

ORIGINAL RESEARCH

Data-analytically derived flexible HbA1c thresholds for type 2 diabetes mellitus diagnostic

Andrew Stranieri*¹, Andrew Yatsko¹, Herbert F. Jelinek^{2,1}, Sitalakshmi Venkatraman^{3,1}

¹Centre for Informatics and Applied Optimisation, Federation University, Ballarat, VIC, Australia

²Centre for Research in Complex Systems and School of Community Health, Charles Sturt University, Albury, NSW, Australia

³Department of HE-Business (IT), NMIT & Melbourne Polytechnic, Preston, VIC, Australia

Received: October 12, 2015

Accepted: December 6, 2015

Online Published: December 16, 2015

DOI: 10.5430/air.v5n1p111

URL: <http://dx.doi.org/10.5430/air.v5n1p111>

ABSTRACT

Glycated haemoglobin (HbA1c) is now more commonly used as an alternative test to the fasting plasma glucose and oral glucose tolerance tests for the identification of Type 2 Diabetes Mellitus (T2DM) because it is easily obtained using the point-of-care technology and represents long-term blood sugar levels. According to WHO guidelines, HbA1c values of 6.5% or above are required for a diagnosis of T2DM. However outcomes of a large number of trials with HbA1c have been inconsistent across the clinical spectrum and further research is required to determine the efficacy of HbA1c testing in identification of T2DM. Medical records from a diabetes screening program in Australia illustrate that many patients could be classified as diabetics if other clinical indicators are included, even though the HbA1c result does not exceed 6.5%. This suggests that a cutoff for the general population of 6.5% may be too simple and miss individuals at risk or with already overt, undiagnosed diabetes. In this study, data mining algorithms have been applied to identify markers that can be used with HbA1c. The results indicate that T2DM is best classified by HbA1c at 6.2% - a cutoff level lower than the currently recommended one, which can be even less, having assumed the threshold flexibility, if additionally to HbA1c being high the rule is conditioned on oxidative stress or inflammation being present, atherogenicity or adiposity being high, or hypertension being diagnosed, *etc.*

Key Words: Decision tree, Incomplete data, Diabetes mellitus, Diagnostic

1. INTRODUCTION

Diabetes Mellitus (DM), by far the second type (T2DM), is one of the world's major chronic diseases anticipated to rise to 439 million people by 2030.^[1] This prevalence motivates the need for effective preclinical screening in order to improve public health. Typically, T2DM is diagnosed by blood plasma glucose criteria, either the fasting plasma glucose (FPG) or the 2-h plasma glucose (2h-PG or TPG) value after the 75-g oral glucose tolerance test (OGTT).^[2] In diagnosing DM clinicians follow guidelines setting limits

for TPG, FPG, or the glycated haemoglobin (HbA1c) levels. The limits for TPG and FPG are known as impaired glucose tolerance (IGT) and impaired fasting glucose (IFG), respectively.^[2] To be exact, IFG and IGT are value bands, either one described as the prediabetic range, with the values below described as normal and above as diabetic in the vicinity of the range.

Evidence emerging from clinical trials demonstrates the effectiveness of targeted population screening in identifying undiagnosed DM. In Europe, a randomized controlled trial

*Correspondence: Andrew Stranieri; Email: a.stranieri@federation.edu.au; Address: Centre for Informatics and Applied Optimisation, Federation University, Ballarat, VIC 3353 Australia.

associated a small, non-significant reduction in the incidence of cardiovascular events and death with the early screening and intensive multifactorial intervention for T2DM.^[3] An Australian study has evaluated an existing in the country protocol for identification of undiagnosed T2DM and recommended the HbA1c assay as a cost effective measure to reduce amount of FPG and TPG testing.^[4]

HbA1c has several advantages to FPG and TPG if the erythrocyte function is normal.^[5-7] HbA1c is a long-term indicator for plasma glucose (PG) levels as it reflects average glycaemia over the preceding 8-12 weeks^[8] and is therefore considered a good surrogate marker for T2DM. Further, HbA1c levels have been shown to be strongly correlated with micro- and macro-vascular complications of T2DM, specifically retinopathy^[9] and cardio-vascular disease.^[10] Also, HbA1c overcomes some of the practical problems associated with obtaining FPG and TPG results: neither fasting is required, nor needs the corresponding test be repeated as frequently as the other two. In addition, HbA1c is more stable, preserving better from the time of collection to the time of assay, and has less specific storage requirements, compared with glucose.

However, there is an incomplete correlation between HbA1c and averaged PG levels, as well as between FPG and TPG. These discrepancies is the biggest contributor to the statistics of undiagnosed DM in the US where approximately 25% of people have borderline results of the blood test for DM.^[11]

According to ADA - the American Diabetes Association^[2] it is important to consider both demographic and clinical features such as gender, age, race/ethnicity, sedentary lifestyle and body mass index (BMI), anaemia/haemoglobinopathies, as well as biomarkers known to be linked to DM in order to determine risk levels of the disease. In addition, certain medications such as atypical antipsychotics increase the risk of DM and associated complications.^[12] This may explain that varied results are obtained world-wide for HbA1c as a clinical marker in T2DM diagnostic.^[8]

Beyond protein glycation, DM progression is linked to markers of endothelial^[13,14] and autonomic nervous system (ANS) dysfunction,^[15,16] inflammatory^[17,18] and oxidative stress^[19,20] biomarkers. Recent literature reports that presence of 8-hydroxy-2'-deoxy-guanosine (8-OHdG) in blood or urine samples is indicative of oxidative stress and deoxyribonucleic acid (DNA) damage due to minor increases in the PG level.^[21,22] In particular, the results of such studies show a statistically significant increase of 8-OHdG levels in prediabetic, and even a greater increase in diabetic participants. Measures of total body fat also play a role in assessing individual T2DM risk levels, though the exact re-

lationship between the weight and health outcomes is not fully understood. A sustained body weight reduction is often accompanied by improved glycaemic control in obese subjects, with estimated 0.75% HbA1c absolute drop for every 10% loss in weight (and so BMI),^[23] which competes with what is achievable by medicinal means.^[24] In this connection, even though some T2DM medicines are more effective, weight loss promises a more permanent solution, while the medicinal treatment may be for the price of weight gain.^[24] Hence, the T2DM risk prediction could be improved if the estimates of body fat were included with existing risk scores. In practice, BMI is used to assess weight, while waist circumference (WC) is regarded as a good indicator of the total body fat. However, a single standard for these metrics cannot be adopted as different classification criteria apply in different population groups. So, the demographics like gender and ethnicity should be included in risk prediction models.^[1]

Threshold levels of HbA1c for the diagnosis of DM vary across jurisdictions but 6.5% (48 mmol/mol) is adopted in the US and Australia.^[2,7] However, it is plausible that, some patients may have HbA1c below the threshold of 6.5% but have levels of other markers that, when coupled with the near miss values for HbA1c, are strongly indicative of T2DM. The objective of the current study was to discover possible markers and respective thresholds that alongside HbA1c could predict T2DM. The approach deployed was drawn from data mining.

“Data mining” or “data analytics” refers to the process of discovering interesting patterns from large amounts of data.^[25] The field has emerged in recent decades from statistics, machine learning and mathematics and involves algorithms for data selection, cleaning, pre-processing, and ultimately for analysing and learning from, typically very large datasets. Clinical repositories that derive from electronic medical records represent an exceptionally rich source of data for mining that has been applied for the discovery of insights into healthcare efficiencies, effective treatments, accurate prognoses and associations that stimulate medical research.^[26] In particular, data mining methods have been successfully applied in systems providing logistic support for DM care and treatment.^[27]

To date, data mining techniques have not been systematically applied for discovery of T2DM marker thresholds. However, data mining techniques are well suited for discovery of ideal cut points for HbA1c under different clinical circumstances in conjunction with other factors that may impact on the HbA1c value. The discovery of sophisticated rules relating HbA1c and emerging biomarker levels, in particular, presents as an interesting clinical model for data mining as

T2DM has crisp cutoff values for diagnosis but a potentially long preclinical period where complications of the disease can already start to manifest. It is evident that the majority of T2DM decision models so far were univariate, having predominantly relied on PG^[9,10] or BMI^[28,29] measures as sole risk factors. Very few models assume a multivariate risk function^[30] or are adaptive in nature.^[31]

The application of data mining techniques for the discovery of markers and threshold values in conjunction with HbA1c requires following elements:

- An appropriate clinical dataset – a dataset that contains data from the same patients on a large number of biomarkers, demographics and other factors;
- A novel data analytical approach – the dataset requires pre-processing to deal with outliers and missing values in addition to mining algorithms that identify combinations of markers and cutoff levels that predict T2DM;
- A format for depicting the combinations of markers and thresholds that is simple enough to be rapidly interpreted by clinicians and researchers.

2. MATERIALS AND METHODS

2.1 Clinical dataset

The dataset used in this study was derived from the Diabetes Screening Complications Research Initiative (DiScRi) conducted at a regional Australian university.^[32] The program implemented all necessary measures to safeguard the research ethics. Particularly, all participants provided informed, written consent. The DiScRi community screening concentrates on DM, Cardiovascular Disease (CVD) and Hypertension (HT) as a triad of diseases, which are specifically identified in the dataset. The screening clinic has been collecting data for longer than ten years now, tracking information such as demographics, socio-economic situation, and clinical test results. The clinical variables include PG, HbA1c; the cholesterol profile: low density lipoprotein (LDL), high density lipoprotein (HDL), total cholesterol (TC) and triglyceride (TG) concentrations in blood; inflammatory and oxidative stress markers from blood and urine; peripheral vascular function parameters; electrocardiogram (ECG) derived variables; blood pressure (BP) measurements, both systolic (SBP) and diastolic (DBP); anthropometrics, including BMI; as well as the medical history. Data on 300 attributes from 2,700 attendances of 900 patients have been collected in recent years. The dataset has been used in several data mining applications for identification of DM^[33,34] and related conditions.^[35,36] The temporal data underwent compression to instantiate patients instead of attendances by calculating longitudinal means or modes on the amount of

available data. Type 1 DM (T1DM) patients were excluded as this study is focused on T2DM.

The final dataset assembled for this study consisted of 99 attributes, of which 93% were incomplete, and 847 patient records, less than 1% complete, with 45% of values missing across all data. In addition, a number of features were generated upon completion using the existing set. The class structure for the main classification problem is taken as T2DM being the diagnostic class (Class 1) and contributing 26% of the data, with the rest assigned to the control class (Class 0). Having cases expert-labelled not only for T2DM but also for CVD and HT - the linked conditions - is crucial for subsequent data completion as approached in this work.

2.2 Novel data analytics

Missing values (MV) are often guessed/imputed for convenience of dealing with complete data when applying classification or other data mining methods, which is a topic we previously discussed in detail.^[33] On a different occasion^[34] we followed an approach on a par with the General Location Model^[37] to set MV in then existing a version of the DiScRi data, which is appropriate for mixed attribute type domains consisting of continuous as well as of categorical features. Values of categorical attributes naturally subdivide data into clusters, allowing to use a small number of well-established features of this type to set MV for all other attributes by using mean or mode for each combination of values of the selector attributes. If a continuous attribute, such as BMI, has to be nominated a selector for influence it exerts on decision making, its range can be subdivided into a number of intervals, adhering to some known thresholds. We recycle these ideas here and propose a method which is a hybridisation of the Hot Deck approach we previously evaluated and advanced.^[33] A difficulty is though presented by some unknown outcomes that cannot be easily predicted, which is expected of classification, because of MV in attributes the decision is supposed to be based on. The Hot Deck approach, which draws similar instances to enter MV by example, allows to overcome this difficulty as the class attribute or any other are treated the same. The class values can be set by a single strongest predictor, subject to its value known, as we did once.^[34] More attributes, even though less strong, can be deployed to enhance decision by ensuring greater similarity between compared instances, which is indeed how the Hot Deck method works. Although previously^[33] we proposed how to make use of this approach for all-categorical data domains, it is most suited for all-continuous data domains. The method advanced in this work is data type insensitive.

Features we use to guide the selection of data subsets to enter MV by subset mean or mode are all well-established in the

sense that they are long-known to be risk factors or indicators of T2DM and conditions linked to it and, incidentally, are well-supplied with values. As previously noted, the DiScRi data provides classification not only in respect of T2DM but also CVD and HT conditions. This is a favourable circumstance because it makes possible to ‘triangulate’ the location that means or modes could be extracted from. Normally, we would have to enjoy a single class attribute. So, the selector set includes: the T2DM, CVD and HT statuses, fasting glucose, age, waist-circumference-to-height ratio (WCHR), SBP lying, diabetic phenotype, joint cardio-vascular and joint cardio-neuropathic facts, graded heart function on results of ECG, combined causes of high BP, and gender. Indeed, the Australian T2DM screening protocol identifies following risk factors:^[4]

- People aged 55+ years;
- People aged 45+ years who have one or more of the following risk factors:
 - Obesity (BMI 30+ kg/m²),
 - First-degree relative with T2DM,
 - Hypertension (SBP 140+ or DBP 90+ mmHg, or taking antihypertensive medications);
- Aboriginal and Torres Strait Islanders aged 35+ years;
- People from high-risk non-English-speaking background groups aged 35+ years;
- People with IGT or IFG;
- People with clinical cardiovascular disease (myocardial infarction, angina, or stroke);
- Women with previous gestational diabetes;
- Women with polycystic ovary syndrome who are obese.

Thus, the selector set has a fair coverage of this list of long-known risk factors and therefore one should expect that at least this segment of data is available, which is indeed observed. The ethnicity aspect is not captured by the selectors as it is not readily accessible in the DiScRi data collection. Although, the rural population is probably less multicultural than urban in Australia. The gender specific conditions are either rare, or less consequential provided the T2DM status is known, and are not reflected in the database. While any additional selector information enhances specificity of the data entry, some omissions are not essential for the data completion in view of the assumed advance knowledge of all T2DM, CVD and HT statuses, adding the specificity at the highest level. This is beyond the domain knowledge and is a part of the formulation itself. Otherwise, the choice of selectors is unconventional in that BMI is replaced with WCHR which we find to be more reliable, and this reflects the contemporary leaning.^[28,29]

The order of selectors stipulates the sequence of MV entry. Firstly, MV including the missing statuses are set in CVD and HT problems, while all T2DM statuses are known. Having obtained the statuses makes MV entry in all other attributes more to the point. The problem which is staged with this purpose is the combined T2DM, CVD and HT classification problem where it is required to predict a combination of the statuses. However, the inversion of this classification problem is actually being solved. The order of selectors is specific to a problem and is different for the CVD, HT and the combined problem of T2DM, CVD and HT. The initial order of selectors can be guessed using the domain knowledge and refined after data completion using the Information Gain (IG) criterion - the higher, the prior - but it is possible to evaluate individual feature IG even from incomplete data.^[33]

The strength of the scheme for the automated data entry derives from the circumstance that the selectors, apart from being well-known, also possess high IG. Particularly, all the mentioned features but gender are on the joint top 20% list where a feature is included if it has a sufficiently high IG at least in one of the mentioned problems, that is, individual T2DM, CVD and HT problems and the combined one. Of course, the statuses are at the very top of IG lists in problems where they are set to be predicted. The advancement of the proposed technique is that the moment a selector has is dropped, due to its value set having been exhausted, and a broader selection is attempted is delayed as much as possible. Selectors with higher IG persist during the data entry, that is, their values stay fixed longer. This, in particular, allows for attracting data more suited for substituting MV from, earlier in the process, and increases confidence of the data entry by drawing on more reliable values previously entered. Once a selector is switched off, it is treated as an ordinary variable, and any MV in it are substituted then; however, the selector comes into play once more when a higher standing selector is exhausted. All selectors are eventually turned off when all their value combinations are encompassed. Any MV that are left out are then substituted from all data.

