Short- and long-term reproducibility of the 1-h 50-g glucose challenge test

GEORGE PHILLIPOV

A study was undertaken to assess the short-term (intrapregnancy) and long-term (interpregnancy) variability of the 1-h 50-g oral glucose challenge test (GCT). Two groups of pregnant women had GCTs in consecutive pregnancies, 1 (n = 77) and 2 (n = 43) years apart. Their results were compared with published results for a group (n = 53) who had GCTs on consecutive days. Robust estimates of the mean error variance (\( \sigma^2_{\text{error}} = \sigma^2_{\text{within-individual}} + \sigma^2_{\text{analytical}} \)) were calculated on log_{10}-transformed data and were for the three groups 0.003995, 0.002603, and 0.0026249 (mg/dL)^2, respectively. There was no significant difference between the group variances, establishing that the short- and long-term reliability of the GCT is comparable. \( \sigma^2_{\text{within-individual}} \) was estimated from the GCT values for 2695 pregnant women tested during the same period and was the main component (67.1%) of the total sample variation (\( \sigma^2_{\text{between-individuals}}/\sigma^2_{\text{population}} \)). Estimates of the population mean, \( \sigma^2_{\text{between-individuals}} \) and \( \sigma^2_{\text{error}} \) were used to compute the probability that an observed GCT value had a true value equal to or greater than the consensus threshold of 7.8 mmol/L (140 mg/dL).

INDEXING TERMS: glucose challenge test • error analysis • pregnancy • gestational diabetes

Screening for gestational diabetes by using a 1-h 50-g oral glucose challenge test (GCT) is a widely accepted procedure.\(^1\) Characterization of an abnormal result requires that an absolute plasma glucose threshold is achieved, ordinarily \( \geq 140 \text{ mg/dL} \) (7.8 mmol/L) \( [1] \). \(^2\) The effect of short- and long-term reproducibility of the GCT on the validity of the threshold, however, has received little attention. Because glucose intolerance is known to increase with advancing gestation \( [2] \), the GCT is standardized within the 24-28 week interval \( [1] \). Sacks et al. \( [3] \) found that, with respect to a glucose threshold of 135 mg/dL, misclassification could occur when the GCT was repeated on consecutive days, and that a value between 95 and 134 mg/dL did not necessarily exclude glucose intolerance.

Although the GCT is known to have marked between-individual variation (\( \sigma^2_{\text{w}} \)), predominantly because of differences in rates of gastrointestinal absorption, the degree of within-individual (\( \sigma^2_{\text{w}} \)) variation has not been examined. This represents a major limitation to assessing the reliability of the GCT as a screening assay for gestational diabetes.

The present study was designed to provide robust estimates of both the short- and long-term error variance \( [\sigma^2_{\text{w}}, \sigma^2_{\text{e}} = \sigma^2_{\text{w}} + \sigma^2_{\text{a}} \text{ (analytical variance)}] \) associated with the GCT for women in the same or consecutive pregnancies and to assess the implications for potential patient misclassification.

Materials and Methods

The subjects were selected from our records of 2708 individual women who, over a 3-year period, had attended the Maternity Outpatient Clinic for a routine GCT between 24 and 28 weeks of gestation. The study met the precepts of the Hospital Ethics Committee. All women who had two pregnancies either 1 or 2 years apart formed the study cohort. The first group comprised 77 women who had a GCT in each of consecutive pregnancies \( \sim 1 \text{ year apart} \). A second group of 43 women had GCTs in consecutive pregnancies \( \sim 2 \text{ years apart} \). For the third group, pregnant women who had GCTs on consecutive days, data were calculated directly from Fig. 1(b) of Sacks et al. \( [3] \), after enlarging the diagram so that the potential error in transcribing data points was minimized to \(<1 \text{ mg/dL} \); only those data pairs \( (n = 53) \) were selected where no ambiguity existed with respect to their identity. The day 1 and day 2 means for the selected subjects \( (n = 53) \) and for the full set \( (n = 80) \) were similar (111 vs 108 mg/dL for day 1 and 110 vs 107 mg/dL for day 2), suggesting that no major selection bias was introduced. Within the total group of pregnant women with two GCTs \( (n = 173) \), 9.2% had one of the two GCT values \( \geq 140 \text{ mg/dL} \), and 5.2% had both values above the threshold.

