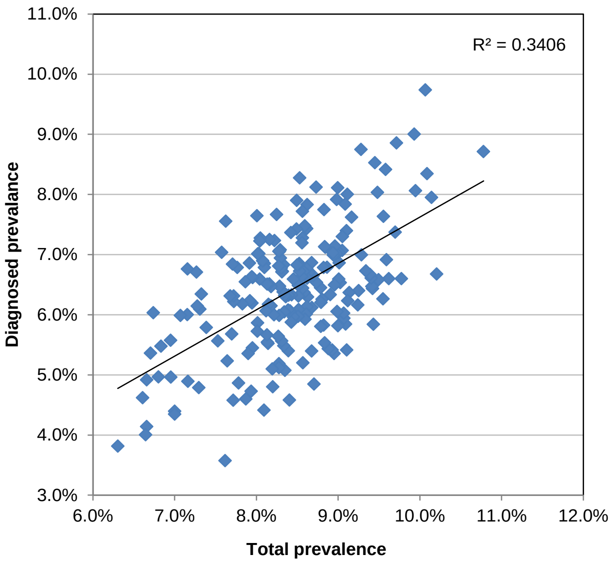
Accounting for deprivation

Deprivation was found to have an independent association with diabetes after adjusting for age, sex and ethnicity (p value <0.0001). However, deprivation was not included as a variable in the regression model due to the nature of the local level data available, ie there are difficulties in reliably estimating deprivation by age, sex and ethnicity by CCG/local authority. A two-stage modelling approach was therefore implemented, combining regression estimates with a deprivation adjustment added at the local level. This is consistent with the APHO and PBS prevalence models which both included an adjustment for deprivation at the local level.

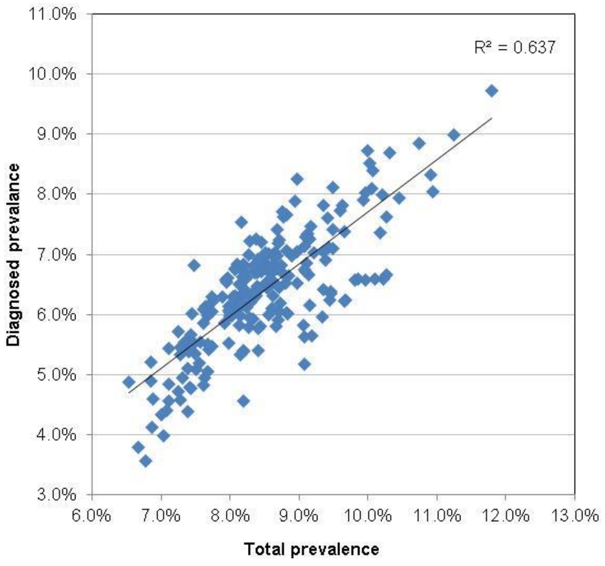
The distribution of the population within each quintile of the indices of multiple deprivations (IMD) 2015 were calculated for each area (local authority and CCG). These data were combined with the relative risks of having diabetes in each quintile nationally to create a deprivation adjustment index. The total diabetes prevalence estimated from the regression model was then multiplied by the deprivation adjustment index to account for the pattern of deprivation within each area. The deprivation adjustment does not alter the total number of people estimated to have diabetes in England but alters their distribution across local authorities and CCGs.

The results of the model output with and without the deprivation adjustment were compared with diagnosed diabetes prevalence recorded in GP registers in the Quality and Outcomes Framework (QOF)[14] in 2014/15 by CCG. The model with deprivation adjustment gave a significantly improved match (graphs 1 and 2).

Graph 1: Without deprivation adjustment



Graph 2: With deprivation



Population data

Estimates of diabetes are created using resident populations and GP registered populations. The 2014 population projections produced by the ONS were used for the resident population estimates. The population projections are available for both local authority and CCG by age and sex. Population projections are produced until the year 2035.

The numbers of patients registered by GP practice in April 2015 were used for the registered population estimates. The population estimates are available for CCG by age and sex. The year-on-year change in population by age, sex and CCG produced by the ONS resident population estimates was applied to the GP practice populations to produce population projections until the year 2035.

The registered population is over two million higher (4% higher) than then resident population. There is wide variation between the resident and registered population by CCG ranging from 16% lower (Nottingham West CCG) to 46% higher (Bradford City CCG). The prevalence estimates for the registered population will therefore differ to the prevalence estimates for the resident population due to overall differences in the population and population mix.