The whole analytics exercise is summed up in Figure 1.

The goal is to produce augmented HbA1c rules for T2DM diagnostic. After taking advantage of the dynamic clustering approach to substitute MV, the next challenge is to induce decision trees to get the general idea of would-be rules. Selected trees are then converted to models which are subsequently optimised for performance.

The decision trees^[38] are produced in the manner of the Decision Branch algorithm previously introduced by us,^[33] although this time we are interested in whole trees, not individual branches. Accordingly, any continuous attributes

are undergone discretisation, which was also covered. The discretisation is only a temporary device to get the idea of a particular tree structure. Various trees are produced by staged-out transformation of the feature-set from where features just next to the root in the current structure are withheld one-by-one. HbA1c is invariably at the root. To guarantee this all other PG related attributes are withheld from the out-set. The resulting trees are then analysed with a view of their structure simplification. Some little promising branches are pruned out; similar branches at a same node are aggregated into a single one by merging adjacent intervals on involved continuous variables. The outcome of this process is creation of a model out of each tree initially produced. The models are essentially data prototypes in the decision tree form, free from the stigma of discretisation. The models are then entered into an optimisation routine and their parameters are tuned for performance. Specifics of the involved calculations are covered where appropriate in the text.

more to the point, which we'd like to find out. Orientation of the threshold, more generally, a decision boundary that in the scheme divides the continuous range of the variable into two parts is also important. The standard rule reads "IF HbA1c > 6.5% THEN T2DM". In the scheme the field with the ball in it symbolises a 'Yes' decision, and the field without a ball - a 'No' decision. If the decision is 'Yes', then the patient being examined has T2DM. The horizontal arrow next to the variable name, in this case HbA1c, is pointed in direction of the value increase.

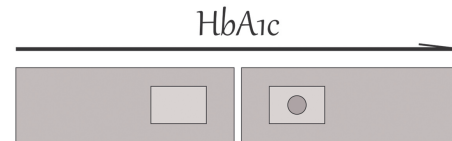


Figure 2. HbA1c-only decision rule by value over a threshold

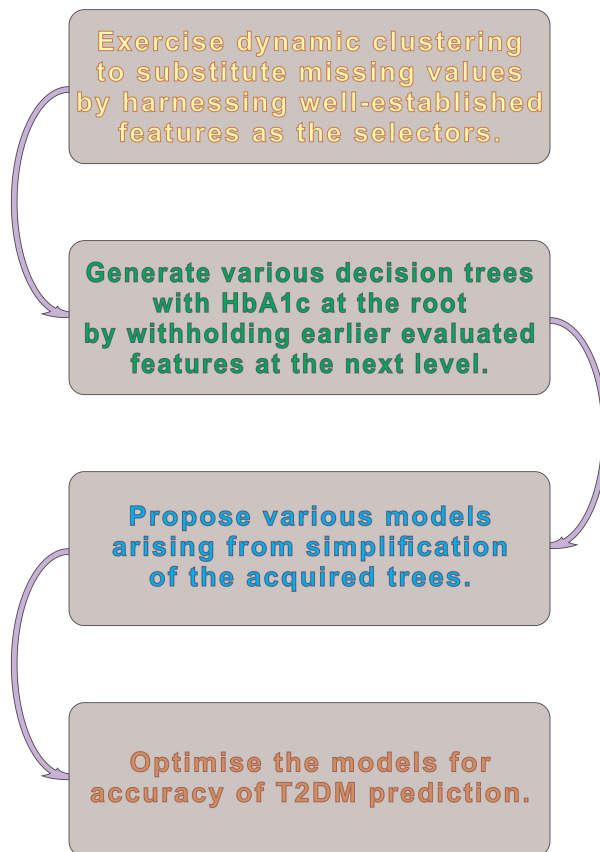


Figure 1. Stages of the analytical exercise

2.3 Representation format

Figure 2 schematises the decision taking based on a single predictor variable and a single threshold. The variable in the focus is HbA1c, and the standard threshold, or cutoff to diagnose T2DM is 6.5%, although a different value may be

A decision rule based on a single variable and a single cut-off may be too 'sharp'. Even though the standing US and Australian recommendations^[7] formally include 6.5% with the 'Yes' branch of the standard HbA1c rule, this exact value or a close one is too ambiguous - a small error can cause a dramatic decision change. Also, the precision of any measurement is always limited. The HbA1c result is usually known to one decimal place, so the gap between 6.4% and 6.6% is formally unaccounted for, if the threshold value of 6.5% is excised. To emphasise this, we always exclude the decision boundary itself from either of the adjacent decision regions. With this in mind, it seems to be more appropriate allowing for an uncertainty gap in the decision rule and then trying to elaborate the decision in the gap.

Figure 3 illustrates a concept decision tree that has been designed to represent more complex rules in a manner that is potentially clinically practical. The idea is to use a second variable with a single threshold in the uncertainty range of the first variable, here HbA1c. For the correction to work, the second variable should be good enough to cover for shortfalls of the first variable in the gap. However, this may be not that restricting because, to be more informative as a set, variables have to be if not independent from each other then different enough. So, assuming all variables have their strong sides when it comes to a decision in certain circumstances, use of a different secondary variable will simply readjust position of the indecision interval on the primary variable. The HbA1c decision uncertainty interval in Figure 3 is distinguished by the vertical arrow to express the necessity to expand into another dimension, meaning that this range of values on the primary variable should be additionally analysed using the secondary one. Otherwise, the same schematics as in Figure

2 applies. The orientation of the decision boundary for the second variable, as depicted, is ‘over’ and is the same as for HbA1c. We discuss another variant of orientation later in the text.

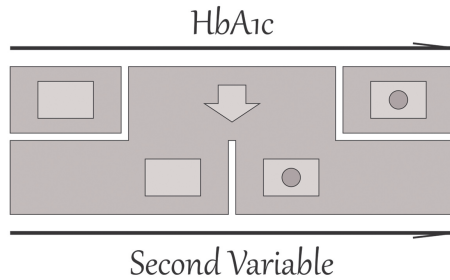


Figure 3. HbA1c decision enhancement with second variable over a threshold

The HbA1c cutoff is thus context sensitive and flexible, as proposed, meaning that different thresholds apply when HbA1c is taken in combination with another variable than if it was taken on its own and that, instead of a single value in the usual sense, such a threshold is represented by a band of values. We sometimes use the terms ‘threshold’ and ‘cutoff’ in that sense. Data mining algorithms are able to identify factors that can be used jointly with HbA1c and generate unique thresholds for a tailored diagnostic of T2DM. The results of actual mining of the DiScRi data are described in the next section.

3. RESULTS

Table 1 parameterises selected decision trees obtained by extension of the standard HbA1c rule as illustrated in Figure 3, listing them in the order of diminishing predictive accuracy. We use a decision tree classifier, as explained, to arrive at the idea of the trees. The trees are then optimised to yield a highest predictive “Yes/No” accuracy of T2DM on the DiScRi data. The actual quantity being maximised is balanced accuracy, that is, simple mean of Classes 1 and 0 individual accuracies, or sensitivity and specificity,^[25] respectively. A rule is run through all of the data and the statistics of successes are accumulated and combined at the end. This is repeated many times for different values of the rule parameters, until an optimal setting shown in the table is found.

If decision changes from ‘No’ to ‘Yes’ in the direction of increase of the auxiliary variable, which is how it appears in Figure 3, the corresponding rule orientation in Table 1 is stated as ‘over’, that is, if a particular reading of the auxiliary variable is over the threshold as in the table, then the decision is ‘Yes’, providing the corresponding value of HbA1c is within the interval of indecision in the table. Decision

is, otherwise, ‘No’ for lower values of HbA1c outside the interval, or if within the interval, for lower values of the extension variable. Decision is ‘Yes’ for higher values of HbA1c outside the interval of indecision. If orientation in the table is stated as ‘under’, this means that decision, based on the corresponding auxiliary variable, changes from ‘No’ to ‘Yes’ in the direction opposite to direction of the variable increase, as shown in Figure 4. Compared with Figure 3 the ball symbolising the ‘Yes’ decision is now found in the opposite field.

Table 1. Enhanced HbA1c rules for T2DM diagnostic using continuous extension variables

HbA1c indecision interval		Auxiliary variable				Balanced Accuracy (%)
From(%)	To(%)	Name	Orientation	Threshold	Unit	
4.92	6.23	8-OHdG	over	982	pg/ml	86.64
5.73	6.22	IL-6	over	64.0	pg/ml	85.63
6.05	6.60	IHR Sample Entropy	under	1.54	nonlinear pro-rata	85.01
5.70	6.22	TC	under	4.03	mmol/L	84.38
6.13	6.58	Protein	over	70.0	g/L	84.37
5.80	6.54	LDL	under	2.74	mmol/L	84.30
5.65	6.22	AIP	over	0.159	nonlinear pro-rata	84.07
6.10	6.58	IHR Approx. Entropy	under	1.27	nonlinear pro-rata	83.82
6.10	6.51	GSH / GSSG	under	7.36	pro-rata	83.56
5.80	6.22	Creatinine	over	87.7	µmol/L	83.28
6.10	6.56	BMI	over	24.4	kg/m ²	83.04
6.06	6.80	TG	over	0.747	mmol/L	83.03
6.00	6.61	WCHR	over	0.562	pro-rata	82.97
6.10	6.19	QRS (over 10 s)	under	114	ms	82.73
6.20	6.81	CRP	over	0	mg/L	82.46
-	-	HbA1c	over	6.20	per-cent	82.46

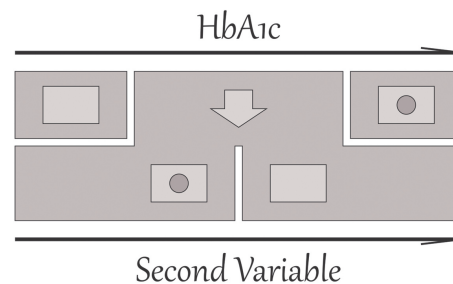


Figure 4. HbA1c decision enhancement with second variable under a threshold

Should an additional condition, expressed verbally but other than the one involving a continuous variable, be met this nonetheless can be modelled using a binary auxiliary variable where ‘0’ corresponds to the state when the condition is not met, and ‘1’ when it is. We could denote ‘0’ and ‘1’ as

‘no’ and ‘yes’, respectively. The outcome of decision making is that kind of variable. Alternatively, we could make use of a continuous auxiliary variable with the threshold of zero where ‘<0’ is equivalent to ‘no’ and ‘>0’ to ‘yes’. Then it is possible to apply the schematics in Figure 3 where the orientation is ‘over’. If it was ‘under’, that would mean that if the additional condition is met the decision is ‘No’, which is how it appears in Figure 4; however, such an occurrence is unconventional in diagnostics. The representation using a binary variable is not limiting, because a range of conditions can be associated with ‘1’ and all other conditions with ‘0’, or vice-versa. So, indeed, any categorical auxiliary variable, with its values distributed in this manner, can be used to elaborate the decision primarily based on HbA1c, the main variable. Having this understanding, the rules involving categorical auxiliary variables are presented in Table 2.

Table 2. Enhanced HbA1c rules for T2DM diagnostic using nominal extension variables

HbA1c indecision interval		Auxiliary variable		Balanced Accuracy (%)
From(%)	To(%)	Name	Values	
6.00	6.64	Feet Reflex	absent, reduced	84.64
5.60	6.22	Ewing Result	definite	84.06
6.10	6.60	HT-Status	yes	83.47
5.80	6.22	CVD-Status	yes	83.11

The rule-sets, or simply rules, schematised in Figures 3 and 4 and parameterised in Tables 1 and 2 should be treated with caution as there are various reasons why interval boundaries may differ from the stated. Of course, firstly, one should allow for variance in results due to choice of selectors and their intervals to guide the data completion. Secondly, the extension clause in a rule applies to a small and sparse subset of data, so any point in or out can bring about a notable change. Thirdly, towards the ‘Yes’ end, essentially due to the overall rarity of T2DM in general population, there are fewer data points to statistically rely on. Above all, interval boundaries for HbA1c and the variables extending the rule cannot precisely be determined because the data sample is finite. Particularly, the accuracy calculation involves counting successes in prediction of possible outcomes. These counts, and so the accuracy, change step-wise.

Extended HbA1c rules may or may not offer an improved accuracy of classification, generally. Either way, any change in accuracy should be measured relatively to the HbA1c indecision interval, which is rule dependent, for proper comparison. Otherwise, any increase may appear to be very modest. A surrogate criterion that can be watched instead when selecting a rule is presented by the indecision range itself, as any

increase in accuracy can be expected to be commensurate with the data it accounts for. Particularly, the middle section may disappear altogether if a rule that is being optimised is not going to be more accurate than the HbA1c-only rule. The QRS (which is an ECG waveform) linked rule at the end of Table 1 is given as an example of this kind, where the HbA1c indecision interval is exclusively narrow, and the rule otherwise is not much different from the HbA1c-only rule setting the benchmark for comparison, which appears on the last line.

Rules depicted in Figures 3 and 4 are fairly simple, it is sufficient to use the ‘over’ or ‘under’ construct to fully express any such rule. More complex rules would include a number of intervals subdividing the HbA1c range with possibly different descendent clauses attached to each interval, with each clause defined by a single variable but is not limited to a single threshold. The structures generated by the decision tree algorithm used to initiate the search inspire, though, simple rules consisting of a single interval exception on HbA1c with a single threshold on the extension variable. The orientation of boundary is often clear from the initial rule produced by the decision tree. However, for many rules in Tables 1 and 3 explained next the orientation is also supported by data-wide considerations as evident from Table 6, that is, how levels of biomarkers and other indicators are generally different between the control and diagnostic classes.

Table 3. T2DM diagnostic rules based on a single predictor

Decision Variable				Accuracy (%)		
Name	Cutoff	Unit	Orientation	Class 0	Class 1	Balanced
HbA1c (optimal)	6.20	per-cent	over	89.58	75.34	82.46
HbA1c (standard)	6.50	per-cent	over	93.75	63.68	78.71
8-OHdG	939	pg/mL	over	99.84	43.50	71.67
LDL	2.70	mmol/L	under	72.44	68.16	70.30
WCHR	0.582	pro-rata	over	62.98	71.30	67.14
Unisex Waist	101	cm	over	80.93	52.02	66.47
AIP	-0.0192	nonlinear pro-rata	over	60.74	71.75	66.24
TC	5.14	mmol/L	under	53.04	78.48	65.76
BMI-1	48.3	kg/m	over	63.14	63.23	63.18
Feet Reflex Score	0	point	over	65.06	60.09	62.58
BMI-2	30.1	kg/m ²	over	73.88	51.12	62.50

4. DISCUSSION

4.1 Contending risks

The rules in Tables 1 and 2 exploit the circumstance that HbA1c is the strongest predictor of T2DM comparing to other features in the tables. In fact, only FPG and PG in general are rivals of HbA1c, and this is unsurprising because all

three feature the glucose content in blood. Table 3 gives an idea of what is achievable when rules based on a single predictor are optimised for accuracy using variables that follow HbA1c on the IG list and also compares the optimal HbA1 rule to the standard one. However, the listing by diminishing balanced accuracy in Table 3 gives a close approximation of this order.

In the literature diagnostic rules with 90+% accuracy are described as 'exceptional', those with 80+% as 'great' and those with 70+% as 'acceptable'^[39] (although the platitudes designating the ranges may be arbitrary). Thus, having optimised the HbA1c rule lifts it out of the obscurity of mere acceptability, which is remarkable. The rule is great but is not exceptional - this corresponds to the role given to HbA1c in diagnosing T2DM, which is an additional check to verify results of FPG and TPG probing:^[8] "An HbA1c of 6.5% is recommended as the cut point for diagnosing diabetes. A value of less than 6.5% does not exclude diabetes diagnosed using glucose tests." Even though we find a better cut point for HbA1c, FPG is more sensitive. Only few other rules, using balanced accuracy for this analysis, are acceptable by the above standard but fall well behind the optimal, or even the standard HbA1c rule. We argue later in the text that LDL and TC cannot be used in usual sense as predictors of T2DM. The 8-OHdG rule is acceptable, but notice the imbalance between the sensitivity and specificity, which are individual accuracies for Classes 1 and 0, respectively. The rule is particularly useless in predicting Class 1 and tells of nature of 8-OHdG which elevation results from the action of reactive oxygen species initiating a chain of free radical conversions which are very unpredictable and impact many established pathways.^[19,20] However the consistency of the 8-OHdG rule may have also been eroded by use of antioxidants. The WCHR, Unisex Waist, and atherogenic index of plasma (AIP) rules are less than acceptable but can be used as a reference point to some extent. Again, the Unisex Waist rule is not balanced, but acceptable in respect of Class 0. We define Unisex Waist as WC for females and WC less by 9.7 cm for males based on the means; however, a different correction is applied in the standard oversize formula (WC of 94 cm or more in men and 80 cm or more in women places them into the increased T2DM risk category),^[12] but the standard is known to vary with jurisdiction, so there may be locality specifics. At the same time, a study of a large, diverse population of T2DM patients in the US reports only 7.7 cm difference in average WC for men and women,^[23] which suggests some other consideration involvement in the mentioned standard. The WCHR and AIP rules are only slightly imbalanced and then in favour of Class 1 where they reach over the acceptability mark. In fact, the results look

alike. Perhaps, this is due to the nature of triglycerides (TG) used in calculation of AIP defined as $\text{Log}_{10}(\text{TG}/\text{HDL})$. The BMI-1, Feet Reflex Score and BMI-2 rules are far from being acceptable, yet the first two are balanced, and the last one is acceptable in respect of Class 0. Here, BMI-2 is BMI in the usual sense, that is, the weight to squared height ratio, whereas BMI-1 is simply the weight (kg) to height (m) ratio.

It is important to recognise that accuracy of the single feature and threshold predictors in Table 3, particularly accuracy of the HbA1c rule can be potentially higher with respect to diagnosis of actual T2DM as opposed to that estimated using PG features. Till recently IFG and IGT were the primary criteria required to determine the T2DM status.

4.2 Intermittent hyperglycaemia

The cutoff for HbA1c indicating transition to T2DM, as found, is somewhat lower than the current recommended level, but this may be due to cohort characteristics. However, the World Health Organization (WHO) made concessions having analysed a substantial body of evidence to date as follows.^[8] "Levels of HbA1c just below 6.5% may indicate the presence of intermediate hyperglycaemia. The precise lower cutoff point for this has yet to be defined, although the ADA has suggested 5.7% - 6.4% as the high risk range. While recognizing the continuum of risk that may be captured by the HbA1c assay, the International Expert Committee recommended that persons with a HbA1c level between 6.0% and 6.5% were at particularly high risk and might be considered for diabetes prevention interventions." The optimal, or 'precise' point of T2DM onset at HbA1c 6.2% that we find for heterogeneous population is clearly within the intervals adopted by both ADA and WHO. Importantly, in combination with inflammatory, oxidative stress and clinical presentation markers a lower HbA1c cutoff associated with T2DM has its own right to apply, which we also report and that may to some extent explain the varied results obtained world-wide for HbA1c as a clinical marker for T2DM.

4.3 Diabetic peripheral neuropathy

To require that Feet Reflex Score was more than zero, as in Table 3, is the same as to verbally condition on Feet Reflex being absent or reduced, as in Table 2. This score and the equivalent expression of reflex weakness is derived from five tests for reflex presence in feet, which extends also to ankles and knees.^[32] One of the five results is a summary of the Semmes-Weinstein monofilament examination, which is by itself a sensitive gage of feet neuropathy, even though more involved.^[40] In any test the reflex is found to be either absent, reduced or present, which scores 1, 0.5 or 0, respectively. To obtain the score the points are added up. Unlike the

score, the reflex weakness is evaluated by taking maximum of the individual test scores and expressing the result in a word accordingly. The tests are intended for detection of Diabetic Peripheral Neuropathy (DPN) in patients, which is a well-known complication of T2DM attributed to chronic hyperglycaemia. DPN is defined as presence of peripheral nerve dysfunction in diabetics after exclusion of other causes. A relationship between glycaemic variability and DPN has been observed in T2DM patients even with well-controlled HbA1c.^[41,42] DPN is a serious condition which can cause ulceration of feet that, if cannot be treated, might require amputation of affected lower limb extremities.^[40] Testing for DPN is theoretically meaningful as the high BP is generally considered a co-factor of T2DM. It creates a favourable environment for glucose to attach to cell receptors. And it is well known that BP in feet is much higher than at the heart level.

4.4 Dyslipidaemia

Dyslipidaemia, characterised particularly by high TG, a component of AIP, accompanies T2DM, but is there a strong connection and how this is related to obesity? It is well known that T2DM increases risk of CVD; however, can the opposite be also true? One study reviews the connection between HbA1c and lipid profiles.^[43] The authors entertain the view of strong connection and mention a report where it was concluded that health education addressing behavioural measures to reduce cardiovascular risk coupled with regular reviews of progress was able to improve HbA1c without any specific, additional medical attention. An educational program of this sort would recommend lifestyle changes with emphasis on dieting and physically activity.^[44] A targeted study of the impact of lifestyle changes involving a number of clinical sites across the US, where one half of participants were required to fulfil dieting and modest physical activity recommendations, and another half was simply educated about the benefits of diet and exercise, has shown a dramatic improvement of participant condition in the intensive lifestyle intervention group after 12 months of the program.^[23] Particularly, notable changes were observed in body weight and HbA1c. All participants were T2DM patients independently medicated as prescribed by their physicians. Engagement in physical activity is generally known to improve prospects of T2DM patients.^[45] However, low carbohydrate diet (LCD) is now considered a mainstream approach to tackle T2DM, although the treatment may be very enduring. For example, a trial demonstrated a significant reduction in HbA1c, TG as well as body weight in the LCD cohort comparing to the control one, both IGT at the baseline, and a significant increase in HDL after 12 months of LCD, with LDL levels also decreased.^[46] The weight loss can be even more profound in the wake of very low calorie

dieting (VLCD) where fats are restricted more than carbohydrates comparing to LCD.^[47] Although the last study does not provide any of the aforementioned characteristics but the body weight, it finds that levels of hormone leptin in VLCD subjects were significantly less than in LCD ones, all obese at the baseline, after 9 months of therapy, which is important for this discussion. Adhering to VLCD for a sufficiently long time makes possible just by itself achieving a full recovery from T2DM.^[48]

Theoretically the connection between dyslipidaemia and hyperglycaemia is plausible. In obesity adipose tissue reaches limits of storing more TG, or fats. Particularly, the tissue emits leptin - a hormone which hypothalamus should interpret as a message of satiety. "Leptin is a hormone made and secreted by adipocytes that acts on the brain to regulate food intake and body weight. It can also act directly and indirectly on peripheral tissues, as almost all tissues express the leptin receptor," according to one interpretation.^[49] However, with T2DM the starving cells must be sending a conflicting message or the reception is only partial. For the same reason, adipocyte ability to take and convert glucose to TG is also limited. With or without T2DM, the issue of the hormone leptin does not seem feasible if biochemically the limit is not reached. Some drug designs actually exploit this circumstance and deal with T2DM if not by enabling the tissue to store more fat then by using up the existing capacity so long this is possible.^[24] Under the T2DM conditions TG absorbed from food do not land in adipose tissue and circulate in blood concealed in soluble chylomicron particles. Some of it is eventually stored in liver and muscles which are not specialised for this. The ectopic deposition of fats is a known feature of T2DM. However, liver has a function of converting fats to glucose which it stores and dispenses as necessary. And it has another function, which is to convert TG to cholesterol. However, this does not immediately happen; instead, TG are packed into very low density lipoprotein particles (VLDL) and dispatched into the blood stream. VLDL is transformed to LDL largely made of cholesterol as necessary. This shows the causal connection between TG and TC, but one has to be sure when examining assay results how the two are accounted for. On the other hand, the glucose metabolism in T2DM is largely replaced by TG oxidation for energy, creating a lower pH environment characteristic of the disease. Incidentally, issue of leptin becomes resolved when it sets off metabolism of fats in cells. Leptin enables energy release from fatty acids (FA) in TG via activation of AMPK, a kinase converting adenosine monophosphate (AMP) back to triphosphate (ATP) when cellular energy levels, particularly in muscles, are low^[49] as follows. "Upon activation, AMPK signals through its downstream substrates

to restore normal energy levels by stimulating processes that generate ATP (such as FA oxidation) and inhibiting those that use ATP (such as TG and protein synthesis).” The activation of AMPK by leptin can either be direct, as explained, or indirect, via signalling from the central nervous system.^[49] While TG is in demand, the acidic environment is harmful and unsustainable without harsh consequences. So, the concentration of TG does not build up, but it is elevated if the disease is progressing slowly. Although, it is known that uncontrolled T2DM may result in a dramatic weight loss within a very short time. This metamorphosis also supports the premise that in obesity the storing capacity of adipose tissue is reached. It is not clear what triggers this release of TG; however, it is credible that lower pH substantially changes the cellular biochemistry. Therefore, at least at early stages or when controlled, obesity, usually gaged with BMI-2, and T2DM are highly correlated. Storage of the life fuel in adipocytes is the mechanism by which T2DM in possibility is delayed. In obesity the limit is reached and the covert T2DM becomes overt.

4.5 Waist-to-height ratio

WCHR appears to have much higher a standing though on the list in Table 3 than BMI-2 and therefore is preferable. The fact that WC is an important indicator of T2DM and a better one than BMI is supported by a large body of evidence. A dedicated to WCHR meta-analysis of results contained in a large number of articles was conducted not long ago.^[29] It was concluded that WCHR (WHtR) is often better than WC and both are consistently better than BMI for a range of conditions, including T2DM (diabetes), CVD and HT. WCHR is unit of measure independent and one can gather that it is gender indiscriminate. The study, though, reports results separately for men and women, and there possibly exists a difference with respect to CVD. The gender differences otherwise are only slight, and the authors also find that WCHR is ethnicity and possibly age invariant. While this meta-analysis^[29] selected only articles where the receiver operating characteristic (ROC) approach was used, another recent meta-analysis^[28] focused on articles where cutoff values were proposed for WCHR and BMI as predictors of several cardiometabolic outcomes to measure relative performance of the metrics. Once again, in the large sample of studies the majority was in favour of WCHR. Our choice of WCHR over BMI as a selector variable in the algorithm for data completion is though explained by the IG advantage of the first. However, the WCHR attribute is much less supplied with values than the BMI measures where weight instead of WC is used. So, we linearly regress WCHR on BMI-1 and use the formula to populate MV in the WCHR attribute. This modified attribute is used as a selector; however, it is not

the actual WCHR. The filled values in the selector attribute impact only how the actual one is clustered, which of course has a lesser footprint than the clustering itself.

The proposed WCHR thresholds corresponding to BMI-2 ones are given in Table 4.

Table 4. Equivalent obesity thresholds for BMI-2 and WCHR

BMI (kg/m ²)	Threshold Meaning	WCHR (pro-rata)
20	Underweight / Normal	0.450
25	Normal / Overweight	0.525
30	Overweight / Obese	0.600

The thresholds were obtained by linearly regressing WCHR on BMI-2 and rounding up the results as in the table. The regression line is only slightly nonlinear, albeit the nonlinearity can be expected theoretically. The threshold for transition from normal to overweight state is particularly closely matching values found for various cardiometabolic conditions in the first meta-analysis:^[29] “Importantly, WHtR offers advantages of a simple boundary value which could be used for men and women, and maybe children, of all ethnic groups ... Analysis of suggested cutoff values from 34 analyses in 16 different papers showed that the mean of proposed boundary values for WHtR, weighted for study size, in men and women, respectively, was 0.52 and 0.53 for diabetes, 0.53 and 0.50 for CVD, 0.50 and 0.50 for HT outcomes ... The mean proposed boundary value (the first cutoff level indicating risk) for WHtR was 0.5.” Note that ROC analysis does not involve calculation of cutoffs in order to compare discriminative powers of different predictors, which is how the issue is approached in studies surveyed in the second meta-analysis.^[28] So, perhaps the thresholds estimated in first one^[29] are indeed on the safe side of respective conditions occurring, and are not optimal. The optimal T2DM onset WCHR level in Table 3 is higher than the normal-to-overweight state change mark in Table 4 but is within the range reported in the second meta-analysis.^[28] Particularly, the median value of 0.56 across 12 studies for “diabetes mellitus” in non-Asian populations is less than, but is comparable to the one in Table 3, and is virtually the same as for early stages of T2DM featured in the clause complementing corresponding HbA1c rule in Table 1. However, meta-analysis of 15 studies performed for Asian populations^[28] found a distinctly lower median value of 0.51 for the assumed T2DM cutoff, which gives WCHR if not ethnical then racial dimension.

