

BMJ Open Real-world evidence studies into treatment adherence, thresholds for intervention and disparities in treatment in people with type 2 diabetes in the UK

Andrew McGovern, William Hinton, Ana Correa, Neil Munro, Martin Whyte, Simon de Lusignan

To cite: McGovern A, Hinton W, Correa A, *et al*. Real-world evidence studies into treatment adherence, thresholds for intervention and disparities in treatment in people with type 2 diabetes in the UK. *BMJ Open* 2016;**6**: e012801. doi:10.1136/bmjopen-2016-012801

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2016-012801>).

Received 24 May 2016
Revised 21 September 2016
Accepted 3 November 2016



CrossMark

Department of Clinical and Experimental Medicine, University of Surrey, Guildford, UK

Correspondence to
Dr Andrew McGovern;
andy@mcgov.co.uk

ABSTRACT

Purpose: The University of Surrey-Lilly Real World Evidence (RWE) diabetes cohort has been established to provide insights into the management of type 2 diabetes mellitus (T2DM). There are 3 areas of study due to be conducted to provide insights into T2DM management: exploration of medication adherence, thresholds for changing diabetes therapies, and ethnicity-related or socioeconomic-related disparities in management. This paper describes the identification of a cohort of people with T2DM which will be used for these analyses, through a case finding algorithm, and describes the characteristics of the identified cohort.

Participants: A cohort of people with T2DM was identified from the Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) data set. This data set comprises electronic patient records collected from a nationally distributed sample of 130 primary care practices across England with scope to increase the number of practices to 200.

Findings to date: A cohort (N=58 717) of adults with T2DM was identified from the RCGP RSC population (N=1 260 761), a crude prevalence of diabetes of 5.8% in the adult population. High data quality within the practice network and an ontological approach to classification resulted in a high level of data completeness in the T2DM cohort; ethnicity identification (82.1%), smoking status (99.3%), alcohol use (93.3%), glycated haemoglobin (HbA1c; 97.9%), body mass index (98.0%), blood pressure (99.4%), cholesterol (87.4%) and renal function (97.8%). Data completeness compares favourably to other, similarly large, observational cohorts. The cohort comprises a distribution of ages, socioeconomic and ethnic backgrounds, diabetes complications, and comorbidities, enabling the planned analyses.

Future plans: Regular data uploads from the RCGP RSC practice network will enable this cohort to be followed prospectively. We will investigate medication adherence, explore thresholds and triggers for changing diabetes therapies, and investigate any ethnicity-related or socioeconomic-related disparities in diabetes management.

Strengths and limitations of this study

- A large cohort of people with type 2 diabetes was defined that has a higher level of data completeness than other similar observational cohorts.
- The cohort comprises a distribution of ages, socioeconomic and ethnic backgrounds, diabetes complications, and comorbidities.
- The network of primary care practices from which data were included comprise volunteer practices which may not be fully representative of the national management and outcomes in people with type 2 diabetes.
- Data analysis using this cohort will be observational and therefore limited by the nature of observational studies.

INTRODUCTION

Type 2 diabetes mellitus (T2DM) is a growing problem worldwide and is associated with an increased risk of cardiovascular disease and mortality.^{1–4} Many glucose-lowering agents have become available in recent years including three new classes of medication: dipeptidase IV (DPP4) inhibitors, glucagon-like peptide 1 (GLP1) receptor agonists and sodium glucose co-transporter 2 inhibitors. With the addition of these newer therapies, the management of T2DM has become more complex.^{5–7} With this increasingly complex array of therapies, it is more important than ever to scrutinise the relative clinical effectiveness of different drug classes in the real-world setting.^{8 9} Selecting optimal second and third line glucose-lowering therapies has become difficult for clinicians as there is limited evidence to support the use of one therapy over another in clinical practice.^{10 11} While the newer agents have demonstrated

blood glucose-lowering efficacy in clinical trials and good cardiovascular safety profiles (and even potential cardiovascular benefit with empagliflozin and liraglutide),^{12–17} their effectiveness in the real-world setting should be monitored, particularly as patient profiles often differ from those included in trials.¹¹

We have identified a large, real-world cohort of people with T2DM which we will use to investigate three important components of diabetes management in clinical practice: (1) medication adherence and persistence, (2) thresholds for changing therapies (in particular treatment intensification), and (3) socioeconomic and ethnic disparities in management:

1. Adherence and persistence are thought to differ markedly between the clinical trial setting and routine clinical practice. Adherence also declines with increasing complexity of treatment.^{18 19} Reduced medication adherence is associated with poorer glycaemic control,^{20–23} increased hospital admissions^{24–27} and increased mortality.^{26 28} Using the cohort described here, we aim to quantify the level of medication adherence to newer and older therapies in the real-world setting.
2. Glucose control in clinical practice is frequently sub-optimal.²⁹ A key factor contributing to this is delay in treatment intensification of glucose-lowering therapy in real-world practice (termed clinical inertia). Despite clinical inertia being identified nearly 10 years ago,^{30 31} the most recent studies do not show any move towards earlier intensification of treatment.³² This may be partially due to patients' and providers' reluctance to initiate injectable therapies,³³ and partially due to concerns that glycaemic targets are not always achieved even with addition of insulin.^{34 35} We also intend to use this cohort to further explore factors which predict treatment intensification and to attempt to identify modifiable clinical factors which may delay or prevent good glycaemic control through treatment intensification. In the UK, the Quality and Outcomes Framework (QOF), an annual reward and incentive programme for general practitioners, sets targets for the management of diabetes (and multiple other conditions and aspects of care).³⁶ We will also identify the impact of the changing QOF glycaemic control targets on treatment intensification.
3. In our preliminary analyses, we have identified some disparities in medication prescriptions³⁷ and medication adherence across socioeconomic and ethnic groups. We intend to use this cohort to explore these potential disparities further.

The cohort described here has been identified using routinely collected data from primary care centres in England. Electronic health records in the UK are coded using Read codes. This system enables coding of information including patient demographics, medical conditions, test results, processes of care and other important healthcare-related data. However, these data have some

limitations, and in diabetes in particular, people are often misclassified or miscoded.^{38 39} QOF targets in primary care, based on clinical codes, have improved clinical coding but difficulties in correct identification of people with diabetes remain. We describe the methods used to minimise the impact of coding limitations and to identify a cohort of people with T2DM. We also describe the characteristics of this T2DM population.

COHORT DESCRIPTION

The cohort comprises all adults with T2DM identified from practices participating in the Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) network. This data set consists of data collected from 130 primary care practices across England, with ongoing recruitment of practices. The RCGP RSC network comprises 1.7% of all primary care practices in England (N=7779).⁴⁰ Only people who were alive and registered with their current practice on 31 July 2015 have been included for analysis. People who were under 18 years old on this date were excluded from the cohort.

We use a two-step informatics ontology-based process to identify people with T2DM. Initially all people with diabetes were identified using a combination of clinical codes (see below). People were then subcategorised into diabetes type 1, type 2 or other, using a seven-step algorithm. We have used all available electronic patient records prior to 31 July 2015 to identify the cohort. Ongoing data uploads from the participating practices will provide prospective follow-up data.

Method for identification and classification of the cohort

Step 1: identification of people with diabetes

People with diabetes were identified using a combination of diagnostic codes, investigation results and medication codes. We have used all commonly recorded, major diagnostic codes, as well as process of care codes for the diagnosis of diabetes (see online supplementary table S1) and excluded codes specific to gestational diabetes, steroid-induced diabetes or other forms of secondary diabetes.

