A glucose meter evaluation co-designed with both health professional and consumer input

Harmony Thompson, Huan Chan, Florence J Logan, Helen F Heenan, Lynne Taylor, Chris Murray, Christopher M Florkowski, Christopher M A Frampton, Helen Lunt

Abstract

Health consumer’s input into assessment of medical device safety is traditionally given either as part of study outcome (trial participants) or during post marketing surveillance. Direct consumer input into the methodological design of device assessment is less common. We discuss the difference in requirements for assessment of a measuring device from the consumer and clinician perspectives, using the example of hand held glucose meters.

Around 80,000 New Zealanders with diabetes recently changed their glucose meter system, to enable ongoing access to PHARMAC subsidised meters and strips. Consumers were most interested in a direct comparison of their ‘old’ meter system (Accu-Chek Performa) with their ‘new’ meter system (CareSens brand, including the CareSens N POP), rather than comparisons against a laboratory standard.

This direct comparison of meter/strip systems showed that the CareSens N POP meter read around 0.6mmol/L higher than the Performa system. Whilst this difference is unlikely to result in major errors in clinical decision making such as major insulin dosing errors, this information is nevertheless of interest to consumers who switched meters so that they could maintain access to PHARMAC subsidised meters and strips. We recommend that when practical, the consumer perspective be incorporated into study design related to medical device assessment.

New clinical measuring devices are usually introduced because they have advantages over the established measuring devices, such as lower cost, greater ease of use and/or greater accuracy and precision. Before introducing new measuring devices, regulatory authorities usually require some form of method comparison assessment.

This article discusses the general principles of method comparison, based on an example using glucose meters. Following the input of consumer representatives, methodology for the current study does however follow a slightly different approach compared to the usual assessment of glucose meters. We also discuss consumer insights into interpretation of study results and how this might influence future study design.

General approach to method comparison

Whilst results from a new device should correlate highly with the established reference method of measurement, a high correlation does not automatically imply good agreement between the two methods. Results can be highly correlated but nevertheless show systematic bias, for example the mean (average) result obtained from one method may differ from the mean result obtained by the second method.
In method comparison studies of measuring devices, assessment for possible bias may be undertaken using Bland Altman plots. On a classical Bland Altman plot, the x axis shows the average value obtained from the two methods under consideration, or alternatively the value obtained from a reference method (“gold standard”).

The difference between these two methods is plotted on the y axis. The Bland Altman 95% limits of agreement as shown in Figures 1a, 2a and 3a is defined as the mean difference ± 1.96, multiplied by the standard deviation of the differences. This represents a measure of the scatter or variability of the paired results, such as paired results from two different meters.

If the Bland Altman plot shows evidence of slight bias, a second problem now exists for the patient and their clinician. Is this bias clinically significant? For example does it result in a change in clinical management such as a change in medications and/or trigger a need for a clinical assessment?

PHARMAC recently implemented a major change to a single supplier of glucose test strips and meters, with exceptions for occasional patients with specific clinical needs. We discuss approaches to assessing patients’ glucose meter performance within the context of this recent switch in glucose meters, undertaken by an estimated 80,000 diabetic patients in New Zealand who want to obtain ongoing supplies of subsidised strips.

Most of patients switching from their ‘old’ meter system, most commonly the Performa (Accu-Chek Performa) meter, onto the ‘new’ CareSens system (CareSens II, CareSens N or CareSens N POP), may not have formulated questions about comparing results from different meter systems in the way we discuss above, but they nevertheless want to know if the performance of their ‘old’ and ‘new’ systems is similar and if not, whether or not these differences are clinically significant.

A visual method for assessing the clinical significance of glucose meter performance relative to the reference method that is intuitively easy to understand and which complements Bland Altman plots, is the Consensus error grid (see Figures 1b and 2b below). In brief, the error grid assumes that minor biases at the extreme ends of glucose values are not clinically important. For example, a broadly similar clinical treatment decision would be made for high glucose values irrespective of whether the value was high or very high.

Conversely a patient suffering a glucose value that is clearly in the hypoglycaemic range should be treated and reassessed, irrespective of the exact value of the measured glucose. A close agreement of the measured capillary glucose value with the reference value is however particularly important at the borderline between the low and normal glucose range, because of the impact on small variations in decision making, in this glucose range.