4.6 HbA1c-only rule variation

The optimal HbA1c-only rule offers a much improved accuracy over the standard one, difficult to eclipse with a modified rule (see Tables 1, 2 and 3). This makes it an easy choice

if not for decision making, then for analysis. It is easier to further discuss the rule optimality if the dependence is visualised. Figure 5 presents results of balanced accuracy evaluation for different cutoff values of HbA1c as a predictor of T2DM. Unlike in all previous evaluation where the entire dataset is treated as a single sample of the conceptual data, we use '5x2' sampling technique to plot the mean accuracy and the standard deviation lines. This technique implies that the data is randomly but proportionally to class memberships is distributed into two equally sized folds, as much as possible, and the rule accuracy is evaluated separately on the resulting samples. The whole procedure is repeated five times. This is a known approach to bias reduction of the perceived result variance attributed to data indiscriminately sampled from an infinitely large, unrestricted pool.^[50] The accuracy is probed for each 0.1 fraction of HbA1c 1% between 5.0% and 7.0%. Notice inflection points between 5.9% and 6.0% and between 6.5% and 6.6%. It is peculiar that near these points there are 'bottlenecks' of increased rule confidence relating a notably reduced standard deviation of the sampled accuracy. This may be a result of a biased interpretation of the existing knowledge about etiology of T2DM. Particularly, the inflection point or a lower maximum past 6.5% is possibly a result of applying the standard HbA1c rule. We will have an additional clue about the other inflection point. The higher maximum at 6.2% as found separately using all data appears to be obscured by the 'veil' of least confidence. This is particularly interesting as this possibly explains the existing discordant opinion about a better cutoff to diagnose T2DM by HbA1c than the current 6.5% and also that additional factors are likely involved.

The presentation given in Figures 5, 6 and 7 that follow is not equivalent but is similar to a ROC chart. Perhaps it is even more revealing: although the sensitivity and specificity are not independently plotted, it is possible to observe the parameter that controls them. The parameter is HbA1c and the balanced accuracy represents a synergistic effect of sensitivity and specificity. Because either that sensitivity is high or specificity is low, or other way around, the effective range for balanced accuracy is between 50% and 100%, not between 0 and 100% as for any of the components.

4.7 Common complications

It is important to understand implications for decision making if the impact of CVD and HT on population is varied, particularly how steady is the optimal threshold. Figure 6 presents results of balanced accuracy evaluation for different cutoff values of HbA1c as a predictor of T2DM, depending on whether CVD and HT are both present or separately absent. The same technique of '5x2' validation is used, al-

though only mean accuracies are shown.

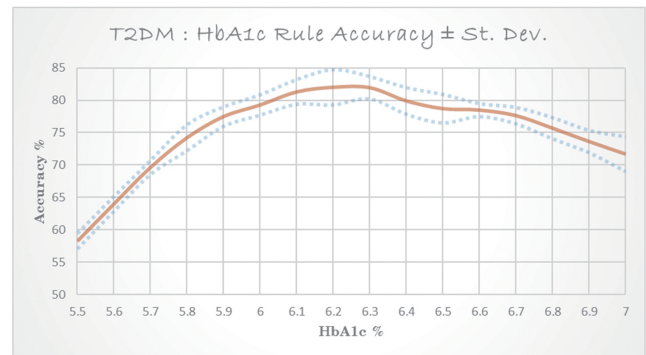


Figure 5. T2DM prediction accuracy and standard deviations for different HbA1c cutoff values

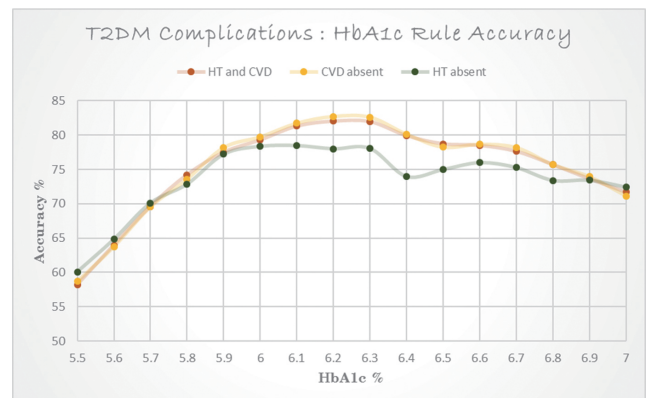


Figure 6. T2DM prediction accuracy for different HbA1c cutoff values in presence or absence of CVD and HT

The profile does not fundamentally change if CVD records are withdrawn. The "CVD absent" line practically overlays the "HT and CVD line" also shown in Figure 5; so much that precisely computed maxima for all of the data of the two profiles given in Table 5 are collocated. This can be expected as CVD is much less consequential for T2DM than HT using the IG criterion. Note, however, that the T2DM status and HbA1c are different variables even though highly correlated. The treatment by statins may have left only a vestige of CVD presence in affected subjects after the condition was initially identified. However, if HT-only content of the data is withdrawn - the "HT absent" line - the profile is much different, particularly the maximum in Table 5 is at a notably lower HbA1c level than for other lines. Surely, HT Status is a stronger predictor than CVD Status; however, incidence of HT in the population is also much higher than that of CVD, so the balance of the data might have been grossly affected. Together, this explains 'sagging' and unevenness of the "HT

absent” profile. The vast and deep drop in accuracy visible in Figure 6 and evident from Table 5 tells of important role of HT presence in determining T2DM outcomes, and is consistent with the theoretical base and experimental findings. Taking into account, though, the uncertainty due to levelling out of the curve about where the maximum might be, and the unlikeness of HT being absent in populations affected by T2DM, no correction of the threshold against the optimal HbA1c-only rule for heterogeneous population is required.

Table 5. T2DM by HbA1c diagnostic optimal cutoff variation in populations differently affected by CVD and HT

Population	HbA1c cutoff (%)	Balanced Accuracy (%)
CVD and HT both present	6.20	82.46
CVD absent, HT present	6.20	83.05
CVD present, HT absent	5.90	79.98

HT is a known circumstance abetting development of T2DM, so much that the precursor of the two: namely, HT in addition to hyperglycaemia and/or dyslipidaemia is recognised as metabolic syndrome (MS) – a distinct condition.^[17,43] However, HT is strongly associated with obesity^[51] and obesity with T2DM.^[28,29,43] In this regard, the little deflection of the “HT and CVD” line where the “HT absent” line joins it in Figure 6 - the inflection point in Figure 5 past 5.9% seems to be more on the part of the actual phenomenon than due to the limited knowledge of the data given by a single set. Certain prejudice may have existed in diagnosing T2DM in HT patients. This is the same, although a less persistent predicament as the one existing in respect of the standard HbA1c rule indiscriminately applied. Despite the substantial sagging of the accuracy profile when HT content is removed from the data, the additional influence HT Status exerts in determining T2DM outcomes is very modest comparing to other variables linked to HbA1c, as seen from Tables 1 and 2. Notice that HT Status interferes with HbA1c in the area where HbA1c is strong by itself, and also that the sagging seen in Figure 6 occurs in accordance with bounds set out in the modified rule. This makes sense if HT merely accelerates HbA1c increase in a prediabetic situation and is not directly involved with the onset of T2DM. However, the sagging is sooner on the part of large amount of data removed; particularly, it engulfs the indecision region. CVD Status in Table 2 has little to add to the HbA1c rule, although indeed it stretches the rule wide. HbA1c levels were found to be useful indicators of lipid levels^[43] and as such may be good for identification of T2DM patients with increased risk of CVD. However, our results show that LDL and TC levels are reduced in T2DM patients instead of being elevated, and we argue later in the

text that this is a result of medicinal intervention targeting LDL. The fact that CVD Status props HbA1c below the optimal cut point is an indirect evidence to this transformation. That is, we should assume that T2DM patients with CVD had initially high LDL and TC but then were treated for both CVD and T2DM and this lead to moderation of HbA1c.

4.8 T2DM disease cluster

T2DM is long being diagnosed by FPG and TPG. Just by itself this reveals the complex nature of the disease made of closely related conditions:^[52] FPG in excess of IFG captures the moment of onset of insulin resistance; TPG in excess of IGT captures the moment of onset of insulin deficiency. The logic requires that IGT was preceding IFG, although this connection is not simple. Therefore, a higher FPG has a connotation of long existing T2DM and in this regard HbA1c is probably more balanced a measure. The standard HbA1c cutoff is well aligned with the IFG range.

A short screening program run in Hong Kong (HK) with the purpose of evaluating applicability for subjects of Chinese ethnicity of the recommended by ADA HbA1c criterion for T2DM diagnostic^[53] gives a unique opportunity to compare results as the authors have tabulated sensitivity and specificity along with presenting the results in the ROC chart form. A comparison of their results with ours is given in Figure 7.

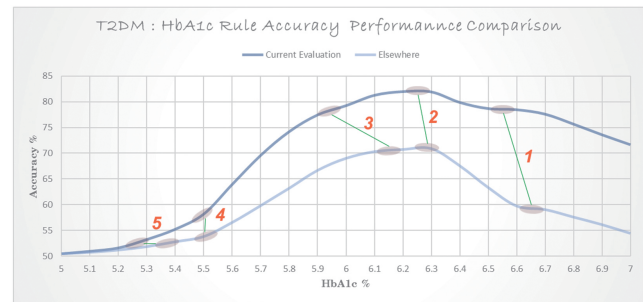


Figure 7. T2DM prediction accuracy for different HbA1c cutoff values in current evaluation and independently reported

The similarity of the two balanced accuracy against HbA1c profiles comes to the fore, as if ours was a magnified copy of theirs, albeit with some qualification. An important aspect of this for us is that the ‘trademark’ features of the accuracy profile, enumerated in the graph, are not a result of fluctuation due to data availability but are typical of the T2DM phenomenon and diagnostic practices. The mission of the study from HK primary healthcare was to prove that HbA1c is not very suitable for T2DM diagnostic among subjects of Chinese ethnicity, based on the observation that IGT is more

prevalent than IFG in this population. This has a bearing on the results which are probably biased. The biased representation is created by the selection of population for screening, even though the numbers are larger than ours. These were IFG persons, some diagnosed with T2DM in the process of screening. Were other glycaemic categories also included - this could have given more support to HbA1c values being analysed for their affinity to T2DM. The caution that ethnic differences can influence HbA1c levels is expressed now and then; however, the veritable truth about the change imparted onto the involved haemoglobin is that it mainly occurs through a non-enzymatic process.^[1]

Despite accuracy in the HK study could be better, it is possible to discern that IGT is probably the best tool to warn about impending T2DM in subjects of Chinese ethnicity - the "golden standard". Indeed, the trademark 'dimple' in the accuracy profile (point 3) is just before the maximum (point 2) on the HbA1c ruler, and the other 'dimple' associated with IFG (point 1) is much ahead of the maximum. The corresponding IGT and IFG 'dimples' in the profile evaluated by us for the Australian rural population are at much lower HbA1c values, both closer to the maximum. This looks as if the entire IGT - IFG range was shifted to better frame the maximum comparing to the other profile. Strangely enough, the maxima of the two profiles are very close - between 6.2% and 6.3%, at least visually, and considering the precision are collocated. We found the best cutoff point at 6.2% and the other study at 6.3%. Note that the curve tracing our result in Figure 7 is the same as the corresponding one in Figures 5 and 6. Recall though that the 6.2% cutoff was found for all of the data - twice larger a sample than used in validation, although invariable. On the graph the point of maximum may appear further up, but note that the summit is elongated and the drawing is approximate. The 6.3% cutoff for the other profile is also evaluated using all their data. However, the chosen point of maximum is just the best one among probed with the step of 0.1% and, considering the continuum, can be less than 6.3%. The measure used to decide the best cutoff point in the study compared with, is equivalent to the balanced accuracy used by us. Even though ethnic and racial specifics may impact sensitivity of the diagnostic, T2DM is a complex phenomenon where insulin resistance and insulin deficiency interplay, despite subpopulations may exist displaying clear signs of one or the other. At the same time, IGT appears to be universally more important than IFG for early detection of T2DM, as further explained.

We previously suspected the 'dimple' at higher values of HbA1c (point 1) to be a result of biased application of the standard rule with the cutoff at 6.5%. This may be still true; however, the other study result is not influenced by using

HbA1c as a diagnostic tool - it was itself evaluated. Both the HbA1c standard and the upper bound of the IFG range are indicators of impending hyperglycaemia and are presumably well aligned with each other. Therefore, we assume that this point is indeed associated with both aforementioned instruments and is situated at higher HbA1c values than the maximum at point 2. Both rules could have been strictly followed. We also suspected the other 'dimple' at lower values of HbA1c (point 3) to be a result of biased association with presence of HT. Now we say it is IGT related. Again, both may be true. The poor glucose clearance in a long run can cause a sustained increase of BP, which we touch on elsewhere in the text. Despite IGT is defined in terms of higher PG levels than in IFG, the postprandial PG increase is transient, and so IGT has to map to lower values of HbA1c than IFG. Rules solely relied on invite a biased perception.

HbA1c between 5.6% and 6.3% is the uncertainty range for T2DM diagnostic determined by the study conducted in HK,^[53] subject to the imposed limitations. The upper bound of this range is essentially the point of maximum of balanced accuracy, as explained. The lower bound, however, is defined basically as a point of 95% sensitivity. This is at odds with how the upper bound is set. At the same time, as seen from Figure 7, the balanced accuracy approaches the 50% mark asymptotically as HbA1c values get less. There is no specific point of minimum to set the lower bound of the range by. Although, there is a point preceding 5.6% where the rate of accuracy growth with HbA1c increases dramatically (point 4). There is probably nothing special about this point, but it is interesting that the growth onwards is steep and linear. This potentially can be used to heuristically define the lower bound of the uncertainty range for HbA1c as a predictor of T2DM. However, to be consistent with the criterion used for the upper bound, the principle of negligible error can be used with balanced accuracy instead of sensitivity. Due to the different range of possible values for balanced accuracy as opposed to sensitivity this translates to the critical level of 52.5% by the same standard as in the other study (point 5). The HbA1c values where the accuracy is just over 52.5% should then be assigned to the uncertainty range. From Figure 7 the corresponding HbA1c level is approximately 5.3% for our result and 5.4% for the result from the other study.

From the angle of this comparison the ethnical argument undermining universality of HbA1c as a tool for T2DM diagnostic does not seem sustainable. All what was required is to shift the entire uncertainty range by 0.1% which is relatively a small correction. It also comes to light that if the standard cutoff is redefined, HbA1c can become an instrument superior to both IGT and IFG, when paired with IGT, because the uncertainty range by HbA1c probably includes all of the IGT

range mapped to HbA1c, and a smaller cutoff value suits better the agenda of early T2DM detection. The proposed region that can be described as the impaired HbA1c range is between 5.3% and 6.2%. However, anti-T2DM medication could have caused a drift of HbA1c values – the concern also voiced by authors of the study compared with.^[53] The pre-diabetic HbA1c range as defined by ADA is between 5.7% (over 5.6) and 6.4% (under 6.5).^[2]

4.9 Diagnostic, monitoring and treatment

One review analyses pros and cons of using HbA1c for diagnosis of T2DM.^[6] The authors note that in ADA guidelines “A1C is indicated as a diagnostic tool alternative but not superior to blood glucose”. That is, health care professionals can choose diagnostic tools best suited to situations they have to deal with. Perhaps the main idea circulated in the article is that there is a difference between diagnostic and monitoring. Particularly the test for IGT is superior for diagnostic than the tests for IFG or impaired HbA1c, which are better for monitoring. The authors give a study example where it was found that the standard HbA1c rule missed more than 60% of cases of early T2DM identified by IGT.^[54] This may explain the jugged out portion of the classification accuracy vs HbA1c profile in Figure 5 in the vicinity of the proper maximum. That is, the standard HbA1c rule or the one based on IFG was often solely relied on to determine the T2DM status.

However, the profile twist may have another explanation. The DiScRi dataset has been accumulated over many years, more than a decade. Although T2DM, CVD and HT are slowly progressing, the screened subjects obviously took certain steps to reverse their ill conditions. A continuous use of certain medicines may keep levels of some biomarkers down. The initial rules generated by the decision tree algorithm are sometimes ambiguous. A biomarker level as in a healthy individual seems to be inconsistent with the person being diagnosed with T2DM. This certainly applies to PG levels that often can be suppressed by insulin injection. One has to choose a particular branch of the initial decision tree and verify its viability through optimisation. Particularly, the 8-OHdG rule is initially equivocal, prompting that causes of this biomarker being high in subjects affected by the disease might have been remedied to some extent. Ultimately, this applies to HbA1c itself. It is possible that the optimality of the threshold is to some extent affected by the data distribution portraying a treated or, sooner, experienced population, conscious of health risks they are exposed to, and taking necessary steps to prevent them. This may be not in the best interests of working out rules for diagnostic of a disease, in general, yet the impartiality requires that the population was taken as is. On the other hand, at the time that many

T2DM medicines strive to achieve HbA1c reduction,^[24] the statins addressing causes of CVD may somewhat increase HbA1c,^[55] so the overall effect is not obvious.

To help further discussion of rules in Tables 1 and 3, means and standard deviations of the involved variables are stated in Table 6 for the two classes of the data. The statistics were extracted from the original, incomplete DiScRi data. The Student’s two-tailed, unpaired *t*-Test was used in order to compare feature means for different cohorts of population. This is a standard statistical procedure; however, the strong assumption of normality of data distribution is made in respect of each variable. The result of the *t*-Test is *p*-value, the smaller the more distinct are the compared means. To be affirmative about the comparison a significance threshold by *p*-value is selected. It is usually set to .05. Table 6 shows the *t*-Test result, in particular.

Table 6. Original data statistics for T2DM diagnostic rule features

Feature	Control Class		Diagnostic Class		<i>t</i> -Test <i>p</i> -value
	Mean	St. Dev.	Mean	St. Dev.	
HbA1c	5.78	0.44	7.09	1.13	.000
LDL	3.15	0.80	2.41	0.94	.000
TC	5.20	0.89	4.50	1.05	.000
WCHR	0.564	0.078	0.621	0.089	.000
AIP	-0.071	0.278	0.102	0.273	.000
Unisex Waist	90.8	12.2	99.8	15.1	.000
BMI-1	46.6	8.5	51.1	9.6	.000
BMI-2	27.7	5.1	30.4	5.7	.000
Foot Reflex Score	0.433	0.746	0.901	1.067	.000
TG	1.36	0.74	1.74	0.84	.000
HDL (mmol/L)	1.43	0.37	1.27	0.36	.000
8-OHdG	256	232	1,190	1,230	.000
IL-6	28.2	22.6	47.2	32.8	.000
LDL / HDL	2.29	0.82	1.97	0.80	.000
Creatinine	74.1	18.6	79.9	22.4	.008
IHR Approx. Entropy	1.18	0.21	1.13	0.24	.023
GSH / GSSG	6.46	4.37	5.35	3.26	.057
Protein	70.8	4.4	71.4	4.7	.331
CRP	2.88	2.57	3.20	4.03	.480
IHR Sample Entropy	1.38	0.32	1.35	0.30	.555
QRS (over 10 s)	102	17	103	16	.854

4.10 Oxidative stress

The results for 8-OHdG and glutathione (GSH) are particularly interesting. The first is a biomarker of oxidative stress due to causes of T2DM and the second is an antioxidant

which function is to offset this stress. Oxidation of the reduced GSH gives rise to a modified species, glutathione disulphide (GSSG). GSSG is supposed to be converted back to GSH by the enzyme glutathione reductase. However, the oxidative stress is also implicated in inhibition of this enzyme.^[56] Thus, GSH levels are expected to drop in T2DM patients and the deficit of glutathione reductase exacerbates this. The accumulation of GSSG makes GSH-to-GSSG ratio even more sensitive a measure of oxidative stress than GSH itself. This is supported by results of previous observations.^[56] Indeed, the GSH-to-GSSG ratio levels are lower in the diagnostic class comparing to the control one, as seen from Table 6, although the distinction between levels is just over the significance threshold of .05 by p -value of the t -Test.

The conditions under which DiScRi data is collected cannot be described as pure whereby only untreated subjects are selected. Providing that, one could probably demonstrate a higher accuracy of T2DM prediction by antagonising the data classes from the start. However, the performance at the 82% accuracy mark of the optimal HbA1c-only rule is by no means poor and is much improved in combination with 8-OHdG, which is responsible for 5% accuracy hike, as seen from Table 1. The product of the two is fortunate because the 8-OHdG-only rule is grossly insensitive, as discussed, but much of this is offset by the sensitive HbA1c-only rule. However, much less known of 8-OHdG than of HbA1c, with missing rates as different as eight and three in ten, respectively. So, the rules in Tables 1 and 3 involving 8-OHdG can be offered only tentatively. Despite the actual numbers are not that small, there might be inconsistencies with respect to fair sampling. We outright dismissed variables with more than 90% of unknown values. Because the class mix is likely to be grossly underrepresented, these variables may appear to be very promising indicators of presence/absence of the disease; however, this cannot be verified. Nonetheless, the power of 8-OHdG as a predictor of T2DM is supported by previous reports.^[22] At least one third of the full content for any other variable in the text is known.

4.11 Inflammation

C-reactive protein (CRP) and interleukin-6 (IL-6) are known to be indicators of inflammation processes associated with T2DM^[17] and are routinely collected for diagnostic of the disease. In brief, the inflammation is caused by expansion of adipose tissue with vasculature development falling behind, which is also known to increase BP. The undersupply of nutrients and constriction of capillaries is causing, particularly, hypoxia in some adipocytes that signal this condition, which in turn is intended to stimulate proliferation of the capillary network. This involves secretion of IL-6. Other response to

the inflammation is lipolysis and FA release by adipocytes themselves. Under hypoglycaemic conditions the FA release is intended to restore energy levels, explaining endocrine down-regulation of insulin production by pancreatic β -cells. However, this is not normal under hyperglycaemic conditions, presenting the conundrum. Macrophages is another source of signalling. Macrophages infiltrate adipose tissue to collect debris from cells or whole cells that disintegrate or become dysfunctional, which is partly self-induced in response to inflammation. More macrophages are attracted by the positive feedback which also involves secretion of IL-6. Thus, elevated IL-6 is one of the signs of increasing adiposity and therefore is a marker of obesity. IL-6 is not only present in obesity but actually abets T2DM by indirectly contributing to insulin resistance,^[17] that is, before the glycation of cellular receptors is effective. Issue of this pro-inflammatory cytokine also stimulates hepatic production of CRP - an acute-phase response agent to inflammation, part of the innate immunity, which contributes to the autoimmune suppression of insulin producing β -cells in pancreas,^[17] and this is not only abetting T2DM but in a long run makes it less and less reversible. And without insulin the cellular glucose uptake is limited. Therefore, both IL-6 and CRP are considered acceptable early markers of T2DM. From the original data statistics in Table 6 the IL-6 level in the T2DM group is significantly higher comparing to the control, and the optimal rule utilizing IL-6 in addition to HbA1c does not contradict this. A study^[21] which has contributed to the DiScRi data, reports a significantly increased inflammation in their prediabetes compared to their control group, represented by elevated levels of both IL-6 and CRP. From Table 6 the increase in the CRP level is not as convincing as in the IL-6 one; however, the diagnostic group in the current research is different - T2DM. The threshold of zero for CRP in Table 1, that is, any inflammation is actually meaningless and it can be observed that the accuracy of the CRP rule is the same, perhaps only marginally better than of the HbA1c-only rule on the next line. Besides, it does not vary mapping of the HbA1c decision regions. Therefore, CRP use in the context of modified HbA1c rule is superfluous. The artefact value of this rule is that despite a substantial span of the indecision interval the modification does not improve on accuracy of the underlying HbA1c-only rule. Depletion of CRP levels may have been brought about by medications protecting function of pancreatic β -cells. In fact, many anti-T2DM drugs attenuate CRP levels, directly or indirectly.^[17] In this connection, while all rules in Table 1 offer mitigations of the HbA1c-only rule, whether the standard or the optimal, in view of the general discussion about their applicability, we would be more interested in modified rules extending

the HbA1c useful range, that is, helping early detection of T2DM. This means smaller ‘from’ values of the indecision interval in Table 1. Particularly, the IL-6 linked rule has this property and also offers a higher accuracy. This confirms utility of IL-6 as a pre-clinical marker of T2DM. There is another consideration to heed when applying rules involving whether IL-6 or CRP. While CRP has been acclaimed as even a better marker for CVD than LDL, the assay is recommended to be repeated if any unrelated sepsis or injury is present, as this can render a response much stronger than is warranted by the CVD decision range.^[18] This concern should equally apply to IL-6, another inflammation marker, even though in the context of T2DM.

4.12 Cardiac autonomic neuropathy

The Ewing Result, encountered in Table 2, is based on three heart rate (HR) and two BP measurements obtained under different conditions.^[57] The Ewing system is used for diagnostic of Cardiac Autonomic Neuropathy (CAN) in patients, which is a complication of DM. HR tests involve the lying-to-standing, the deep breathing, and the exhalation against a closed airway (the Valsalva manoeuvre) exercises. BP are evaluated in the hand grip, and the lying-to-standing situations. Of each quantity it is decided whether it should be considered normal, borderline or abnormal, and the results are combined to assess presence of CAN. It is regarded ‘absent’ - if all tests are normal or one is borderline, ‘early’ - if one HR is abnormal or two are borderline, ‘definite’ - if two or more HR are abnormal, ‘severe’ - if additionally to ‘definite’ one or two BP are abnormal or two are borderline; or else the result is ‘atypical’.^[36] It has been reported^[58] that HbA1c variability is positively associated with presence of CAN, identified using the Ewing battery of tests. The longitudinal aspect of the DiScRi data can be exploited to some extent to evaluate variability of HbA1c in connection to CAN. This was not attempted presently. CAN, regarded a complication of DM, can actually precede hyperglycaemia by many years.

Instantaneous (given a moment of time) HR (IHR) estimated from ECG recordings can be used for diagnostic of various health abnormalities in patients. Utility of Sample Entropy, a particular measure engaging IHR for identification of CAN was previously demonstrated.^[16] Sample Entropy “was developed to reduce the bias caused by the self matching in approximate entropy which is a mathematical approach to quantifying the complexity and regularity of a system,” as claimed.^[16] So, we shall assume the two measures are similar; however, only Sample Entropy is discussed in the concerned paper^[16] where an expression for it can be found. Both IHR entropy measures are featured in Table 1. The rules

have ‘under’ orientation corresponding to the original data statistics for these variables in Table 6, although a desired clearance between the two classes is not guaranteed. In the study behind the paper^[16] data for a small number of T2DM patients with and without CAN was analysed to evaluate various IHR measures. CAN was diagnosed using a single component of the Ewing system out of the consideration of data availability. The measure relied on was lying-to-standing HR, admissibility of which as a sole indicator of CAN was supported by previous independent analyses. Despite the study is small, the authors conclude, with references to similar results elsewhere, that their “. . . observations demonstrate that SampEn decreases in CAN+ patients relative to CAN- diabetic subjects.” They also engage the decision tree approach to estimate a Sample Entropy threshold to determine CAN. Our effort to have any missing entries filled, particularly in all variables the Ewing system draws on, pays with the ability of relying on all data, and we find the threshold to be at 1.54 which is very close to 1.56 reported in the mentioned study.^[16] Of course, the rule in Table 1 is just an enhanced HbA1c rule for T2DM diagnostic. It is only because of the similarity between our and the other party results for IHR Sample Entropy we can hypothesise that CAN is involved and that at least some of CAN is explained by T2DM. The utility of the IHR Sample Entropy is also acknowledged in another study concerned with CAN diagnostic in a mixed population of T2DM subjects and healthy controls.^[15] The analysis was performed on a sample of DiScRi data, and the first study may as well have contributed to project. However, there is a number of ways of calculating the quantity.^[35] We assume that the same standard to how it is done applied in both studies.^[15,16] IHR Approximate Entropy is not notionally different from IHR Sample Entropy, and so qualitatively same result is expected. It appears though that IHR Sample Entropy is much worse on its own in determining T2DM outcomes, as seen from Table 6, which correlates with other results obtained for CAN diagnostic in the DiScRi space.^[35] The rules involving these variables, or even more to the point Ewing Result previously discussed, basically claim that if it is not an overt T2DM then it is CAN, which positions CAN within the T2DM network of realisations.

