Using surrogate markers in primary electronic patient record systems to confirm or refute the diagnosis of diabetes

Ashkan Bagheri MEng
Medical Student

Ahmed Sadek BSc MBBS DPhil
Academic Foundation Year 2 Doctor

Tom Chan RGN RMN MSc PhD
Senior Lecturer
St George’s, University of London, London, UK

Kamlesh Khunti MBChB MD FRCGP
Professor of Primary Care, University of Leicester, UK

Simon de Lusignan MBBS MSc MD(Res) FRCGP FBCS CITP
Reader in General Practice and Biomedical Informatics, St George’s, University of London, London, UK

ABSTRACT

Background  UK primary care records are computerised and these records are used for both research and quality improvement. However, there is disparity in the prevalence of diabetes found in epidemiological studies compared with that reported through the UK’s national quality improvement scheme.

Objective  To investigate how non-diagnostic computer data could be used to identify, confirm or refute prevalent cases of people with diabetes.

Method  We carried out a literature review to identify the most accurate non-diagnostic markers. For each type of diabetes we focused on four broad areas; demographic details, biochemical markers, clinical features and therapeutic strategies. Sample markers were tested by calculating their positive predictive value (PPV) and sensitivity (Sn) and their ability to differentiate between types of diabetes.

Results  Biochemical markers were useful in identifying cases of diabetes but not in differentiating between types of diabetes as the same plasma glucose criterion is used to diagnose Type 1, Type 2, and ‘other’ types of diabetes; the lack of a ‘fasting’ qualifier blunts the use of this marker.

Auto-immune markers were the most accurate in identifying Type 1 diabetes but are not recorded frequently in primary care.

Clinical features of diabetes and therapeutic strategies were of some use – however, without time sequence data are difficult to interpret.

Raised plasma glucose (PG), and glycated haemoglobin (HbA1c), had useful PPV but low Sn. When PG was more than 7.0 and less than 11.1 mmol/l, PPV equaled 77.8% and Sn 48%; and when PG was 11.1, PPV equaled 92% and Sn 17%. For an HbA1c of more than 6.5%, PPV was 89% and Sn 73.3%, and for an HbA1c of more than 8, PPV was 92% and Sn 26%.

A person with a combination of aged under 30 years and prescribed insulin has an 84% PPV of Type 1 diabetes; if they also have a BMI <30 kg/m2 the PPV increases to 88%. A person age over 45 years and with a BMI >30 kg/m2 has a 5.3% PPV of Type 2 diabetes; if they are also hypertensive the PPV is 30%; Asian ethnicity increases PPV to 44%.

Conclusion  Non-diagnostic data has the potential to confirm or refute the diagnosis of diabetes and identify its type.

Keywords: algorithms, computer systems, diabetes mellitus, family practice, medical informatics, medical record systems, computerised, vocabulary, controlled
Introduction

Diabetes mellitus is one of the greatest health challenges facing the UK, with an estimated prevalence of approximately 4%. However, according to data from the Quality and Outcomes Framework (QOF; financially incentivised quality targets), the prevalence of diabetes is 3.7% (NHS Information Centre).

Many different diagnostic labels have been used to classify diabetes: the terms insulin dependent diabetes mellitus (IDDM) and non-insulin dependent diabetes mellitus (NIDDM) were used by the World Health Organization (WHO) in 1985. This was updated in 1999, when the NIDDM and IDDM terms were dropped and a new classification introduced (Table 1).

The majority of UK general practices are computerised and key data coded as recommended by the Royal College of General Practitioners and the British Medical Association. This rich pool of data provides an opportunity to use non-diagnostic computer codes not only to identify unrecorded cases of diabetes but also to classify the type of diabetes according to the WHO 1999 classification.

This paper reports the feasibility of creating algorithms able to confirm or refute both the diagnosis of diabetes and the type of diabetes.

Method

We conducted a literature review using Medline to identify articles about the diagnosis of diabetes. The literature review focused on three main areas. We first looked at biochemical markers such as fasting plasma glucose (FPG) and oral glucose tolerance tests (OGTT) since these are the foundation of accurate diagnostic testing of diabetes. Clinical features and associated conditions were another area that identified various risk factors such as age, BMI, ethnicity etc. which are linked to specific types of diabetes. The therapeutic strategies currently employed in diabetes care were also investigated.

Putative algorithms were then generated using the following surrogate biochemical, clinical and therapeutic markers; age, insulin prescription, anti-obesity prescription, BMI, polyuria, plasma glucose level and plasma HbA1c level. Specificity (Sn) and Positive Predictive Value (PPV) analysis was performed to assess whether the algorithm, generated on the basis of surrogate markers, was robust in identifying diagnosed cases of diabetes from a known dataset.