**Patient feedback suggests that the “new” meters read higher than their “old” meters**

During the current transition period from the ‘old’ to the ‘new’ meter systems, anecdotal reports emerged from patients and their clinicians that the new meter was reading higher than the old meter. This was generating anxiety amongst some individuals and/or their caregivers.
Two consumer representatives from a local diabetes lay society therefore co-designed the below study, together with clinical researchers. They requested that the capillary glucose reading from the ‘old’ Performa meter be used as the reference value, thereby allowing a direct comparison of Performa capillary results with those of the CareSens N POP meter/ CareSens N strip system. The study has Health and Disability Ethics Committee (New Zealand) approval; HDEC number 12/STH/22/AM01.

**Combined clinician and consumer approach to defining methodology and interpreting results**

The methodology used to prepare samples was similar to that described previously, with the exception that antecubital fossa venous samples were placed into plasma separator tubes and spun immediately after collection, thereby minimising pre-analytical glycolysis in the laboratory measured venous plasma glucose samples. Two meters of each type (Performa, CareSens N POP) were used, thus a total of four capillary samples were obtained from each patient. Meters were rotated during the assessment period to eliminate time dependent analytical bias. Participants were recruited from diabetes outpatients and 105 patients took part in the study (see Table 1 below).

**Table 1. Characteristics of study population (N=105)**

<table>
<thead>
<tr>
<th>Gender</th>
<th>%</th>
<th>Number of Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>59%</td>
<td>62</td>
</tr>
<tr>
<td>Female</td>
<td>41%</td>
<td>43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Median age, years (range)</th>
<th>54 (19 -88)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Diabetes type</th>
<th>%</th>
<th>Number of Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>32%</td>
<td>34</td>
</tr>
<tr>
<td>Type 2</td>
<td>68%</td>
<td>71</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>%</th>
<th>Number of Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>European</td>
<td>76%</td>
<td>80</td>
</tr>
<tr>
<td>Maori</td>
<td>12%</td>
<td>13</td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>4%</td>
<td>4</td>
</tr>
<tr>
<td>Asian</td>
<td>7%</td>
<td>7</td>
</tr>
<tr>
<td>Other</td>
<td>1%</td>
<td>1</td>
</tr>
</tbody>
</table>

Haematocrit range was 0.34–0.51, thus haematocrit values fell within the manufacturers’ recommended range for both brands of glucose meter.
Figures 1 and 2 present results in a traditional way, by comparing the venous plasma reference value against the Performa and the CareSens N POP capillary values, respectively. The mean and 95% CI (confidence intervals) for [capillary – venous plasma] glucose difference were as follows: The Performa mean (95% CI) was -0.25 (-0.04 to -0.46) mmol/L and the corresponding value for the CareSens was +0.26 (+0.43 to +0.08).
There was therefore evidence of minor systematic bias only, in this aspect of the experiment. These findings are not inconsistent with earlier local meter assessments, as subtle differences in findings between studies are to be expected, in relation to slight differences in participant characteristics between studies, including differences in glucose range.

Of greater relevance to consumers is the comparison of the first Performa capillary value obtained from each patient with the equivalent CareSens N POP value, shown in Figures 3a and 3b.

These findings confirm consumers’ experience that the CareSens N strips read higher than the Performa (see the horizontal line in the Bland Altman plot of Figure 3a, which shows that the mean capillary glucose [CareSens-Performa] difference is 0.59mmol/L) with the 95% confidence interval for this point estimate spanning from 0.42mmol/L to 0.77mmol/L.

The current (2013) ISO standards for glucose meter assessment consider glucose values less than 5.6mmol/L to be sufficiently critical in clinical decision making, that the glucose differences between the reference and test glucose values should be small (±0.83mmol/L).

We therefore undertook a secondary analysis looking at the capillary-venous difference in the eleven participants with laboratory glucose results <5.6mmol/L. Mean (95% CI) difference between the first glucose readings from the two meters (N=11) was 0.38 (0.09 to 0.67), compared to 0.59 (0.42 to 0.77) for the group as a whole (N=105, see previous paragraph). The glucose values obtained from the two meter types therefore appears to show closer agreement at lower glucose values.

The CareSens’s ‘over-read’ relative to the Performa meter is however unlikely to result in major clinically significant errors of self-management (see Figure 3b; Consensus grid). Summarising the results illustrated in these three Figures, the Performa tended to read slightly low and the CareSens read slightly high, when compared to plasma venous glucose, thereby producing a summative error when looking at the direct comparison of the Performa with the CareSens meter/strip systems (Figure 3a).

Figure 3. Comparison of Performa against CareSens (first capillary sample)
Does one brand of meter/strips give results that, on repeat testing, seem more variable than those of the alternative brand of glucose meter?

The above question is of interest to both clinicians and consumers and is answered by measuring the CV (co-efficient of variation %), which is the ratio of the standard deviation of the repeat measurements (SD) to the mean ($\bar{x}$), calculated as $[SD \div \bar{x}] \times 100\%$. A lower CV is associated with greater precision of measurement, i.e. less variability. The % CVs for the paired capillary values for the two Performa samples and CareSens samples were 5.3% and 5.3% respectively.

