Short-term Variability in Measures of Glycemia and Implications for the Classification of Diabetes

Elizabeth Selvin, PhD, MPH; Ciprian M. Crainiceanu, PhD; Frederick L. Brancati, MD, MHS; Josef Coresh, MD, PhD

Background: Short-term variability in measures of glycemia has important implications for the diagnosis of diabetes mellitus and the conduct and interpretation of epidemiologic studies. Our objectives were to characterize the within-person variability in fasting glucose, 2-hour glucose, and hemoglobin A1c (HbA1c) levels and to assess the impact of using repeated measurements for classification of diabetes.

Methods: We analyzed repeated measurements from 685 fasting participants without diagnosed diabetes from the National Health and Nutrition Examination Survey III Second Examination, a substudy conducted from 1988 to 1994 in which repeated examinations were conducted approximately 2 weeks after the original examination.

Results: Two-hour glucose levels had substantially more variability (within-person coefficient of variation [CVw], 16.7%; 95% confidence interval [CI], 15.0 to 18.3) compared with either fasting glucose (CVw, 5.7%; 95% CI, 5.3 to 6.1) or HbA1c (CVw, 3.6%; 95% CI, 3.2 to 4.0) levels. The proportion of persons with a fasting glucose level of 126 mg/dL or higher (to convert to millimoles per liter, multiply by 0.0555) on the first test who also had a second glucose level of 126 mg/dL or higher was 70.4% (95% CI, 49.8% to 86.2%). Results were similar using the 2-hour glucose cutoff point of 140 mg/dL or higher. The prevalence of undiagnosed diabetes using a single fasting glucose level of 126 mg/dL or higher was 3.7%. If a second fasting glucose level of 126 mg/dL or higher was used to confirm the diagnosis (American Diabetes Association guidelines), the prevalence decreased to 2.8% (95% CI, 1.5% to 4.0%), a 24.4% decrease.

Conclusions: We found high variability in 2-hour glucose levels relative to fasting glucose levels and high variability in both of these relative to HbA1c levels. Our findings suggest that studies that strictly apply guidelines for the diagnosis of diabetes (2 glucose measurements) may arrive at substantially different prevalence estimates compared with studies that use only a single measurement.

Arch Intern Med. 2007;167(14):1545-1551

©2007 American Medical Association. All rights reserved.
8 hours at either the first or repeated NHANES III examination (n = 1394), those who reported a prior physician diagnosis of diabetes (n = 53), and those who were missing plasma glucose or HbA1c values (n = 28). After these exclusions, 685 participants remained for analysis.

Oral glucose tolerance tests (OGTTs) were performed only for persons aged 40 to 74 years (n = 374). A random assignment was made before conducting the OGTT to determine who should receive a morning examination (after an overnight fast). The subsample of persons who received a morning examination most closely conforms to the World Health Organization criteria for OGTTs to identify diabetes. Thus, in addition to these exclusions, persons who had afternoon or evening examinations were excluded from OGTT analyses (n = 35). Twenty-seven individuals were also missing or had incomplete 2-hour glucose measurements. All analyses of 2-hour glucose measurements are limited to this smaller sample of individuals with valid OGTT data (n = 312).

**MEASUREMENT OF FASTING AND 2-HOUR GLUCOSE AND HbA1c LEVELS**

The NHANES III Second Examination was conducted approximately 2 weeks after the first examination by trained personnel following the same standardized protocols. The NHANES III examinations included a 2-hour 75-g OGTT in adults 40 to 75 years old. After the fasting blood specimen was obtained by venipuncture, participants received a 75-g glucose drink, and a second venipuncture was performed approximately 2 hours after the first blood collection. Glucose was measured at a central laboratory using a hexokinase assay (Cobas Mira; Roche, Basel, Switzerland). The analytic (method) coefficients of variation (CVs) for this assay ranged from 1.6% to 3.7%. The HbA1c measurements were performed in the same laboratory using the Diamat high-performance liquid chromatography assay (Bio-Rad Laboratories, Hercules, California) standardized to the Diabetes Control and Complications Trial reference method. The analytic CVs for this assay ranged from 1.1% to 3.1%. Because variant hemoglobins can interfere with HbA1c measurement by the Diamat assay, samples suspected of containing abnormal hemoglobin levels were analyzed using affinity chromatography. Detailed information on data collection and laboratory procedures in the NHANES III are available elsewhere.

