

# Identifying undiagnosed diabetes: cross-sectional survey of 3.6 million patients' electronic records

Tim A Holt, David Stables, Julia Hippisley-Cox, Shaun O'Hanlon and Azeem Majeed

## ABSTRACT

### Background

Around 1% of the UK population has diabetes that is either undiagnosed or unrecorded on practice disease registers.

### Aim

To estimate the number of people in UK primary care databases with biochemical evidence of undiagnosed diabetes. To develop simple practice-based search techniques to support early recognition of diabetes.

### Design of study

Cross-sectional survey of 3 630 296 electronic records.

### Setting

Four hundred and eighty UK practices contributing to the QRESEARCH database.

### Method

Electronic searches to identify people with no diabetes diagnosis in one of two categories (A and B), using the most recently recorded blood glucose measurement: random blood glucose level  $\geq 11.1$  mmol/l or fasting blood glucose level  $\geq 7.0$  mmol/l (A); either a random or a fasting blood glucose level  $\geq 7.0$  mmol/l (B). An additional outcome measure was the proportion of the population with at least one blood glucose measurement in the record.

### Results

The number (percentage) identified in category A was 3758 (0.10% of the total population); the number in category B was 32 785 (0.90%). Projected to a practice of 7000 patients, around eight patients have biochemical evidence of undiagnosed diabetes, and 68 have results suggesting the need for further follow-up. One-third of people aged over 40 years without diabetes have a blood glucose measurement in the past 2 years in their record.

### Conclusion

People with possible undiagnosed diabetes are readily identifiable in UK primary care databases through electronic searches using blood glucose data. People with borderline levels, who may benefit from interventions to reduce their risk of progression to diabetes, can also be identified using practice-based software.

### Keywords

blood; clinical informatics; diabetes; diagnosis; glucose; screening.

## INTRODUCTION

The global rise in diabetes prevalence poses major challenges for healthcare providers.<sup>1</sup> Symptoms of type 2 diabetes develop insidiously, so that a significant proportion of patients are undiagnosed, even in countries with a well-developed healthcare infrastructure. Up to 1% of the population of England is thought to have undiagnosed or unrecorded diabetes,<sup>2</sup> and is not receiving access to well-organised systematic care and follow-up.<sup>3</sup> In addition, people with impaired glucose regulation are likely to be excluded from the structured follow-up offered to those with diabetes,<sup>4</sup> even though they benefit from interventions to reduce the risk of progression to diabetes,<sup>5-7</sup> and to control cardiovascular risk. Therefore, there is a good case for identifying individuals with evidence of undiagnosed diabetes or impaired glucose regulation.

General practice records might assist in the identification of individuals most likely to have undiagnosed diabetes, but this approach depends largely on the availability of relevant, coded data. In a recent study in the Netherlands, hypertension, cardiovascular disease, and lipid metabolism disorders were identified as the most useful risk factors available in the electronic records.<sup>8</sup> Family history is an important factor but was not coded

*TA Holt, MRCP, FRCGP, clinical lecturer, Health Sciences Research Institute, University of Warwick, Coventry.*

*D Stables, MB ChB, medical director; S O'Hanlon, MA, MRCP, MRCGP, clinical design director, EMIS, Fulford Grange, Leeds. J Hippisley-Cox, PhD, professor of clinical epidemiology and general practice, Centre for Population Studies, Division of Primary Care, University of Nottingham. A Majeed, PhD, professor of primary care, Department of Primary Care and Social Medicine, Imperial College, London.*

### Address for correspondence

Dr Tim A Holt, Health Sciences Research Institute, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK. E-mail: tim.holt@warwick.ac.uk

Submitted: 3 August 2007; Editor's response: 25 October 2007; final acceptance: 19 December 2007.

©British Journal of General Practice 2008; 58: 192-196.