Estimates of ethnicity by CCG/local authority and ten-year age bands and sex were derived from Hospital Episode Statistics (HES) admissions between 2012/13 to 2014/15 for patients of known ethnicity. Patients with unknown ethnicity were excluded. The proportion of admissions by ‘South Asian and black’ and ‘white, mixed, other’ by 10-year age bands and sex and local area were applied to the population estimates by age and sex. It was assumed that the HES admissions data was representative of the ethnic population of the area and that unknown ethnicity was missing at random.

Comparisons were made between model output produced using ethnic proportions calculated from the HES data with those from the 2011 census. There was little difference to the overall prevalence estimate for England, a difference of 0.04%, with the majority of local authority prevalence estimates within 0.1% of each other and nearly all within 0.2%. The remainder were within 0.5%. The largest differences were observed primarily in the London area and could be due to actual differences in the ethnic populations, although this cannot be verified. Recent research has supported the usefulness of ethnicity data from UK-based primary and secondary care[15,16] while acknowledging the presence of bias due to missing or incorrect ethnicity data. The use of HES data provides a mechanism to continue to provide ethnicity estimates that are more timely than the 10-year census data.

Model application

The model was applied to each distinct sub-population in a CCG/local authority (as defined by the input variables) to calculate a probability of having diabetes for individuals in that group. The probability was then multiplied by the size of the sub-population and the resulting numbers accumulated to give the estimated number of people with diabetes for that CCG/local authority. This was then multiplied by the deprivation adjustment and then divided by the total population aged 16 years and over to give the prevalence estimates by CCG/local authority.

References

1 Health and Social Care Information Centre. National Diabetes Audit 2012-2013: report 1: care processes and treatment targets. [Internet]. Leeds: Health and Social Care Information Centre; 2014 [cited 2016 May 20] Available from: http://www.hscic.gov.uk/catalogue/PUB149702

2 Health Survey for England, UK data service https://discover.ukdataservice.ac.uk/series/?sn=2000021

3 Population Projections Unit. 2012-based subnational population projections: local authorities in England mid-2012 to mid-2037. London: Office for National Statistics; 2014

4 Number of patients registered at a GP practice – April 2015 The Health and Social Care Information Centre

5 APHO diabetes prevalence model for England www.yhpho.org.uk//resource/view.aspx?RID=81126

6 Forouhi NG, Merrick D, Goyder E, Ferguson BA, Abbas J, Lachowycz K and Wild SH Diabetes prevalence in England 2001 – estimates from an epidemiological model Diabet Med 2005; 23: 189-197

7 Health Survey for England – 2013: Methods and documentation www.hscic.gov.uk/catalogue/PUB16076

8 World Health Organization. Use of glycated haemoglobin. Geneva: World Health Organisation; 2011 [cited 2015 Jul 15]. Available from: www.who.int/diabetes/publications/diagnosis_diabetes2011/en

9 Personal communication, Anne Connelly, Research Director, NatCen Social Research

10 International Diabetes Federation: www.idf.org/about-diabetes/risk-factors

11 Scottish Diabetes Survey Monitoring Group Scottish Diabetes Survey 2014 [Internet]. Edinburgh: NHS Scotland; 2014 [cited 2016 May 20]. Available from: http://www.diabetesinscotland.org.uk/Publications/SDS2014.pdf

12 Tabak AG, Herder C, Rathmann W, Brunnder E J et al. Prediabetes: a high-risk state for diabetes development. Lancet. 2012 Jun 16; 379(9833):2279-90.

13 Morris DH, Khunti K, Achana F, Srinivasan et al. Progression rates from HbA1c 6.0-6.4% and other prediabetes definitions to type 2 diabetes: a meta-analysis. Diabetologia. 2013 Jul; 56(7):1489-93.

14 Quality and Outcomes Framework (QOF), 2014/15, Copyright © 2015, re-used with the permission of the Health and Social Care Information Centre. All rights reserved

15 Rohini Mathur, Completeness and usability and ethnicity data in UK-based primary care and hospital databases, Journal of Public Health

16 CL Saunders, GA Abel, A EL Tuabi, F Ahmed, G Lyratzopoulous, Accuracy of routinely recorded ethnic group information compared to self-reported ethnicity: evidence from the English Cancer Patient Experience Survey, BMJ Open