**DEFINITIONS OF DIABETES AND IMPAIRED GLYCEMIC STATES**

To assess the impact of measurement variability on the classification of diabetes and impaired glycemic states, we applied current diagnostic definitions to these data. We compared estimates using 1, both, and the mean of the measurements from the 2 examinations for the diagnosis of diabetes. That is, we estimated the prevalence of undiagnosed diabetes using a fasting glucose level of 126 mg/dL or higher (to convert to millimoles per liter, multiply by 0.0555) and compared prevalence estimates based on a fasting glucose level of 126 mg/dL or higher (1) at examination 1, (2) at examination 2, (3) at both examination 1 and examination 2, (4) at either examination 1 or examination 2, and (5) as determined by the mean of examinations 1 and 2 (ie, measurement at examination 1 plus measurement at examination 2 divided by 2). We also examined the reliability of current cutoff points for undiagnosed diabetes, impaired fasting glucose, and impaired glucose tolerance by examining the proportion of persons who had abnormal test results on the first test who also had abnormal test results on the second test using the current clinically relevant cutoff points of 100 mg/dL or higher, 126 mg/dL, and 200 mg/dL for fasting glucose and 100 mg/dL, 140 mg/dL, and 200 mg/dL for 2-hour glucose.

---

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Examinations 1, Mean (SD)</th>
<th>Examinations 2, Mean (SD)</th>
<th>Mean Difference Between Examinations 1 and 2, Mean (SD)</th>
<th>95% Limits of Agreementa CVw (%) (95% CI)b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasting glucose, mg/dL</td>
<td>98.3 (20.6)</td>
<td>98.3 (20.4)</td>
<td>-0.03 (8.3)</td>
<td>-16.2 to 16.2</td>
</tr>
<tr>
<td>2-h Glucose, mg/dL</td>
<td>132.3 (66.6)</td>
<td>130.5 (66.7)</td>
<td>1.86 (29.4)</td>
<td>-55.8 to 59.5</td>
</tr>
<tr>
<td>HbA1c, %</td>
<td>5.4 (0.8)</td>
<td>5.4 (0.8)</td>
<td>0.03 (0.3)</td>
<td>-0.51 to 0.57</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; CVw, within-person coefficient of variation; HbA1c, hemoglobin A1c.

a Mean ± 1.96 × SD.

b The CVw was calculated using the root mean square approach.
STATISTICAL ANALYSES

We compared fasting glucose, 2-hour glucose, and HbA1c values obtained during the first NHANES III examination with those obtained at the second. Differences were calculated as examination 1 minus examination 2, the within-person coefficients of variation (CVw, s) were calculated as the square root of the within-subject variance divided by the mean squared, and their confidence intervals (CIs) were obtained using bootstrap methods. Scatterplots were used to graphically present the association among fasting glucose, 2-hour glucose, and HbA1c levels at the 2 examinations and the covariation of these measurements. Plots developed by Bland and Altman were generated to display the difference (examination 1 minus examination 2) against the mean of the measurements from the 2 examinations. We calculated prevalence according to different criteria for the diagnosis of diabetes and their corresponding 95% CIs. We also compared the absolute and percentage differences in prevalence, depending on the criteria imposed (eg, single measurement, 2 measurements, or mean of 2 measurements). The 95% CIs for the differences in prevalence were computed for the paired observations using nonparametric bootstrap methods. Statistical analyses were conducted using Stata statistical software, version 8.2 (Stata Corp, College Station, Texas) and R (http://www.r-project.org/).