DOI: 10.3399/bjgp08X277302

consistently and so could not be included. In a community-based study in New Zealand, randomly selected householders were invited for casual blood glucose testing and risk assessment, followed by oral glucose tolerance testing in selected cases.<sup>9</sup> The results suggested that glucose measurements themselves are superior to traditional risk factors alone as a means of identifying groups for formal diabetes screening, particularly in Europeans. A general practice-based targeting strategy based on this principle but utilising existing blood glucose data would depend on their availability in electronic records, and this has not previously been measured in current UK practice. The volume of such data is set to rise following the guidelines of the Joint British Societies, which recommend opportunistic blood glucose measurement in everyone over 40 years undergoing cardiovascular risk assessment.<sup>10</sup>

This study was undertaken to estimate the number of people in the UK with biochemical evidence of undiagnosed diabetes, and the number justifying retesting to clarify their diabetes status, identifiable by existing electronic blood glucose data in primary care. The study was designed to test the utility of electronic searches as a means of targeting this group, and to measure the availability of blood glucose data to support them. This approach has the advantage that the existence of raised blood glucose measurements in a patient's record provides justification for further testing that is independent of the argument for or against population screening for diabetes.

## METHOD

### Study design

A population-based, cross-sectional survey using version 11 of the UK QRESEARCH database was undertaken.<sup>11</sup> This database contains the anonymised electronic healthcare records of over 9 million patients ever registered with 499 general practices throughout the UK. The information recorded on the database includes patient demographics (year of birth, sex, socioeconomic data derived from the UK 2001 census), characteristics (height, weight, smoking status), symptoms (if coded electronically), clinical diagnoses, consultations, referrals, prescribed medication, and results of investigations. Biochemistry results (including electronically-coded blood glucose values) are available that are now transmitted directly into the patients' notes from the local laboratory. The database has been validated by comparing birth and death rates, consultation rates, and prevalence and mortality rates with other data sources including the General Household Survey<sup>12</sup> and the General Practice Research Database.<sup>13</sup> Practices were included in the analysis if they had complete data transmission until at least 1 June 2006.

## How this fits in

The availability of blood glucose data in UK general practice records offers a new opportunity for identifying individuals requiring testing for diabetes. Most practices have patients with biochemical evidence of undiagnosed diabetes, who are identifiable by simple computer searches. As part of practice-based software, these searches may allow such individuals to be identified and followed up during routine care.

Patients were included if they were registered with the practice on 1 June 2006.

### Diagnostic definitions

The current diagnostic criteria for diabetes of the World Health Organization (WHO) were used in this study.<sup>14</sup> A fasting plasma glucose level  $\geq 7.0$  mmol/l on two occasions is diagnostic of diabetes in a patient with no symptoms. In a clearly symptomatic patient, a single raised random plasma glucose level of  $\geq 11.1$  mmol/l can be used to make the diagnosis.

### Patient subsets and code groups

The following subsets of patients were identified according to information in their electronic health record (codes available from the authors):

- patients with a computer-recorded diagnosis of diabetes;
- patients with a computer diagnosis of impaired glucose tolerance or impaired fasting glycaemia;
- patients with at least one random glucose measurement or one fasting glucose measurement; and
- patients with computer-recorded evidence of a normal glucose tolerance test.

### Search strategies

Two alternative electronic search strategies were applied, examining the most recent blood glucose measurement available in the record. Both strategies initially identified patients who do not have a diagnosis of diabetes (or impaired glucose tolerance) and have not had a diagnosis of diabetes excluded by a normal glucose tolerance test, or as indicated by the use of an appropriate Read code, as detailed in the Results section.

#### Strategy A

Patients were included in the results if the most recent blood glucose measurement was a fasting level  $\geq 7.0$  mmol/l or a random level of  $\geq 11.1$  mmol/l.

#### Strategy B

Patients were included in the results if the level (fasting or random) of the most recent blood glucose

**Table 1. Percentages of people with diagnosed diabetes and those identified by strategies A and B in the study.**

Sex	Age band, years	Prevalence of diagnosed diabetes per 100	Percentage of study population identified using Strategy A	Percentage of study population identified using Strategy B
Female	<15	0.15	0.00	0.03
	15-24	0.40	0.01	0.17
	25-44	1.01	0.04	0.56
	45-64	4.05	0.11	0.99
	65-74	9.91	0.22	1.84
	≥75	10.39	0.34	3.57
Male	<15	0.15	0.00	0.03
	15-24	0.42	0.01	0.13
	25-44	1.23	0.04	0.36
	45-64	5.85	0.20	1.30
	65-74	13.53	0.34	2.44
	≥75	13.95	0.36	3.45
Total	All ages	3.54	0.10	0.90

**RESULTS**

**Study population and prevalence of diagnosed diabetes**

A total of 480 general practices met the inclusion criteria for the study. There were 3.63 million patients registered, of whom 50.43% were female, and 8.03% were aged 75 years or over.

