Continuous glucose and exhaled breath analysis in the Intensive Care Unit

Leopold, J.H.

Citation for published version (APA):
Chapter 3

Point Accuracy and Reliability of an Interstitial Continuous Glucose Monitoring Device in Critically Ill Patients: A Prospective Study


Critital Care 2015, 9
ABSTRACT

**Introduction**: There is need for continuous glucose monitoring in critically ill patients. The objective of this trial was to determine the point accuracy and reliability of a device designed for continuous monitoring of interstitial glucose levels in intensive care unit patients.

**Methods**: We evaluated point accuracy by comparing device readings with glucose measurements in arterial blood using blood gas analyzers. Analytical and clinical accuracy was expressed in Bland–Altman plots, glucose prediction errors, and Clarke error grids. We used a linear mixed model to determine which factors affect the point accuracy. In addition, we determined the reliability, including duration of device start-up and calibration, skips in data acquisition, and premature disconnections of sensors.

**Results**: We included 50 patients in whom we used 105 sensors. Five patients from whom we could not collect the predefined minimum number of four consecutive comparative blood draws were excluded from the point accuracy analysis. Therefore, we had 929 comparative samples from 100 sensors in 45 patients (11 [7–28] samples per patient) during 4,639 hours (46 [27–134] hours per patient and 46 [21–69] hours per sensor) for the accuracy analysis. Point accuracy did not meet the ISO14971 standard for insulin dosing accuracy, but improved with increasing numbers of calibrations, and was better in patients who did not have a history of diabetes. Out of 105 sensors, 60 were removed prematurely for a variety of reasons. The device start-up time was 49 [43–58] minutes. The number of skips in data acquisition was low, resulting in availability of real-time data during 95 [89–98]% of the connection time per sensor.

**Conclusion**: The point accuracy of a device designed for continuous real-time monitoring of interstitial glucose levels was relatively low in critically ill patients. The device had few down times, but one third of the sensors were removed prematurely because of unresolved sensor–or device related problems.

**Trial registration number**: Netherlands Trial Register www.trialregister.nl NTR3827. Registered 30 January 2013.

INTRODUCTION

Handheld blood glucose meters or department-based blood gas analyzers are currently the preferred methods to measure blood glucose levels in intensive care unit (ICU) patients [1, 2]. These intermittent glucose monitoring techniques have variable accuracies [3], but foremost lack useful trending because of the interval between consecutive measurements. Continuous glucose monitoring (CGM) is suggested to increase practicalities and safety of insulin titration in ICU patients [4, 1], in particular when targeting normal or near-normal blood glucose levels when hypoglycemic episodes can be expected [5–13]. Glucose oxidase technique-based interstitial CGM devices have been used...
before in diabetic patients outside the ICU setting [14]. It is uncertain, however, whether interstitial CGM devices are point accurate in critically ill patients [1]. An altered relationship between blood and interstitial fluid glucose levels during critical illness could affect the point accuracy of interstitial CGM to reflect the blood glucose level [15]. Several interstitial CGM sensor systems originally designed for non–ICU patients have been tested in the ICU setting in recent years [16–28]. Medtronic MiniMed (Northridge, CA) developed the Sentrino Continuous Glucose Management System, an interstitial CGM device that was especially designed for use in critically ill patients. This device was improved from previous models by creating the processor cable and pole-mounted monitor, and by four sensing elements designed to increase responsiveness to glucose changes and to limit the influence from drug interactions. The aim of this study was to test its point accuracy and reliability in a mixed medical–surgical ICU. We hypothesized that the device would provide an accurate reflection of the blood glucose level in ICU patients treated according to a local guideline for blood glucose control targeting blood glucose levels between 90–144 mg/dL. In addition, we determined its reliability, including duration of the device start-up, the need for calibration, skips in data acquisition and number and reasons for premature disconnections.

**METHODS**

**Study design and informed consent**

This was an investigator–initiated observational trial. The Institutional Review Board of the Academic Medical Center (Amsterdam, The Netherlands) approved the study protocol (study ID: NL41498.018.12). Medtronic MiniMed provided three devices for the duration of the trial and the necessary sensors, but had no influence on study design or study reporting. Patients or next of kin had to provide written informed consent before start of any study–related procedure.

