

# Are Population-Based Diabetes Models Useful for Individual Risk Estimation?

Barry G. Saver, MD, MPH, J. Lee Hargraves, PhD, and Kathleen M. Mazor, EdD

**Background:** Predictive models are increasingly used in guidelines and informed decision-making interventions. We compared predictions from 2 prominent models for diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) outcomes model and the Archimedes-based Diabetes Personal Health Decisions (PHD) model.

**Methods:** Ours was a simulation study comparing 10-year and 20-year model predictions for risks of myocardial infarction (MI), stroke, amputation, blindness, and renal failure for representative test cases.

**Results:** The Diabetes PHD model predicted substantially higher risks of MI and stroke in most cases, particularly for stroke and for 20-year outcomes. In contrast, the UKPDS model predicted risks of amputation and blindness ranging from 2-fold to infinitely higher than the Diabetes PHD model. Predictions for renal failure all differed by more than 2-fold but in a complicated pattern varying by time frame and specific risk factors. Relative to their predictions for white men, the UKPDS model predicted much lower MI and stroke risks for women and Afro-Caribbean men than the Diabetes PHD model did for women and black men. A substantial majority of the Diabetes PHD point estimates fell outside of the UKPDS outcomes model's 95% CIs.

**Conclusions:** These models produced markedly different predictions. Patients and providers considering risk estimates from such models need to understand their substantial uncertainty and risk of misclassification. (J Am Board Fam Med 2011;24:399–406.)

**Keywords:** Diabetes, Medical Decision-Making

The prevalence of diabetes among the American population has increased dramatically in the last 2 decades.<sup>1</sup> Diabetes increases the risks of multiple adverse health outcomes and is frequently accompanied by other cardiovascular risk factors. Historically, diabetes treatment has been very glucocentric,<sup>2,3</sup> despite the fact that premature morbidity

and mortality in diabetes are more strongly affected by cardiovascular risk factors than glucose control.<sup>4–7</sup>

If patients are to make informed choices about where to focus their efforts for risk reduction, access to accurate, personal estimates of risk and effects of risk reduction activities is a logical first step. The 2 most readily available models that can be used to predict risks of multiple outcomes for persons with diabetes and model the effects of risk factor modification are the United Kingdom Prospective Diabetes Study (UKPDS) outcomes model<sup>8</sup> and the Diabetes Personal Health Decisions (PHD) tool, available from the American Diabetes Association<sup>9</sup> and based on the Archimedes Model.<sup>10</sup> We report here a comparison of output from these 2 models for a set of test cases that we undertook when trying to decide which model to use in a study of communication with diabetic patients about their multiple risks and risk reduction options.

This article was externally peer reviewed.

Submitted 28 January 2011; revised 13 March 2011; accepted 1 April 2011.

From the Department Family Medicine and Community Health (BGS, JLH) and the Department of Medicine (KMM), University of Massachusetts Medical School, Worcester.

**Funding:** Support for this research was provided by Robert Wood Johnson Foundation grant no. 63824.

**Prior presentation:** This work was presented at the 38<sup>th</sup> annual meeting of the North American Primary Care Research Group, Seattle, WA, 13–17 November 2010.

**Conflict of interest:** none declared.

**Corresponding author:** Barry Saver, Department of Family Medicine and Community Health, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655 (E-mail: Barry.Saver@umassmed.edu).

## Methods

We obtained permission from the American Diabetes Association to use the Diabetes PHD website for research purposes, and we received a research license to use the UKPDS outcomes model, version 1.2, from the University of Oxford.

We constructed sets of sample cases to obtain predictions from each model, as outlined in Table 1. The cases for each model were identical to the extent possible given that each model uses some information the other does not. Starting with the male and female base cases, we created additional cases varying one factor at a time from the base case.

The UKPDS outcomes model allows the researcher to specify risk factor values for up to the subsequent 40 years, or the program can model them based on the current values (and project regression toward the mean). Here, we report models in which we forced the values to remain constant for 40 years, which tends to increase estimates of

benefits from risk factor control by preventing regression to the mean. However, the differences in the estimates were modest and there was no other way to model quitting smoking without relapse. The Diabetes PHD model allows risk factors to be changed and then immediately displays changes in risk projections; how it models risk factor changes mathematically has not been publicly reported. In the UKPDS outcomes model, probability estimates are generated by incorporating a random number seed to reflect the inherent randomness in the occurrence of any event for an individual. To produce stable estimates, the developers provide an option to run multiple simulations and average the results to obtain more stable estimates, a technique referred to as “Monte Carlo simulation.” We specified 10,000 Monte Carlo simulation loops because this produced stable estimates. The UKPDS outcomes model also will produce 95% CIs around its estimates using a bootstrapping approach based on sampling with replacement (taking multiple ran-

