Predicting blood glucose levels from HbA_{1C} is not the same as predicting HbA_{1C} from blood glucose levels: a common methodological misunderstanding.

Daniel D. Reidpath[a,b], Pascale Allotey[a,b], Mark R. Diamond[c]

[a] *South East Asia Community Observatory (SEACO), Monash University*
[b] *Global Public Health, School of Medicine and Health Sciences, Monash University Malaysia*
[c] *Department of Psychology, University of Tasmania*

Abstract

A methodological problem with the conversion of HbA_{1C} to blood glucose level and visa versa has slipped under our noses. The prediction equations on which the conversions are based assume that predicting HbA_{1C} from blood glucose is the same as predicting blood glucose from HbA_{1C}. The problem is pervasive and can be seen on the websites of the American Diabetes Association, Diabetes UK, the US, National Glycohemoglobin Standardization Program, on commercial sites, and in research. Using a secondary analysis of publicly available outpatient data we illustrate why predicting one from the other is not the same as the reverse. The implications for clinical care, research, and patient self management are highlighted.

Keywords: HbA_{1C}, Blood glucose, Prediction, HbA_{1C}, Regression

SEACO Research Note: *October 20, 2015*

doi:10.4225/03/56239BD13C5E0

Introduction

National diabetes associations, commercial providers, and researchers incorrectly assume that the same basic equation can be used to convert back and

*For correspondence please write to Daniel D. Reidpath

Email addresses: daniel.reidpath@monash.edu (Daniel D. Reidpath), pascale.allotey@monash.edu (Pascale Allotey), mark.diamond@utas.edu.au (Mark R. Diamond)

forth between measures of HbA_{1C} and blood glucose levels. This assumption is demonstrably wrong.

The measurement of glycaemia is a pillar of diabetic management, and central to diabetes research. The specific measure of interest may be the average blood glucose levels, the variability in blood glucose levels, or blood glucose levels at points in time. Each of these measures has independent clinical value,[1, 2, 3, 4, 5] and in combination develop a complete picture of glycaemic control. Obtaining all the measures is usually impractical, however, and one measure is often used to predict another.

A test of glycolsylated haemoglobin levels (HbA_{1C}), for example, is often used as an estimate of average blood glucose levels, and overall glycaemic control.[6] The rate of haemoglobin (Hb) glycation—the bonding of a glucose molecule to an Hb molecule—is a function of plasma glucose concentration, with higher plasma glucose levels associated with higher rates of Hb glycation.[6, 7] Once glycation has occurred, it is not reversed for the life of the glycated cell. As a consequence, measuring HbA_{1C} levels may be used as an estimate of average blood glucose levels over the two to three months prior to testing.[6] It is important to note here that the correspondence between average blood glucose levels and HbA_{1C} is imperfect and significant inter- and intra-individual variation occurs. [7, 8, 9, 10]

Notwithstanding the variation, researchers have developed prediction equations for the estimation of average blood glucose levels based on HbA_{1C} test results, and the estimation of HbA_{1C} levels based on blood glucose test results.[11, 12] The immediate intuition is that these prediction equations should be simple algebraic transpositions of each other. If I have an equation to estimate (average) blood glucose levels from HbA_{1C} results, I should be able to transpose the equation to derive a measure of HbA_{1C} from (average) blood glucose measures. National diabetes support associations including the *American Diabetes*

Association;[1] *Diabetes UK*;[2] the US, *National Glycohemoglobin Standardization Program* (NGSP);[3] as well as commercial vendors[4] provide online "conversion" facilities of HbA_{1C} and blood glucose values for diabetes management. A quick exploration of the facilities reveals that they all rely on an algebraic transposition of an underlying equation for moving back and forth between the HbA_{1C} levels and average blood glucose levels.

There is a significant flaw in the assumption that the same basic equation can be used to move back and forth between HbA_{1C} and blood glucose, and it arises from a misunderstanding about the statistical regression techniques used to develop the prediction equations. We demonstrate the problem with publicly available blood glucose and HbA_{1C} data and discuss the implications.