**Study population**

Patients were recruited between October 2012 and February 2014 in a 30–bed mixed medical–surgical ICU of a large university hospital (the Academic Medical Center). Patients were eligible for inclusion if they were aged ≥ 18 years and had an anticipated life expectancy > 96 hours. Patients were excluded from participation if they had a platelet count < 30 x 10⁹/L, had participated in a trial testing an investigational product or treatment within the past 30 days, were pregnant, or had a suspected or diagnosed medical condition, which in the opinion of the investigators prevented the patient from completing the study.

**Glucose control**

ICU nurses performed glucose control with insulin, following a local guideline for blood glucose control targeting a blood glucose level between 90–144 mg/dL [29]. Insulin titration adjustments were based on sliding scales. The local
guideline for blood glucose control dictated nurses to perform blood glucose measurements at least every four hours, and more frequently if blood glucose levels were out of range or were expected to change rapidly. For details, see the electronic supplement. During conduct of the study, ICU nurses were not allowed to change insulin infusion rate based on the readings by the investigational device. However, they were allowed to perform additional blood glucose measurements if the device suggested rapid changes in the glucose level, or when there was a trend towards hypoglycemia.

The investigational device
The disposable glucose sensors of the device were glucose oxidase–based; two probes, each with two sensing elements. The individual measurement results were combined and displayed on the device monitor every minute. The signal was transmitted through the processor cable to the monitor. It was single-patient single-use sensor, which could be used for up to 72 hours. The processor cable was reusable.

The sensor was inserted into the subcutaneous tissue using two parallel introducer needles. The two needles automatically retracted when the introducer hub was pulled away from the sensor base; the sensor probes remained in the subcutis. Each new sensor needed calibration using blood glucose levels after insertion and initialization, and after 1 hour and 2 hours; thereafter, repeated calibrations were performed every 8 hours.

Study procedures
Sensors were inserted into the subcutis of the thigh. Successive sensors could be used for 72 hours, depending on length of stay in the ICU, but never longer than for 30 days. Arterial blood glucose levels were measured using RapidLab 1265 blood gas analyzers (Siemens Healthcare Diagnostics, The Hague, The Netherlands), which were used for calibrations of the device. Not only did ICU nurses provide the mandatory calibration blood glucose levels, but also the routinely obtained blood glucose levels (i.e., blood glucose measurements which were not requested by the device for calibrations, but were taken by the nurses as dictated by the local guideline for blood glucose control) were entered into the device as well. Therefore, these measurements were also used for calibrations of the device. If the device displayed a message requesting an additional non-routine calibration to resolve a sensor performance issue (i.e., a ‘Poor Sensor Signal’ alert), the nurses were permitted to disregard manufacturer recommendation and remove sensors rather than enter the requested calibration.

Each day, the place of insertion was photographed and inspected for redness, bruises, and swelling. In case the patient was awake we questioned the patient...
whether it was painful. Every item could be scored as ‘none’, ‘minor’ or ‘major’.

Power calculation
We intended to enroll 50 patients to assess accuracy of the CGM device. With 50 patients we expected to have at least 40,000 subcutaneous CGM device results and at least 1,200 blood glucose level measurements with the RapidLab 1265. Based on previous studies testing point accuracy we assumed to have a sufficiently high number of paired samples to enable evaluation of the point accuracy of the device.

Analysis plan
The glucose data collected with each new sensor were downloaded from the device after use in a patient; the arterial blood glucose levels were downloaded from the patient data management system. The arterial blood glucose levels in the patient data management system were compared with the entries for calibrations into the device. In case of an entry error, defined as a difference between the arterial blood glucose level in the patient data management system and the calibration entry > 9 mg/dL, the correct blood glucose level was used in the accuracy analysis. The subsequent pairs, though, were excluded from the accuracy analysis, since these were influenced by the preceding entry.

For reporting point accuracy we used analytical and clinical accuracy measures, i.e., Bland–Altman plot with bias and limits of agreement (bias ± 1.96 x standard deviation of the bias) [30], glucose prediction errors, and Clarke error grid analyses [31]. According to ISO criteria, 95% of the paired measurements should be within the glucose prediction error criteria; general consensus is that 95% of the values should be in zones A and 5% in zones B of Clarke error grid analyses. Finally, we expressed the linearity between the device glucose results and blood glucose results by the Pearson correlation coefficient and coefficient of determination, R2.