<table>
<thead>
<tr>
<th>Table 1 WHO Classification of Diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
</tr>
<tr>
<td>Type 1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Type 2</td>
</tr>
<tr>
<td>Other specific types</td>
</tr>
<tr>
<td>Genetic</td>
</tr>
<tr>
<td>Drug or chemical induced</td>
</tr>
<tr>
<td>Unknown</td>
</tr>
<tr>
<td>Gestational</td>
</tr>
<tr>
<td>Not strictly diabetes</td>
</tr>
<tr>
<td>Impaired glucose metabolism</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Results

Type 1 diabetes mellitus (T1DM)

Biochemical markers
Plasma glucose level testing is the gold standard in diabetes diagnosis and is based on the WHO criteria:

- random venous plasma glucose = 11.1 mmol/l or
- fasting plasma glucose (no caloric intake for at least eight hours) = 7.0 mmol/l or
- venous plasma glucose two hours after ingestion of 75 g oral glucose load (OGTT) = 11.1 mmol/l.

T1DM is primarily due to auto-immune mediated β-cell destruction in the pancreas, often accompanied by ketoacidosis. There are many markers that can serve as evidence of this auto-immune process such as islet cell autoantibodies (ICAs), autoantibodies to insulin (IAAs) and autoantibodies to glutamic acid decarboxylase (GAD). One or more of these antibodies is present in 85% to 90% of people with T1DM presenting with fasting hyperglycaemia. Unfortunately there are no Read codes currently in use to record such data.

One biochemical marker that is particularly useful in the identification of T1DM is C-peptide, a by-product from the conversion of proinsulin to insulin. Ketoacidosis is a hallmark of uncontrolled diabetes, in particular T1DM where deficient insulin results in cells being starved of glucose and the body providing an alternative energy source via the metabolism of amino acids and fatty acids to produce ketones. Despite an increasing number of ketoacidosis cases now being reported in Type 2 diabetes (T2DM).

Clinical characteristics
Young age and absence of obesity are two clinical characteristics of T1DM and are used by clinicians to help differentiate from T2DM. In T1DM polyuria, polydipsia and weight loss generally arise over a short period of time compared with T2DM.

T1DM is associated with various other auto-immune diseases such as Crohn’s disease, coeliac disease, Grave’s disease, Hashimoto’s thyroiditis, Addison’s disease and pernicious anaemia due to auto-immune atrophic gastritis. T1DM is present in approximately 5% of first-degree relatives, transferred predominantly through the major histocompatibility complex (MHC). Five-byte Read codes are available to indicate the presence of a family history of diabetes but none are specific to T1DM or other types of diabetes.

Therapy
Insulin remains the main treatment for T1DM. Oral hypoglycaemic drugs are generally not indicated in T1DM with many now being solely used for the treatment of T2DM. However, the use of metformin treatment alongside insulin has increased recently as it has been shown to reduce the daily requirement of insulin needed to achieve adequate glycaemic control.

Type 2 diabetes mellitus (T2DM)

Biochemical markers
Of types of diabetes, T2DM is by far the most common in the UK with some models suggesting it accounts for 92% of all people with diabetes. Its underlying pathology is predominantly insulin resistance and diagnosis is based on plasma glucose using WHO criteria.

Clinical features
Age and BMI are two of the biggest factors when considering the likelihood of T2DM. The American Diabetes Association (ADA) has identified several risk factors associated with T2DM which could be used as potential markers when searching through clinical records:

- age over 45 years
- BMI more than 30 kg/m²
- family history of diabetes (i.e. parents or siblings with diabetes)
- race/ethnicity (e.g. Asian/black/Hispanic)
- previously identified impaired fasting glycaemia (IFG) or impaired glucose tolerance (IGT)
- history of gestational diabetes mellitus (GDM) or delivery of a baby weighing more than 4.5 kg
- hypertension (140/90 mmHg in adults)
- polycystic ovary syndrome (PCOS)
- Acanthosis nigricans (AN)
- history of vascular disease.

Therapy
Treatment of T2DM starts with lifestyle modifications using diet and exercise; next, with the use of oral hypoglycaemic drugs. First line treatment is often with metformin which can be used in combination with insulin secretagogues such as sulphonylureas. T2DM patients do not need insulin for survival; however, many use it to control their blood glucose levels, often alongside oral hypoglycaemic drugs.
Differentiating between Type 1 and Type 2 diabetes mellitus

By collating all the different characteristics of T1DM and T2DM, an algorithm for differentiating between them was developed:

(i) Determine type of diabetes based on age and therapy:
   - if age<30 AND insulin treated – T1DM probable
   - if age>30 AND oral hypoglycaemic treated – T2DM probable.

(ii) If no therapy is recorded, then patients below 30-years old should be classed as T1DM unless two or more of the following markers are present:
   - BMI>30 kg/m² at presentation
   - no episodes of ketoacidosis
   - black or Asian origin
   - presence of AN, PCOS.

(iii) Markers suggestive of T1DM diabetes:
   - presentation with polyuria, polydipsia, weight loss
   - these markers are more significant if occurring over shorter period of time
   - presence of other auto-immune diseases.

(iv) Markers suggestive of T2DM diabetes
   - presence of serum C-peptides
   - anti-obesity drugs.

Single gene type of diabetes

The main type of diabetes falling in this category is commonly known as maturity onset diabetes of the young (MODY syndrome). This condition is autosomal dominant and can be caused by mutations in any of six different genes expressed in the pancreas (MODY 1–6). Diagnosis of this type is based on the WHO criteria and presentation of symptoms is often mild. Treatment is with insulin and/or oral hypoglycaemic drugs but this depends on the exact underlying mutation. Owing to its relatively low prevalence, many cases of MODY are initially assumed to be more common forms of diabetes; T1DM if the patient is young and not overweight and T2DM if the patient is overweight. However, MODY can be screened for by the absence of auto-immune markers, persistent low requirement of insulin (e.g. less than 0.5 u/kg/day) and absence of T2DM risk factors (e.g. obesity, metabolic syndrome etc.).

Other types of diabetes

Diabetes mellitus can be a secondary consequence of another underlying pathological process. Diagnosis of these types is based on the WHO criteria but their clinical characteristics and treatment are very much dependent on the underlying pathology. These cases of diabetes are quite rare in clinical practice making their detection difficult. One example is diagnosis of diabetes before six months of age, as less than 1% of Type 1 diabetes is diagnosed before this age.

Impaired glucose metabolism

Biochemical markers

Impaired glucose metabolism consists of two pre-diabetic states, namely impaired fasting glycaemia (IFG) and impaired glucose tolerance (IGT). It should be noted that IFG and IGT are not clinical entities in their own right but rather risk factors for the future development of diabetes. They represent an intermediate stage in the diabetes disease process where plasma glucose and OGTT measurements are above normal yet not high enough to infer diagnosis of diabetes mellitus. IGT and IFG diagnosis is based on the World Health Organization criteria summarised in Table 2. These criteria were used by Shaw et al to create an algorithm to differentiate between IFG, IGT and diabetes as shown in Figure 1.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>WHO criteria for the diagnosis of IGT, IFG and diabetes mellitus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fasting plasma glucose (mmol/l)</td>
</tr>
<tr>
<td>Normal</td>
<td>4.0–6.0</td>
</tr>
<tr>
<td>IFG</td>
<td>6.1–6.9</td>
</tr>
<tr>
<td>IGT</td>
<td>&lt;6.1</td>
</tr>
<tr>
<td>Diabetes</td>
<td>=7.0</td>
</tr>
</tbody>
</table>

* Venous plasma glucose two hours after ingestion of 75 g oral glucose load (OGTT)
Clinical features
Impaired glucose tolerance often precedes T2DM, hence metabolic syndrome traits are strongly associated with it. Other T2DM clinical features are also relevant, notably age, obesity and Asian ethnic origin.

Therapy
Treatment of IGT is predominantly through lifestyle modification, including both dietary change and routine exercise. Therapy may include metformin or antiobesity drugs (though check local drug licensing arrangements first), both of which have been demonstrated to decrease the risk of Type 2 diabetes development.16

Gestational diabetes mellitus (GDM)
Gestational diabetes mellitus is defined as glucose intolerance that appears or is first recognised during pregnancy and disappears after parturition.17 Women with known diabetes who become pregnant do not have GDM but rather ‘diabetes and pregnancy’ and should be recorded accordingly during pregnancy. Studies have shown patients with GDM are at an increased risk of developing Type 2 diabetes in the future.18

Biochemical markers
Despite large amounts of research into GDM there is still no consensus on its prevalence, diagnosis and management. In the UK many NHS trusts use the Scottish Intercollegiate Guideline Network19 criteria for GDM diagnosis:
- fasting plasma glucose > 5.5 mmol/l or
- two-hour plasma glucose = 9.0 mmol/l.
Glycosuria is unreliable in diagnosing GDM since 73%20 of pregnant women with glycosuria are not diagnosed positive for GDM.