**Table 1. Case Characteristics for the United Kingdom Prospective Diabetes Study and Diabetes Personal Health Decision Models**

Characteristics	UKPDS	Diabetes PHD
Baseline		
Age (years)	60	60
Sex	Male and female	Male and female
Race/ethnicity	White	Non-Hispanic white
Age at diagnosis of diabetes (years)	50	50
Weight	100 kg	220 lb
Height	1.75 m	5' 9"
Total/HDL cholesterol	6/1 mmol/L	232/39 mg/dL (LDL estimated at ~165 mg/dL by program)
Systolic blood pressure (mm Hg)	160	160 (diastolic unknown)
Glycosylated hemoglobin (%)	9	9
Smoker	Yes	Yes
Complications or pre-existing conditions	None	None
Other factors	Risk factors at diagnosis equal to current levels	No medications; asymptomatic with unknown blood glucose level at diagnosis; sedentary; no family history of diabetes; started smoking at age 18; sedentary; sees doctor at least twice a year for checkups but no regular foot or eye exams
Varied		
Cholesterol	Total/HDL: 4/1 mmol/L	LDL: 96 mg/dL
Systolic blood pressure (mm Hg)	130	130
Glycosylated hemoglobin (%)	7	7
Smoker	Ex-smoker	Stopped
Weight	80 kg	176 lb
Race/ethnicity	Afro-Caribbean, Asian-Indian	Black, Asian, Hispanic

UKPDS, United Kingdom Prospective Diabetes Study; PHD, personal health decisions; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

dom samples from the data, allowing an observation to be sampled as many times as it is randomly chosen and examining the distribution of predictions obtained to estimate variance around the mean).<sup>11</sup> We estimated CIs for the UKPDS outcomes model estimates with 999 bootstraps and evaluated whether the Diabetes PHD estimates fell within these limits as one measure of agreement between the models.

We tabulated 10- and 20-year outcome probability estimates from each system for myocardial infarction (MI), stroke, limb amputation, blindness in one eye, and renal failure. The Diabetes PHD website produces graphical output, not numerical tables. However, hovering the mouse over a spot on the plots displays probabilities that change with each year when following the curve. We used this approach to obtain 10- and 20-year risk estimates.

## Results

Table 2 shows the 10- and 20-year risk estimates obtained from both models. For MI risk, several patterns were apparent. Diabetes PHD produced higher risk estimates for MI than the UKPDS outcomes model, with the differences larger at 20 years than 10 years. The point estimate changes for changes in different risk factors did not always parallel, eg, the UKPDS model estimated similar risk reductions for lowering all risk factors other than weight, which it projected as having almost no effect, whereas the Diabetes PHD model estimated the greatest reduction coming from smoking cessation and weight loss as yielding similar benefits to other risk factor reduction except at 20 years for a white man. The UKPDS model predicted substantially lower MI risks for women than men, whereas the Diabetes PHD male-female differences were much more modest. Relative to the white male base case, the UKPDS model predicted a substantially lower MI risk for the Afro-Caribbean man, reflecting an observed finding of the UKPDS trial.<sup>12</sup> However, for the black man, the Diabetes PHD model predicted a similar MI risk. The Diabetes PHD point estimates fell outside the UKPDS outcomes model 95% CIs for all cases except the 10-year MI risks for white men who quit smoking or lost weight. The different model predictions are perhaps more easily appreciated in Figure 1, where we plotted the 10- and 20-year MI risk data for the white female cases, including the UKPDS boot-

strap 95% CIs. In all but 2 cases, the 10-year Diabetes PHD predictions exceeded the 20-year UKPDS predictions.

For stroke risk, again the Diabetes PHD model always produced higher risk estimates than the UKPDS outcomes model. For all of the female cases and the 20-year estimates for men, the Diabetes PHD estimates were more than twice the UKPDS outcomes model estimates—in some cases, 3 to 4 times larger. Only the 10-year Diabetes PHD risk estimate for a white man who lost weight fell within UKPDS 95% CIs. For stroke, both models agreed that relative risk reduction was greatest for lowering blood pressure, followed closely by smoking cessation and weight loss.

In contrast, the UKPDS outcomes model always predicted at least 2-fold higher risks of amputation and blindness than did the Diabetes PHD model. Absolute risks for amputation and blindness were much smaller than for MI and stroke, ranging from 0% to 6%. Except for the 20-year risk for the white female base case, all of the Diabetes PHD point estimates for amputation in the female cases fell below the UKPDS 95% CIs, whereas the Diabetes PHD estimates were within the UKPDS CIs for a number of the male cases. For blindness, in all cases the Diabetes PHD estimates fell below the UKPDS 95% CIs.