In a post–hoc analysis we also report point accuracy according to the recently published consensus recommendations [1]. In this round–table meeting of ICU experts in blood glucose control it was recommended to always report the mean absolute relative difference (MARD) when testing a CGM device, where MARD values should be < 14%; values > 18% should be considered to represent poor accuracy [32]. We added the MARD as a post–hoc analysis. Furthermore, we analyzed the point accuracy following the recently published surveillance error grid [33]. For more details see the electronic supplement.

We also reported reasons for early disconnection, defined as the removal of a sensor before 72 hours. For details, see the electronic supplement. The time between calibrations using an incorrect glucose value entry and the next calibration was extracted from the total connection time of the device. Definitions of the metrics used to assess device reliability, including those suggested by
recent consensus recommendations [1] are described in the electronic supplement.

**Statistical analysis**
We reported data as means (± SD) or medians [IQR] where appropriate. In order to be considered for the statistical analysis, each patient needed to have at least 4 comparative blood glucose results for accuracy analysis. However, the excluded patients remained included in the reliability analysis. In a post–hoc analysis, we used a linear mixed model to determine which variables influence the accuracy of the device. In addition, we stratified the accuracy results by diabetic status. For detailed description of this model, see the electronic supplement. Analyses were performed using R (version: 2.15.1; R Foundation for Statistical Computing, Vienna, Austria).

**RESULTS**

**Patients and sensors**
We included 50 patients. In total, we used 105 sensors (median of 1 [1–3] sensor per patient) with a total connection time of 4,639 hours (median of 46 [27–134] hours per patient and median of 46 [22–69] hours per sensor). Five patients from whom we could not collect the minimum number of four consecutive comparative blood draws were excluded from the point accuracy analysis. A CONSORT diagram is provided in Figure 1. Patient characteristics and metrics of glucose control are shown in Tables 1 and 2.

We could not inspect the insertion site of 3 sensors. Major bruises were observed in 3 out of 102 inspected insertion sensor sites; 10 minor bruises were seen. Major redness of the skin was observed in 7 out of 102 insertion sensor sites, and minor redness in 6 out of 102 insertion sites. Swelling of the skin was never seen, and none of the conscious patients mentioned pain at the sensor insertion site.

**Point accuracy**
We collected 929 comparative samples (11 [7–28] samples per patient). Bland–Altman plot, glucose prediction error grid, and Clarke error grid are presented in Figure 2. The surveillance error grid is presented in figure 3. The Pearson correlation coefficient was 0.81; the R2 was 0.65. The MARD was 14.8%.

Fifty–eight percent of the device results were within 12.5% of the arterial blood glucose results (or within 10 mg/dL for results < 99 mg/dL) and 75% were within 20% of the arterial blood glucose results.

In the linear mixed model, only history of diabetes (P = 0.02) and number of calibrations per sensor (P = 0.04) affected the absolute difference between blood
glucose and device result. Per each new calibration the absolute difference decreased by 1.4% (standard error of 0.006%), meaning that the sensor performance increased. The effect of a history of diabetes was larger, since an increase by 34.3% (standard error of 13.0%) in the absolute difference was found when comparing patients with a history of diabetes compared to patients without diabetes. In addition, we stratified the accuracy results by diabetic status, results are shown in figure S1 and table S2. For detailed results of multivariate random intercept model, see the electronic supplement.

Reliability of the CGM device

Start–up time after placing a new sensor was 49 [43–58] minutes. The number of skips in data acquisition was low, resulting in availability of real-time data during 95 [89–98]% of the connection time per sensor. Table 3 summarizes reliability metrics of the investigational device.

Out of 105 sensors, 60 were removed before 72 hours; reasons and connection times of sensors are shown in table 4. Out of 105 sensors, 42 were removed before 72 hours after insertion for other reasons than ICU discharge or death; and 36 sensors were removed because of an unresolved ‘Poor Sensor Signal’–alert or a device error – 19 with no attempt to resolve.