A total of 128 421 patients were identified with a computer-recorded diagnosis of diabetes, giving an overall prevalence of 3.54%. The median practice prevalence of diabetes is 3.60%. Table 1 shows the age-sex-specific prevalence rates for a diagnosis of diabetes.

**Exclusions from the target population**

Of the 3.51 million people without a computer-recorded diagnosis of diabetes, 0.27% were excluded from the target population because of a diagnosis of impaired glucose tolerance or impaired fasting glycaemia, a code for ‘diabetes resolved’ (for example, after pregnancy), or a code for ‘diabetes excluded’. As shown in Figure 1, there were 3.49 million people in the target population for further analysis.

**Availability of blood glucose data**

Overall, 1.05 million people (30% of the target population) had at least one computer-recorded blood glucose value, with considerable variation between practices. This figure rises to two-thirds of those aged 65 years or older. Some of these measurements were taken over 10 years ago, but in the past 2 years approximately half of all patients aged over 65 years, one-third of those aged over 40 years, and one-fifth of the total population have undergone blood glucose estimation. The proportion of blood glucose measurements that were clearly fasting was 15.9%. In 22.2% they were reported using a random code, but in 62% of measurements it was not clear from the code whether the sample was fasting or random. For all cases that were not clearly reported as fasting, it was assumed that the sample was random, but some of these measurements may have in fact been taken on fasting samples.

**Primary outcomes**

Using strategy A, 3800 patients were identified from the target population whose last glucose measurement was suggestive of diabetes (that is, a random value of ≥11.1 mmol/l or a fasting value ≥7.0 mmol/l).

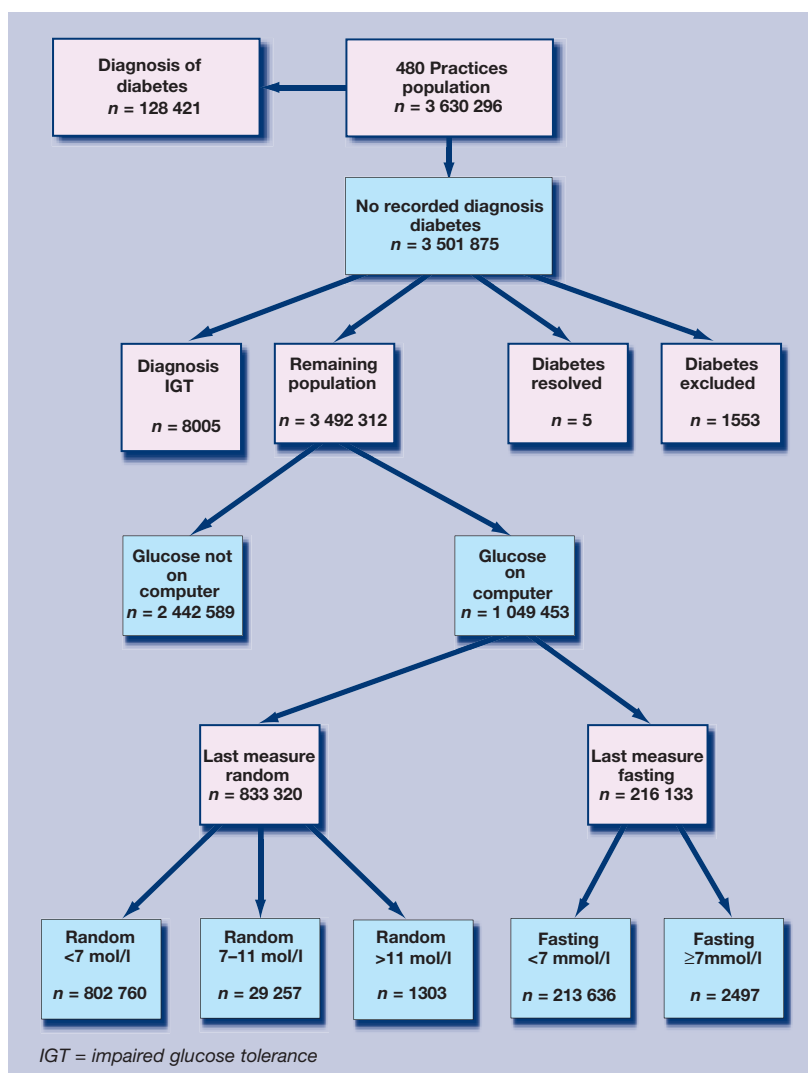
Of those identified, only 1.3% had diabetes excluded on the basis of a subsequent normal oral glucose tolerance test, leaving 3758 patients requiring further investigation or follow-up.