Investigation results included both blood glucose and glycated haemoglobin (HbA1c) measurements. Glucose 'Read codes' were stratified by test provenance (fasting, random, glucose tolerance test, unknown or other) using a categorisation system that we have previously described.⁴¹ Individuals were included if they had two or more recorded glucose values which met the diagnostic criteria for diabetes (≥ 7.0 mmol/L for fasted samples and ≥ 11.1 mmol/L for random samples; samples taken at 2 hours, postglucose load, as part of a glucose tolerance test, or where the provenance of the test was unknown).⁴² Individuals were also included if they had two or more measurements of HbA1c above the diagnostic threshold; ≥ 48 mmol/mol (6.5%).⁴³ Individuals with a single laboratory measurement consistent with diabetes were not included unless they met one of the other

criteria for diabetes. Finally, people were also included in the diabetes cohort if they had been prescribed oral hypoglycaemic agents (except metformin) or if they had been prescribed insulin (on at least two occasions) or GLP1 analogues.

Step 2: classification of diabetes

Once identified as having diabetes, we categorised people as either type 1 diabetes mellitus (T1DM) or T2DM using the algorithm shown in figure 1. Given the occasional inaccuracy in clinical recording of diabetes type, the first three steps of this categorisation algorithm allows over-riding of diagnostic codes where clinical characteristics are highly likely to indicate a specific type of diabetes:

Step 1: If the person had insulin initiated before age 35 and no oral agent use ever recorded (except metformin), they are categorised as having T1DM.

Step 2: If on multiple oral hypoglycaemic agents, the person is categorised as T2DM.

Step 3: If on one or more oral hypoglycaemic agent, excluding metformin, for more than 12 months and no history of insulin use, the person is categorised as T2DM.

If none of these criteria are met, then step 4 searches for diagnostic codes with no evidence of conflict; if all clinically entered codes are consistent with either T1DM or T2DM, then the person is categorised accordingly. Occasionally, people are coded as having both T1DM and T2DM and so would be considered to have a conflicting code type. If there are no recorded diagnostic codes or codes are conflicting, then the algorithm proceeds through steps 5–7:

Step 5: If there is no documented insulin use in the entire clinical record, then the person is categorised as T2DM.

Step 6: If prescribed oral hypoglycaemic agents for more than 12 months, then the person is categorised as T2DM.

Step 7: If the body mass index (BMI) measured nearest to the time of diagnosis is ≥ 25 kg/m², then the person is categorised as T2DM.

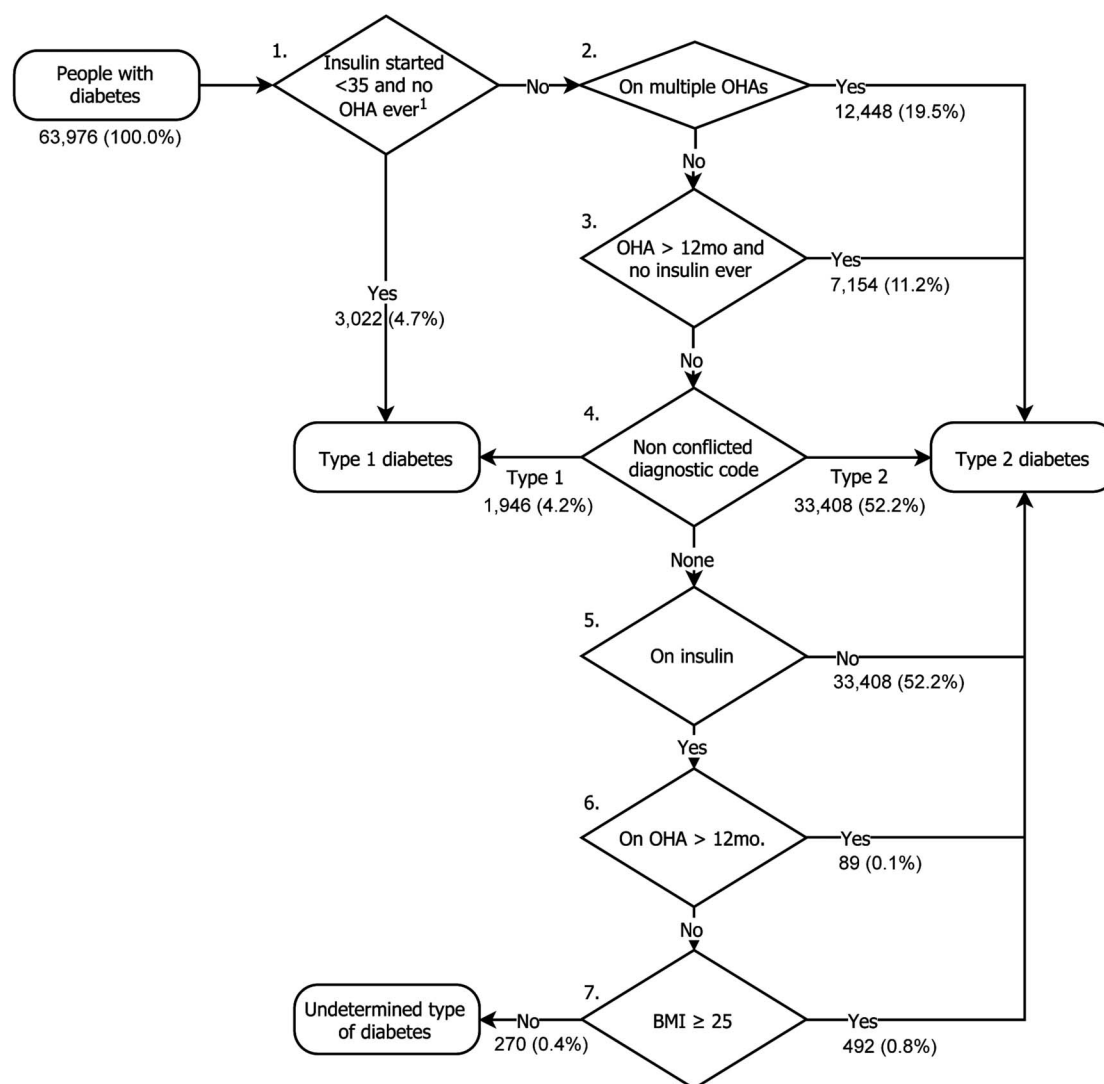


Figure 1 Seven-step algorithm for categorising type of diabetes. ¹Excluding prescriptions for metformin. BMI, body mass index; OHA, oral hypoglycaemic agent.

If the type of diabetes is not identified by these steps, then the type is categorised as 'undetermined'. People with an undetermined diabetes type will not be included in any subsequent analyses. We report the absolute rate of diabetes in the population and age-gender direct standardised rate, adjusted using the 2011 UK census.⁴⁴

Method for describing the cohort

We describe the demographic, social and clinical characteristics of the identified cohort. The demographics of people included in the cohort are reported using descriptive statistics; age (as of 31 July 2015), gender and ethnicity. Ethnicity recording in primary care is limited with between 44.5% and 75.4% ethnicity recording in diabetes populations.⁴⁵ Ethnicity identification was therefore extended using an informatics ontology which uses a combination of ethnicity codes and codes which infer ethnicity (such as area-specific languages).⁴⁶ All inferential codes were checked for consistency with definitive ethnicity codes (>0.85 positive predictive value

for ethnic group). We are due to report this ethnicity identification algorithm elsewhere (results under peer review). The distribution of deprivation scores is also reported. This was calculated at the point of data extraction, from patient postcode, using the index of multiple deprivation (IMD).⁴⁷ Higher values for IMD score represent a greater level of socioeconomic deprivation.