DISCUSSION

We determined the point accuracy and reliability of a device specifically designed for continuous real–time monitoring of interstitial glucose levels in critically ill patients. The analytic point accuracy of the device was low in a typical cohort of patients from a mixed medical–surgical ICU, according to ISO criteria and consensus recommendations. The clinical point accuracy was low according to Clarke error grid analysis, but better according to surveillance error grid analysis. The device had few down–times, but one–third of the sensors were removed prematurely because of sensor– or device–related problems. The present findings are in line with results from a previous trial testing the same device in cardiac surgery patients [34]. In that study the mean absolute relative difference was 12.2% with 95% real-time data. Similar results come from studies testing other devices for interstitial glucose monitoring that were originally designed for use in non–critically ill patients. Those studies were performed in cardiac surgery patients [21, 24, 35], surgery patients [26], patients with neurologic emergencies [27] and non-surgical patients [16, 22, 25], with only two reporting more favorable accuracy results [21, 22]. Taken together, these data suggest that point accuracy of interstitial glucose monitoring cannot replace blood glucose level measurements.

In contrast to our findings, a previous publication by Brunner et al. suggests a better point accuracy of another interstitial CGM device in critically ill patients [18]. This report combined data of two separate trials in medical ICU patients...
The tested device in that study was from the same manufacturer, though not specifically designed for use in critically ill patients. In addition, the sensor was used for up to 72 hours and never replaced. One important difference with the present study was that the sensors were placed exclusively under the skin of the abdomen in patients included in these two trials. In most other trials sensors were inserted under the skin of the abdomen [16, 18, 22, 24, 26, 28], thigh [25, 26] or shoulder [21]. Reported point accuracies do not suggest superiority of one of these sites. Certainly, there could be other unknown and unreported factors that could have resulted in the differences in performance.

We performed a mixed linear model to determine which factors could have influenced the point accuracy of the tested sensor. Rank order of measurement and presence of a history of diabetes affected the accuracy. The finding that rank order of measurement improved sensor performance is not new [18], and certainly not surprising; more calibrations may always increase accuracy of a sensor. A history of diabetes was the most important variable influencing point accuracy, which deteriorated sensor performance with 34%. As yet, this effect remains unexplained. It could be that microcirculation alteration in diabetic patients affects interstitial glucose level. However in previous studies with interstitial devices, diabetes was not found to be significantly associated with poor sensor accuracy in critically ill and cardiac surgery patients [16, 18, 28]. Moreover, in a recent study in cardiac surgery patients, an impaired microcirculation did not affect accuracy of two interstitial glucose sensors from two different manufactures [28]. The difference found between patients with and without diabetes might also be related to glucose variability. Patients with diabetes will have more glucose variability compared to patients without diabetes. Thereby when focusing on percent difference a greater disparity could be found when comparing variability differences.

It should be stressed that we compared interstitial glucose measurements with glucose levels in arterial blood samples, which is far from comparable. Indeed, the interstitial glucose level is dependent on several factors other than the blood glucose level, such as the speed of glucose diffusion from blood to interstitial spaces, as well as the rate of glucose uptake by subcutaneous cells [37]. Importantly, these factors are not constant, particularly in critically ill patients. Furthermore, there is a time lag between interstitial glucose and blood glucose measurement [37]. Studies suggested that the interstitial glucose level decreases before the blood glucose decreases [37, 38], though this was not confirmed in other studies [39]. It is probably very difficult, if not impossible, to correct for factors causing a difference between interstitial and arterial blood glucose levels. Moreover, it is unknown whether differences between arterial and interstitial glucose levels are physiological. Nevertheless, subcutaneous glucose monitoring could have advantages. One
potential advantage is that continuous monitoring of interstitial glucose levels which enables detection of trends in the blood glucose level [32]. This could allow earlier responses to a rise or a decline of the blood glucose level. In both cases, knowledge of the direction of the trend may be more valuable than the exact blood glucose level.

It is clear that the tested device can never replace blood glucose measurements. First, initial calibrations are always necessary, as are calibrations every 8 hours thereafter. As nurses were allowed to perform additional blood glucose measurements, and as we asked them to insert the values into the investigational device monitor where they were used for additional calibrations, the number of calibrations in this study was higher than mandated. In fact, this could have improved the accuracy of the investigational device: it is possible that with fewer calibrations point accuracy becomes worse.