RESULTS

STUDY POPULATION

Table 1 lists the selected characteristics of the overall study population and separately for the subset of persons with valid OGTT data (aged 40-75 years). The mean±SD number of days between the first and second examinations was 17±8, and the mean fasting time was approximately 14 hours at both visits. The study sample was approximately 50% male and racially and ethnically diverse. Among those persons who underwent the OGTT, the mean±SD interval between the glucose drink and the second blood measurement was 120±7 minutes, as per the NHANES study protocol.

COMPARISON OF SHORT-TERM VARIABILITY IN FASTING AND 2-HOUR GLUCOSE AND HbA1c LEVELS

Table 2 gives the summary statistics for fasting glucose, 2-hour glucose, and HbA1c levels at each visit, the differences between visits (examination 1 minus examination 2), and the CVw, s. The mean of the differences was small for all measures, but the CVw, s were substantially different, with nonoverlapping CIs. Two-hour glucose levels had substantially more variability (CVw, 16.7%; 95% CI, 15.0 to 18.3) compared with either fasting glucose (CVw, 5.7%; 95% CI, 5.3% to 6.1%) or HbA1c (CVw, 3.6%; 95% CI, 3.2% to 4.0%). Figure 1 displays the correlation matrix and corresponding Pearson correlation coefficients (r) for the first and second measurements for each measure. The relation between fasting and 2-hour glucose levels was strong (r=0.80 at examination 1 and examination 2) but appeared to have a steeper association at higher values (non-
Figure 2 displays the scatterplots of the first and second fasting glucose, 2-hour glucose, and HbA1c measurements on separate plots with gridlines at clinically relevant cutoff points. The Pearson correlations for the scatterplots are indicated in the lower right-hand corners (note that the correlations for fasting glucose and HbA1c levels in this figure are slightly different than those presented in the correlation matrix because the matrix is made up of only those persons with complete OGTT results). The diagonals formed by the grid show the number of persons who tested positive on both tests using cutoff points of 100, 126, and 200 mg/dL for the fasting glucose level and 100, 140, and 200 mg/dL for the 2-hour glucose level. Because the HbA1c test is not currently recommended for use as a diagnostic test for diabetes, we compared cutoff points of 6.1%, 6.5%, and 7.0%, which were shown to be clinically relevant in previous studies.\(^8\,^9\) The proportion of persons with a fasting glucose level of 100 mg/dL or higher on the first test who also had a second fasting glucose level of 100 mg/dL or higher was 78.0% (95% CI, 71.8% to 83.4%). The comparable proportions for a fasting glucose level of 126 mg/dL or higher and 200 mg/dL were 70.4% (95% CI, 49.8% to 86.2%) and 100.0% (95% CI, 63.1% to 100.0%), respectively. However, in this general population, the number of persons with 2 fasting glucose measurements of 200 mg/dL or higher was small (n=8). By contrast, the proportion of persons with a 2-hour glucose level of 100 mg/dL or higher on the first test who also had a second 2-hour glucose level of 100 mg/dL or higher was 82.8% (95% CI, 77.4% to 87.4%). The corresponding proportions for 2-hour glucose cutoff points of 140 and 200 mg/dL were 72.0% (95% CI, 62.1% to 80.5%) and 72.0% (95% CI, 52.8% to 87.3%).

Using cutoff points of 6.1%, 6.5%, and 7.0% for HbA1c, the proportions were 89.0% (95% CI, 80.2% to 94.9%), 83.3% (95% CI, 67.2% to 93.6%), and 100.0% (95% CI, 76.8% to 100.0%), respectively.

The differences in within-person variability (reliability) among fasting glucose, 2-hour glucose, and HbA1c levels are also evident from Bland-Altman plots (Figure 3). The plots compare the differences between the measurements (examination 1 minus examination 2) against the mean of the 2 measurements. The differences were normally distributed for all measurements. In Figure 3, the differences in fasting glucose and 2-hour glucose levels are plotted on the same scale (−100 to +100 mg/dL), showing clearly the greater variability (lower reproducibility) of the 2-hour glucose measurements. For ease of comparison, the differences in HbA1c levels were plotted on an equivalent scale (from −5% to +5%) using the premise that a difference of 100 mg/dL of fasting glucose is roughly equivalent to a 5–percentage point difference in HbA1c levels (mean plasma glucose in milligrams per deciliter = [35.6 × HbA1c] − 77.3).\(^{10}\) The horizontal lines on the Bland-Altman plots represent the mean of the differences and the corresponding 95% limits of agreement (±1.96 × SD) as presented in Table 2. The
dotted vertical gray lines indicate current cutoff points for the diagnosis of diabetes: 126 mg/dL for fasting glucose and 200 mg/dL for 2-hour glucose. A cutoff point of 6.1% for HbA1c was used for comparison. Using these definitions to define an abnormal test result, the different markers in the Bland-Altman plots indicate those individuals who had an abnormal test result at only 1 examination (plus signs), those who had an abnormal test result at both examinations (squares), and those who did not have abnormal test results at either examination (circles).

**PERSISTENCE OF DIABETES MELLITUS**

Table 3 gives the prevalence estimates of undiagnosed diabetes according to different criteria and the differences (absolute and percentage) in prevalence when using multiple tests compared with a single test only. In our study population, the prevalence of diabetes based on a single fasting glucose level of 126 mg/dL or higher was 3.9% (95% CI, 2.5% to 5.4%) at the first examination and 3.5% (95% CI, 2.1% to 4.9%) at the second examination (mean prevalence based on a single measurement, 3.7%). If a second (confirmatory) fasting glucose level of 126 mg/dL or higher (ie, examination 1 and examination 2) was used to determine a diagnosis of diabetes, the prevalence decreased to 2.8% (95% CI, 1.5% to 4.0%), a 24% decrease (95% CI, −40.5% to −10.8%).

The prevalence of undiagnosed diabetes using the 2-hour glucose values was 9% (95% CI, 5.6% to 12.2%) at both the first and second examinations. If 2 abnormal results were required, the prevalence decreased substantially to 6.7% (95% CI, 3.9% to 9.5%), a 26% decrease (95% CI, −41.1% to 16.7%). For diabetes defined using a fasting or a 2-hour glucose measurement, prevalence was increased by roughly 25% when an abnormal test result on either examination 1 or examination 2 was considered diagnostic. The mean provided a reliable estimate and did not have a large effect on changes in prevalence; the percentage change was nonsignificant for both fasting and 2-hour glucose measurement. For example, diabetes prevalence defined by fasting glucose using the mean of the values from examination 1 and examination 2 was increased to 3.9%, a nonsignificant 5.4% increase (95% CI, −8.1% to 18.9%) compared with a single measurement.

**SENSITIVITY ANALYSIS IN PERSONS WITH VALID 2-HOUR GLUCOSE MEASUREMENTS**

Because 2-hour glucose measurement was available for only a subsample of individuals, we conducted sensitivity analyses, limiting all our analyses to those individuals with valid OGTT data (n=312). In this subsample, our estimates of variability, including the CVs and percentage changes in prevalence, for fasting glucose and HbA1c measurements using single vs multiple measurements were not appreciably altered (data not shown).

Our results provide information on the reliability of a single glucose measurement (fasting or 2-hour glucose) for the classification of undiagnosed diabetes. A single fasting glucose level of 126 mg/dL or higher was a fairly reliable indicator of diabetes, with 70% repeatability.
Higher levels of glucose were extremely reliable, with an estimated 100% reliability for a fasting glucose level of 200 mg/dL or higher. In contrast, a single 2-hour glucose level was much less reliable, with substantially higher intraday individual variation during the approximately 2-week period than either fasting glucose or HbA1c levels.