The precision shown by the two different meter/strip systems was therefore both small and also similar. The first capillary value obtained from each of the Performa and CareSens N strips (i.e. Performa/CareSens comparison) had a CV of 6.8% demonstrating that the variability in results obtained using a strip from each of the two different meter systems was about the same as comparing two strips from the same system.

Study limitations—the consumer’s perspective

Although the study described above utilises standard meter assessment methodology, the consumer representatives identified several limitations.

Firstly, it was undertaken by a research team familiar with meter assessment, thus it lacks a real world element to it. For example, the likelihood of inadvertent end user error in field conditions such as problems associated strip fill functionality (over- or underfill), although likely to be small, was not formally assessed.

Secondly, it does not assess clinical questions around meter features related to usability that are especially important to the end user, such as use at colder temperatures.

Thirdly, although the research team attempted to select and obtain written consent from patients at risk of hypoglycaemia, we were unable to obtain many results from asymptomatic hypoglycaemic patients, thus there are limited comparisons in the critical glucose range of <5.6mmol/L.

Arguably, it would have been informative to include patients with plasma glucose levels in the hypoglycaemic range, but there is no single glucose value that defines hypoglycaemia. Samples with low glucose values could have been obtained by various alternative means, for example by delaying laboratory analysis of samples so that glucose values are lowered due to pre-analytical glycolysis, or by including samples from fasting healthy volunteers.

The relationship between prandial status and capillary and venous plasma glucose is complex and there is no evidence that the behaviour of samples that have been manipulated or are from healthy volunteers, mimics the behaviour of those with established diabetes. For this reason, we decided to collect samples only from patients with established diabetes.
Another study limitation identified by consumers is that combining the findings from 105 patients undergoing a single assessment does not inform patients about the theoretical possibility that occasional individuals might run with measured glucose values that are persistently much higher (or lower) with one meter/strip system compared to another, because of the possibility that interfering substances in their blood interact differently with the different enzymes systems used by different brands of meter/strip systems. For this reason and also so that familiarity and confidence can be developed with a new meter system that gives subtly different results, many patients undergoing the process of meter change-over have chosen to undertake a personal series of duplicate tests on the ‘old’ and ‘new’ meters, ‘just to be sure’.

Some of the meter assessment limitations identified by consumers might be best addressed using a program of active, structured post marketing surveillance designed to monitor the specific concerns mentioned above, with findings being linked to a mechanism for device co-development with the manufacturer in situations where this is both possible and appropriate.

**Conclusion**

In conclusion, undertaking a method comparison study using a study design that was clinically meaningful from a consumer perspective demonstrated that patients’ perceptions of meter performance were correct; the ‘new’ CareSens N POP meter read higher than the ‘old’ Performa meter. This systematic bias is however, unlikely to result in major errors in clinical decision making.

This study also highlights the need to develop future study methodologies that provide a high proportion of glucose samples in the hypoglycaemic range, which directly reflect the physiology of samples from those with diabetes.

The need to spend the New Zealand health dollar wisely may require more changes in measurement technologies. We recommend that if method comparison studies of measurement devices are to be presented in a way that is meaningful to consumers, study design is best undertaken in conjunction with the consumer.

**Competing interests:** Helen Lunt and Chris M Florkowski have previously undertaken meter validation studies on behalf of PHARMAC.

**Author information:** Harmony Thompson, 5th Year Medical Student, University of Otago, Christchurch; Huan K Chan, Endocrinology Registrar, Diabetes Centre, Christchurch; Florence J Logan, Research Nurse, Diabetes Centre, Christchurch; Helen F Heenan, Study Co-ordinator, Diabetes Centre, Christchurch; Lynne Taylor, Manager, Diabetes Christchurch; Chris Murray, Diabetes Youth Canterbury Representative, Diabetes Christchurch; Christopher M Florkowski, Clinical Biochemist and Diabetologist, Canterbury Health Laboratories, Christchurch; Christopher M A Frampton, Biostatistician, Department of Medicine, University of Otago, Christchurch; Helen Lunt, Diabetologist, Diabetes Centre, Christchurch

**Acknowledgement:** Harmony Thompson undertook this study as part of a University of Otago Christchurch summer studentship.

**Correspondence:** Dr Huan Chan, Diabetes Centre, 550 Hagley Avenue, PO Box 4710, Christchurch, New Zealand. Fax: +64 (0)3 3640171; email: huan.chan@cdhb.health.nz
References:


