Do We Need the Oral Glucose Tolerance Test to Identify Future Cases of Type 2 Diabetes?

In our recent article (1), in which we presented a predicting model based on readily available clinical variables for identifying individuals at high risk of future type 2 diabetes, we concluded with the words: ". . . we hope that this report will stimulate other researchers with suitable databases to evaluate (similar) prediction models. . . " The first group to take up this challenge is McNeely et al. (2), whose report appears in this issue of *Diabetes Care*. The results of their analyses in Japanese Americans raise the interesting possibility that the predicting model performs better in younger than older individuals.

A major difference between the Japanese-American cohort and our San Antonio cohort was the older age range of the former (34–75 years, mean ~52 vs. 25–64, mean 43.5). One reason why 2-h glucose might play a larger role in predicting future diabetes in older subjects is that the other risk factors for diabetes, notably, lipids and blood pressure, are also risk factors for cardiovascular disease. Thus, the older the cohort, the greater the likelihood that individuals with risk factors will fail to return for follow-up glucose tolerance testing as a result of having developed either cardiovascular morbidity or mortality. It thus becomes less likely that these same risk factors will emerge as predictive of diabetes. The finding by McNeely et al. that BMI is a stronger risk factor in individuals <55 years of age supports this concept. Based on their suggestion, we have reanalyzed our data and have also observed an interaction in the same direction between age and BMI. McNeely et al. also observed an interaction between age and HDL, although we did not.

All of the above points notwithstanding, however, it must be acknowledged that the fact that only eight individuals in the study by McNeely et al. died of a cardiovascular cause, and that the cardiovascular risk factor status at baseline was no worse among those who failed to return for follow-up than among those who did return, weakens the argument that selective survival fully accounts for the differences between our two studies.

McNeely et al. have analyzed their data separately in individuals who were above and below 55 years of age at baseline. The younger group, it seems to us, constitutes the most appropriate comparison group for the San Antonio cohort, since their age range (34–55 years) is contained within the age range of the San Antonio cohort (25–64 years). It is to be expected that a predicting model will perform less well in an independent validation dataset than in the dataset in which it was originally developed, since there is always at least some tendency for models to be "overfit" to the data from which they were derived. Despite this expectation, the performance of the model, as judged by the area under the receiver operator characteristic (ROC) curve, was actually somewhat better for young Japanese Americans after 5–6 years of follow-up than for San Antonio residents (area under the ROC curve: 89.6 vs. 84.3%), although, as expected, it was slightly worse after 10 years of follow-up (area under the ROC curve: 80.7 vs. 84.3%). As expected, the areas under the ROC curves increased when McNeely et al. reestimated the model parameters using their own data (area under the ROC curve: 89.7 vs. 89.6% and 82.7 vs. 80.7% after 5–6 and 10 years of follow-up, respectively). We consider, however, that the use of the original model parameters developed using the San Antonio data provide a more rigorous test of the external validity of the model in an independent dataset.

In young Japanese Americans, the predicting model performed slightly better than 2-h glucose after 5–6 years of follow-up (area under the ROC curve: 89.6 vs. 85.1%) and slightly worse after 10 years of follow-up (area under the ROC curve: 80.7 vs. 82.9%), although

neither of these differences were statistically significant. However, similar performance by the predicting model and the 2-h glucose test should be counted as an advantage for the former, since use of the model avoids the cost and inconvenience of an oral glucose tolerance test. We have recently called attention to the fact that if one values 2 h of a person's time at the average U.S. hourly wage ($13.70) and applies this figure to the population for whom the American Diabetes Association currently recommends screening (3), the indirect cost of screening with a glucose tolerance testing exceeds $3 billion (4). Although people do not necessarily explicitly calculate the value of their time in monetary terms, they nevertheless place an implicit value on it; it is this cost, we would contend, that constitutes a principal obstacle to the widespread adoption of the oral glucose tolerance test. Of course, this analysis does not address the potential benefits of screening with oral glucose tolerance tests that could conceivably exceed $3 billion. In view of the costs, however, we would argue that it is incumbent upon the advocates of this test to demonstrate its offsetting benefits. Moreover, as we have pointed out elsewhere (4), any analysis of the benefits of the oral glucose tolerance test should take account of the fact that these benefits accrue only to those whose high risk status is not uncovered by other, simpler means. To address these issues, efforts are currently underway to develop models based on readily available clinical variables that will identify individuals with a high likelihood of having either impaired glucose tolerance (IGT) or diabetes that is diagnosable only by the 2-h glucose value, i.e., individuals in whom oral glucose tolerance testing may be warranted.