Prediction of renal failure presented a more complex pattern of differences. The UKPDS model predicted higher 10-year rates for men and women than did the Diabetes PHD model, but the highest predicted risk for any case was 3%, whereas the Diabetes PHD model predicted a 0% risk for many cases. Except for lowering blood pressure in the white female case, all of the Diabetes PHD estimates fell below the UKPDS 95% CIs. In contrast, the Diabetes PHD model generally predicted notably higher 20-year rates except in the cases of lowering glycosylated hemoglobin (HbA1c) to 7.0% and losing 44 lb (20 kg), for which it predicted complete or almost complete elimination of the risk of renal failure; the UKPDS outcomes model estimated almost no risk reduction. The Diabetes PHD estimates fell within the UKPDS 95% CIs for the male and female cases with lowered blood pressure and weight loss, plus the female case with lowered HbA1c, all cases for which the Diabetes PHD model predicted greatly reduced risks.

**Table 2. Comparison of Predicted Probabilities from the United Kingdom Prospective Diabetes Study Outcomes Model and Diabetes Personal Health Decision Model**

Case	Estimated Probability of Outcome Over the Subsequent 10 Years																	
	MI			Stroke			Amputation			Blindness			Renal failure					
	UKPDS (95% CI)	PHD	PHD	UKPDS (95% CI)	PHD	PHD	UKPDS (95% CI)	PHD	PHD	UKPDS (95% CI)	PHD	PHD	UKPDS (95% CI)	PHD	PHD			
<b>Male</b>																		
White base case	0.35 (0.28-0.42)	*	0.43	0.12 (0.06-0.17)	*	0.21	0.03 (0.01-0.05)	>	*	0.01	0.03 (0.02-0.04)	>	*	0.01	0.02 (0.01-0.04)	>	*	0.00
Lower cholesterol	0.25 (0.19-0.30)	*	0.37	0.11 (0.06-0.16)	*	0.21	0.03 (0.01-0.05)	>	*	0.00	0.03 (0.02-0.04)	>	*	0.01	0.02 (0.00-0.04)	>	*	0.00
Lower BP	0.28 (0.23-0.33)	*	0.38	0.06 (0.04-0.09)	<	0.13	0.02 (0.01-0.03)	>	*	0.01	0.03 (0.02-0.04)	>	*	0.01	0.01 (0.00-0.02)	>	*	0.00
Lower HbA1c	0.28 (0.22-0.34)	*	0.41	0.10 (0.05-0.14)	*	0.18	0.01 (0.00-0.02)	>	*	0.01	0.02 (0.01-0.02)	>	*	0.01	0.02 (0.00-0.04)	>	*	0.00
Quit smoking	0.27 (0.22-0.32)	*	0.29	0.09 (0.05-0.13)	*	0.13	0.03 (0.01-0.05)	>	*	0.01	0.03 (0.02-0.04)	>	*	0.01	0.02 (0.01-0.04)	>	*	0.00
Loss weight	0.34 (0.28-0.41)	*	0.35	0.12 (0.06-0.17)	*	0.14	0.03 (0.01-0.05)	>	*	0.01	0.03 (0.02-0.04)	>	*	0.01	0.02 (0.01-0.04)	>	*	0.00
Afro-Caribbean/black†	0.12 (0.04-0.19)	<	0.41	0.13 (0.07-0.19)	*	0.20	0.03 (0.01-0.05)	>	*	0.01	0.03 (0.02-0.04)	>	*	0.00	0.02 (0.01-0.04)	>	*	0.00
Asian-Indian/Asian†	0.35 (0.28-0.41)	*	0.45	0.12 (0.06-0.17)	*	0.20	0.03 (0.01-0.04)	>	*	0.01	0.03 (0.02-0.04)	>	*	0.00	0.02 (0.00-0.04)	>	*	0.01
Latino	—		0.42	—		0.21	—			0.01	—			0.01	—			0.01
<b>Female</b>																		
White base case	0.18 (0.13-0.23)	*	0.26	0.08 (0.04-0.12)	<	0.20	0.03 (0.01-0.05)	>	*	0.01	0.03 (0.02-0.04)	>	*	0.00	0.02 (0.01-0.04)	>	*	0.00
Lower cholesterol	0.11 (0.08-0.15)	<	0.28	0.06 (0.03-0.10)	<	0.20	0.03 (0.01-0.05)	>	*	0.00	0.03 (0.02-0.04)	>	*	0.00	0.03 (0.01-0.04)	>	*	0.00
Lower BP	0.13 (0.09-0.16)	*	0.26	0.04 (0.02-0.06)	<	0.1	0.02 (0.01-0.03)	>	*	0.00	0.03 (0.02-0.03)	>	*	0.00	0.01 (0.00-0.02)	>	*	0.00
Lower HbA1c	0.14 (0.10-0.18)	<	0.31	0.06 (0.03-0.10)	<	0.15	0.01 (0.01-0.02)	>	*	0.00	0.02 (0.01-0.03)	>	*	0.00	0.02 (0.01-0.04)	>	*	0.00