The key baseline clinical characteristics of the cohort were also defined (using an ontological approach where appropriate) and are also reported; smoking status, alcohol use, BMI, HbA1c, blood pressure, cholesterol and renal function (estimated glomerular filtration rate). Smoking status and alcohol use are reported using the most recently recorded values. Similarly, BMI, HbA1c, blood pressure and cholesterol measurements are reported using the most recently recorded value. Where data on BMI were missing, we have calculated BMI from the most recent weight and height measurements. We report the number of people with missing data for all variables.

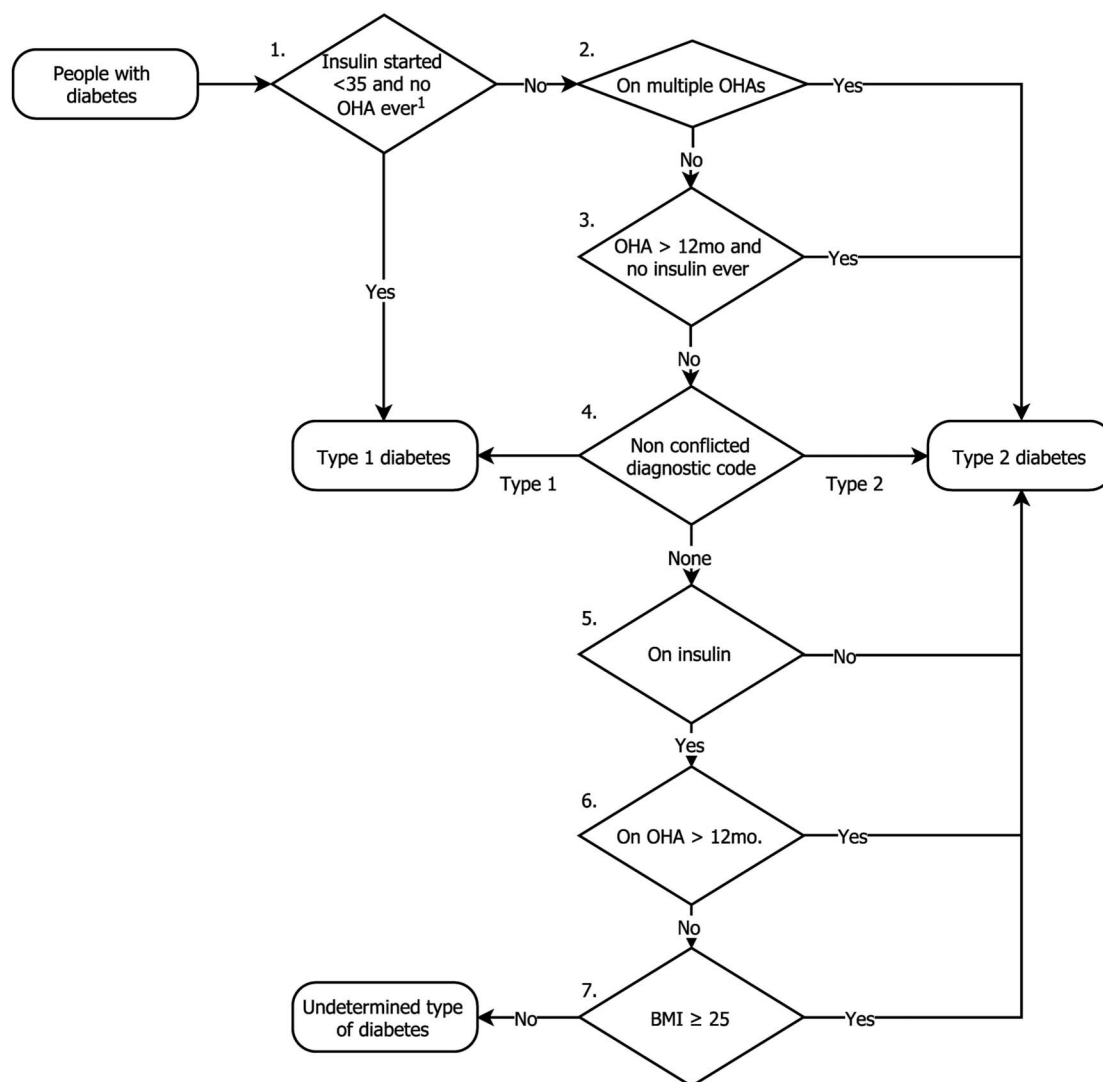


Figure 2 Seven-step algorithm for categorising type of diabetes. The number of people classified at by each step is shown. ¹Excluding prescriptions for metformin. BMI, body mass index; OHA, oral hypoglycaemic agent.

Table 1 The characteristics of the complete adult population, adults with T1DM (for comparison) and adults with T2DM

Characteristic	All adults n (%*) or mean (SD) (N=1 015 867)	Adults with T1DM n (%*) or mean (SD) (N=4478)	Adults with T2DM n (%*) or mean (SD) (N=58 717)
Age (years)	47.8 (18.9)	42.3 (15.5)	66.2 (14.1)
Female	518 456 (51.0)	2075 (46.3)	26 706 (45.4)
Ethnicity recorded	733 024 (72.2)	3542 (79.1)	48 199 (82.1)
White	637 288 (62.7)	3208 (71.6)	40 380 (68.8)
Asian	54 632 (5.4)	186 (4.1)	5127 (8.7)
Black	22 705 (2.2)	85 (1.9)	1858 (3.2)
Mixed	9278 (0.9)	40 (0.9)	483 (0.8)
Other	9121 (0.9)	23 (0.5)	351 (0.6)
IMD recorded	959 767 (94.5)	4217 (94.2)	56 522 (96.3)
IMD score†	18.9 (15.4)	19.8 (16.0)	20.5 (16.3)
Smoking status recorded	785 051 (93.0)	3590 (99.4)	50 890 (99.3)
Never	462 574 (54.8)	1995 (55.2)	20 587 (40.2)
Current	322 477 (38.2)	865 (23.9)	7485 (14.6)
Ex-smoker	171 570 (20.3)	1595 (44.1)	30 303 (59.1)
Alcohol use recorded	689 913 (67.9)	3938 (87.9)	54 808 (93.3)
None	163 476 (16.1)	1018 (22.7)	19 499 (33.2)
Within limits	297 526 (29.3)	1373 (30.7)	14 176 (24.1)
Excess	228 911 (22.5)	1547 (34.5)	21 133 (36.0)
Duration of diabetes recorded	–	4478 (100.0)	58 717 (100.0)
<4 years	–	943 (21.1)	18 513 (31.5)
4≤x<7 years	–	654 (14.6)	11 516 (19.6)
7≤x<10 years	–	760 (17.0)	11 321 (19.3)
≥10 years	–	2121 (47.4)	17 092 (29.1)
HbA1c measured	–	4288 (95.8)	57 501 (97.9)
HbA1c (mmol/mol)	–	69.8 (20.9)	56.1 (16.5)
BMI measured	803 995 (79.1)	4377 (97.7)	57 557 (98.0)
BMI (kg/m ²)	26.6 (5.9)	26.2 (6.2)	31.4 (6.6)
Blood pressure measured	847 028 (83.4)	4478 (99.0)	58 379 (99.4)
SBP (mm Hg)	126.6 (15.4)	126.1 (14.5)	131.4 (14.4)
DBP (mm Hg)	76.0 (9.7)	74.1 (9.2)	75.0 (9.4)
Total cholesterol measured	77 120 (7.6)	723 (16.1)	8025 (13.7)
Total cholesterol (mmol/L)	5.3 (1.1)	4.6 (1.1)	4.9 (1.3)
LDL cholesterol measured	424 756 (41.8)	3482 (77.8)	51 336 (87.4)
LDL cholesterol (mmol/L)	3.2 (1.0)	2.6 (0.9)	3.0 (1.1)
eGFR recorded	586 052 (57.7)	4054 (90.5)	57 404 (97.8)

*Percentage values shown are of the column total.

†A higher score equates to a higher level of deprivation although the scale is non-linear.

BMI, body mass index; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; HbA1c, glycated haemoglobin; IMD, index of multiple deprivation; LDL, low-density lipoprotein; SBP, systolic blood pressure; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

Duration of diabetes at the start of follow-up is also reported. This has been defined as the time since the first diabetes defining event. These events comprise: first diabetes diagnostic code, first investigation result consistent with diabetes or date of first prescription of glucose-lowering medication. We use the duration of diabetes to identify a recent onset population (those with a diagnosis of T2DM for <2 years) and report the characteristics of this subset separately. This subset of newly diagnosed people with T2DM will enable a prospective analysis of current trends in early management.

Clinical codes were also used to define the presence of comorbidities and diabetes complications; hypertension, atrial fibrillation, angina, myocardial infarction,

congestive heart failure, stroke, transient ischaemic attack, chronic kidney disease (stages 3–5), current renal replacement therapy, dementia, depression, previous amputation, peripheral neuropathy, retinopathy (all stages), peripheral vascular disease, rheumatoid arthritis, chronic liver disease (all causes) and non-alcoholic fatty liver disease (NAFLD).

Cohort characteristics

A total of 1 260 761 people were included from 130 primary care practices. In total, 1 015 867 were aged 18 or older on 31 August 2015. From these, 63 465 (6.25%) people were identified as having any type of diabetes. The majority of these (58 717; 91.8%) were categorised as most likely to have T2DM, and a minority T1DM

(4478; 7.62%) or not categorised (270; 0.43%) (figure 2). The crude prevalence of T2DM in the adult population was 5.78% and direct standardised prevalence (using the 2011 UK census⁴⁴) similar at 5.78% (95% CI 5.72% to 5.82%).

The mean age of the T2DM cohort was 66.2 (SD 14.1) years and 26 706 (45.4%) were female (table 1). The age and gender distribution is shown in figure 3. The crude prevalence of cardiovascular comorbidities was higher in adults with T2DM than T1DM; however, the age and gender standardised rates were similar (tables 2 and 3, respectively). A higher prevalence of diabetes complications (peripheral vascular disease, peripheral neuropathy, retinopathy and chronic kidney disease) were seen in people with T1DM after age and gender standardisation (table 3).

The crude prevalence of T2DM varied considerably by ethnicity: white 6.34%, Asian 9.38%, black 8.18%, mixed 5.21%, other 3.85% and no ethnicity identified 3.72%. After adjusting for age and gender, the prevalence of T2DM remained highly variable across ethnicity groups: white 5.91% (95% CI 5.85% to 5.96%), Asian 16.99% (16.48% to 17.52%), black 13.84% (13.11% to 14.65%), mixed 11.81% (10.62% to 13.46%), other 9.25% (8.10% to 10.65%) and no ethnicity identified 3.74% (3.67% to

3.81%). Clinical characteristics also vary considerably by ethnicity in people with T2DM (table 4).

Metformin was the most commonly prescribed diabetes medication in people with T2DM (table 5). In total, 12 284 (20.9%) people had no diabetes medication prescribed. The subset of people with recent onset (<2 years) had a lower prevalence of complications than the complete cohort of people with T2DM (table 6).

DISCUSSION AND FINDINGS TO DATE

A large cohort (N=58 717) of adults with T2DM has been identified from the RCGP RSC population. The characteristics of this cohort are described here and compared with those of the complete adult population and a population of people with T1DM. We have identified a large burden of macrovascular and microvascular disease in our T2DM cohort and a broad range of therapies in use. This baseline analysis demonstrates high levels of data completeness in the diabetes population for a range of variables including ethnicity, smoking status, alcohol use, HbA1c, BMI, blood pressure, low-density lipoprotein cholesterol and renal function. This is higher than that reported for other T2DM

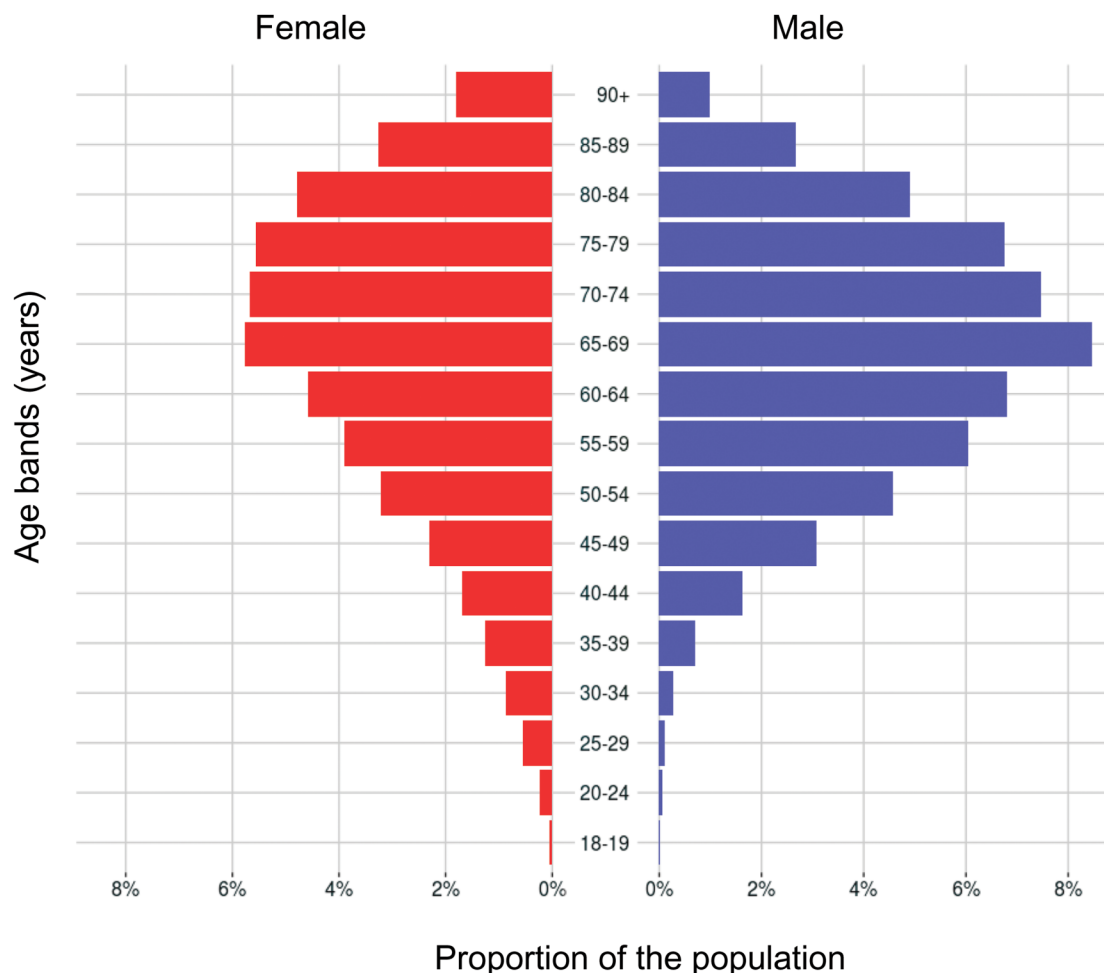


Figure 3 The age and gender profile of the cohort of people with T2DM. T2DM, type 2 diabetes mellitus.

Table 2 The number of people with common comorbidities of the complete adult population, adults with T1DM (for comparison), and adults with T2DM

Comorbidities	All adults n (%) (N=1 015 867)	Adults with T1DM n (%) (N=4478)	Adults with T2DM n (%) (N=58 717)
Hypertension	164 502 (16.2)	967 (21.6)	33 467 (57.0)
Atrial fibrillation	25 209 (2.5)	53 (1.2)	5253 (8.9)
Angina	17 837 (1.8)	83 (1.9)	4688 (8.0)
Myocardial infarction	11 227 (1.1)	73 (1.6)	2697 (4.6)
Congestive heart failure	11 216 (1.1)	61 (1.4)	3146 (5.4)
Stroke	12 120 (1.2)	50 (1.1)	2366 (4.0)
Transient ischaemic attack	9870 (1.0)	28 (0.6)	1838 (3.1)
Chronic kidney disease	41 054 (4.0)	285 (6.4)	10 585 (18.0)
Renal replacement therapy	1163 (0.1)	56 (1.3)	312 (0.5)
Dementia	9964 (1.0)	19 (0.4)	1883 (3.2)
Depression	95 941 (9.4)	657 (14.7)	6927 (10.7)
Amputation	1799 (0.2)	65 (1.5)	472 (0.8)
Peripheral neuropathy	7170 (0.7)	326 (7.3)	3847 (6.6)
Retinopathy	24 698 (2.4)	2667 (59.6)	21 396 (36.4)
Peripheral vascular disease	6820 (0.7)	98 (2.2)	2172 (3.7)
Rheumatoid arthritis	8519 (0.8)	40 (0.9)	1028 (1.8)
Chronic liver disease*	10 651 (1.0)	80 (1.8)	2638 (4.5)
NAFLD	6530 (0.6)	54 (1.2)	1989 (3.4)

*All liver disease including NAFLD.

NAFLD, non-alcoholic fatty liver disease; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

populations identified from other large UK-based population databases; such as the QResearch database and the Clinical Practice Research Database (CPRD).⁴⁵

Given that this T2DM cohort has a high level of data completeness, a long duration of follow-up, a range of socioeconomic and ethnic groups represented, and covers people with a range of diabetes medications and with a range of diabetes-related complications, it is ideally suited for a number of different population-based studies. The planned analyses include medication adherence and persistence, thresholds for changing therapies (in particular treatment intensification), and socioeconomic and ethnic disparities in management and outcomes.

The crude prevalence of most comorbidities was higher in people with T2DM than those with T1DM. Similar trends have been reported in a large UK-based population.⁴⁸ However, most diabetes complications (peripheral neuropathy, retinopathy and amputations) were more prevalent in people with T1DM. In contrast, age-adjusted and gender-adjusted rates of comorbidities were similar between the groups. After adjustment, only hypertension remained more prevalent in T2DM, whereas the prevalence of peripheral neuropathy, retinopathy and peripheral vascular disease remained higher in people with T1DM.

Ethnic differences in diabetes prevalence are well established.⁴⁹ Our data show that the T2DM prevalence has increased in all ethnicity groups when compared with UK data reported 10 years ago.⁵⁰ In addition, our data demonstrate substantial variation in clinical characteristics across these groups.

Strengths and limitations

We have recently reported that the RCGP RSC population is broadly representative of the national population albeit with some oversampling of 25–44 years old, and less deprived people.⁵¹ In addition to the strengths already discussed, this subcohort of people with T2DM has the advantages of large size and of a high frequency of data uploads which enable prospective follow-up. Data are uploaded to the RCGP RSC database weekly.

An additional strength of the cohort is of high-quality recording of infectious diseases (primary care practices which comprise the RCGP RSC receive regular feedback on the quality of their recording of infectious diseases, which should ensure a high standard of recording). This has already enabled analysis of infection rates in elderly people with T2DM.⁵² Furthermore, many of the UK primary care pay-for-performance targets for people with T2DM include recording and monitoring of disease indicators relevant to diabetes such as HbA1c, peripheral neuropathy and renal function, which have led to the high level of data completeness reported here.

Limitations include the observational nature of the data and some residual incompleteness of recording of variables. Incomplete data recording in people with diabetes is disproportionately more common in people with poor outcomes.⁵³ The network is also comprised of volunteer practices, which may be better at managing diabetes and related conditions or at recording these conditions than other primary care practices. Indeed we have previously reported that the included practices score more highly on pay-for-performance targets.⁵¹

Table 3 Crude and age-standardised and gender-standardised prevalence (compared with the 2011 population census) for common comorbidities of the complete adult population, adults with T1DM (for comparison), and adults with T2DM

Comorbidities	All adults (N=1 015 867)		Adults with T1DM (N=4478)		Adults with T2DM (N=58 717)	
	Crude prevalence (%)	Adjusted prevalence (95% CI) (%)	Crude prevalence (%)	Adjusted prevalence (95% CI) (%)	Crude prevalence (%)	Adjusted prevalence (95% CI) (%)
Hypertension	16.2	16.1 (16.1 to 16.2)	21.6	28.1 (25.9 to 30.9)	57.0	34.6 (33.9 to 35.9)
Atrial fibrillation	2.48	2.42 (2.39 to 2.46)	1.18	2.00 (1.44 to 3.38)	8.94	3.47 (3.34 to 4.50)
Angina	1.76	1.73 (1.70 to 1.75)	1.85	2.70 (2.09 to 4.11)	7.98	3.49 (3.31 to 4.53)
Myocardial infarction	1.11	1.09 (1.07 to 1.11)	1.63	2.32 (1.74 to 3.70)	4.59	2.23 (2.08 to 3.36)
Congestive heart failure	1.11	1.08 (1.06 to 1.10)	1.36	2.62 (1.75 to 4.29)	5.36	2.37 (2.19 to 3.43)
Stroke	1.19	1.17 (1.15 to 1.19)	1.11	1.85 (1.33 to 3.21)	4.03	1.73 (1.65 to 2.78)
Transient ischaemic attack	0.97	0.95 (0.93 to 0.97)	0.63	1.01 (0.65 to 2.31)	3.13	1.31 (1.24 to 2.37)
Chronic kidney disease	4.23	4.15 (4.11 to 4.19)	6.47	10.8 (9.26 to 12.9)	18.7	7.67 (7.42 to 8.72)
Renal replacement therapy	0.11	0.11 (0.11 to 0.12)	1.25	1.21 (0.90 to 2.45)	0.53	0.33 (0.28 to 1.47)
Dementia	0.98	0.94 (0.92 to 0.96)	0.42	1.19 (0.68 to 2.61)	3.21	1.13 (1.07 to 2.19)
Depression	9.44	9.50 (9.44 to 9.56)	14.7	14.4 (13.2 to 16.2)	11.8	14.7 (13.9 to 16.1)
Amputation	0.17	0.17 (0.17 to 0.18)	1.45	1.80 (1.33 to 3.12)	0.80	0.41 (0.36 to 1.53)
Peripheral neuropathy	0.71	0.70 (0.68 to 0.71)	7.28	9.38 (8.12 to 11.3)	6.55	3.44 (3.30 to 4.48)
Retinopathy	2.43	2.42 (2.39 to 2.46)	59.6	62.8 (60.1–66.0)	36.4	23.6 (22.9 to 24.9)
Peripheral vascular disease	0.67	0.66 (0.64 to 0.68)	2.18	3.42 (2.70 to 4.89)	3.70	1.53 (1.46 to 2.59)
Rheumatoid arthritis	0.84	0.84 (0.83 to 0.86)	0.89	1.32 (0.89 to 2.65)	1.75	1.08 (0.99 to 2.15)
Chronic liver disease*	1.05	1.06 (1.03 to 1.07)	1.78	1.66 (1.29 to 2.91)	4.49	4.70 (4.25 to 5.86)
NAFLD	0.64	0.64 (0.63 to 0.66)	1.21	1.07 (0.79 to 2.31)	3.39	3.80 (3.37 to 4.97)

*All liver disease including NAFLD.

NAFLD, non-alcoholic fatty liver disease; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

Table 4 Clinical characteristics of people with T2DM (N=58 717) by ethnicity group

Characteristic	White n (%) or mean (SD) (N=40 380)	Asian n (%) or mean (SD) (N=5127)	Black n (%) or mean (SD) (N=1858)	Mixed n (%) or mean (SD) (N=483)	Other n (%) or mean (SD) (N=351)	Not recorded n (%) or mean (SD) (N=12 128)
Age (years)	67.5 (13.6)	58.4 (14.4)	59.6 (13.7)	60.2 (14.3)	57.5 (13.4)	66.6 (14.3)
IMD recorded	38 949 (96.5)	4888 (95.3)	1780 (95.8)	463 (95.9)	338 (96.3)	10 104 (96.1)
IMD score*	19.4 (16.0)	27.6 (17.6)	30.8 (14.5)	24.8 (16.3)	25.9 (15.5)	19.4 (15.8)
Smoking status recorded	40 380 (99.8)	5127 (99.7)	1858 (99.9)	483 (99.6)	351 (99.7)	10 518 (97.6)
Never	12 320 (30.5)	2812 (54.8)	1047 (56.4)	236 (48.9)	157 (44.7)	4015 (38.2)
Current	5413 (13.4)	564 (11.0)	167 (9.0)	63 (13.0)	53 (15.1)	1225 (11.6)
Ex-smoker	22 574 (55.9)	1737 (33.9)	642 (34.6)	182 (37.7)	140 (39.9)	5028 (47.8)
Alcohol use recorded	38 640 (95.7)	4950 (96.5)	1787 (96.2)	466 (96.5)	333 (94.9)	8632 (82.1)
None	11 295 (28.0)	3435 (67.0)	1027 (55.3)	218 (45.1)	193 (55.0)	3331 (31.7)
Within limits	10 284 (25.5)	726 (14.2)	428 (23.0)	118 (24.4)	68 (19.4)	2552 (24.3)
Excess	17 061 (42.3)	789 (15.4)	332 (17.9)	130 (26.9)	72 (20.5)	2749 (26.1)
Duration of diabetes						
<4 years	11 856 (29.4)	1936 (37.8)	708 (38.1)	193 (40.0)	140 (39.9)	3680 (35.0)
4≤x<7 years	8039 (19.9)	988 (19.3)	393 (21.2)	97 (20.1)	80 (22.8)	1919 (18.2)
7≤x<10 years	8107 (20.1)	905 (17.7)	295 (15.9)	84 (17.4)	67 (19.1)	1863 (17.7)
≥10 years	12 378 (30.7)	1298 (25.3)	462 (24.9)	109 (22.6)	64 (18.2)	3056 (29.1)
HbA1c measured	39 687 (98.3)	5054 (98.6)	1819 (97.9)	472 (97.7)	342 (97.4)	10 127 (96.3)
HbA1c (mmol/mol)	55.9 (16.3)	57.3 (16.8)	58.7 (20.6)	56.4 (17.5)	57.0 (19.0)	56.0 (16.4)
BMI measured	39 898 (98.8)	5062 (98.7)	1819 (97.9)	480 (99.4)	344 (98.0)	9954 (94.6)
BMI (kg/m ²)	31.9 (6.6)	28.4 (5.3)	31.3 (6.5)	30.4 (6.1)	29.8 (5.6)	31.5 (6.6)
Blood pressure measured	40 271 (99.7)	5104 (99.6)	1844 (99.2)	483 (100.0)	351 (100.0)	10 326 (98.2)
SBP (mmHg)	128.2 (19.6)	120.0 (22.6)	123.5 (24.6)	126.2 (18.5)	121.0 (23.3)	118.0 (28.5)
DBP (mmHg)	78.0 (17.1)	82.0 (21.1)	86.6 (23.6)	78.3 (16.5)	83.4 (21.8)	89.2 (26.7)

*A higher score equates to a higher level of deprivation although the scale is non-linear.

BMI, body mass index; DBP, diastolic blood pressure; HbA1c, glycated haemoglobin; IMD, index of multiple deprivation; SBP, systolic blood pressure; T2DM, type 2 diabetes mellitus.

Table 5 Proportion of people with T2DM (n=58 717) prescribed medications in the past 1, 3 and 10 years prior to 31 July 2015

Medication	Prescribed in past 1 year n (%)	Prescribed in past 3 years n (%)	Prescribed in past 10 years n (%)
Metformin	37 045 (63.1)	39 872 (67.9)	43 179 (73.5)
Thiazolidinediones	1826 (3.1)	2642 (4.5)	5989 (10.2)
Sulfonylureas	15 547 (26.5)	17 696 (30.1)	20 819 (35.5)
Meglitinides	174 (0.3)	257 (0.4)	541 (0.9)
DPP-4 inhibitors	6941 (11.8)	8462 (14.4)	9453 (16.1)
α-glucosidase inhibitors	98 (0.2)	146 (0.2)	295 (0.5)
SGLT2 inhibitors	1267 (2.2)	1267 (2.2)	1267 (2.2)
GLP1 analogues	1930 (3.3)	2497 (4.3)	2842 (4.8)
Insulin	8403 (14.3)	8806 (15.0)	9442 (16.1)

DPP-4, dipeptidase IV; GLP1, glucagon-like peptide 1; SGLT2, sodium glucose co-transporter 2; T2DM, type 2 diabetes mellitus.

Finally, while in our experience, the proportion of people with T2DM moving practices is generally small, anyone who moves to a practice which is not part of the RCGP RSC network will be lost to follow-up.