Predictions from regression

We use data derived from 349 diabetic out-patients each contributing a single HbA_{1C} measure and a single random blood glucose measure [5].[13] We contrast the regression equation for predicting HbA_{1C} levels from blood glucose levels, and the regression equation for predicting blood glucose levels from HbA_{1C} levels. The code for running the analyses in the R statistical environment are available as a supplementary material.[6][14]

To minimise visual clutter, Figure 1 shows the derived HbA_{1C} and blood glucose measures taken from 20 of the 349 patients selected at random. The solid black line running through the graph shows the ordinary least squares regression line for the prediction of HbA_{1C} (mmol/mol) from the patients' blood glucose (mmol/l). The equation for the line is calculated so that it minimises the sum of the squared vertical distances (deviations) between each of

[1] `http://professional.diabetes.org/glucosecalculator.aspx`
[2] `http://www.diabetes.co.uk/hba1c-to-blood-sugar-level-converter.html`
[3] `http://www.ngsp.org/convert1.asp`
[4] `http://www.pendiq.com/en/blood-glucose-calculator`
[5] (`http://lib.stat.cmu.edu/datasets/hba1c_bloodGlucose.dat`)
[6] `https://gist.github.com/dreidpath/76bc76a1cecb1d454801`

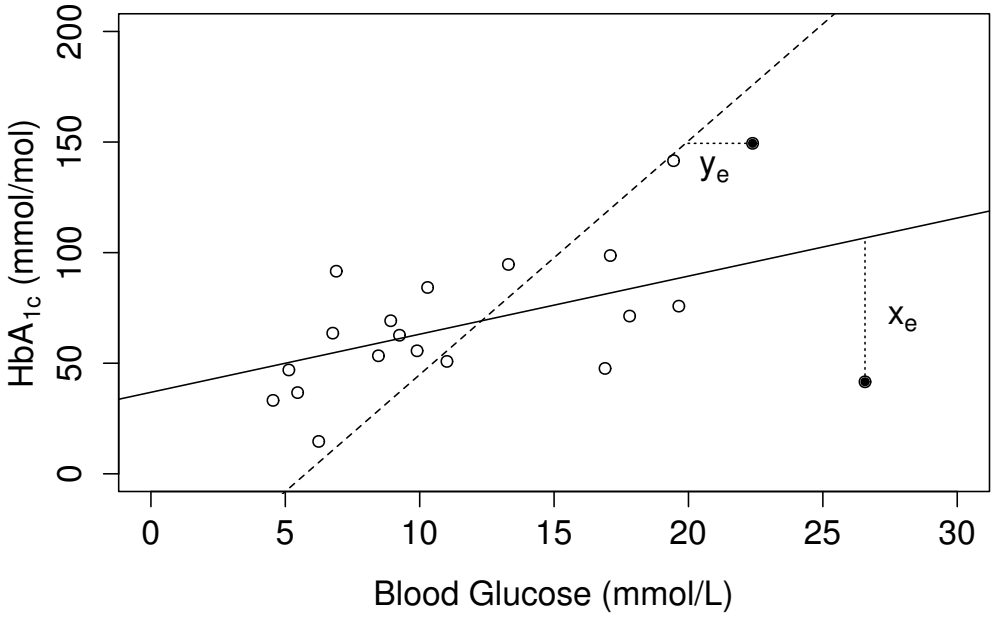

Figure 1 The relationship between the HbA_{1C} (mmol/mol) and random blood glucose (mmol/L) from 20 of 349 randomly selected out-patients. The solid black line shows the regression line for blood glucose predicting HbA_{1C}. The vertical line – x_e – shows an example of the error between one of the points and the solid prediction line. The dashed black line shows the regression line for HbA_{1C} predicting blood glucose. The horizontal line – y_e – shows an example of the error between one of the points and the dashed prediction line.

the 20 points and the line. That is, the sum of the squared errors between the x-values (blood glucose) and the line predicting the y-values (HbA_{1C}) is minimised. Any other line (i.e., any other prediction equation) would have a greater sum of squared errors. The vertical dotted line marked x_e shows the error between the actual data and the prediction line for one of the twenty data points. Now consider the steeper sloped, dashed line. This shows the regression line for the opposite prediction of blood glucose from HbA_{1C}. This time the equation for the line is calculated so that it minimises the sum of the squared

horizontal distances between each of the points and the line. That is, the sum of the squared errors between the y-values ($\mathrm{HbA_{1C}}$) and the line predicting the x-values (blood glucose) is minimised. The horizontal dotted line marked y_e shows the error between the actual data and the prediction line for one of the twenty points.

The minimisation of the squared errors between the line and the data is the optimisation problem solved by linear regression. However, finding the line that minimises the sum of the squared y_e's is not the same problem as finding the line that minimises the sum of the squared x_e's. Although the prediction equations are not the same, the fit of the two models to the data (i.e., the variance explained by the models) is always identical — in this case $R^2 = .21$.