Quit smoking	0.14 (0.10-0.17)	*	0.23	0.06 (0.04-0.09)	<	0.12	0.03 (0.01-0.05)	>	*	0.00	0.03 (0.02-0.04)	>	*	0.00	0.02 (0.00-0.04)	>	*	0.00
Loss weight	0.17 (0.13-0.22)	*	0.26	0.08 (0.04-0.12)	<	0.13	0.03 (0.01-0.05)	>	*	0.01	0.03 (0.02-0.04)	>	*	0.00	0.02 (0.00-0.04)	>	*	0.00
Afro-Caribbean/black†	0.05 (0.01-0.08)	<	0.30	0.09 (0.04-0.13)	<	0.17	0.03 (0.01-0.05)	>	*	0.01	0.03 (0.02-0.04)	>	*	0.01	0.02 (0.01-0.04)	>	*	0.00
Asian-Indian/Asian†	0.17 (0.13-0.22)	*	0.34	0.08 (0.04-0.12)	<	0.22	0.03 (0.01-0.05)	>	*	0.00	0.03 (0.02-0.04)	>	*	0.01	0.02 (0.01-0.04)	>	*	0.01
Latino	—		0.33	—		0.16	—			0.01	—			0.01	—			0.01
<b>Estimated Probability of Outcome Over the Subsequent 20 Years</b>																		
<b>Male</b>																		
White base case	0.52 (0.43-0.61)	*	0.70	0.19 (0.10-0.27)	<	0.46	0.05 (0.01-0.08)	>	*	0.02	0.04 (0.02-0.06)	>	*	0.02	0.04 (0.00-0.08)	>	*	0.14
Lower cholesterol	0.38 (0.30-0.47)	*	0.66	0.18 (0.09-0.26)	<	0.49	0.05 (0.01-0.09)	>	*	0.01	0.05 (0.03-0.06)	>	*	0.01	0.04 (0.00-0.08)	>	*	0.12
Lower BP	0.43 (0.35-0.51)	*	0.65	0.10 (0.06-0.15)	<	0.35	0.03 (0.01-0.05)	>	*	0.02	0.04 (0.02-0.06)	>	*	0.01	0.02 (0.00-0.03)	>	*	0.03
Lower HbA1c	0.44 (0.35-0.52)	*	0.69	0.16 (0.09-0.24)	<	0.44	0.02 (0.00-0.04)	>	*	0.01	0.03 (0.01-0.04)	>	*	0.01	0.04 (0.00-0.08)	>	*	0.00
Quit smoking	0.44 (0.37-0.52)	*	0.54	0.16 (0.09-0.22)	<	0.36	0.05 (0.02-0.09)	>	*	0.01	0.05 (0.03-0.07)	>	*	0.02	0.05 (0.00-0.09)	>	*	0.14
Loss weight	0.51 (0.42-0.60)	*	0.66	0.18 (0.10-0.27)	<	0.38	0.05 (0.01-0.08)	>	*	0.01	0.04 (0.02-0.06)	>	*	0.02	0.04 (0.00-0.08)	>	*	0.14
Afro-Caribbean/black†	0.21 (0.09-0.33)	<	0.67	0.22 (0.12-0.32)	<	0.45	0.05 (0.01-0.09)	>	*	0.02	0.05 (0.03-0.07)	>	*	0.01	0.04 (0.00-0.09)	>	*	0.14
Asian-Indian/Asian†	0.52 (0.43-0.61)	*	0.71	0.19 (0.10-0.28)	<	0.48	0.04 (0.01-0.08)	>	*	0.02	0.04 (0.02-0.06)	>	*	0.01	0.04 (0.00-0.08)	>	*	0.24
Latino	—		0.72	—		0.48	—			0.02	—			0.02	—			0.12
<b>Female</b>																		
White base case	0.29 (0.21-0.37)	*	0.60	0.14 (0.07-0.21)	<	0.46	0.05 (0.01-0.09)	>	*	0.02	0.05 (0.03-0.07)	>	*	0.00	0.04 (0.00-0.08)	>	*	0.17
Lower cholesterol	0.19 (0.13-0.25)	<	0.52	0.12 (0.05-0.18)	<	0.46	0.05 (0.01-0.09)	>	*	0.00	0.05 (0.03-0.07)	>	*	0.00	0.05 (0.00-0.09)	>	*	0.14
Lower BP	0.22 (0.16-0.28)	<	0.49	0.07 (0.03-0.10)	<	0.31	0.03 (0.01-0.05)	>	*	0.00	0.05 (0.03-0.07)	>	*	0.00	0.01 (0.00-0.04)	>	*	0.03
Lower HbA1c	0.23 (0.16-0.30)	<	0.56	0.11 (0.05-0.17)	<	0.39	0.02 (0.01-0.04)	>	*	0.00	0.03 (0.02-0.04)	>	*	0.00	0.05 (0.00-0.09)	>	*	0.15
Quit smoking	0.24 (0.18-0.30)	*	0.44	0.11 (0.07-0.16)	<	0.35	0.06 (0.02-0.10)	>	*	0.01	0.05 (0.03-0.07)	>	*	0.00	0.04 (0.00-0.09)	>	*	0.00
Loss weight	0.28 (0.21-0.36)	*	0.48	0.13 (0.06-0.20)	<	0.38	0.05 (0.02-0.09)	>	*	0.01	0.04 (0.02-0.06)	>	*	0.00	0.04 (0.00-0.08)	>	*	0.00
Afro-Caribbean/black†	0.09 (0.03-0.16)	<	0.56	0.15 (0.07-0.23)	<	0.44	0.05 (0.01-0.09)	>	*	0.02	0.05 (0.03-0.07)	>	*	0.02	0.05 (0.00-0.09)	>	*	0.12
Asian-Indian/Asian†	0.28 (0.21-0.36)	<	0.62	0.14 (0.07-0.21)	<	0.52	0.05 (0.02-0.09)	>	*	0.00	0.05 (0.03-0.07)	>	*	0.02	0.05 (0.00-0.09)	>	*	0.31
Latino	—		0.61	—		0.47	—			0.02	—			0.02	—			0.17

