Evaluation of a Dipstick Test for Glucose in Urine

Jørn Dyerberg,1 Lissi Pedersen,2 and Ole Aagaard2

As an example of qualitative tests, a dipstick analysis for glucose in urine has been tested for the influence of modifying factors on the test result. Two different types of dipsticks were examined, "Clinistix" and "S-Gluko-test." Used according to manufacturer's instructions, the latter is more sensitive and selective. By multivariable analysis the following variables were examined: urine samples, inter- and intra-analyst, exposure to light, and dipstick batch. The first three contributed significantly to the total variation in results, inter-specimen variation being the most important. With knowledge of the frequency of testing urines with a given glucose concentration and the probability of the result at that concentration, an expression of the probability of the glucose content of a urine sample can be obtained. Even with the tests of the type examined having a sensitivity and specificity exceeding 95%, 14 of 100 patients suspected of having diabetes mellitus on the basis of a dipstick examination will be found to have a urinary glucose concentration of <2 mmol/liter. These figures were found when the prevalence of urines with a glucose concentration exceeding 2 mmol/liter was 17.5%.

Additional Keyphrases: analysis of variance • screening

Tests that give either a positive or a negative result, the so-called qualitative tests, make up a quite considerable part of the diagnostic process. Such tests are used extensively in the clinical chemical field, and as the binary result decides whether the patient will be further examined and possibly treated, it is necessary to know the efficiency and level of discrimination of the basic analyses, as well as what effect modifying factors will have on the test results. Furthermore, these analyses must be controlled, just as quantitative analyses are.

Qualitative tests performed with "dipsticks" are widely used because of their simplicity. However, in our opinion these tests are often performed without sufficient knowledge of the possibilities of misinterpretation and without the necessary measures being taken against modifying factors. We have tried here to illustrate (a) which modifying factors may influence results of such tests, and (b) what precautions must be taken.

As an example of these tests we have chosen the test for glucose in urine, because this test is one of the most frequently performed. Also, glucose is a well-defined substance that can be quantitated specifically.

Material

We used dipsticks from Ames Co., Miles Lab. Ltd., England ("Clinistix") and from Boehringer Mannheim Corp., Waldhof, Germany ("S-Gluko-test"); both were kindly supplied by the manufacturers.

Experimental Design and Results

Factors Affecting Color Change

For both dipsticks, interpretations are based on a change in color to be observed at a fixed time after immersing the dipsticks in urine. Clinistix gives a positive reading when the color changes from red to blue after 10 s. S-Glukotest changes from yellow to green after 30 s. The sensitivity is given as 5–6 mmol/liter for Clinistix and as 2–3 mmol/liter for S-Glukotest. This accounts for the initial color change of the dipstick. Although the manufacturers say it is possible to semiquantify a positive result by the intensity of the color obtained, we used only the results from the initial color change.

The purpose of this part of the study was to establish the effect of modifying factors on the total variation of the color change: variation among samples (urine factor), interperson variation (laboratory technician factor), intra-person variation (time factor), among batches (batch factor), and finally, variation of light (environmental factor).

The effect of the above factors on color change has been examined by multivariate analysis, in which the interaction terms are presumed to be negligible and where the factors have been picked at random—following a so-called extended Greek Roman square (I). Each factor is present in the square at 13 levels. The test plan is illustrated in detail in Figure 1. By such analysis it is possible to decide the relative contribution of a particular factor to the total variation of the system. Any combination of two factors will be encountered only once, and the effect of a particular factor can be compared at different levels because all the other factors make the same contribution at a given level, thus neutralizing each other.

---

1 Clinical Chemical Department, Aalborg Hospital North, Box 365, 9100 Aalborg, Denmark.
2 Clinical Chemical Department, Aalborg Hospital South, Box 365, 9100 Aalborg, Denmark.

Received Sept. 3, 1975; accepted Nov. 12, 1975.
**The urine factor** is represented by 13 night urines selected at random from nondiabetic inpatients. First, the actual glucose content of the urine was determined with an LKB 8600 by use of a hexokinase reagent kit from Boehringer Mannheim (cat. No. 159931). The concentration range was found to be 0–0.6 mmol/liter.

Next, glucose was added to the urine samples, creating 10 different concentrations of the same urine sample, ranging from 0–10 mmol/liter, with intervals of 1 mmol/liter. The samples were then tested in random order.

**The laboratory technician factor** is represented by 13 laboratory technicians with various degrees of experience in reading Clinistix, but who were not familiar with S-Glukostest.

**The time factor** indicates the order in which the urine samples were tested by each laboratory technician.

**The environmental factor** is represented by 13 different locations with various combinations of daylight, neon light, and incandescent lamps.

**The batch factor** is represented by 13 different batch numbers.

The calculations were based on the average concentrations of the last negative and the following positive reading (Figure 2). In the case of dubious readings, the procedure was as described in Figure 2.

The test was performed twice, first with Clinistix and then with S-Glutokset, under identical conditions. Between the tests the urine samples were kept frozen at −60 °C.
The influence of the factors could now be analyzed, the median point of color change calculated, and the results from the two squares compared by a Wilcoxon test, assuming gaussian-distributed rank sums.

Results

Results of the Variance Analyses

The log-normal distribution was used for the variance analysis of the two test squares. The test results appear in Table 1. Table 2 (a and b) shows the results of the F-test. Evidently, not only the urine factor but also the laboratory technician and environmental factors contribute significantly to the variation in the point of color change. Clinistix read after 10 s and S-Glukotest after 30 s show a median point of color change of 2.8 and 1.8 mmol/liter, respectively. The "95–90" tolerance interval for these points of change was calculated to 0.5–0.8 mmol/liter for Clinistix and 0.5–4.5 mmol/liter for S-Glukotest. Owing to the discrete figure distribution, the tolerance intervals were calculated according to the nonparametric tolerance method (2).

Figure 3 shows the distribution of positive/negative answers from 2 × 1690 single results from the test squares, summed up for each of the 10 concentration intervals.

The points of change established by the two test squares, compared by the Wilcoxon test, proved that the sensitivity of Clinistix was significantly poorer than that of S-Glukotest. The rank sum for Clinistix: 36 159 (P < 0.001). Consequently, we extended the time for reading the Clinistix to 30 s during the succeeding tests, thereby increasing the sensitivity.

As already mentioned, we have arrived at the results presented so far by "translating" the positive/negative answers to a urine glucose concentration. For this we used the matrix shown in Figure 2. Dubious results were obtained only with Clinistix. Among each of the 169 points of change there were 24 "doubtful" cases, four "false-positive" cases, and seven "false-negative" cases. All the results obtained by S-Glukotest were "normal" (Figure 2). This difference in occurrence of dubious answers is statistically significant (P < 0.01) according to the binomial distribution.
Level and Ability of Discrimination under Favorable Analytical Conditions

As 13 randomly selected urine samples were not considered representative of all the urine samples tested in a clinical chemical laboratory, and as it was desirable to determine more accurately the discrimination level and ability of the dipsticks, 77 urine samples were enriched with glucose and tested under the most favorable conditions, i.e., by skilled laboratory technicians in the best possible light (neon light and incandescent lamps). For this part of the test, dipsticks of the same batch number were used, and the urine samples were tested before the other analyses that day. The urine samples were collected from non-diabetic inpatients chosen without conscious bias, and the glucose content was determined by the hexokinase method. Their glucose concentrations ranged from 0 to 0.9 mmol/liter. The urine samples were enriched with glucose so that six concentrations were prepared from each sample, ranging from 0 to 6 mmol/liter with intervals of 1 mmol/liter, and tested in random order.

As stated above, the point of color change occurred at a considerably higher glucose concentration for Clinistix than for S-Glukotest, but as this difference was lessened by reading the Clinistix after 30 s—and not after 10 s as recommended by Ames—both dipsticks were read after 30 s during this experiment.

Figure 4 shows the distribution of positive/negative results from 77 night urines enriched with glucose at 1 mmol/liter intervals; both Clinistix and S-Glukotest were read after 30 s.

A comparison of Figures 3 and 4 shows the improved sensitivity when Clinistix is not read until after 30 s.

Interpretation of a Dipstick Result

We used the hexokinase method to measure glucose in urine samples from inpatients, received in the laboratory for glucose screening during one week.

By combining the information about a probable positive/negative dipstick result in the different concentration intervals with the probability of finding urines in the same intervals, we are able to express the value of the information obtained from a dipstick examination, i.e., to predict the probable glucose content of the urine sample. The expression is given by Bayes' formula (3):

$$P(C_i \leq C < C_{i+1}|\text{neg. statement}) =$$
$$\frac{P(\text{neg. statement}|C_i \leq C < C_{i+1}) 	imes P(C_i \leq C < C_{i+1})}{\Sigma_i P(\text{neg. statement}|C_i \leq C < C_{i+1}) 	imes P(C_i \leq C < C_{i+1})}$$

where $C_i = 0, 1, 2, 3, \ldots, n$ mmol glucose/liter, and $C_{i+1} = 1, 2, 3, 4, \ldots, (n + 1)$ mmol glucose/liter.

Figure 5 shows the distribution of the glucose concentrations of urines received in the laboratory during one week, as measured by the hexokinase reaction.

Table 3 shows the results when Bayes’ formula was used to calculate the probability of a urinary glucose value being within certain concentration limits when a negative dipstick reading has been obtained.

The above may picture the capacity of the dipstick method too optimistically because of the relatively large number of normal urines used for the calculation. Thus we have included the same figures in a table of sensitivity and specificity (Table 4).
Table 3. The Probability of Finding the Urine Glucose Content within a Certain Concentration Interval when the Dipstick Reading is Negativea

<table>
<thead>
<tr>
<th>Interval (mmol/liter)</th>
<th>Clinistix Negativeb</th>
<th>S-Glukotest Negativeb</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C_i &lt; C &lt; C_i + 1 )</td>
<td>( P(C_i &lt; C &lt; C_i + 1) )</td>
<td>( P(C_i &lt; C &lt; C_i + 1) )</td>
</tr>
<tr>
<td>( 0 \leq C &lt; 1 )</td>
<td>0.953</td>
<td>0.957</td>
</tr>
<tr>
<td>( 1 \leq C &lt; 2 )</td>
<td>0.043</td>
<td>0.041</td>
</tr>
<tr>
<td>( 2 \leq C &lt; 3 )</td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td>( 3 \leq C &lt; 4 )</td>
<td>0.0004</td>
<td>0.0001</td>
</tr>
<tr>
<td>( C \geq 4 )</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

a For meaning of symbols, see text
b Both dipsticks read after 30 s

Table 4. Outline of Specificities at a Selected Discrimination Level of 2 mmol/litera

<table>
<thead>
<tr>
<th>Interval</th>
<th>&lt; 2 mmol</th>
<th>&gt; 2 mmol</th>
<th>( \Sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Clinistix</td>
<td>79.8</td>
<td>80.2</td>
</tr>
<tr>
<td></td>
<td>S-Glukotest</td>
<td>79.6</td>
<td>79.8</td>
</tr>
<tr>
<td>Positive</td>
<td>Clinistix</td>
<td>2.7</td>
<td>19.8</td>
</tr>
<tr>
<td></td>
<td>S-Glukotest</td>
<td>2.9</td>
<td>20.2</td>
</tr>
<tr>
<td>( \Sigma )</td>
<td>82.5</td>
<td>17.5</td>
<td>100.0</td>
</tr>
</tbody>
</table>

a Figures are percent of total

Discussion

The \( F \)-test shows that much of the variation in test readings for both types of dipsticks is attributable to inter-sample variation. The lowest positive urine and the highest negative urine had glucose concentrations of 0.3 and 9.3 mmol/liter, respectively.

Like other authors (4), we experienced a considerably smaller sensitivity when testing glucose in urine than when experimenting with glucose in water, probably because urine contains a number of compounds that tend to inhibit the analytical reactions applied in the experiment (4, 5), particularly certain drugs (4). We have not studied the effect of drugs further, as we judge it unwarranted and find it simpler to determine the glucose content of drug-containing urines by quantitative methods.

Unlike the urine factor, the other factors tested can be modified. Among these factors we experienced a significant difference between the laboratory technicians, which illustrates the need for very careful instruction before even as simple a test as the dipstick test is done.

The environment factor has a weak but significant influence on the dipstick readings; it is necessary to do the tests under good and consistent lighting.

However, we found no significant difference between results for various batches of dipsticks or between one particular laboratory technician at different times. We conclude that the two brands of dipsticks are prepared from reasonably consistent compounds and that one laboratory technician can justifiably do analyses in series of up to 130.