4.13 ECG waveforms

Various guidelines on DM prevention and treatment advise,^[44] that “high blood sugar puts a strain on the heart,” which is a general principle, and intuitively this may lead to an elevated BP. Glucose is much bigger a molecule than water, it is well soluble in the latter, and so increased concentration of the former should also elevate viscosity of the blood plasma. Remarkably, the Ewing system for diagnostic of CAN exploits the HR and BP criteria despite the very

different nature of CAN. Yet from the biological perspective higher PG levels can lead to free radical damage of ANS in patients with inadequately controlled T2DM. Indeed, hyperglycaemia causes increased HR in observations.^[15] Is this effect, though, discernable in ECG unaided? The obvious choice of feature to examine first is the QRS complex, a waveform so denoted - the main impulse, distinct on a normal ECG. However, the original data statistics in Table 6 render such an optimism groundless, with the ten seconds average QRS duration levels being no different in the T2DM cohort comparing to the control. Some indirect evidence supporting this also exists. Particularly, glycaemic control has been shown to be a marker for ventricular tachycardia and sudden cardiac death, independent of the QRS interval obtained from electrocardiology investigations, especially in high-risk people with DM.^[59] Indeed, QRS has a very narrow foothold in the modified HbA1c rule in Table 1. Shorter QRS intervals can be expected due to HR increase and vice-versa. This lines up with the 'under' orientation of the rule in Table 1. Neither this contradicts other available evidence suggesting that a weakened heart will show longer than normal QRS intervals. Particularly, it was reported that the cutoff for abnormal QRS intervals is 100 ms and prolongation of QRS (120+ ms) occurs in 14% to 47% of heart failure patients.^[60] Despite this consistency, the modified HbA1c rule featuring QRS in Table 1 offers no real advantage over, and indeed is not much different from the HbA1c-only rule in the same table.

4.14 Glycation and dehydration

Although protein (the total serum content of) as a parameter is not closely watched in DM literature, its elevation is generally indicative of some chronic condition leading to dehydration - both is true in respect of T2DM. The main part of serum protein content - albumin is thought to be instrumental in cellular fluid retention. This is partially explained by the relatively small size of this protein, so that produced in large quantities it helps to maintain the plasma osmotic pressure, as reviewed.^[61] The homeostasis otherwise requires that the excessive glucose was flushed out, which is indeed what is observed in DM and for what it gets its name - siphoning out of the sweet. Hence, albumin is BP regulator in DM. What is also peculiar about specifically albumin in serum is that it serves as a transporter for free FA, probably due to their insolubility otherwise. This is important for this discourse as FA were identified as a factor of insulin deficiency.^[17] The advent of free FA in obesity and T2DM is due to inflammation, as discussed, and so albumin might be there to meet the increased FA concentration, in the first place. The second aspect of raised protein levels in T2DM patients is how it is measured. Any protein in serum is exposed to

glycosylation under DM conditions, just like haemoglobin. So, even at the same molar concentration in serum, protein is gaining more weight per mol in DM subjects comparing to unaffected ones.^[11] Indeed, the protein concentration in Table 1 is expressed as weight per volume and the rule is appropriate. However, the trend to protein content elevation in serum at T2DM in Table 6 is not strongly expressed, perhaps due to the additional variance brought about by individual differences and the unspecified protein content being measured. Note that HbA1c is expressed differently, as a relative concentration which varies less. Also, when examining serum protein levels one should be concerned with how the data was obtained. Different labs contributing to the same database may use different methods, which of course applies to any parameter being measured. However, it may appear being purposeful to apply generic and therefore less costly assay methods to capture non-specific protein content, which may be then ignorant of the effect of glycation. Nevertheless, there exist assays that directly measure glycated protein content:^[11] "The fructosamine assay is the most widely used alternative. It is a measure of glycated serum protein, which incorporates glycated albumin. Methods are also available for the quantification of glycated albumin specifically." Albumin is particularly attractive for its deemed high potential as a marker for DM:^[1,61] it has a shorter half-life (21 days) compared with haemoglobin, so it could be used as a shorter-term glycaemic control; it is not easily altered by an abnormal haemoglobin metabolism; the amount of non-enzymatic glycation of albumin is nine times more than in the case of HbA1c; glycation of albumin is ten times faster than of haemoglobin.

4.15 Cholesterol profiles and lipid lowering drugs

Various guidelines on T2DM treatment target reduction of LDL and TC.^[12,44] Here, no reduction is required because it is already in place, Table 6, to which rules in Tables 1 and 3 do not contradict. Once again, this is likely a result of the awareness. The rationale is as follows: CVD complicates DM (DM increases CVD risk^[39]); CVD can be treated, and so the negative impact of DM compensated by reducing LDL and TC. A medical practitioner would include a lipid lowering therapy into the treatment plan because CVD is life threatening and DM is not. On the other hand, the lipid lowering drugs (LLD) are known to be effective, especially at early stages of CVD. LLD may be not optional when complementing T2DM medications that increase CVD risk, particularly those with a side effect of weight gain.^[24] The authors of a study mentioned in connection to 8-OHdG^[22] use essentially the same data as in the current research, although selectively, to contrast their research purposes, and incidentally the data is much better known, with "missing data rate" no more than

20%, which we assume applies to 8-OHdG. Apart from confirming the 8-OHdG result they also note the impact of LLD: “Medication use may have contributed to some of the results noted in our study. However, there was no significant difference in the use of antihypertensive medication between the two groups. Statin use was higher in the IFG group and may have contributed to the borderline normal levels of LDL in both the groups but despite their reported antioxidant effects, the IFG group had significantly higher 8-OHdG. . . . 19.4% of the control group used statins compared with 37.5% of the IFG group.” Another study using a sample of the DiScRi data actually acknowledges the TC level reduction:^[62] “Both total cholesterol and HDL-C were significantly lower in the IFG group, while triglycerides and LDL-C showed no difference between the groups.” It was also noted that “. . . LDL-C does not play a major role in diabetes disease progression.” In this study the use of statins was claimed to be nearly twice as much, significantly more frequent in the IFG than in the control group. Yet another study gets qualitatively the same result as ours, although the LLD connection is played down:^[21] “Unexpectedly, the prediabetic subjects were characterized by a slightly improved blood lipid profile. A possible explanation for that phenomenon is that even though the differences in statin use between the two groups were only small, this could still have had an influence on the blood lipid outcomes of the prediabetes group, since statins lower LDL and increase HDL levels”. Again, at least some of the data acquired in that study is accessed in the current research. This authenticates our result, although not by independent reporting. However, general improvement of DM care outcomes, particularly drop in HbA1c, LDL and TC levels was concurrently and independently observed in NSW, Australia.^[63] The authors write: “Significant improvements in HbA1c, total Cholesterol, HDL and LDL (excluding Triglycerides) levels were also found between 2000 and 2004 which corresponded with implementation of the Diabetes SIP” (*i.e.* Screening Incentive Payments). However, the overall reduction trend pertaining to TC is evident there since the beginning of the data stream in 1995. In this instance urban NSW population was studied. For the reference, the DiScRi program is run in rural NSW since 2002. Australia-wide, having analysed the official statistics, the diabetes-related death rate has declined by 20% from 1997 to 2010, which is traceable to improvements in lipid levels, blood pressure and blood glucose levels.^[64]

4.16 Cholesterol level factors

It is on the public record that through this time statins have seen rise to the heights of best profit generators pharmaceutical companies worldwide ever new. While LLD marketing by the companies should not be discounted, there are

stronger factors the significant reduction of LDL and TC in populations can be linked to, apart from the LLD efficiency. Perhaps one of them is rise of Asian markets and the effort by Australian meat producers, embracing the philosophy of free trade, to meet the demand, even at expense of affordability for the domestic consumer. So, this channel of external cholesterol supply could have been economically constricted. Surprisingly, the role of meat consumption in development of CVD and T2DM is inadequately covered in the literature,^[65] although the dwindling evidence that exists suggests “. . . a possible negative impact of processed pork on glucose-insulin metabolism and a possible positive impact of pork intake on waist circumference and HDL-C.” We previously noted the success of LCD and VLCD treatments with emphasis on dietary proteins.^[46,47] Meat is a major source of proteins; however, it comes with cholesterol. As much as the cholesterol in HDL is desirable, it is undesirable in LDL from the perspective of CVD and T2DM treatment. Another socio-economic factor of scale is general aging of the population, seeing the pension age risen several times in the recent past. According to one report^[66] “. . . the Australian Age Pension eligibility age for women has increased from 60 in 1995 at the rate of six months every two years during the past decade, reaching the constant eligibility age of 65 for males in 2014. From 2017, the pension qualifying age for the two genders will start rising again from 65 years old by six months every two years until reaching 67 years of age in 2023.” This would ensure that working elders stay physically active and personally satisfied – the factors considered significantly reducing risks of T2DM,^[45] while age is a CVD risk factor.^[39] Physically active T2DM patients, studied on one occasion,^[45] had a better health-related quality of life (HRQL) comparing to inactive ones, but it was quickly degrading with progressing obesity. HRQL in this study accounted not only for various physical but also social, emotional and psychological abilities. The correlation between physical and mental domains of HRQL does not preclude the possibility of T2DM being independently reversible through improvement of mental wellbeing. The pension age hikes is an extraordinary development because the age did not change in Australia since 1910^[66] and is, of course, a result of general improvements in economy and health care. It is also known that reduced LDL and TC levels are characteristic of advanced age (70+ years),^[43,67] which may be due to the general slow-down of enzymatic activity and endocrine decline, let alone the unfavourable survival rate in this cohort. Unlike a one-time evaluation effort,^[43] the other study^[67] is a longitudinal one, which allowed the authors to conclude that the reduction is not explained by non-survival, neither it is weight related and otherwise lessens the probability of CVD events. All partici-

pants on any medication modifying cholesterol levels were excluded from the analysis. Explaining the mechanism the authors concur with others that:^[67] “The observed decline in cholesterol levels might reflect inflammatory responses to infections, which increase with age, or age-related reduced cholesterol absorption from the intestine.”

However, the scenario implicating medication is very probable even though in the current research the medicinal aspect of the DiScRi data is not being pursued: from Table 6 it is evident that LDL, the main target of statin medication, is less by 23% in the T2DM cohort comparing to the control - it is far less than statins are known to achieve in trials.^[68,69] The awareness may have also improved, not only the medication, and also the taken for granted access to information. There are behavioural measures that are very effective and the general population just is not educated enough to take full advantage of few simple wisdoms.^[44] In view of the explained, the LDL and TC extended HbA1c rules in Table 1 should be seen as artefacts rather than valid rules for diagnostic of T2DM. However, to be impartial one should not discard the existing evidence that older patients prescribed statins can develop T2DM after long use, found by examining results of a many large trials,^[55] although the risk is small: “Meta-regression showed that risk of development of diabetes with statins was highest in trials with older participants, but neither baseline body-mass index nor change in LDL-cholesterol concentrations accounted for residual variation in risk. Treatment of 255 (95% CI 150-852) patients with statins for 4 years resulted in one extra case of diabetes.” That is, comparing to the same number of participants on placebo and in excess of the DM pro-rata number inherent in the population to whom statins are recommended, which extends to the placebo cohort.

4.17 Experimental setup

However, any misconception about perceived LDL and TC levels in different categories of subjects should be avoided.

Firstly, in one-time studies the information how participants are medicated is often unreported. For example, one study finds a positive substantial correlation of HbA1c with TC but a faltering one with LDL.^[70] In another study the positive correlation is found for both TC and LDL; however, the levels in low and high HbA1c categories are less than 5% different for both TC and LDL.^[43] Actually, in the last study the low and high categories are set well apart, separated by a medium HbA1c range (6.0% - 9.0%); and TC and LDL levels dip before they surface, which may represent patients successfully treated with LLD known to be especially effective at early stages of CVD. Both studies and another one^[71] are one-time, the volunteers are not patients undergoing a

treatment and how they are medicated is unknown, possibly poorly.

Secondly, some efforts are small^[70] and this is even acknowledged.^[71] In the last study a positive substantial correlation was found for HbA1c with both TC and LDL. By comparison, the amount of data in another mentioned study,^[43] where only a small increase, if any, is detected in TC and LDL levels with HbA1c, is more than ten times the amount of data in either of the two.^[70,71]

Thirdly, studies are narrowly focused. For example, a number of studies are interested in T2DM patients only.^[43,68-71] No healthy cohort is included in any of the studies to compare with. Perhaps this parsimony is not groundless. It is difficult to assemble a healthy cohort because requirements might be very specific. A T2DM-free cohort is not the healthy one. Particularly, in the current research all subjects may independently experience CVD and HT. Any comparison involving marker levels in different categories of subjects should be correctly aligned. Even within the same T2DM cohort understanding of what is low and high HbA1c level may differ. For instance, these subcategories in one mentioned study^[71] (below or above 7.0%) do not match those in another mentioned study.^[43] Also, the population cohort being studied may be loosely defined. For example, it is improper to compare results for T2DM and T1DM populations or treat them as if they were same. The etiology of T2DM is very different. Perhaps in respect of the lipoprotein profiles a T1DM population should be included with the healthy cohort.

Also, it is worth nothing that LDL, HDL, TC and TG contents together or individually can be assayed by different methods. LDL is often only estimated using the other three quantities; however, even the expression for it may differ.^[43,70] All four quantities can be directly assayed.^[71]

Unlike all above-mentioned, one study^[72] compares diabetic (T2DM) participants to non-diabetic counterparts, so it implements exactly the same sampling approach as in the current research. The use of statins was accounted for, although it was not a trial, and the authors report the following: “Rate of hypertension, dyslipidaemia, and thereby systolic blood pressure and statin use were significantly higher in the diabetic group, while the serum levels of cholesterol, HDL-C and LDL-C were significantly higher in the control group.” While the non-diabetic participants were almost not using statins (4%), the T2DM patients were heavily relying on them (42%). The LDL levels were down by 17%, TC levels down by 14% and HDL levels down by 13% in the T2DM cohort comparing to controls. All lipid related concentrations were directly measured, which adds to the credibility of the results. Although the study is small, it is noteworthy

that we observe exactly the same trend, and the numbers are close to what we get. From Table 6 the average LDL is less by 23%, TC by 13% and HDL by 11% in the population with T2DM compared to that without. We, of course, do not know in this instance the exact numbers of statin users in either category of individuals screened under provisions of DiScRi, only suspect it was high in the T2DM cohort. Similarly, there is a study in the DiScRi space where T2DM subjects are compared to healthy controls, and where the cholesterol profiles were examined, even though the study is seemingly unrelated, concerned with CAN.^[15] The use of unspecified anti-cholesterol medicines was there higher by T2DM patients (68%) than by healthy subjects (25%). Accordingly, the cholesterol levels were down in the T2DM cohort comparing to controls as follows: LDL by 23%, TC by 14%, and HDL by 14%. This is indeed very close to what we get.