The short-term intraday variation in both fasting glucose and 2-hour glucose levels has important implications for screening for diabetes and for estimating prevalence from epidemiologic data. Current guidelines from the American Diabetes Association recommend confirmation of hyperglycemia at a second visit.1 On the basis of a recent analysis of the 1999-2002 NHANES, the prevalence of undiagnosed diabetes in the general US population 20 years or older is estimated to be 2.8% (95% CI, 2.3% to 3.5%), which corresponds to 5.8 million individuals.11 Epidemiologic studies that obtain blood samples, including large cross-sectional surveys from which prevalence estimates are derived (NHANES is just 1 example), almost exclusively use a single fasting glucose measurement for the classification of diabetes and impaired glycemic states. Our results suggest that if a confirmatory glucose measurement was required 2 weeks later, the prevalence of undiagnosed diabetes in the United States would decrease by approximately 24%, from 5.8 million to 4.4 million. Our results, however, do not change the interpretation of epidemiologic studies of the prevalence of diagnosed diabetes or the numerous studies that document trends, including the epidemic increases in diabetes around the world.

The phenomenon of regression dilution, when a single measurement instead of the mean of multiple measurements across time are used to represent a variable, is well known.12-14 Short-term biological variability or imprecision in measurement (random error) will increase the prevalence of a condition when a continuous measure is used to classify individuals above or below a threshold (diabetes and glucose as well as hypertension and blood pressure are just 2 examples). This phenomenon applies broadly to situations in which persons are classified by a threshold and occurs because the mean value of a repeated measurement in the high (or low) subgroup will be closer to the mean of the original group. Single measurements in the presence of high imprecision can result in substantial misclassification of individuals. With regard to prediction of subsequent events, regression dilution will bias the association toward the null (ie, the observed association is weaker than the true association in the absence of measurement error). The use of both multiple blood pressure measurements and reporting of mean blood pressure in clinical practice, clinical trials, and other research studies is a common approach to help minimize misclassification that results from the unavoidable intraday individual variation among patients. Using multiple measurements and/or taking the mean of 2 or more glucose measurements would achieve a similar reduction in misclassification in diagnosing diabetes but is not commonly done in either clinical practice or research.

The high within-person variability in 2-hour glucose measurements has been well documented,15-18 but previous studies have not explicitly compared the reliability of a single measurement (vs 2 measurements) or quantified and compared the effect of variability in measures of glycemia, including 2-hour glucose, on the prevalence of undiagnosed diabetes. Our analysis also confirms previous studies that have shown that HbA1c tracks much better in individuals across time compared with fasting glucose measurements.19-21 The short-term variability in single glucose measurements suggests that epidemiologic studies may substantially underestimate the true association of fasting hyperglycemia with diabetic complications. Because HbA1c measurement is much less variable, we would expect stronger associations for HbA1c with clinical outcomes compared with a single fasting mea-

### Table 3. Prevalence Estimates According to Different Criteria for the Diagnosis of Diabetes and Impaired Glycemic States and Comparison With Prevalence Based on a Single Measurement