Figure 2 shows the $\mathrm{HbA_{1C}}$ and blood glucose data for all 349 patients with the respective regression lines. It is evident that the two prediction lines are quite different. Although the blood glucose data are based on a single random blood glucose test there is a readily apparent correspondence between $\mathrm{HbA_{1C}}$ levels and random blood glucose levels. In both cases the predictors are significant (p¡.0001), and they account for a little over 40% of the variance ($R^2 = .44$; $r = .67$). It is, however, only at the point of intersection of the two lines (marked with a solid black square) that the prediction of $\mathrm{HbA_{1C}}$ levels from blood glucose levels will agree with the prediction of blood glucose levels from $\mathrm{HbA_{1C}}$ levels: Blood glucose 13.2 mmol/L $\approx$ $\mathrm{HbA_{1C}}$ 80 mmol/mol. The further away from the point of intersection, the greater the error that will arise by assuming the relationship between the two measures are a function of a simple algebraic transposition. One practical question that arises from this analysis is, "how big a problem is it?" An estimate, even a biased estimate may still be useful — a "ball park" figure. Often exact answers are not required, just a general idea. Within this data set we could easily calculate the error (maybe the root mean squared error (*RMSE*)) between the $\mathrm{HbA_{1C}}$ levels predicted with the correct equation and actual $\mathrm{HbA_{1C}}$ levels; and those predicted with the incorrect equation and actual $\mathrm{HbA_{1C}}$ levels.

This would, unfortunately, only get us part way towards understanding the

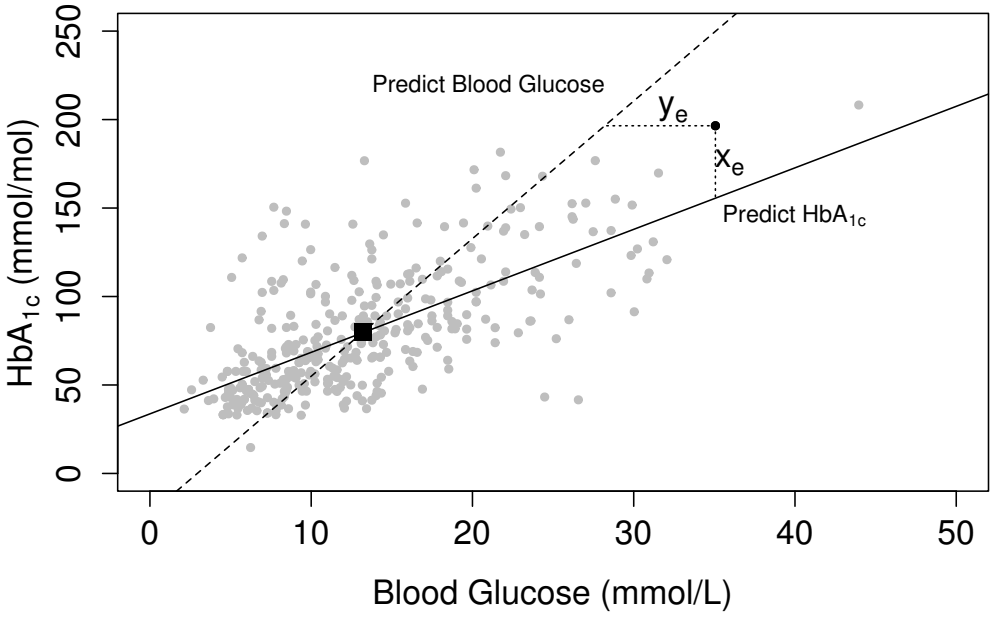

Figure 2 The relationship between the HbA_{1C} (mmol/mol) and random blood glucose (mmol/L) from 349 out-patients. The solid black line shows the regression line for blood glucose predicting HbA_{1C}. The dashed black line shows the regression line for HbA_{1C} predicting blood glucose. The point of intersection is highlighted.

magnitude of the problem. We cannot know what level of error would arise from the misapplication of the published (web-available) equations used by the *American Diabetes Association*, the NGSP, etc. It is also further complicated because the level of error in the sample used to develop the prediction equations maybe quite different from the level of error one could expect when applying it in a different population. Consider, for example, a study of people with high blood glucose levels, for whom we want to estimate $\mathrm{HbA_{1C}}$ levels. The *RMSE* for that population would be much greater than the *RMSE* for a population with blood glucose levels that lay nearer to the intersection between the two prediction lines.

The limitation of this analysis lies directly in the data we had available to explore the ideas. The web applications that diabetes organisations have made available were for converting between $\mathrm{HbA_{1C}}$ and average blood glucose. We did not have access to *average* blood glucose data, but random blood glucose data. It is likely, therefore that the difference between the two prediction lines would not be as great for the diabetes organisations' approach.