The 50-g glucose (anhydrous) load was administered to nonfasting women as a carbonated lime-flavored beverage (300

Received May 15, 1995; accepted November 7, 1995.

1 Nonstandard abbreviations: GCT, glucose challenge test; \( \sigma^2_{\text{w}} \), between-individuals variation; \( \sigma^2_{\text{w}} \), within-individual variation; \( \sigma^2_{\text{e}} \), error variance; \( \sigma^2_{\text{a}} \), analytical variance; \( \mu \), population mean; \( \sigma^2_{\text{p}} \), population variance; and OGTT, oral glucose tolerance test.

2 To convert GCT values from mg/dL to mmol/L, divide by 18.

Endocrine and Diabetes Service, The Queen Elizabeth Hospital, Woodville, South Australia 5011, Australia. Fax 61-8-222-6021; e-mail geophysill@edam.com.au.
mL). Blood for glucose estimation was collected by antecubital venipuncture into tubes containing sodium fluoride and analyzed on a Yellow Springs 23AM (Yellow Springs Instruments, Yellow Springs, OH) glucose analyzer. The total assay variability at a plasma glucose concentration of 88 mg/dL was 1.4%.

Differences between the group means were tested for statistical significance by the Mann–Whitney U-test, after checking for outliers by using Reed’s criterion [4]. The statistical significance between successive measurements for variables within groups was assessed by the paired t-test. The population mean ($\mu$) and variance ($\sigma^2_p$) were estimated from the 2708 GCT values remaining after log$_{10}$ transformation and elimination of 13 outliers, all above the 99.59 percentile, to produce a gaussian distribution of data (skewness = 0.067, kurtosis = –0.147); $\sigma^2_p = \sigma^2_b + \sigma^2_e$. The test–retest variance for the patients was calculated as one-half of the square of the difference between the successive log$_{10}$-transformed GCT values.

A robust estimate of short-term (intrapregnancy) and long-term (interpregnancy) $\sigma^2_e$ was based on the bootstrap procedure, a computationally intensive statistical method that does not require standard distribution assumptions about well-known test statistics [5, 6]. In the present work, 5000 bootstrap samples were randomly drawn (with replacement) from each of the three groups’ set of variances, such that the number of selections equaled the number of observations; the respective mean was calculated for each of these 5000 sets, and the estimated mean was taken as the median of the bootstrap cumulative distribution function. The standard error was calculated as half the central 67% interval of values in the distribution of bootstrap means. As a further extension to ensure that the estimation of the mean was not affected by extreme values, for each bootstrap sample selected, values more than $\pm 1.5$ SDs from the initially calculated mean were eliminated and a new “trimmed” mean was determined [5].

In Bayesian statistics, the true GCT value for a subject is unknown but has an assumed prior distribution with mean $\mu$ and variance $\sigma^2_p$ for a specified population. Results for a subject’s repeat tests will vary over time due to $\sigma^2_e$. This allows calculation of the respective confidence interval for a single measurement as $R \cdot x_{obs} + (1 - R) \cdot \mu \pm z \cdot \sqrt{(R \cdot \sigma^2_e)}$, where $R = \sigma^2 / \sigma^2_p$ is the reliability of the observed measurement and $z$ is the Z-statistic [7, 8]. Thus knowledge of GCT values in the given population allows the Bayesian approach to produce a tighter confidence interval than obtained by the classical approach ($x_{obs} \pm z \cdot \sqrt{\sigma^2_e}$). However (as noted by one of the reviewers of this paper), although the Bayesian approach on average is unbiased and more precise, there will be rare occasions when the original observation corresponds to the individual’s true mean and adjusting towards the population mean will result in a less accurate estimation.