In older Japanese Americans, the 2-h glucose value appears to outperform the predicting model, as judged by the area under the ROC curve (79.2 vs. 59.9% after 5–6 years of follow-up and 79.3 vs.

72.9% after 10 years of follow-up). The superior performance of the 2-h glucose value was highly statistically significant for the 5- to 6-year follow-up ($P < 0.001$), but not statistically significant for the 10-year follow-up. Thus, it may be that the 2-h glucose value rather than the predicting model should be the test of choice for older subjects. Before accepting this conclusion, however, we feel compelled to call attention to the unusually high prevalence of IGT in the Japanese-American cohort (38.3%), along with the exceedingly high sensitivity of IGT in this cohort for identifying future cases of diabetes (80–90%). The corresponding prevalence in the San Antonio cohort is only 14.0%, and the sensitivity is 50.9%. In a report of a recent international conference in which seven major diabetes epidemiology studies were summarized, the prevalence of IGT ranged from 8.3 to 18.8% and the sensitivity from 31.5 to 62.5% (5). Thus, it would seem that the Japanese-American cohort studied by McNeely et al. was unusual, both in terms of its high prevalence of IGT and the high sensitivity of IGT for predicting future diabetes. Thus, the superiority of 2-h glucose as a predictor of future diabetes in older individuals will need to be confirmed in other populations with a more typical prevalence and sensitivity of IGT than the present Japanese-American cohort.

It is interesting that McNeely et al. chose to test the sensitivity and specificity of the San Antonio predicting model at cut points corresponding to cost-benefit ratios of 1:2 and 1:4, i.e., assuming the cost of false negatives (missed diagnoses) to be two or four times higher than the cost of false positive diagnoses. McNeely et al. do not state whether these ratios are intended to represent the discounted costs and benefits, but assuming they are, it should be noted that false negatives create only a future harm, i.e., a failure to prevent type 2 diabetes and, presumably, its complications, whereas false positive diagnoses can create a present harm, including possible adverse psychological effects as well as adverse effects on employment, medical insurance, etc. While these harms are perhaps less likely to occur with a diagnosis of IGT as opposed to diabetes itself, it should be noted that some are now urging that IGT be defined as a disease entity, which could increase the risk of harm to those who are falsely labeled with this condition. Although it has now been shown that type 2 diabetes can be prevented by behavioral and pharmacological interventions, given the difficulties of achieving this result, the future benefit in many cases may never be realized (hence, the need for discounting). Moreover, it remains to be demonstrated that this benefit can be translated into improvements in genuine health outcomes such as reduced cardiovascular disease, renal dialysis, mortality, etc., at a reasonable cost. Given these uncertainties, it is doubtful that the cost of a false negative is in fact double, let alone quadruple, the cost of a false positive. Formal cost benefit analyses, currently underway, are needed to address this issue more rigorously.

Now that McNeely et al. have broken the ice by evaluating our predicting model in their cohort, we conclude as we began with the words: ". . . we hope that this report will stimulate other researchers with suitable databases to evaluate (similar) prediction models . . . " (1).

MICHAEL P. STERN, MD
KEN WILLIAMS, MS
STEVEN M. HAFFNER, MD, MPH

From the Division of Clinical Epidemiology, Department of Medicine, University of Texas Health Science Center, San Antonio, Texas.
Address correspondence to Dr. Michael P. Stern, 7703 Floyd Curl Dr., San Antonio, TX 78229-3900. E-mail: stern@uthscsa.edu.

●●●●●●●●●●●●●●●●●●●●●●●●

References
1. Stern MP, Williams K, Haffner SM: Identification of individuals at high risk of type 2 diabetes: do we need the oral glucose tolerance test? *Ann Int Med* 136:575–581, 2002
2. McNeely MJ, Boyko EJ, Leonetti DL, Kahn SE, Fujimoto WY: Comparison of a clinical model, the oral glucose tolerance test, and fasting glucose for prediction of type 2 diabetes risk in Japanese Americans. *Diabetes Care* 26:758–763, 2003
3. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus: Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 20:1183–1197, 1997
4. Stern MP, Fatehi P, Williams K, Haffner SM: Predicting future cardiovascular disease: do we need the oral glucose tolerance test? *Diabetes Care* 25:1851–1856, 2002
5. Unwin N, Shaw J, Zimmet P, Alberti G: Impaired glucose tolerance and impaired fasting glycaemia: the current status on definition and intervention. *Diabet Med* 19:708–723, 2002