4.18 Atherogenic index

A previously mentioned study^[22] sought a connection between DNA damage in endothelium due to DM, of which 8-OHdG is an indicator, and TG. It is well known, though, that TG is up and HDL is down in T2DM patients, which is a part of dyslipidaemia accompanying T2DM.^[71] In this regard, the connection between TG and WCHR should not be oversighted, which was earlier discussed. Therefore, AIP given as $\text{Log}_{10}(\text{TG}/\text{HDL})$ is a sensitive measure of T2DM.^[73] The atherogenicity of AIP is on the part of HDL - the lacking "good cholesterol" to unsettle build-up of the "bad cholesterol" due to LDL, but the excessive TG, which is a feature of T2DM, also increases the propensity to CVD in T2DM subjects, which explains the index utility for diagnostics of both CVD and T2DM.^[71,73] There is no contradiction here from the data point of view (see Table 6). However, the HDL level may have been inadvertently impacted by the LLD use, as proposed, although this conflicts with the statement in a previously mentioned paper,^[21] weakly supported by results, that statins lower LDL but increase HDL. Other sources claim that there are good drugs to lower LDL cholesterol, but there are no drugs that can raise HDL cholesterol effectively. Indeed, modus-operandi of the statin medication is to trick the metabolism into destroying free LDL particles in serum instead of recycling cholesterol in them.^[74] Incidentally, this explains why statins are especially effective at early stages of CVD. Therefore, both LDL and TC are expected to fall. However, the HDL level is not guaranteed to rise, if only relatively to LDL - the cholesterol is shared by LDL and HDL. Although, effectively, LLD covers for HDL. The pattern observed by us and quantified above is no different. By comparison, LDL and TC levels reported elsewhere^[68] in subjects intensively treated for T2DM, which included statins

for LLD with some exception but also antihypertensives and anticoagulants, were down by about 2/5 and 1/5 from charted results, respectively, comparing to their opposite numbers on conventional therapy by other means, while HDL was not reported. Another study was specifically looking at effects of atorvastatin in T2DM subjects without a history of CVD.^[69] The full effect was evident in months rather than years, so the trial had to be discontinued sooner than planned. Both the placebo and the atorvastatin groups were otherwise similarly medicated. The lipid marker levels at the end of the trial in the atorvastatin group comparing to the placebo group were down by about 2/5 and 1/5 for LDL and TC, respectively (remarkably, the same as the above^[68]), while HDL levels were no different in the two groups, as charted. Hence, the potential of LLD, statins in particular, is much higher than the possible effect of treatment with LLD seen in our results, which are therefore realistic.

4.19 Creatinine

A similar situation exists in respect of creatinine featured in Table 1. An increase of creatinine content in blood is characterising muscle work. This is unlikely to happen in DM because of the undersupply of glucose to muscles. Indeed, one study finds lower serum creatinine levels in T2DM subjects.^[75] However, this is not what the rule linked to creatinine projects and, from Table 6, the levels are confidently higher in the T2DM cohort than in the control one. Creatinine in plasma is a waste product which level is routinely used for estimation of Glomerular Filtration Rate (GFR), so that if creatinine is raised this may be a sign of renal incapacity.^[12] This explains that antihypertensive medication may elevate plasma creatinine levels. Indeed, in a study-contributor to the DiScRi data^[21] "... the percentage of subjects on antihypertensives was remarkably higher in the prediabetes group (35.8%) compared to the control group (5.2%) ..." Instead, the contingent investigated in the first study^[75] was that of non-diabetic office workers who did not have T2DM at the beginning of the trial, were not medicated, and some 10% of whom had developed T2DM over the trial period which lasted few years. It was hypothesised that a population exposed to T2DM is characterised by atrophy of musculature - that is, not only by accumulation of fat. However, the cellular energetics can also play a part.

5. CONCLUSION

The test for HbA1c is easily administered and can be used for estimation of long term blood glucose average level required for identification of the T2DM. However, many HbA1c trials were inconsistent suggesting that a single threshold to determine T2DM may be too simple for the diverse population. There exist markers that alone are not good, clear

indicators of T2DM but were surmised to be useful when paired with HbA1c, particularly near the internationally recommended threshold of 6.5%. The current research deploys novel data analytics to discover co-markers and co-marker related HbA1c thresholds that enhance the predictive ability inherent in HbA1c in identifying T2DM. Particularly, it provides a novel insight into interactions of oxidative stress markers, inflammatory cytokines and lipids with HbA1c in T2DM classification.

Data mining, or data analytics, has been deployed with clinical data for numerous purposes in recent decades including enhancement of diagnoses, improvement of prognoses, early identification of adverse drug reactions, and a variety of insurance and efficiency related objectives. However, one of the main challenges for data mining in healthcare is dealing with missing values, noise and outliers. To deal with many unknowns it makes sense attracting any additional knowledge from the domain of diagnostics when pre-processing the data. Getting results in the decision tree form offers simple rules that enlighten interpretation of the phenomenon in hand. Rules obtained from data with missing values may lead to a discovery if they reveal rare indicators endowed with a high potential. To our knowledge, data mining has not been previously deployed to discover thresholds of a main marker and revised thresholds when combined with a co-marker, also identified via analytics.

The application of data mining in this way requires a dataset where very many plausible markers are collected from the same patients. The dataset used in the current research is derived from a diabetes and cardiovascular screening unit data collection run in recent years in Australia, currently listing 2,700 participant attendance records on 300 attributes. Sophisticated methods for imputing missing values were deployed and decision tree induction was used in combination with optimisation to find precise HbA1c thresholds for relevant co-markers. A new visual representation of the thresholds and co-markers was developed in order to facilitate interpretation of proposed rules.

A number of rules were extracted this way from the data and optimised for performance. This set includes the HbA1c-only rule using a single cutoff level for diagnostic of T2DM, and others that extend this rule by using a second variable. This allows to mitigate the HbA1c based decision depending on patient circumstances and expands the space for decision making in proximity of the cutoff. It has been confirmed that the optimal HbA1c cut point for diagnostic of T2DM is lower than the currently recommended 6.5%, what earlier studies imply, and was estimated to be at the level of 6.2%. Moreover, the useful range of HbA1c can be extended using

additional variables identified in the literature as associates of T2DM. This study presents for the first time objective positive results of emerging blood biomarkers for classification of T2DM. 8-OHdG, GSH-to-GSSG ratio and the pro-inflammatory cytokine IL-6 are oxidative stress markers that have been previously implicated in the pathophysiology of T2DM, especially in the prediabetes stage. This further emphasises the utility of the adopted data mining approach.

However, HbA1c is individually much stronger than any of the extension variables, so that many composite rules only slightly lower the optimal threshold. Nonetheless, the rules offer an improved accuracy of T2DM classification. Such are HbA1c rules extended by BMI and WCHR. However, it was found that individually WCHR, or the waist-circumference-to-height ratio, is a much stronger predictor of T2DM than the conventional BMI, adding to the existing evidence in support of this effect. WCHR or BMI are attractive for their simplicity but do not improve much on the optimal threshold for HbA1c. At the same time, the related dyslipidaemia parameter AIP propagates the HbA1c-only rule below the optimal value.

The proposed rules cannot perform worse than the HbA1c-only rule because of the applied optimisation, and it was found that the optimal HbA1c-only rule is more than 80% accurate. However, due to the small size of the region where the modification is applied only a modest improvement of the overall rule performance can be expected. Despite this, the primary interest in diagnostics is posed by early identification of the possible demise, and any increase in accuracy of the diagnosis should be appreciated closer to the point in the parametrical space, given by individual circumstances, where the transition to the ill state is likely to occur. However, medication may much change the fabric of this space. Particularly, the trend exhibited by lipoproteins changes to opposite the usual perception of their levels being higher in T2DM subjects comparing to healthy people. This may be a feature of the data but is partially confirmed by independent reports. A future study should involve the medicinal component of the database.

The standard rule does not account for well-controlled T2DM where HbA1c values are in the normal range. However these records were retained in the dataset to reflect more closely the population sample. In addition, the HbA1c cutoff level, as defined by WHO guidelines, was originally set based on the presence of diabetic retinopathy and is thus linked to a complication of DM and not the occurrence of DM. However, DM may also manifest in other conditions such as DPN or CAN. That is, removing those with DM and HbA1c above the task force 6.5% would still leave some DM contingent

explained by other features. In this light, it seems to be more appropriate to regard DM an unobservable variable influencing others that are observable. It is not that DM can be independently diagnosed and then any complications, rather it is seen in complications. More narrowly, this all applies to T2DM.

A limitation of the proposed rules is that HbA1c is combined with one other variable, which can be seen as inconclusive. A different variable teamed with HbA1c may inform the opposite. A generic approach to consolidation of all evidence is presented by feature construction where the feature-set is extended using binary features generated from the rules and then the problem of conflict resolving is reduced to the

conventional classification one. A feature of the DiScRi data that could have given a lesser edge to our results is that, once a T2DM diagnosis was initially arrived at for a patient, it is assumed to be correct regardless of the PG level that may go up or down. Both these incongruences are mainly due to the data scarcity. The role of additional biomarkers in prediction or classification of T2DM mandates further exploration with carefully selected population samples, ideally from a larger database.

ACKNOWLEDGEMENTS

The authors express their sincere gratitude to all volunteers who partook in the DiScRi program conducted at CSU.

REFERENCES

- [1] Hare MJ, Shaw JE, Zimmet PZ. Current controversies in the use of haemoglobin A1c. *Journal of Internal Medicine*. 2012; 271(3): 227-36. <http://dx.doi.org/10.1111/j.1365-2796.2012.02513.x>
- [2] American Diabetes Association. Standards of medical care in diabetes: Classification and diagnosis of diabetes. *Diabetes Care*. 2015; 38(S1): 8-16. <http://dx.doi.org/10.2337/dc15-S005>
- [3] Griffin SJ, Borch-Johnsen K, Davies MJ, *et al.* Effect of early intensive multifactorial therapy on 5-year cardiovascular outcomes in individuals with type 2 diabetes detected by screening (ADDITION-Europe): A cluster-randomised trial. *Lancet*. 2011; 378(9786): 156-67. [http://dx.doi.org/10.1016/S0140-6736\(11\)60698-3](http://dx.doi.org/10.1016/S0140-6736(11)60698-3)
- [4] Colagiuri S, Hussain Z, Zimmet P, *et al.* Screening for type 2 diabetes and impaired glucose metabolism: The Australian experience. *Diabetes Care*. 2004; 27(2): 367-71. PMID:14747215. <http://dx.doi.org/10.2337/diacare.27.2.367>
- [5] Saudek CD, Herman WH, Sacks DB, *et al.* A new look at screening and diagnosing diabetes mellitus. *Journal of Clinical Endocrinology and Metabolism*. 2008; 93(7): 2447-53. <http://dx.doi.org/10.1210/jc.2007-2174>
- [6] Bonora E, Tuomilehto J. The pros and cons of diagnosing diabetes with A1C. *Diabetes Care*. 2011; 34(2): 184-90. <http://dx.doi.org/10.2337/dc11-s216>
- [7] d'Emden MC, Shaw JE, Colman PG, *et al.* The role of HbA1c in the diagnosis of diabetes mellitus in Australia. *Medical Journal of Australia (MJA)*. 2012; 197(4): 220-1. <http://dx.doi.org/10.5694/mja12.10988>
- [8] World Health Organization. Use of glycosylated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus. *Abbreviated Report*. 2011.
- [9] Colagiuri S, Lee CM, Wong TY, *et al.* Glycemic thresholds for diabetes-specific retinopathy: Implications for diagnostic criteria for diabetes. *Diabetes Care*. 2011; 34(1): 145-50. <http://dx.doi.org/10.2337/dc10-1206>
- [10] Selvin E, Steffes MW, Zhu H, *et al.* Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. *New England Journal of Medicine*. 2010; 362(9): 800-11. PMID:20200384. <http://dx.doi.org/10.1056/NEJMoa0908359>
- [11] CDC - Centers for Disease Control and Prevention. Self-reported prediabetes and risk-reduction activities — United States, 2006. *Morbidity and mortality weekly report (MMWR)*. 2008; 57(44): 1203-5.
- [12] Deed G, Ackermann E, Newman R, *et al.* General practice management of type 2 diabetes. 2014. Royal Australian College of General Practitioners.
- [13] Thorand B, Baumert J, Chambless L, *et al.* Elevated markers of endothelial dysfunction predict type 2 diabetes mellitus in middle-aged men and women from the general population. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2006; 26: 398-405. <http://dx.doi.org/10.1161/01.ATV.0000198392.05307.aa>
- [14] Tan KCB, Chow WS, Ai VH, *et al.* Advanced glycation end products and endothelial dysfunction in type 2 diabetes. *Diabetes Care*. 2002; 25: 1055-9. <http://dx.doi.org/10.2337/diacare.25.6.1055>
- [15] Tarvainen MP, Laitinen TP, Lipponen JA, *et al.* Cardiac autonomic dysfunction in type 2 diabetes - effect of hyperglycemia and disease duration. *Frontiers in Endocrinology*. 2014; 5(130). <http://dx.doi.org/10.3389/fendo.2014.00130>
- [16] Khandoker AH, Jelinek HF, Palaniswami M. Identifying diabetic patients with cardiac autonomic neuropathy by heart rate complexity analysis. *BioMedical Engineering OnLine*. 2009; 8(3). <http://dx.doi.org/10.1186/1475-925X-8-3>
- [17] Badawi A, Klip A, Haddad P, *et al.* Type 2 diabetes mellitus and inflammation: Prospects for biomarkers of risk and nutritional intervention. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*. 2010; 3: 173-86. <http://dx.doi.org/10.2147/DMSOT.T.99089>
- [18] Yeh ETH, Willerson JT. Coming of age of C-reactive protein: Using inflammation markers in cardiology. *Circulation*. 2003; 107: 370-1. <http://dx.doi.org/10.1161/01.CIR.0000053731.05365.5A>
- [19] de-M-Bandeira S, da-Fonseca LJS, da-S-Guedes G, *et al.* Oxidative stress as an underlying contributor in the development of chronic complications in diabetes mellitus. *International Journal of Molecular Sciences*. 2013; 14(2): 3265-84. <http://dx.doi.org/10.3390/ijms14023265>
- [20] Giacco F, Brownlee M. Oxidative stress and diabetic complications. *Circulation Research*. 2010; 107: 1058-70. <http://dx.doi.org/10.1161/CIRCRESAHA.110.223545>
- [21] Maschirow L, Khalaf K, Al-Aubaidy HA, *et al.* Inflammation, coagulation, endothelial dysfunction and oxidative stress in prediabetes — Biomarkers as a possible tool for early disease detection