Using strategy B, 33 057 people were identified

**Figure 1. Number and percentage of the study population in each diagnostic category.**

measurement was ≥7.0 mmol/l.

The prevalence of each outcome was also determined by 5-year age-sex band.



where the most recent blood glucose value (random, fasting, or non-specifically coded) was  $\geq 7.0$  mmol/l. This figure fell to 32 785 once those with a normal glucose tolerance test had been excluded. This represents just under 1% of the entire study population and approximately 3.5% of all patients aged over 65 years.

Table 1 shows the age–sex breakdown of patients identified using strategies A and B.

### Time since last recording

Figure 2 shows the time periods since the abnormal glucose measurement according to strategies A and B. In some cases the last recorded value was relatively recent, so the individual may still be in the usual process of follow-up and no other action would be required. The distributions indicate that this was more likely to be the case for strategy A individuals, whose results may be more obviously significant. But in over one-third of the 'A' patients and half of the 'B' patients, the last recorded value was more than 1 year ago. In around one-fifth of the 'A' patients and nearly one-third of the 'B' patients, it was in excess of 2 years.

### Variation between practices

Table 2 shows the mean, median, and interquartile ranges of the numbers identified in the two categories in the 480 practices, demonstrating a wide range of values obtained. The majority of practices (440/480) had at least one strategy A patient identified, and all but one had at least one strategy B patient.

## DISCUSSION

### Summary of main findings

A significant proportion of the population has undergone blood glucose measurement, and the use of simple electronic searches allows us to identify people requiring clarification of their glucose tolerance. The numbers identified varied between practices. The majority (440/480) of practice databases include patients with evidence of undiagnosed diabetes based on the most recent blood glucose measurements (strategy A). Some of these may be known to have diabetes but are not on the practice diabetes register, and are therefore unlikely to receive well-organised, systematic care and follow-up. A proportion of the blood glucose levels reported by laboratories using random or non-specific codes may in fact have been taken from fasting samples.

All but one of the 480 practices in this study's sample included people whose most recent blood glucose level probably requires further follow-up according to current guidelines (strategy B).<sup>10</sup> This would involve a review of symptoms suggestive of