Collaboration

We encourage collaboration with researchers interested in pursuing research questions regarding the

identification, management, natural history or trends in people with diabetes which can be addressed using this large observational data set either prospectively or retrospectively. The database can be made available to researchers on a case-by-case basis and is subject to ethical approval for each study by the National Health Service (NHS) Research Ethics Committee. The policy of the RCGP is that patient-level data must remain

Table 6 The characteristics of people with recent onset (<2 years) T2DM

Characteristic	Adults with recent onset T2DM n (%) (N=10 105)
Age (years)	62.0 (15.8)
Female	4872 (48.2)
HbA1c measured	9616 (95.2)
HbA1c (mmol/mol)	53.1 (15.6)
BMI measured	9507 (94.1)
BMI (kg/m ²)	31.6 (6.9)
Blood pressure measured	9896 (97.9)
SBP (mm Hg)	124.5 (22.2)
DBP (mm Hg)	82.0 (20.1)
Hypertension	4414 (43.7)
Atrial fibrillation	827 (8.2)
Angina	637 (6.3)
Myocardial infarction	424 (4.2)
Congestive heart failure	452 (4.5)
Stroke	334 (3.3)
Transient ischaemic attack	263 (2.6)
Chronic kidney disease	1132 (11.2)
Renal replacement therapy	31 (0.3)
Dementia	276 (2.7)
Depression	1163 (11.5)
Amputation	37 (0.4)
Peripheral neuropathy	191 (1.9)
Retinopathy	1031 (10.2)
Peripheral vascular disease	219 (2.2)
Rheumatoid arthritis	168 (1.7)
Chronic liver disease*	394 (3.9)
NAFLD	297 (2.9)

*All liver disease including NAFLD.

BMI, body mass index; DBP, diastolic blood pressure; HbA1c, glycated haemoglobin; NAFLD, non-alcoholic fatty liver disease; SBP, systolic blood pressure; T2DM, type 2 diabetes mellitus.

within the secure network at the University of Surrey. Direct access for interested researchers can be arranged subject to the information governance standards of the University of Surrey, alternatively aggregated data can be provided. We encourage interested researchers to attend the short courses on how to analyse primary care data offered by the university biannually.

CONCLUSION

We have devised methods to optimise the identification of diabetes and diabetes subtype from the routine data collected. The large population of people with T2DM identified from the RCGP RSC cohort provides an excellent opportunity to use real-world evidence to explore the clinical use and impact of newer diabetes therapies in clinical practice, and observation of the clinical management of T2DM in practice. We encourage collaborative use of this data set with interested researchers.

Acknowledgements The authors would like to thank Dr Filipa Ferreira, project manager, for generous support throughout this project and for initiating the

research collaboration; Rachel Byford and Barbara Arrowsmith, SQL developers, for their help with database management and data extraction; and all the participating practices and patients for providing the data for this cohort. The authors would also like to thank the funders Eli Lilly and Company.

Contributors AM led the analysis and drafting of the manuscript. AM and SdL conceived the paper. WH, AC, NM and MW supported the analysis. All the authors have contributed to the final manuscript.

Funding Eli Lilly and Company. Grant number (10.13039/100004312).

Competing interests The RCGP RSC is primarily funded by Public Health England and the development of the RCGP RSC data set is supported by surveillance work funded by Public Health England. The subcohort of people with T2DM identified from the RCGP RSC has been developed for a number of planned studies funded, as part of the University of Surrey-Lilly Real World Evidence (RWE) projects, by Eli Lilly and Company. AM, WH, MW and SdL are funded by Eli-Lilly and Company. SdL and AC have undertaken research funded by GlaxoSmithKline. NM has received fees for serving as a speaker, a consultant or an advisory board member for Allergan, Bristol-Myers Squibb-Astra-Zeneca, GlaxoSmithKline, Eli Lilly, Lifescan, MSD, Metronic, Novartis, Novo Nordisk, Pfizer, Sankio, Sanofi, Roche, Servier, Takeda. MW has received speaker fees from Astra-Zeneca.

Ethics approval Health Research Authority.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The RCGP RSC data set can be accessed by bona fide researchers on a case-by-case basis. Ethical approval by the NHS Research Ethics Committee is needed for data requests to be considered. Aggregated data tables may be created from the source data to allow specific analyses for approved research and surveillance projects. Researchers wishing to directly analyse the patient-level anonymised data will be required to complete information governance training and work on the data from the secure servers at the University of Surrey. Patient-level data cannot be taken out of the secure servers at the University of Surrey. The authors encourage interested researchers to attend the short courses on how to analyse primary care data offered by the university twice a year.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Kannel WB, McGee DL. Diabetes and cardiovascular risk factors: the Framingham study. *Circulation* 1979;59:8–13.
2. Sarwar N, Gao P, Seshasai SR, *et al.* Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet* 2010;375:2215–22.
3. Danaei G, Finucane MM, Lu Y, *et al.* National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants. *Lancet* 2011;378:31–40.
4. Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus—present and future perspectives. *Nat Rev Endocrinol* 2012;8:228–36.
5. The National Institute for Health and Care Excellence (NICE). Type 2 diabetes in adults: management. London: NICE 2008. <https://www.nice.org.uk/guidance/cg66>.
6. The National Institute for Health and Care Excellence (NICE). Type 2 diabetes: The management of type 2 diabetes. London: NICE 2009. <https://www.nice.org.uk/guidance/cg87>.
7. National Institute for Health and Care Excellence. Dapagliflozin in combination therapy for treating type 2 diabetes: NICE technology appraisals [TA288]. 2013. <http://www.nice.org.uk/Guidance/ta288> if needed
8. Kenney JT Jr. Managing the evolving complexity of pharmacologic treatment: comparative effectiveness research, pharmacoeconomic