Our trial knows several strengths and weaknesses. Strengths include the fact that we were able to use the sensors for several days in the participating patients. Moreover, we used accurate blood gas analyzer measurements for comparisons, as well as for the calibrations. Furthermore, we were able to test the device in a typical mixed medical-surgical ICU. Weaknesses include the small sample size and the single-center design of the trial. Furthermore, we did not collect as many samples as we expected. A more important limitation of our trial, though, is the fact that the vast majority of blood glucose levels were in a narrow range, in particular preventing us to draw firm conclusions regarding accuracy in the hypoglycemic range. While the ICU nurses were not allowed to change insulin infusion rates, they could have anticipated hypoglycemia by performing new blood glucose measurements earlier than dictated by the local guideline for blood glucose control, allowing them to respond earlier to e.g., hypoglycemia. Still, some hypoglycemic events occurred, probably because not all nurses were paying attention to the readings of the investigational device. In addition, nurses could have noted that its point accuracy was not always good, so they could have mistrusted the device readings. Finally, we cannot exclude the possibility that hypoglycemia can occur even with the use of CGM. The later will be subject of planned trial. An accuracy analysis limitation was that the assessment focused on percent difference comparisons between the continuous sensor and discrete reference points, evaluated to standards meant for discrete measurements for dosing. Another important limitation is that trend accuracy was not evaluated. Trending is the most interesting endpoint, but mandates very short intervals (i.e., a short as 15 minutes) between blood glucose reference measurements [32, 40]. Trend accuracy should and will be evaluated in future studies.

Notably, length of stay in ICU and sensor connection time was far from similar. This was caused by the fact that sensors could not be used before informed
consent was obtained. Thus we may have missed an important phase of glucose control (i.e., the first day or days of stay in the ICU). In addition, one third of the sensors were removed before sensor life ended, because of sensor – or device related factors. This is an important problem for the reliability of the device. However, nurses did not always attempt to solve sensor – or device related problems that could have been solved. During conduct of the trial they were always allowed to remove the sensor because of 'Poor Sensor Signal'–alerts or recurrent alarms. With increasing device–specific experience it could be that there are fewer, early removals.

CONCLUSIONS

The point accuracy of a device designed for continuous real–time monitoring of the interstitial glucose level did not meet the ISO15197 standard or the recent consensus guidance for discrete glucose measurement for dosing when used on critically on critically ill patients admitted to a mixed medical–surgical ICU. While this device is not a replacement for current blood gas analyzer–measurements, a real–time system may be used for trend guidance on timely reference measurement for insulin adjustment. The device had few down times, but one third of the sensors were removed prematurely because of unresolved sensor – or device–related problems.

References


12: R120.


FIGURES

Patients screened
N = 790

Excluded:
- discharge < 24 hours     N = 531
- No Informed Consent     N = 61
- life expectancy < 96 hours     N = 42
- missed     N = 40
- platelet count < 30     N = 21
- doctor exclusion     N = 16
- not willing to participate     N = 15
- in other trial     N = 5
- pregnant     N = 3
- no monitor available     N = 2
- age < 18 years     N = 2
- readmission already included     N = 20

Included patients
N = 50

Comparative samples < 4     N = 5

Included in the point accuracy analysis
N = 45

Figure 1.
CONSORT diagram of the study.

Figure 2.
Bland–Altman plot with bias and limits of agreement (bias ± 1.96 x standard deviation of the bias)[30], glucose prediction errors, and Clarke error grid analyses[31].
Figure 3.
Surveillance error grid with risk scores.
# TABLES