Conclusion

It is wrong to treat the estimation of $\mathrm{HbA_{1C}}$ from blood glucose as the same problem as the estimation of blood glucose from $\mathrm{HbA_{1C}}$. In the absence of data, and knowledge of the study or patient population, it is hard to know the impact of the error. Fortunately it is easily corrected.

References

[1] C.-C. Lin, C.-P. Yang, C.-I. Li, C.-S. Liu, C.-C. Chen, W.-Y. Lin, K.-L. Hwang, S.-Y. Yang, T.-C. Li, Visit-to-visit variability of fasting plasma glucose as predictor of ischemic stroke: competing risk analysis in a national cohort of Taiwan Diabetes Study, BMC medicine 12 (2014) 165, 00000 PMID: 25255837. doi:10.1186/s12916-014-0165-7.

[2] T. S. Temelkova-Kurktschiev, C. Koehler, E. Henkel, W. Leonhardt, K. Fuecker, M. Hanefeld, Postchallenge plasma glucose and glycemic spikes are more strongly associated with atherosclerosis than fasting glucose or HbA1c level, Diabetes Care 23 (12) (2000) 1830–1834, 00521 PMID: 11128361.

[3] M. Muggeo, G. Zoppini, E. Bonora, E. Brun, R. C. Bonadonna, P. Moghetti, G. Verlato, Fasting plasma glucose variability predicts 10-year survival of type 2 diabetic patients: the Verona Diabetes Study, Diabetes Care 23 (1) (2000) 45–50, 00235 PMID: 10857967.

[4] D. L. Trence, I. B. Hirsch, Motherhood, apple pie, hemoglobin A(1c), and the DCCT, Endocrine practice: official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists 18 (1) (2012) 78–84, 00005 PMID: 22336443.

[5] G. A. Nichols, S. Joshua-Gotlib, S. Parasuraman, Glycemic Control and Risk of Cardiovascular Disease Hospitalization and All-Cause Mortality, Journal of the American College of Cardiology 62 (2) (2013) 121–127. doi:10.1016/j.jacc.2013.04.031.

[6] K. Makris, L. Spanou, A. Rambaouni-Antoneli, K. Koniari, I. Drakopoulos, D. Rizos, A. Haliassos, Relationship between mean blood glucose and glycated haemoglobin in Type 2 diabetic patients, Diabetic Medicine 25 (2) (2008) 174–178, 00015. doi:10.1111/j.1464-5491.2007.02379.x.

[7] R. M. Cohen, R. S. Franco, P. K. Khera, E. P. Smith, C. J. Lindsell, P. J. Ciraolo, M. B. Palascak, C. H. Joiner, Red cell life span heterogeneity in hematologically normal people is sufficient to alter HbA1c, Blood 112 (10) (2008) 4284–4291. doi:10.1182/blood-2008-04-154112.

[8] E. S. Kilpatrick, A. S. Rigby, S. L. Atkin, Variability in the Relationship between Mean Plasma Glucose and HbA1c: Implications for the Assessment of Glycemic Control, Clinical Chemistry 53 (5) (2007) 897–901. doi:10.1373/clinchem.2006.079756.

[9] S. A. Chalew, J. M. Hempe, R. McCarter, Clinically Significant Disagreement between Mean Blood Glucose and Estimated Average Glucose in Two Populations: Implications for Diabetes Management, Journal of diabetes science and technology (Online) 3 (5) (2009) 1128–1135.

[10] J. M. Hempe, A. A. Soros, S. A. Chalew, Estimated Average Glucose and Self-Monitored Mean Blood Glucose Are Discordant Estimates of Glycemic Control, Diabetes Care 33 (7) (2010) 1449–1451. doi:10.2337/dc09-1498.

[11] F. C. F. Otieno, L. Ng'ang'a, M. Kariuki, Validity of random blood glucose as a predictor of the quality of glycaemic control by glycated haemoglobin in out-patient diabetic patients at Kenyatta National Hospital, East African Medical Journal 79 (9) (2002) 491–495.

[12] D. M. Nathan, J. Kuenen, R. Borg, H. Zheng, D. Schoenfeld, R. J. Heine, A1c-Derived Average Glucose Study Group, Translating the A1c assay into estimated average glucose values, Diabetes care 31 (8) (2008) 1473–1478, 00799. doi:10.2337/dc08-0545.

[13] O. F. Daramola, Assessing the validity of random blood glucose testing for monitoring glycemic control and predicting HbA1c values in type 2 diabetics at Karl Bremer hospital, Thesis, Stellenbosch : Stellenbosch University (Dec. 2012).
URL http://scholar.sun.ac.za/handle/10019.1/80458

[14] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2015).
URL https://www.R-project.org/