Clinical characteristics
Many clinical characteristics and risk factors are associated with GDM. However, it is important to assess the strengths of these associations, as the more relevant ones must have a higher weighting in any algorithm used to identify women with GDM. The strong and weak risk factors for GDM are listed below.

---

**Figure 1** Algorithm for the diagnosis of IFG, IGT and diabetes mellitus (adapted from Shaw et al, 2003)
The strong risk factors:
- age > 30 years
- family history of diabetes in a first degree relative
- obesity – BMI >30 kg/m² at booking
- Asian/black ethnic origin
- previous history of GDM.

The weak risk factors:
- previous macrosomia – >4.5 kg (or >90th centile for gestational age)
- previous unexplained stillbirth
- polyhydramnios
- large for gestational age (LGA) foetus in current pregnancy
- history of PCOS.

**Therapy**

Treatment of GDM is initially through a regime of diet and exercise. For most patients this is sufficient to control their blood glucose levels and reduce the risk of complications for both mother and baby. If diet and exercise are not successful then insulin therapy may be considered. The ADA has recommended guidance on the cut-off values at which insulin should be initiated:
- fasting plasma glucose > 5.8 mmol/l and
- two-hour plasma glucose > 7.2 mmol/l.

The use of oral hypoglycaemic drugs is not currently recommended as most of these agents can pass through the placenta and safety issues have not yet been fully resolved.

**Algorithm generation and testing**

A summary of the findings are shown in Table 3. There is potential for non-diagnostic codes to confirm or refute diagnoses. Exemplar tests have been carried out and are described below.

The use of plasma glucose levels alone as a non-diagnostic marker to identify diabetic patients was specific and had a high predictive value. The use of plasma glucose levels greater than 7.0 mmol/l (11.1 mmol/l as per National Institute for Health and Clinical Excellence (NICE) guidelines) to identify known diabetic cases had Sn of 48% (17%), specificity of 99.5% (99.9%) and PPV of 78% (89.9%). Similarly the use of HbA1c levels as surrogate markers was statistically robust at identifying cases of pre-diagnosed diabetics. Levels of HbA1c greater than 6.5 (and 8 as per WHO guidelines) had an Sn of 73.3% (26%), specificity of 99.6% (99%) and PPV of 89% (92%).

The use of a combination of these surrogate markers in series to generate algorithms for the identification of Type 1 and Type 2 diabetics also resulted in significant serial predictive values. The identification of Type 1 diabetics on the basis of an insulin prescription and an age under 30 had a positive predictive value of 84%, and the addition of a BMI of less than 30 kg/m² resulted in a PPV of 88%. Similarly for Type 2 diabetes use of age greater than 45, BMI greater than 30 kg/m² and the presence of hypertension gave a PPV of 30% in identifying previously diagnosed patients. The addition of the codes for black or Asian race increased the positive predictive value to 44%.

**Discussion**

T1DM and T2DM cannot be differentiated through blood glucose levels since they are diagnosed according to the same WHO criteria. Auto-immune markers provide evidence of the pancreatic β-cell destruction underlying T1DM. Unfortunately, these are rarely measured in general practice. We therefore used age, therapy, ethnicity, and obesity to differentiate between the two conditions.

Age and therapy are initially used to distinguish between Type 1 and Type 2 diabetes as they are the strongest differential indicators. Other less specific markers for distinguishing between T1DM and T2DM are used in the final steps of the algorithm. The success of using therapy to distinguish between types of diabetes was mixed. Initial testing of the algorithms suggest plasma glucose and HbA1c levels have high specificity and PPV in identifying diabetic patients.

**Implications of findings**

A refined algorithm might identify unrecorded cases of diabetes and help differentiate the type of diabetes. Such algorithms could be used by practices to explore the gap between their QOF and expected prevalence of diabetes.