- for rural screening. *Clinical Chemistry*. 2015; 48(9): 581-5. <http://dx.doi.org/10.1016/j.clinbiochem.2015.02.015>
- [22] Al-Aubaidy H, Jelinek HF. Oxidative stress and triglycerides as predictors of subclinical atherosclerosis in prediabetes. *Redox Report*. 2014; 19(2): 87-91. <http://dx.doi.org/10.1179/1351000213Y.0000000080>
- [23] Pi-Sunyer X, Blackburn G, Brancati FL, *et al*. Reduction in weight and cardiovascular disease risk factors in individuals with type 2 diabetes: One-year results of the Look AHEAD trial. *Diabetes Care*. 2007; 30(6): 1374-83. <http://dx.doi.org/10.2337/dc07-0048>
- [24] Apovian CM. Management of diabetes across the course of disease: Minimizing obesity-associated complications. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*. 2011; 4: 353-69. <http://dx.doi.org/10.2147/DMSO.S24022>
- [25] Cios KJ, Pedrycz W, Swiniarski RW, *et al*. Data mining: A knowledge discovery approach. 2007. Springer.
- [26] Mullins IM, Siadat MS, Lyman J, *et al*. Data mining and clinical data repositories: Insights from a 667,000 patient data set. *Computers in Biology and Medicine*. 2006; 36(12): 1351-77. <http://dx.doi.org/10.1016/j.combiomed.2005.08.003>
- [27] Aljumah AA, Ahamad MG, Siddiqui MK. Application of data mining: Diabetes health care in young and old patients. *Journal of King Saud University - Computer and Information Sciences*. 2013; 25(2): 127-36. <http://dx.doi.org/10.1016/j.jksuci.2012.10.003>
- [28] Savva SC, Lamnisos D, Kafatos AG. Predicting cardiometabolic risk: Waist-to-height ratio or BMI — A meta-analysis. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*. 2013; 6: 403-19. <http://dx.doi.org/10.2147/DMSO.S34220>
- [29] Ashwell M, Gunn P, Gibson S. Waist-to-height ratio is a better screening tool than waist circumference and BMI for adult cardiometabolic risk factors: Systematic review and meta-analysis. *Obesity Reviews*. 2012; 13: 275-86. <http://dx.doi.org/10.1111/j.1467-789X.2011.00952.x>
- [30] Meng XH, Huan YZ, Rao DP, *et al*. Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. *Kaohsiung Journal of Medical Sciences*. 2013; 29(2): 93-9. <http://dx.doi.org/10.1016/j.kjms.2012.08.016>
- [31] Burstein F, De-Silva D, Jelinek H, *et al*. Multivariate data-driven decision guidance for clinical scientists. 29-th International Conference on Data Engineering (ICDE), Workshop on Data-Driven Decision Support and Guidance Systems. 2013: 193-9. IEEE. <http://dx.doi.org/10.1109/ICDEW.2013.6547449>
- [32] Jelinek HF, Wilding C, Tinely P. An innovative multi-disciplinary diabetes complications screening program in a rural community: A description and preliminary results of the screening. *Australian Journal of Primary Health*. 2006; 12(1): 14-20. <http://dx.doi.org/10.1071/PY06003>
- [33] Jelinek HF, Yatsko A, Stranieri A, *et al*. Diagnostic with incomplete nominal/discrete data. *Artificial Intelligence Research*. 2015; 4(1): 22-35. <http://dx.doi.org/10.5430/air.v4n1p22>
- [34] Bagirov A, Yatsko A, Stranieri A, *et al*. Feature selection using misclassification counts. 9-th Australasian Data Mining Conference (AusDM); Conferences in Research and Practice in Information Technology (CRPIT). 2011; 121: 51-62.
- [35] Cornforth D, Jelinek HF, Tarvainen M. A comparison of nonlinear measures for the detection of cardiac autonomic neuropathy from heart rate variability. *Entropy*. 2015; 17(3): 1425-40. <http://dx.doi.org/10.3390/e17031425>
- [36] Stranieri A, Abawajy B, Kelarev A, *et al*. An approach for Ewing test selection to support the clinical assessment of cardiac autonomic neuropathy. *Artificial Intelligence in Medicine*. 2013. <http://dx.doi.org/10.1016/j.artmed.2013.04.007>
- [37] Schafer JL. Analysis of incomplete multivariate data. 1997. Chapman and Hall.
- [38] Quinlan R. C4.5: Programs for machine learning. 1993. Morgan Kaufmann.
- [39] Lloyd-Jones DM. Cardiovascular risk prediction: Basic concepts, current status, and future directions. *Circulation Research* 2010; 121(15): 1768-77. <http://dx.doi.org/10.1161/CIRCULATIONAHA.109.849166>
- [40] Feng Y, Schlösser FJ, Sumpio BE. The Semmes Weinstein monofilament examination as a screening tool for diabetic peripheral neuropathy. *Journal of Vascular Surgery*. 2009; 50: 675-82. <http://dx.doi.org/10.1016/j.jvs.2009.05.017>
- [41] Bansal D, Gudala K, Muthyala H, *et al*. Prevalence and risk factors of development of peripheral diabetic neuropathy in type 2 diabetes mellitus in a tertiary care setting. *Journal of Diabetes Investigation*. 2014; 5(6): 714-21. <http://dx.doi.org/10.1111/jdi.12223>
- [42] Xu F, Zhao LH, Su JB, *et al*. The relationship between glycemic variability and diabetic peripheral neuropathy in type 2 diabetes with well-controlled HbA1c. *Diabetology and Metabolic Syndrome*. 2014; 6(1): 139. <http://dx.doi.org/10.1186/1758-5996-6-139>
- [43] Khan HA. Clinical significance of HbA1c as a marker of circulating lipids in male and female type 2 diabetic patients. *Acta Diabetologica*. 2007. <http://dx.doi.org/10.1007/s00592-007-0003-x>
- [44] Wagner J, Lacey K, Chyun D, *et al*. Development of a questionnaire to measure heart disease risk knowledge in people with diabetes: The heart disease fact questionnaire. *Patient Education and Counseling*. 2005; 58: 82-7. <http://dx.doi.org/10.1016/j.pec.2004.07.004>
- [45] Eckert K. Impact of physical activity and bodyweight on health-related quality of life in people with type 2 diabetes. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*. 2012; 5: 303-11. <http://dx.doi.org/10.2147/DMSO.S34835>
- [46] Maekawa S, Kawahara T, Nomura R, *et al*. Retrospective study on the efficacy of a low-carbohydrate diet for impaired glucose tolerance. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*. 2014; 7: 195-201. <http://dx.doi.org/10.2147/DMSO.S62681>
- [47] Rolland C, Hession M, Broom I. Effect of weight loss on adipokine levels in obese patients. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*. 2011; 4: 315-23. <http://dx.doi.org/10.2147/DMSO.S22788>
- [48] Plourde G. Reversal of type 2 diabetes mellitus in an obese man: Role of dietary modification and physical activity. *Clinical Nursing Studies*. 2013; 1(2): 38-42. <http://dx.doi.org/10.5430/cns.v1n2p38>
- [49] Coughlan KA, Valentine RJ, Ruderman NB, *et al*. AMPK activation: A therapeutic target for type 2 diabetes? *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*. 2014; 7: 241-53. <http://dx.doi.org/10.2147/DMSO.S43731>
- [50] Dietterich T. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*. 1998; 10: 1895-1923. <http://dx.doi.org/10.1162/089976698300017197>
- [51] Hall JE. The kidney, hypertension, and obesity. *Hypertension*. 2003; 41(2): 625-33. <http://dx.doi.org/10.1161/01.HYP.0000052314.95497.78>
- [52] Faerch K, Hulman A, Solomon TPJ. Heterogeneity of pre-diabetes and Type 2 diabetes: Implications for prediction, prevention and treatment responsiveness. *Current Diabetes Reviews*. 2016; 12: 30-41. <http://dx.doi.org/10.2174/1573399811666150416122903>
- [53] Yu EYT, Wong CKH, Ho SY, *et al*. Can HbA1c replace OGTT for the diagnosis of diabetes mellitus among Chinese patients with

- impaired fasting glucose? *Family Practice*. 2015; 32(6): 631-8. <http://dx.doi.org/10.1093/fampra/cmz077>
- [54] Pajunen P, Peltonen M, Eriksson JG, *et al.* HbA1c in diagnosing type 2 diabetes in impaired glucose tolerance: The Finnish diabetes prevention study. *Diabetic Medicine*. 2010; 28: 36-42. <http://dx.doi.org/10.1111/j.1464-5491.2010.03183.x>
- [55] Sattar N, Preiss D, Murray HM, *et al.* Statins and risk of incident diabetes: A collaborative meta-analysis of randomized statin trials. *Lancet*. 2010; 375: 735-42. [http://dx.doi.org/10.1016/S0140-6736\(09\)61965-6](http://dx.doi.org/10.1016/S0140-6736(09)61965-6)
- [56] Jelinek HF, Al-Aubaidy HA, Maschirow L, *et al.* Glutathione: Glutathione sulfide redox imbalance in early impaired fasting glucose. *Cardiology and Angiology: An International Journal*. 2014; 2(4): 223-9. <http://dx.doi.org/10.9734/CA/2014/11389>
- [57] Ewing DJ, Martyn CN, Young RJ, *et al.* The value of cardiovascular autonomic function tests: 10 years experience in diabetes. *Diabetes Care*. 1985; 8: 491-8. <http://dx.doi.org/10.2337/diacare.8.5.491>
- [58] Jun J, Jin SM, Baek J, *et al.* The association between glycemic variability and diabetic cardiovascular autonomic neuropathy in patients with type 2 diabetes. *Cardiovascular Diabetology*. 2015; 14(1): 70. <http://dx.doi.org/10.1186/s12933-015-0233-0>
- [59] Chen-Scarabelli C, Scarabelli TM. Suboptimal glycemic control, independently of QT interval duration, is associated with increased risk of ventricular arrhythmias in a high-risk population. *Pacing and clinical electrophysiology (PACE)*. 2006; 29(1): 9-14. PMID:16441711. <http://dx.doi.org/10.1111/j.1540-8159.2006.00298.x>
- [60] Kashani A, Barold SS. Significance of QRS complex duration in patients with heart failure. *Journal of the American College of Cardiology*. 2005; 46(12): 2183-92. <http://dx.doi.org/10.1016/j.jacc.2005.01.071>
- [61] Arasteh A, Farahi S, Habibi-Rezaei M, *et al.* Glycated albumin: An overview of the in vitro models of an in vivo potential disease marker. *Journal of Diabetes and Metabolic Disorders*. 2014; 13(49). <http://dx.doi.org/10.1186/2251-6581-13-49>
- [62] Jelinek HF, Jamil DA, Al-Aubaidy HA. Impaired fasting glucose and 8-iso-prostaglandin F2a in diabetes disease progression. *British Journal of Medicine and Medical Research*. 2014; 4(33): 5229-37. <http://dx.doi.org/10.9734/BJMMR/2014/11147>
- [63] Taggart J, Wan Q, Davies GP, *et al.* A longitudinal analysis of type 2 diabetes data from the Macarthur Division of General Practice. UNSW Research Centre for Primary Health Care and Equity Report 2006. University of New South Wales.
- [64] Twigg SM, Wong J. The imperative to prevent diabetes complications: A broadening spectrum and an increasing burden despite improved outcomes. *Medical Journal of Australia (MJA)*. 2015; 202(6): 300-6. <http://dx.doi.org/10.5694/mja14.01234>
- [65] Stettler N, Murphy MM, Barraij LM, *et al.* Systematic review of clinical studies related to pork intake and metabolic syndrome or its components. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*. 2013; 6: 347-57. <http://dx.doi.org/10.2147/DMSO.S51440>
- [66] Mavromaras K, Richardson S, Zhu R. Age pension age eligibility, retirement and health outcomes in Australia. NILS Working Paper Series. 2015; 201. National Institute of Labour Studies; Flinders University, Adelaide, Australia.
- [67] Ferrara A, Barrett-Connor E, Shan J. Total, LDL, and HDL cholesterol decrease with age in older men and women: The rancho Bernardo study 1984-1994. *Circulation*. 1997; 96: 37-43. <http://dx.doi.org/10.1161/01.CIR.96.1.37>
- [68] Gaede P, Vedel P, Larsen N, *et al.* Multifactorial intervention and cardiovascular disease in patients with type 2 diabetes. *New England Journal of Medicine*. 2003; 348(5): 383-93. PMID:12556541. <http://dx.doi.org/10.1056/NEJMoa021778>
- [69] Colhoun HM, Betteridge DJ, Durrington PN, *et al.* Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the collaborative atorvastatin diabetes study (CARDS): Multicentre randomised placebo-controlled trial. *Lancet*. 2004; 364: 685-96. [http://dx.doi.org/10.1016/S0140-6736\(04\)16895-5](http://dx.doi.org/10.1016/S0140-6736(04)16895-5)
- [70] Singh G, Kumar A. Relationship among HbA1c and lipid profile in Punjabi type 2 diabetic population. *Journal of Exercise Science and Physiotherapy*. 2011; 7(2): 99-102. <http://dx.doi.org/10.18376//2011/v7i2/67614>
- [71] Bodhe C, Jankar D, Bhutada T, *et al.* HbA1c: Predictor of dyslipidemia and atherogenicity in diabetes mellitus. *International Journal of Basic Medical Sciences and Pharmacy (IJBMS)*. 2012; 2(1): 25-7.
- [72] Bagherzadeh A, Nejati-Afkham A, Tajallizade-Khoob Y, *et al.* Association of cardiac autonomic neuropathy with arterial stiffness in type 2 diabetes mellitus patients. *Journal of Diabetes and Metabolic Disorders*. 2013; 12(55). <http://dx.doi.org/10.1186/2251-6581-12-55>
- [73] Dobiashova M. Atherogenic index of plasma Log(Triglycerides/HDL-Cholesterol): Theoretical and practical implications. *Clinical Chemistry*. 2004; 50(7): 1113-5. <http://dx.doi.org/10.1373/clinchem.2004.033175>
- [74] Wang CY, Liu PY, Liao JK. Pleiotropic effects of statin therapy: Molecular mechanisms and clinical results. *Trends in Molecular Medicine*. 2007; 14(1): 37-44. <http://dx.doi.org/10.1016/j.molmed.2007.11.004>
- [75] Harita N, Hayashi T, Sato KK, *et al.* Lower serum creatinine is a new risk factor of type 2 diabetes: The Kansai healthcare study. *Diabetes Care*. 2009; 32(3): 424-6. <http://dx.doi.org/10.2337/dc08-1265>