<table>
<thead>
<tr>
<th></th>
<th>N=50 All included patients</th>
<th>N=45 Patients included in the point accuracy analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age – years, median [IQR]</td>
<td>65 [56–72]</td>
<td>65 [55–72]</td>
</tr>
<tr>
<td>Male gender, number (%)</td>
<td>25 (50%)</td>
<td>24 (53%)</td>
</tr>
<tr>
<td>Race, number (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Caucasian</td>
<td>45 (90%)</td>
<td>40 (89%)</td>
</tr>
<tr>
<td>– Black</td>
<td>4 (8%)</td>
<td>4 (9%)</td>
</tr>
<tr>
<td>– Asian</td>
<td>1 (2%)</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>BMI – kg/m², median [IQR]</td>
<td>24.7 [22.4–27.6]</td>
<td>24.4 [22.2–27.3]</td>
</tr>
<tr>
<td>Admission diagnosis, number (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Medical</td>
<td>31 (62%)</td>
<td>26 (58%)</td>
</tr>
<tr>
<td>– Emergency surgery</td>
<td>11 (22%)</td>
<td>11 (24%)</td>
</tr>
<tr>
<td>– Planned surgery</td>
<td>8 (16%)</td>
<td>8 (18%)</td>
</tr>
<tr>
<td>Planned admission, number (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– No diabetes</td>
<td>39 (78%)</td>
<td>34 (76%)</td>
</tr>
<tr>
<td>– Diabetes, unknown treatment</td>
<td>2 (4%)</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>– Diabetes treated with insulin</td>
<td>4 (8%)</td>
<td>4 (9%)</td>
</tr>
<tr>
<td>– Diabetes treated with oral agents</td>
<td>5 (10%)</td>
<td>5 (12%)</td>
</tr>
<tr>
<td>ICU mortality, number (%)</td>
<td>11 (22%)</td>
<td>10 (22%)</td>
</tr>
<tr>
<td>Hospital mortality, number (%)</td>
<td>15 (30%)</td>
<td>14 (31%)</td>
</tr>
</tbody>
</table>

Table 1.
Patient Characteristics. Abbreviations: APACHE, Acute Physiology and chronic Health Evaluation; BMI, body mass index; IQR, interquartile range; ICU, intensive care unit; LOS, length of stay; SAPS, Simplified Acute Physiology Score;
Number of measurements 929
Mean blood glucose level per patient – mg/dL (median [IQR]) 132 [125–148]
Standard deviation of blood glucose level per patient – mg/dL (median [IQR]) 24 [16–33]
Number of measurements per patient (median, [IQR]) 11 [7–29]
Severe Hypoglycemia ≤ 40 mg/dL – measurements, number (%) 3 (0.3%)
Severe Hypoglycemia ≤ 40 mg/dL – patients, number (%) 2 (4.4%)
Mild hypoglycemia 41 – 70 mg/dL – measurements, number (%) 15 (1.6%)
Mild hypoglycemia 41 – 70 mg/dL – patients, number (%) 7 (15.6%)
Mild hyperglycemia 150 – 179 mg/dL – measurements, number (%) 163 (17.5%)
Mild hyperglycemia 150 – 179 mg/dL – patients, number (%) 35 (77.8%)
Severe hyperglycemia > 180 mg/dL – measurements, number (%) 111 (11.9%)
Severe hyperglycemia > 180 mg/dL – patients, number (%) 19 (42.2%)

Table 2.
Measurements. Data considers all paired measurements and result of blood gas analyzer is shown; Abbreviations: IQR, interquartile range
<table>
<thead>
<tr>
<th></th>
<th>Per patient</th>
<th>Per sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of sensors used</td>
<td>–</td>
<td>105</td>
</tr>
<tr>
<td>Number of sensors used, median [IQR]</td>
<td>1 [1–3]</td>
<td>–</td>
</tr>
<tr>
<td>Total connection time in hours, median [IQR]</td>
<td>46.2 [26.8–134.2]</td>
<td>45.8 [21.1–69.1]</td>
</tr>
<tr>
<td>Initialization time in minutes, median [IQR]</td>
<td>Median 34 [34–34.5]</td>
<td>34 [34–35]</td>
</tr>
<tr>
<td>Real-time data in hours</td>
<td>42.3 [23.1–130.3]</td>
<td>41.4 [20.6–64.0]</td>
</tr>
<tr>
<td>Percentage of real-time data, median [IQR]</td>
<td>94.1 [88.9–97.1]</td>
<td>94.6 [88.7–97.9]</td>
</tr>
<tr>
<td>Time of skips in data acquisition in hours, median [IQR]</td>
<td>4.3 [1.2–9.1]</td>
<td>2.6 [0.6–5.4]</td>
</tr>
<tr>
<td>Time of skips in data acquisition in hours caused by poor sensor signal, median [IQR]</td>
<td>0 [0–1.0]</td>
<td>0 [0–0.2]</td>
</tr>
<tr>
<td>Time of skips in data acquisition in minutes caused by other reasons, median [IQR]</td>
<td>3.3 [0.9–8.4]</td>
<td>2.0 [0.4–3.7]</td>
</tr>
<tr>
<td>Percentage of time of skips in data acquisition, median [IQR]</td>
<td>5.9 [2.9–11.1]</td>
<td>5.4 [2.1–11.3]</td>
</tr>
<tr>
<td>Percentage of time of skips in data acquisition in caused by poor sensor signal, median [IQR]</td>
<td>0 [0–0.7]</td>
<td>0 [0–0.3]</td>
</tr>
<tr>
<td>Percentage of time of skips in data acquisition caused by other reasons, median [IQR]</td>
<td>4.2 [2.3–8.0]</td>
<td>3.8 [1.5–8.0]</td>
</tr>
<tr>
<td>Number of calibrations, median [IQR]</td>
<td>14 [9–34]</td>
<td>12 [7–16]</td>
</tr>
<tr>
<td>Number of mandated calibrations, median [IQR]</td>
<td>8 [4–20]</td>
<td>6 [4–8]</td>
</tr>
</tbody>
</table>