**Results**

The overall findings for the three groups of pregnant women are summarized in Table 1. First, there was no significant difference, with respect to the gestational age at which GCTs were performed, for consecutive pregnancies in either group 1 or group 2 (paired $t$-test, $P > 0.6$). Second, the average gestational age at time of testing for all groups compared closely with the mean of 26.4 ± 2.2 weeks estimated for all women on record ($n = 2708$).

There was also no significant difference between the GCT means for the three groups (Mann–Whitney $U$-test, $P > 0.35$). Within each group, no statistically significant change occurred between the test–retest 1-h plasma glucose results (paired $t$-test, $P > 0.09$).

The estimated values of log$_{10}\sigma^2_e$ show that the short-term (group 3, intrapregnancy) and long-term (groups 1 and 2, interpregnancy) GCT reproducibility is comparable (Table 1); this is confirmed by the lack of significant difference of variances between the three groups (Mann–Whitney $U$-test, $P > 0.46$). O’Sullivan and Mahan [9], testing the variability of the 3-h 100-g oral glucose tolerance test (OGTT) during three consecutive pregnancies, also found that the variation due to time was minimal. Accordingly, the three groups could be combined ($n = 173$) and the overall log$_{10}\sigma^2_e$ was calculated to be $0.0033204 \pm 0.00035$. Moreover, the mean GCT value was shown to be independent of the respective variance (Spearman $r = –0.13$, $P > 0.1$).

The population mean ($\mu$) and variance for the normally distributed log$_{10}$GCT values ($n = 2695$) were calculated as 2.00710 (102 mg/dL) and 0.0100952, respectively; $\sigma^2_b$ was therefore estimated to be 0.0067748 and the coefficient of reliability ($R$) for repeat measurements 0.671. These parameters were in turn used to calculate the percentage probability that the true GCT value, for any observed value ($x_{obs}$), would be equal to or greater than the 140 mg/dL threshold by solving the equations described above for the Z-statistic (Fig. 1).

**Discussion**

The results show that the short-term (intrapregnancy) and long-term (interpregnancy) variabilities of the GCT, for the same individual, are comparable, if the test is done at about the same gestational age (Table 1). Furthermore, $\sigma^2_e$, rather than $\sigma^2_p$, is the main contributor to total population variance, as indicated by the coefficient of reliability (0.671). In contrast, O’Sullivan and Mahan [9] reported that the reliability of determination for women undergoing a 3-h 100-g OGTT in three successive

<table>
<thead>
<tr>
<th>Table 1. Characteristics and results for repeat GCTs in women during the same and consecutive pregnancies.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>n</td>
</tr>
<tr>
<td>Interval between tests</td>
</tr>
<tr>
<td>Age, years*</td>
</tr>
<tr>
<td>Weeks of gestation*</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1-h glucose, mg/dL*</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Log$_{10}\sigma^2_e$ (mg/dL)$^2$</td>
</tr>
</tbody>
</table>

* Data from Sacks et al. [3].

* Mean ± SD.
In conclusion, the results of this study suggest that $\sigma^2_p$ for the GCT is similar during the same and subsequent pregnancies, as long as 2 years apart. Furthermore, the reliability of GCT is significantly greater than observed for the corresponding 3-h 100-g OGTT used to diagnose glucose intolerance [1]. From the estimates of $\mu$, $\sigma^2_p$, and $\sigma^2_e$, there is only a low probability that a GCT $\leq 125$ mg/dL will have a true value $\geq 140$ mg/dL in either short- or long-term repeat measurements. Lowering the 140 mg/dL threshold to 125 mg/dL would generate 8.9% more diagnostic OGTTs at this institution, where the proportion of women having a GCT result $\geq 140$ mg/dL is 9.8%.

The help and assistance of S.K. James and C.J. Seaborn in the database compilation is gratefully acknowledged.

References