However, the similarity between two dipsticks ends here, as illustrated in Figure 3 and by the Wilcoxon test. We conclude that when the two are used according to the manufacturers’ instructions, S-Glukotest is more sensitive and gives a more consistent interpretation.

Values in the literature vary greatly on the subject of the physiological limits of glucose in urine. For example, Renschler et al. (6) fix the upper limit at 1.67 mmol/liter, and Pileggi and Szustkiewicz (7) at 0.55 mmol/liter. The point of change obtained by S-Glukotest is closer to these values than those obtained by Clinistix. By increasing the time before reading the Clinistix to 30 s, we improved its sensitivity to compare with that of S-Glukotest, both now showing an average point of change between 1 and 2 mmol/liter and a range between 1 and 4 mmol/liter (Figure 4).

Another reason for preferring a sensitive method of excluding urines that are normal as far as glucose concentration is concerned is that we have a quantitative method of differentiating all positive tests and thus establishing “false positive” results.

Table 3 states the probable glucose concentration in any particular urine sample when the dipstick test is negative. Note that the calculation has been based on the distribution of urines in Figure 5, which is based on local data and not necessarily universally representative. When a result is negative there is a greater than 95% probability that the glucose concentration is less than 1 mmol/liter and more than 99% probability that it is less than 2 mmol/liter.

The above information seems to justify use of the dipsticks for screening. One must bear in mind, however, that of urine samples with a glucose content of 2 to 3 mmol/liter, Clinistix may give a “negative” answer for 15% and S-Glukotest for 9%. For urines with a glucose content between 3 and 4 mmol/liter, 4% and 2%, respectively, may be so overlooked (Figure 4). This is true even if the sensitivity of Clinistix is improved by reading after 30 s.

On the other hand, the predictive value of the tests is found to be 86%; i.e., 14% of subjects with a “positive” result will have a urinary glucose concentration of less than 2 mmol/liter (Table 4). Suppose that all untreated diabetics have a urinary glucose concentration above and all nondiabetics a concentration below this value—which of course is not quite true—14 of 100 patients suspected to have diabetes mellitus on the basis of a dipstick examination will in fact be healthy.
Generally, we should choose the type of analysis with the highest level of discrimination between well and sick in relation to the consequences of a positive/negative result, and with the greatest ability to discriminate; i.e., with the sharpest change at the appropriate decision value.

Of course, it would be better specifically to measure the glucose in all urine samples, thereby avoiding "false-positive" and "false-negative" results. However, there may be conditions under which only the simple screening test can be done.

One method for increasing the certainty of the decision in the case of a doubtful result is to increase the urine glucose content of the specimen by 2 mmol/liter and observe whether the doubtful reaction, checked at 30 s, now becomes clearly positive. If it is not, the reason may be that inhibitory substances are present, and the glucose must then be checked by other methods.

In any case the consequences of a "false-positive" or "false-negative" result must be considered before choosing the decision value. For example, when a urine sample is tested for glucose, the consequence of a "false-positive" result will usually be only economically detrimental (the cost of a quantitative determination of glucose in urine), plus the fact that the patient may be inconvenienced by an oral glucose-tolerance test. It is quite a different matter with some other qualitative analyses. For instance, to refute a positive result of a dipstick test for protein in urine, one must do several quantitative protein analyses and possibly expose the patient to extensive and trying tests (urography, for example).

When dipsticks are used to screen urines for glucose, there should be quality control of the test system. Because of the great influence of urine composition on dipstick results, we suggest that each day a random series of urines be tested, prepared from a pool of negative urines enriched with glucose to concentration intervals of 1 mmol/liter. When changing to a new batch of dipsticks, results from this batch should be compared with those for old batch by comparing the fraction of positive results obtained when analyzing 20 times at random a control urine with a glucose content corresponding to 0.1 and 0.9 of positive results, respectively.

On the basis of our results we cannot recommend the relatively crude dipstick examination for glucose in urine for quantitative purposes. Differentiating subtle graduations in color intensity at a fixed time will be susceptible to even greater errors than those observed in the "yes or no" situation. As a matter of fact such an attempt was made by the technicians during the study; the results were such that any statistical analysis was impossible. Quick and reliable methods for measuring glucose concentration in urine are at hand. Clinical decisions that imply therapeutic consequences (e.g., alterations in insulin dosage) should be based on information having a validity that justifies these decisions.

References