**Table 2. Variation in results among practices.**

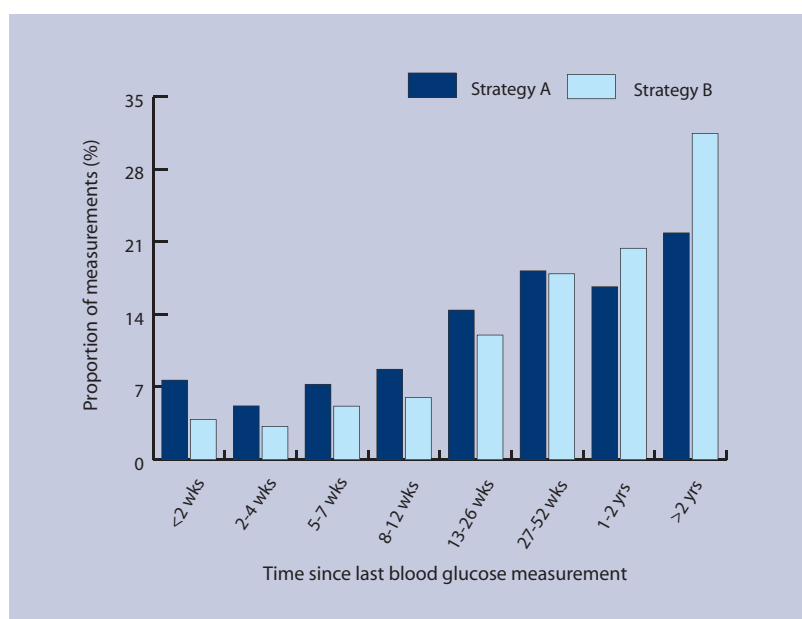
	Strategy A	Strategy B
Mean number of patients per practice	8	68
Median number of patients per practice	5	57
Interquartile range	2–10	30–92
Maximum	67	344
Number of practices (out of 480) with no patients identified	40	1

diabetes, clarification over whether the sample was fasting or random, and/or investigation with a fasting glucose or oral glucose tolerance test.

### Strengths and limitations of the study

This is a large population-based study utilising routinely collected data from general practices all over the UK. The age–sex structure of the study population is similar to national estimates, although the QRESEARCH population is slightly older. The prevalence of diagnosed diabetes in this study (3.54%) is very similar to the national prevalence for all practices in England (QMAS prevalence for 2006 3.60%),<sup>15</sup> suggesting QRESEARCH is likely to be representative of practices for case ascertainment, screening, diagnosis, and electronic recording of diabetes. QRESEARCH uses data exclusively from EMIS (Egton Medical Information Systems) practices. EMIS provides clinical software to approximately 59% of general practices in the UK. The national recommended Read codes used to transmit results to practices from hospital laboratories are the same for EMIS and non-EMIS practices.<sup>16</sup> Through the cross-sectional design of this study it was not possible to determine the

**Figure 2. Proportion of blood glucose measurements identified by strategies A and B according to time interval since the measurement.**



proportion of patients identified who were subsequently diagnosed with diabetes or impaired glucose regulation.

### Comparison with existing literature

This is the first nationwide survey to investigate the potential of practice-based searches for the detection of undiagnosed diabetes. The Diabetes Audit and Research in Tayside Scotland (DARTS) study identified some individuals with isolated hyperglycaemia in healthcare databases who were not on diabetes registers.<sup>17</sup> This study involved just eight general practices, and took place in a small region of Scotland in 1996, when the prevalence of diabetes was significantly lower than today. It predated the introduction of the Quality and Outcomes Framework in 2004 (which promotes the maintenance of electronic diabetes registers),<sup>3</sup> the widespread establishment of laboratory links for electronic data transmission, and the JBS2 report,<sup>10</sup> which established a policy for the follow-up of borderline blood glucose levels and actively encourages the testing of blood glucose in the over-40-year population, through cardiovascular risk assessments.

### Implications for future research and clinical practice

The proportions in these categories in which diabetes or impaired glucose regulation is confirmed on further testing will be investigated by the present authors, using a sample from current primary care, to clarify the usefulness of the software in the identification of undiagnosed diabetes. Further research is also planned to investigate the obstacles to follow-up of borderline blood glucose levels. These might include lack of clarity among practitioners over diagnostic thresholds, failure of laboratories to specify random or fasting results, problems with practice follow-up systems, or lack of patient concordance with invitations for retesting. It can be concluded that there is considerable scope in the UK for using electronic health records to identify people for recall and further assessment, thereby assisting in early detection of diabetes and impaired glucose regulation. Since the completion of this project, EMIS has introduced new software into all its practices for flagging up individuals in the 'A' and 'B' categories, to assist in their identification and follow-up during routine care.