**Limitations of the method**

Some codes recording blood glucose levels do not specify whether they relate to fasting, random or OGTT samples, making electronic patient record data hard to interpret.
<table>
<thead>
<tr>
<th>Type 1</th>
<th>Biochemistry</th>
<th>Clinical features</th>
<th>Therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blood glucose levels: FPG = 7.0 mmol/l OR two-hour PG = 11.1 mmol/l</td>
<td>Young age (&lt;30 years of age) Acute onset of symptoms Polyuria, polydipsia, weight loss</td>
<td>Insulin Metformin used occasionally</td>
</tr>
<tr>
<td></td>
<td>Autoimmune antibodies: ICA, IAA and anti-GAD positive C-peptide positive Ketoacidosis episodes may be reported Glycosuria</td>
<td>Family history of diabetes Presence of associated autoimmune diseases: Grave's disease, Hashimoto's thyroiditis, Addison's disease, Crohn's disease, coeliac disease</td>
<td></td>
</tr>
<tr>
<td>Type 2</td>
<td>Blood glucose levels: FPG = 7.0 mmol/l OR two-hour PG = 11.1 mmol/l</td>
<td>Older age (&gt;25 years of age) Obese (BMI &gt;30 kg/m²) High prevalence in black, Asian and Hispanic ethnic groups Family history of diabetes Associated conditions: PCOS, AN History of IFG, IGT, GDM Metabolic syndrome</td>
<td>Oral hypoglycaemic drugs: metformin insulin secretagogues thiazolidinediones Anti-obesity drugs: Orlistat Lifestyle changes: diet, exercise Insulin used occasionally</td>
</tr>
<tr>
<td></td>
<td>Autoimmune antibodies: ICA, IAA and anti-GAD negative C-peptide negative Ketoacidosis episodes usually absent Glycosuria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single gene</td>
<td>Blood glucose levels: FPG = 7.0 mmol/l OR two-hour PG = 11.1 mmol/l</td>
<td>Absence of obesity (BMI &lt;30 kg/m²) Absence of metabolic syndrome traits Non-black/Asian background</td>
<td>Persistent low requirement of insulin (&lt;0.5 u/kg/day)</td>
</tr>
<tr>
<td>Other</td>
<td>Blood glucose levels: FPG = 7.0 mmol/l OR two-hour PG = 11.1 mmol/l</td>
<td>Diagnosis before six months of age</td>
<td>Insulin Oral hypoglycaemic drugs</td>
</tr>
<tr>
<td>Impaired glucose metabolism</td>
<td>Blood glucose levels IGT: FPG &lt;6.1 mmol/l AND two-hour PG 7.8–11.1 mmol/l Blood glucose levels IFG: FPG 6.1–6.9 mmol/l AND two-hour PG &lt;7.8 mmol/l</td>
<td>Presence of IFG Older age (&gt;25 years of age) Obese (BMI &gt;30 kg/m²) High prevalence in black, Asian and Hispanic ethnic groups</td>
<td>Lifestyle changes: diet, exercise Metformin to prevent development to Type 2 diabetes</td>
</tr>
<tr>
<td>GDM</td>
<td>Blood glucose levels FPG ≥5.5 mmol/l OR two-hour PG ≥9.0 mmol/l Glycosuria</td>
<td>Older age (&gt;30 years of age) Obese (BMI &gt;30 kg/m²) High prevalence in black, Asian and Hispanic ethnic groups Family history of diabetes History of GDM, baby with macrosomia, stillbirth, polyhydramnios Large for gestational age (LGA) foetus in current pregnancy</td>
<td>Diet and exercise for most cases of GDM Insulin where FPG &gt;5.8 mmol/l AND two-hour PG &gt;7.2 mmol/l Oral hypoglycaemic drugs not recommended due to possible safety issues</td>
</tr>
</tbody>
</table>
Another limitation of the algorithms described is that they treat the various incorporated markers equally. Other studies investigating screening algorithms have prioritised different markers according to their predictive value and assigned each marker a clinical score. The algorithms ignored diagnostic codes, which were investigated separately.30

Finally, incomplete records make it impossible to use time sequence analysis (e.g. was metformin started before insulin?) to help distinguish T1DM from T2DM.

**Call for further research**

The algorithms and criteria presented in this report could be combined with scoring systems to prioritise markers, in addition to using a wider range of clinical data.

**Conclusion**

Harnessing the power of computerised patient data provides an opportunity to improve the reporting of diabetes in primary care. The ambiguity over how diabetes patients should be classified has made it difficult for general practitioners to apply specific diagnostic codes to certain cases and has contributed to the under-reporting of diabetes cases. The use of non-diagnostic computer codes has the potential to close the gap between reported cases of diabetes and the epidemiological prevalence of the disease.

**REFERENCES**

CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

Simon de Lusignan
Reader in General Practice and Biomedical Informatics
Division of Community Health Sciences
St George’s, University of London
London SW17 0RE
UK
Tel: +44 (0)20 8725 5661
Email: slusigna@sgul.ac.uk

Accepted July 2009