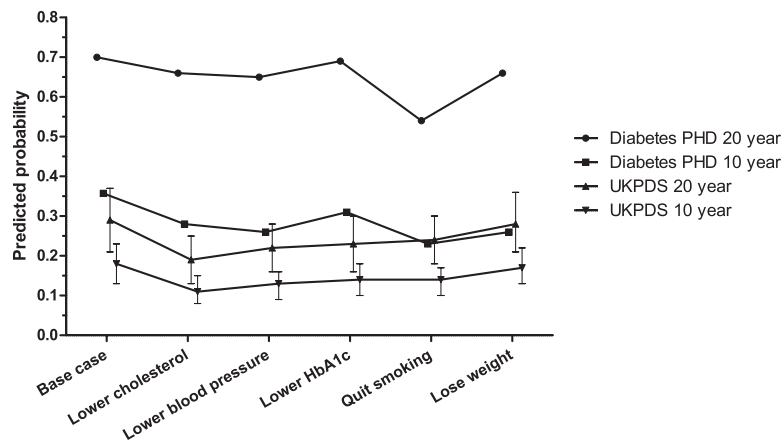
Differences between smaller and larger predictions of at least 100% are flagged with < or > to show the direction of the difference. Diabetes PHD predictions were obtained on April 10, 2010 and July 8, 2010.

\*Diabetes PHD prediction lies outside the 95% CI produced by the UKPDS outcomes model.

†First labels (Afro-Caribbean and Asian-Indian) apply to UKPDS columns and second labels (black and Asian) apply to Diabetes PHD columns. Note that these populations are different, but persons in the United States or the United Kingdom attempting to use the model developed in the other country might be tempted to use the comparison group under the assumption that they should be comparable.

UKPDS, United Kingdom Prospective Diabetes Study outcome model; PHD, Personal Health Decision outcome model; MI, myocardial infarction; BP, blood pressure; HbA1c, glycosylated hemoglobin.

**Figure 1. Ten- and 20-year United Kingdom Prospective Diabetes Study (UKPDS) outcome model and Diabetes Personal Health Decisions (PHD) outcome model predictions of myocardial infarction for white women.**



## Discussion

We found that these 2 sophisticated, well-known models produced moderately to extremely divergent risk and risk reduction estimates. Qualitatively, the risk estimates were generally consistent with current evidence and guidelines; cardiovascular risks substantially exceeded other risks and received substantial benefit from better cardiovascular risk control. Modeling of some risk factors (eg, weight, HbA1c) and outcomes (renal failure) clearly differed between the 2 models, both qualitatively as well as quantitatively.

One might argue that qualitative consistency with existing evidence is good enough to justify presenting such predictions to patients, but the spurious precision of such numerical estimates is likely to lead to a false belief in their accuracy. The models offer conflicting suggestions about the benefits of weight reduction. Neither attempts to quantify effects of changes in diet and exercise, 2 of the factors most directly under patients' control, other than that they are reflected in changes in the measured risk factors. Some of the changes, such as the dramatic reductions in the risk of renal failure predicted by the Diabetes PHD model for lowering HbA1c and losing weight, are probably too speculative to present to patients. However, a recent study of general practitioners reported that they felt computerized models producing estimates for multiple outcomes would be helpful<sup>13</sup> and, in some settings, use of predictive models is already a standard practice, such as in the US National Cholesterol Education Program's ATP-III guidelines<sup>14-16</sup> and the US Preventive Service Task Force's rec-

ommendations on the use of aspirin for primary prevention.<sup>17</sup> Both of these use the Framingham model even though it has been shown to be a poor estimator of cardiovascular risk in a variety of populations.<sup>18-22</sup>

Our comparisons of these 2 predictive models do not let us say whether one is better than another; we have no gold standard providing "correct" probability estimates. We did not seek to map out agreement for a wide range of all factors; we merely hoped to compare the models for a limited set of plausible cases. We started with base cases having suboptimal risk factor control, if risk factors are all at target levels, because one would not use such risk calculators to discuss potential benefits of risk factor reduction with patients. The UKPDS model produces a random number seed when generating projections, and the Diabetes PHD model estimates vary modestly between runs, but varying these did not appreciably affect our findings. We do not know for certain that the 2 models defined MI and stroke exactly the same way, though minor differences in definition could not account for the magnitude of the differences in risk predictions. The creators of the UKPDS outcomes model make clear that one cannot assume it is valid for populations other than those included in the UKPDS trial, particularly different racial/ethnic groups.<sup>11</sup> Although this is an issue for comparing predictions for black versus Afro-Caribbean persons or Asian-Indian versus Asian-American persons in the United States, it is not obvious why predictions should differ widely for UK versus US non-His-

panic white men and women, whether in terms of baseline risks or effects of risk reduction activities.

The UKPDS outcomes model is based on observational data from 3642 participants in the UKPDS and has been shown to predict the population incidence of the modeled outcomes in the UKPDS participants accurately.<sup>8</sup> The Diabetes PHD is based on a sophisticated model that has shown an excellent ability to predict the aggregate outcomes of a number of large randomized, controlled trials,<sup>23</sup> indicating good “calibration-in-the-large”<sup>24</sup> for those situations. Both models have evidence of reasonable calibration-in-the-large when predicting MI and stroke and response to lipid lowering with a statin,<sup>25</sup> but information about discrimination was not provided in that publication. However, calibration-in-the-large does not mean a model is appropriate for individual predictions. Lemeshow and colleagues<sup>26</sup> pointed out some years ago that several models for predicting intensive care unit mortality had very similar discrimination and calibration yet produced quite divergent risk estimates for many individuals, and they suggested that the limitations of the models were too great to be useful for individual decision making. Stern<sup>27</sup> recently summarized a number of studies comparing predictive models and found the same issues of poor agreement among equally valid models for individual-level predictions.

We have found no published studies reporting external validation of individual-level predictions of either of these models for any outcomes. All studies of generalizability have examined aggregate predictions. The Archimedes diabetes model was developed using data from studies including the UKPDS and specifically used UKPDS data in modeling retinopathy and nephropathy; it has been shown to model the annual proportion of persons in the UKPDS experiencing an MI very accurately, as well as accurately predicting the proportion of UKPDS participants experiencing retinopathy and nephropathy; however, this study provided no data on the accuracy of prediction at the individual level.<sup>28</sup> The Archimedes model showed reasonable discrimination in predicting the development of diabetes in one validation study,<sup>29</sup> though this is a different outcome than we evaluated here. The Framingham equation and a variety of other cardiovascular risk models have been shown to perform poorly for persons with diabetes,<sup>18</sup> and, in general, the performance of the Framingham equa-

tion has been found to vary widely among different populations.<sup>30</sup> In a prospective cohort study of patients with newly diagnosed type 2 diabetes in the United Kingdom, the UKPDS risk engine, which is less sophisticated than the outcomes model, showed modest discrimination and poor calibration when predicting coronary heart disease (CHD) events.<sup>19</sup> This was also the case for Chinese<sup>31</sup> and Dutch/German<sup>32</sup> cohorts. A study among Australian diabetics found the UKPDS risk engine had good calibration and discrimination for predicting stroke but performed poorly for coronary heart disease, whereas the Framingham equation did poorly for both outcomes.<sup>33</sup> Berger et al,<sup>34</sup> in a review and comparison of 6 different cardiovascular risk calculators, noted that optimal decision making would require both good risk assessment and good estimates of risks of treatment side-effects, which are often unavailable. Mohan et al<sup>35</sup> pointed out the differing cardiovascular risk estimates for 3 test cases in relation to the US Preventive Services Task Force’s guideline on the use of aspirin for primary prevention.

A recent review by Sheridan et al<sup>36</sup> found evidence for a small reduction in estimated CHD risk from repeated presentation of risk information to patients. However, whether the actual risk estimates or simply repeatedly bringing up CHD risk may lead to this surrogate outcome is not known. For example, a recent study found that presenting spirometric “lung age” to subjects increased smoking cessation, regardless of whether the test revealed impaired function or not.<sup>37</sup> Two studies published subsequent to the Sheridan review found no improvement in risk factor control from interventions providing information about personal risks and benefits to diabetic patients.<sup>38,39</sup>

## Conclusions

Our findings and those of others indicating frequently poor performance and agreement by a variety of predictive models raise the question of whether such predictions are accurate enough to justify a prominent role in treatment guidelines and decision making by patients and providers. At the population level, it may be appropriate to suggest a treatment based on some level of risk. At the individual level, such a wide range of uncertainty means that many persons whose estimated risks are near any cutoff will be misclassified. Mathematically,

individual-level predictions must have a much greater degree of uncertainty than population means. Others have outlined methods for assessing predictive model performance, including discrimination, calibration, generalizability, and estimation of net benefit.<sup>40,41</sup> Until a model has been shown to have sufficient individual-level predictive accuracy for members of the population under consideration, we believe it should be used circumspectly, with the risk projections treated as “ballpark estimates”—perhaps appropriate as one factor considered by patients and providers making informed, patient-centered treatment choices, but not appropriate for guidelines and quality measures to use with specific numerical cutoffs.

## References

- Centers for Disease Control and Prevention. Number (in millions) of civilian/noninstitutionalized persons with diagnosed diabetes, United States, 1980–2009. 10 March 2011. Available at: <http://www.cdc.gov/diabetes/statistics/prev/national/figpersons.htm>. Accessed 18 May 2011.
- Brown LC, Johnson JA, Majumdar SR, Tsuyuki RT, McAlister FA. Evidence of suboptimal management of cardiovascular risk in patients with type 2 diabetes mellitus and symptomatic atherosclerosis. *CMAJ* 2004;171(10):1189–92.
- Grant RW, Cagliero E, Murphy-Sheehy P, Singer DE, Nathan DM, Meigs JB. Comparison of hyperglycemia, hypertension, and hypercholesterolemia management in patients with type 2 diabetes. *Am J Med* 2002;112(8):603–9.
- UK Prospective Diabetes Study Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet*. 1998;352(9131):837–53.
- UK Prospective Diabetes Study Group. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. *BMJ* 1998;317(7160):703–13.
- Adler AI, Stratton IM, Neil HA, et al. Association of systolic blood pressure with macrovascular and microvascular complications of type 2 diabetes (UKPDS 36): prospective observational study. *BMJ* 2000;321(7258):412–9.
- Hansson L, Zanchetti A, Carruthers SG, et al. Effects of intensive blood-pressure lowering and low-dose aspirin in patients with hypertension: principal results of the Hypertension Optimal Treatment (HOT) randomised trial. HOT Study Group. *Lancet* 1998;351(9118):1755–62.
- Clarke PM, Gray AM, Briggs A, et al. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia* 2001;47(10):1747–59.
- American Diabetes Association. Diabetes PHD. Available at: <http://www.diabetes.org/diabetesphd/default.jsp>. Accessed 12 April 2010.
- Eddy DM, Schlessinger L. Archimedes: a trial-validated model of diabetes. *Diabetes Care* 2003;26(11):3093–101.
- University of Oxford Diabetes Trials Unit, Health Economics Research Centre. UKPDS Outcomes Model User Manual, Version 1.3. Available at: <http://www.dtu.ox.ac.uk/outcomesmodel/UKPDSOutcomesManual.pdf>. Accessed 18 May 2011.
- Stevens RJ, Kothari V, Adler AI, Stratton IM. The UKPDS risk engine: a model for the risk of coronary heart disease in type II diabetes (UKPDS 56). *Clin Sci (Lond)* 2001;101(6):671–9.
- Muller-Riemenschneider F, Holmberg C, Rieckmann N, et al. Barriers to routine risk-score use for healthy primary care patients: survey and qualitative study. *Arch Intern Med* 2010;170(8):719–24.
- Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *JAMA* 2001;285(19):2486–97.
- National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* 2002;106(25):3143–421.
- Grundy SM, Cleeman JI, Merz CN, et al. Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. *Circulation* 2004;110(2):227–39.
- US Preventive Services Task Force. Aspirin for the prevention of cardiovascular disease: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2009;150(6):396–404.
- Coleman RL, Stevens RJ, Retnakaran R, Holman RR. Framingham, SCORE, and DECODE risk equations do not provide reliable cardiovascular risk estimates in type 2 diabetes. *Diabetes Care* 2007;30(5):1292–3.
- Guzder RN, Gatling W, Mullee MA, Mehta RL, Byrne CD. Prognostic value of the Framingham cardiovascular risk equation and the UKPDS risk engine for coronary heart disease in newly diagnosed type 2 diabetes: results from a United Kingdom study. *Diabet Med* 2005;22(5):554–62.
- Fiscella K, Tancredi D, Franks P. Adding socioeconomic status to Framingham scoring to reduce dis-