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Prevalence, % (95% CI)</th>
<th>Absolute Difference (95% CI)</th>
<th>Percentage Difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes by fasting glucose (≥126 mg/dL) (n = 685)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examination 1 only</td>
<td>3.9 (2.3 to 5.4)</td>
<td>Reference = 3.7</td>
<td>Reference = 0.0</td>
</tr>
<tr>
<td>Examination 2 only</td>
<td>3.5 (2.1 to 4.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both examination 1 and examination 2</td>
<td>2.8 (1.5 to 4.0)</td>
<td>−0.9 (−1.5 to −0.4)</td>
<td>−24.3 (−40.5 to −10.8)</td>
</tr>
<tr>
<td>Either examination 1 or examination 2</td>
<td>4.7 (3.1 to 6.3)</td>
<td>1.0 (0.5 to 1.5)</td>
<td>27.0 (13.5 to 40.5)</td>
</tr>
<tr>
<td>Mean (examination 1 and examination 2)/2</td>
<td>3.9 (2.6 to 4.6)</td>
<td>0.2 (−0.3 to 0.7)</td>
<td>5.4 (−8.1 to 18.9)</td>
</tr>
<tr>
<td>Diabetes by 2-h glucose (≥200 mg/dL) (n = 312)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examination 1 only</td>
<td>9.0 (5.6 to 12.2)</td>
<td>Reference = 9.0</td>
<td>Reference = 0.0</td>
</tr>
<tr>
<td>Examination 2 only</td>
<td>9.0 (5.6 to 12.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both examination 1 and examination 2</td>
<td>6.7 (3.9 to 9.5)</td>
<td>−2.3 (−3.5 to −1.1)</td>
<td>−25.6 (−38.9 to −12.2)</td>
</tr>
<tr>
<td>Either examination 1 or examination 2</td>
<td>11.2 (8.0 to 14.7)</td>
<td>2.2 (1.1 to 3.5)</td>
<td>24.4 (12.2 to 38.9)</td>
</tr>
<tr>
<td>Mean (examination 1 and examination 2)/2</td>
<td>8.7 (5.5 to 11.8)</td>
<td>−0.3 (−1.4 to 0.8)</td>
<td>−3.3 (−15.6 to 8.9)</td>
</tr>
</tbody>
</table>

Abbreviation: CI, confidence interval.

*SI conversion factor: To convert glucose to millimoles per liter, multiply by 0.0555.*

*Difference from the mean of the prevalence estimates based on a single glucose measurement at examination 1 and examination 2.*
measurement. Recent reports suggest HbA1c is a strong predictor of incident cardiovascular disease.22–24 Studies are needed to understand the impact of accounting for this documented short-term variability on risk factor associations in epidemiologic research.

The NHANES III Second Examination is a unique resource to examine short-term variability in laboratory measurements and one of the largest studies of its kind in a population-based setting. One strength of this study was the availability of 2-hour glucose measurements in addition to fasting glucose and HbA1c levels, which allowed us to directly compare the within-person variability for all 3 of these measurements. This analysis also benefited from the rigorous methods of the NHANES III; measurements were obtained by trained personnel using standardized protocols at both examinations. Restricting our study population to persons who were fasting at both examinations reduced the analytic sample size. However, we believed it was important to ensure that people were fasting to standardize the glucose measurements and replicate the relevant clinical setting.

This analysis documents high variability in 2-hour glucose measurements relative to fasting glucose measurements and high variability in both of these relative to HbA1c measurements. Our results also suggest that studies strictly apply clinical guidelines for the diagnosis of diabetes (2 glucose measurements) may arrive at different prevalence estimates compared with studies that use only a single measurement of glucose.

Accepted for Publication: March 14, 2007.
Correspondence: Elizabeth Selvin, PhD, MPH, Department of Epidemiology and the Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins Bloomberg School of Public Health, 2024 E Monument St, Suite 2-600, Baltimore, MD 21287 (Iselvin@jhsph.edu).

Author Contributions: Study concept and design: Selvin, Crainiceanu, Brancati, and Coresh. Acquisition of data: Selvin. Analysis and interpretation of data: Selvin, Crainiceanu, Brancati, and Coresh. Drafting of the manuscript: Selvin and Crainiceanu. Critical revision of the manuscript for important intellectual content: Selvin, Crainiceanu, Brancati, and Coresh. Statistical analysis: Selvin and Crainiceanu. Obtained funding: Selvin. Study supervision: Crainiceanu and Brancati.

Financial Disclosure: None reported.
Funding/Support: Drs Selvin (grant K01 DK076595) and Brancati (grant K24 DK62222) were supported by the National Institutes of Health/National Institute of Diabetes and Kidney Diseases.

REFERENCES