- data analyses, and other decision support tools. *Am J Manag Care* 2012;18(Suppl 10):S234–9.
9. McGovern A, Hinchliffe R, Munro N, *et al.* Basing approval of drugs for type 2 diabetes on real world outcomes. *BMJ* 2015;351:h5289.
 10. Esposito K, Ceriello A, Giugliano D. Does personalized diabetology overcome clinical uncertainty and therapeutic inertia in type 2 diabetes? *Endocrine* 2013;44:343–5.
 11. Naci H, Lehman R, Wouters OJ, *et al.* Rethinking the appraisal and approval of drugs for type 2 diabetes. *BMJ* 2015;351:h5260.
 12. White WB, Cannon CP, Heller SR, *et al.* Alogliptin after acute coronary syndrome in patients with type 2 diabetes. *N Engl J Med* 2013;369:1327–35.
 13. Zannad F, Cannon CP, Cushman WC, *et al.* Heart failure and mortality outcomes in patients with type 2 diabetes taking alogliptin versus placebo in EXAMINE: a multicentre, randomised, double-blind trial. *Lancet* 2015;385:2067–76.
 14. Scirica BM, Bhatt DL, Braunwald E, *et al.* Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. *N Engl J Med* 2013;369:1317–26.
 15. Green JB, Bethel MA, Armstrong PW, *et al.* Effect of Sitagliptin on cardiovascular outcomes in type 2 diabetes. *N Engl J Med* 2015;373:232–42.
 16. Zinman B, Wanner C, Lachin JM, *et al.* Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med* 2015;373:2117–28.
 17. Marso SP, Daniels GH, Brown-Frandsen K, *et al.* Liraglutide and cardiovascular outcomes in type 2 diabetes. *N Engl J Med* 2016;375:311–22.
 18. Iglay K, Cartier SE, Rosen VM, *et al.* Meta-analysis of studies examining medication adherence, persistence, and discontinuation of oral antihyperglycemic agents in type 2 diabetes. *Curr Med Res Opin* 2015;31:1283–96.
 19. Donnan PT, MacDonald TM, Morris AD. Adherence to prescribed oral hypoglycaemic medication in a population of patients with type 2 diabetes: a retrospective cohort study. *Diabet Med* 2002;19:279–84.
 20. Aikens JE, Piette JD. Longitudinal association between medication adherence and glycaemic control in type 2 diabetes. *Diabet Med* 2013;30:338–44.
 21. Kim N, Agostini JV, Justice AC. Refill adherence to oral hypoglycemic agents and glycemic control in veterans. *Ann Pharmacother* 2010;44:800.
 22. Wu CH, Wong MC, Wang HH, *et al.* The eight-item Morisky Medication Adherence Scale (MMAS-8) score was associated with glycaemic control in diabetes patients. *Hypertension* 2014;64(Wu C.H.; Li H.W.; Hui E.M.; Lam A.T.) Dept of Family Medicine, New Territories East Cluster, Hosp Authority, Hong Kong, Hong Kong.
 23. Rozenfeld Y, Hunt JS, Plauschinat C, *et al.* Oral antidiabetic medication adherence and glycaemic control in managed care. *Am J Manag Care* 2008;14:71–6.
 24. Lau DT, Nau DP. Oral antihyperglycemic medication nonadherence and subsequent hospitalization among individuals with type 2 diabetes. *Diabetes Care* 2004;27:2149–53.
 25. Zhu VJ, Marrero DG, Tu W, *et al.* Oral hypoglycemic agent adherence and hospitalization among patients with type 2 diabetes: a call for enhanced guidelines. *Pharmacoepidemiol Drug Saf* 2012;21(Zhu V.J.) Biomedical Informatics, Regenstrief Institute, Indianapolis, United States):29.
 26. Hong JS, Kang HC. Relationship between oral antihyperglycemic medication adherence and hospitalization, mortality, and healthcare costs in adult ambulatory care patients with type 2 diabetes in South Korea. *Med Care* 2011;49:378–85.
 27. Heaton PC, Tundia NL, Luder HR. US emergency departments visits resulting from poor medication adherence: 2005-07. *J Am Pharm Assoc* 2013;53:513–19.
 28. Currie CJ, Peyrot M, Morgan CLL, *et al.* The impact of treatment noncompliance on mortality in people with type 2 diabetes. *Diabetes Care* 2012;35:1279–84.
 29. Valensi P, Benroubi M, Borzi V, *et al.* The IMPROVE study—a multinational, observational study in type 2 diabetes: baseline characteristics from eight national cohorts. *Int J Clin Pract* 2008;62:1809–19.
 30. Rubino A, McQuay LJ, Gough SC, *et al.* Delayed initiation of subcutaneous insulin therapy after failure of oral glucose-lowering agents in patients with type 2 diabetes: a population-based analysis in the UK. *Diabet Med* 2007;24:1412–18.
 31. Calvert MJ, McManus RJ, Freemantle N. Management of type 2 diabetes with multiple oral hypoglycaemic agents or insulin in primary care: retrospective cohort study. *Br J Gen Pract* 2007;57:455–60.
 32. Khunti K, Wolden ML, Thorsted BL, *et al.* Clinical inertia in people with type 2 diabetes: a retrospective cohort study of more than 80,000 people. *Diabetes Care* 2013;36:3411–17.
 33. Peyrot M, Rubin RR, Lauritzen T, *et al.* Resistance to insulin therapy among patients and providers: results of the cross-national Diabetes Attitudes, Wishes, and Needs (DAWN) study. *Diabetes Care* 2005;28:2673–9.
 34. Yki-Järvinen H, Kauppinen-Mäkelin R, Tiikkainen M, *et al.* Insulin glargine or NPH combined with metformin in type 2 diabetes: the LANMET study. *Diabetologia* 2006;49:442–51.
 35. Hayward RA, Manning WG, Kaplan SH, *et al.* Starting insulin therapy in patients with type 2 diabetes: effectiveness, complications, and resource utilization. *JAMA* 1997;278:1663–9.
 36. Quality and Outcomes Framework. QOF 2015/16 results. NHS Digital 2016. <http://qof.digital.nhs.uk/>.
 37. McGovern A, Hinton W, Van Vlymen J, *et al.* Real world evidence on the prescribing trends of glucagon-like peptide-1 agonists in UK primary care. *Diabet Med* 2016;33:165.
 38. de Lusignan S, Khunti K, Belsey J, *et al.* A method of identifying and correcting miscoding, misclassification and misdiagnosis in diabetes: a pilot and validation study of routinely collected data. *Diabet Med* 2010;27:203–9.
 39. de Lusignan S, Sadek N, Mulnier H, *et al.* Miscoding, misclassification and misdiagnosis of diabetes in primary care. *Diabet Med* 2012;29:181–9.
 40. Health & Social Care Information Centre. Quality and Outcomes Framework—prevalence, achievements and exceptions reports, England, 2014-15, 2015. <http://content.digital.nhs.uk/catalogue/PUB18887/qof-1415-annex3-DQandFAQ.pdf>
 41. McGovern AP, Fieldhouse H, Tippu Z, *et al.* Glucose test provenance recording in UK primary care: was that fasted or random? *Diabet Med* 2016;Published online first.
 42. World Health Organization. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation. WHO, 2006. http://www.who.int/diabetes/publications/diagnosis_diabetes2006/en/
 43. World Health Organization. Use of glycated haemoglobin (HbA1c) in diagnosis of diabetes mellitus: abbreviated report of a WHO consultation. WHO, 2011. http://www.who.int/diabetes/publications/diagnosis_diabetes2011/en/
 44. Office for National Statistics. 2011 Census: Office for National Statistics. 2011. <https://www.ons.gov.uk/census/2011census>
 45. Hippisley-Cox J, Coupland C. Development and validation of risk prediction equations to estimate future risk of blindness and lower limb amputation in patients with diabetes: cohort study. *BMJ* 2015;351:h5441.
 46. Tippu Z, Liyanage H, Correa A, *et al.* Ontologies to improve the identification of ethnicity in people with type 2 diabetes. *Diabet Med* 2016;33:197–80.
 47. Department for Communities and Local Government. The English Indices of Deprivation. 2015. <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015>, 2015
 48. McAlpine RR, Morris AD, Emslie-Smith A, *et al.* The annual incidence of diabetic complications in a population of patients with type 1 and type 2 diabetes. *Diabet Med* 2005;22:348–52.
 49. McBean AM, Li S, Gilbertson DT, *et al.* Differences in diabetes prevalence, incidence, and mortality among the elderly of four racial/ethnic groups: whites, blacks, Hispanics, and Asians. *Diabetes Care* 2004;27:2317–24.
 50. Oldroyd J, Banerjee M, Heald A, *et al.* Diabetes and ethnic minorities. *Postgrad Med J* 2005;81:486–90.
 51. Correa A, Hinton W, McGovern A, *et al.* Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. *BMJ Open* 2016;6:e011092.
 52. McGovern AP, Hine J, de Lusignan S. Infection risk in elderly people with reduced glycaemic control. *Lancet Diabetes Endocrinol* 2016;4:303–4.
 53. McGovern AP, Rusholme B, Jones S, *et al.* Association of chronic kidney disease (CKD) and failure to monitor renal function with adverse outcomes in people with diabetes: a primary care cohort study. *BMC Nephrol* 2013;14:198.