- parities in coronary risk assessment. *Am Heart J* 2009;157(6):988–94.
21. Tzoulaki I, Liberopoulos G, Ioannidis JP. Assessment of claims of improved prediction beyond the Framingham risk score. *JAMA* 2009 ;302(21):2345–52.
  22. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* 2008;336(7659):1475–82.
  23. Archimedes Inc. Accuracy and validations. San Francisco, CA. Available at: <http://archimedesmodel.com/sites/default/files/Archimedes-Model-Validations-0511.pdf>. Accessed 6/8/11.
  24. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med* 2004; 23(16):2567–86.
  25. The Mount Hood 4 Modeling Group. Computer modeling of diabetes and its complications: a report on the Fourth Mount Hood Challenge Meeting. *Diabetes Care* 2007;30(6):1638–46.
  26. Lemeshow S, Klar J, Teres D. Outcome prediction for individual intensive care patients: useful, misused, or abused? *Intensive Care Med* 1995;21(9): 770–6.
  27. Stern RH. The discordance of individual risk estimates and the reference class problem. Ann Arbor, MI: CVC Cardiovascular Medicine, University of Michigan. Available at: <http://arxiv.org/pdf/1001.2499v1>. Accessed 18 May 2011.
  28. Eddy DM, Schlessinger L. Validation of the archimedes diabetes model. *Diabetes Care* 2003;26(11): 3102–10.
  29. Stern M, Williams K, Eddy D, Kahn R. Validation of prediction of diabetes by the archimedes model and comparison with other predicting models. *Diabetes Care* 2008;31(8):1670–1.
  30. Brindle P, Beswick A, Fahey T, Ebrahim S. Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: a systematic review. *Heart* 2006;92(12):1752–9.
  31. Yang X, So WY, Kong AP, et al. Development and validation of a total coronary heart disease risk score in type 2 diabetes mellitus. *Am J Cardiol* 2008; 101(5):596–601.
  32. van Dieren S, Peelen LM, Nothlings U, et al. External validation of the UK Prospective Diabetes Study (UKPDS) risk engine in patients with type 2 diabetes. *Diabetologia* 2011;54(2):264–70.
  33. Davis WA, Colagiuri S, Davis TM. Comparison of the Framingham and United Kingdom Prospective Diabetes Study cardiovascular risk equations in Australian patients with type 2 diabetes from the Fremantle Diabetes Study. *Med J Aust* 2009;190(4): 180–4.
  34. Berger JS, Jordan CO, Lloyd-Jones D, Blumenthal RS. Screening for cardiovascular risk in asymptomatic patients. *J Am Coll Cardiol* 2010;55(12):1169–77.
  35. Mohan AV, Mohan CP, Balaban R. Responses to USPSTF guideline on aspirin for prevention of cardiovascular disease. *Ann Intern Med* 2009;151(8): 587–8.
  36. Sheridan SL, Viera AJ, Krantz MJ, et al. The effect of giving global coronary risk information to adults: a systematic review. *Arch Intern Med* 2010;170(3): 230–9.
  37. Parkes G, Greenhalgh T, Griffin M, Dent R. Effect on smoking quit rate of telling patients their lung age: the Step2quit randomised controlled trial. *BMJ* 2008;336(7644):598–600.
  38. Mann DM, Ponieman D, Montori VM, Arciniega J, McGinn T. The Statin Choice decision aid in primary care: a randomized trial. *Patient Educ Couns* 2010;80(1):138–40.
  39. O'Connor PJ, Sperl-Hillen J, Johnson PE, Rush WA, Crain AL. Customized feedback to patients and providers failed to improve safety or quality of diabetes care: a randomized trial. *Diabetes Care* 2009; 32(7):1158–63.
  40. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;130(6):515–24.
  41. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21(1):128–38.