### Ethical approval

QRESEARCH is approved by Trent MREC and all studies are reviewed by the QRESEARCH scientific committee (03/4/021)

### Competing interests

Julia Hippisley-Cox, David Stables and Shaun O'Hanlon are

unpaid directors of QRESEARCH, which is a collaborative venture between the University of Nottingham and EMIS (commercial supplier of computer systems to 59% of general practices in the UK). David Stables and Shaun O'Hanlon both work full-time for EMIS. Publication of this paper could lead to an increased awareness of the scope of the database for research. QRESEARCH undertakes research commissioned by government organisations including the Health and Social Care Information Centre, National Audit Office, Disability Rights Commission, Health Protection Agency, and the Department of Health. There are no other competing interests

### Acknowledgements

We thank EMIS practices who contribute data to QRESEARCH, and EMIS for expertise in creating and maintaining QRESEARCH.

### Discuss this article

Contribute and read comments about this article on the Discussion Forum: <http://www.rcgp.org.uk/bgjp-discuss>.

### REFERENCES

- Zimmet P, Alberti KG, J Shaw. Global and societal implications of the diabetes epidemic. *Nature* 2001; **414**(6865): 782–787.
- NHS Health and Social Care Information Centre. *National Diabetes Audit 2003/4*. 2005-IC-0025. London and Leeds: NHS Information Centre, 2007. <http://www.ic.nhs.uk/our-services/improving-patient-care/the-national-clinical-audit-support-programme-ncas/audit-reports/diabetes> (accessed 7 Feb 2008).
- British Medical Association. Quality and Outcomes Framework. Summary of indicators — clinical domain. <http://www.bma.org.uk/ap.nsf/Content/qof06--sumclinical#DiabetesMellitus> (accessed 3 Jan 2008).
- Unwin N, Shaw J, Zimmet P, Alberti KG. Impaired glucose tolerance and impaired fasting glycaemia: the current status on definition and intervention. *Diabet Med* 2002; **19**(9): 708–723.
- Lindström J, Ilanne-Parikka P, Peltonen M, on behalf of the Finnish Diabetes Prevention Study Group. Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. *Lancet* 2006; **368**(9548): 1673–1679.
- Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 2002; **346**(6): 393–403.
- Ratner RE. The Diabetes Prevention Program Research. An update on the Diabetes Prevention Program. *Endocr Pract* 2006; **12** Suppl 1: 20–24.
- Klein Woolthuis EP, de Grauw WJ, van Gerwen WH, et al. Identifying people at risk for undiagnosed type 2 diabetes using the GP's electronic medical record. *Fam Pract* 2007; **24**(3): 230–236.
- Simmons D, Thompson CF, Engelgau MM. Controlling the diabetes epidemic: how should we screen for undiagnosed diabetes and dysglycaemia? *Diabet Med* 2005; **22**(2): 207–212.
- Joint British Societies (JBS2). Joint British Societies' guidelines on prevention of cardiovascular disease in clinical practice. *Heart* 2005; **91**(suppl 5): v1–52.
- Hippisley-Cox J, Stables D, Pringle M. QRESEARCH: a new general practice database for research. *J Informatics Prim Care* 2004; **12**(1): 49–50.
- Office for National Statistics. The General Household Survey. [http://www.statistics.gov.uk/ssd/surveys/general\\_household\\_survey.asp](http://www.statistics.gov.uk/ssd/surveys/general_household_survey.asp) (accessed 3 Jan 2008).
- The General Practice Research Database (GPRD). <http://www.gprd.com/home/> (accessed 3 Jan 2008).
- World Health Organization. *Definition, diagnosis and classification of diabetes mellitus and its complications: report of a WHO consultation*. WHO/NCD/NCS 99.2. Geneva: World Health Organization, 1999.
- The National Health Service Information Centre. Online GP practice results database. <http://www.qof.ic.nhs.uk/> (accessed 3 Jan 2008).
- Connecting for Health. Pathology reports messaging — pathology bounded code list. [http://www.connectingforhealth.nhs.uk/systemsandservices/pathology/docs/copied/bounded\\_code\\_list](http://www.connectingforhealth.nhs.uk/systemsandservices/pathology/docs/copied/bounded_code_list) (accessed 3 Jan 2008).
- Morris AD, Boyle DI, MacAlpine R, et al. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. DARTS/MEMO Collaboration. *BMJ* 1997; **315**(7107): 524–528.