Table 3. Device reliability. Abbreviations: IQR, interquartile range
Total number of sensors used 105
Total number of sensors removed < 72 hours 60

<table>
<thead>
<tr>
<th>Number of sensors</th>
<th>Percentage of sensors removed &lt; 72 hours</th>
<th>Percentage of total number of sensors used</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient–related factors:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Discharge &lt; 72 hours after insertion</td>
<td>14</td>
<td>(23%)</td>
</tr>
<tr>
<td>– Death, 72 hours after insertion</td>
<td>4</td>
<td>(7%)</td>
</tr>
<tr>
<td><strong>Sensor– or device– related factors:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Accidental removal of sensor</td>
<td>6</td>
<td>(10%)</td>
</tr>
<tr>
<td>– Poor sensor signal (19 had no attempt to resolve)</td>
<td>34</td>
<td>(57%)</td>
</tr>
<tr>
<td>– Device error</td>
<td>2</td>
<td>(3%)</td>
</tr>
</tbody>
</table>

**Duration of sensors in place (hours), median [IQR]**

| – All sensors | 46 [21–69] |
| – Sensor that were removed < 72 hours | 22.1 [13.7–35.3] |
| – Sensors that were removed < 72 hours because of patient– related factors | 27.0 [21.3–41.47] |
| – Discharge < 72 hours after insertion | 24.7 [21.0–42.6] |
| – Death, 72 hours after insertion | 30.4 [26.5–34.8] |
| – Sensors that were removed < 72 hours because of patient– related factors | 18.9 [9.9–31.7] |
| – Accidental removal of sensor | 29.4 [20.7–30.7] |
| – Poor sensor signal | 19.4 [11.3–32.8] |
| – Device error | 8.0 [7.4–8.5] |

**Table 4.**
Sensors removed < 72 hours
SUPPLEMENTARY INFORMATION

The local guideline for glucose control
ICU nurses performed glucose control with insulin, following a local guideline for blood glucose control targeting a blood glucose level between 90–144 mg/dL [29]. According to this local guideline, insulin infusion was started when the blood glucose level was > 144 mg/dL. Insulin titration adjustments were made based on sliding scales. The guideline advised to stop insulin infusion and to give boluses of dextrose only when the blood glucose level declined to < 61 mg/dL. Insulin infusion was exclusively given intravenously and continuously; boluses of insulin were only allowed when the blood glucose level was > 360 mg/dL; subcutaneous insulin boluses were never allowed.
The local guideline also dictated to perform blood glucose measurements at least every four hours, but more frequently if blood glucose levels were out of range or rapidly changing. Typically, blood glucose levels were measured more frequently at the start of insulin titration, and in cases of an increased risk of hypoglycemia. Blood glucose levels used for insulin adjustment were measured in arterial blood samples using RapidLab 1265 blood gas analyzers (Siemens Healthcare Diagnostics, The Hague, The Netherlands), located in the ICU. The results were automatically downloaded to the patient data management system (MetaVision®, iMDsoft, Tel Aviv, Israel), which was present at every ICU bed.

Methods to calculate point accuracy
For reporting point accuracy we used glucose prediction errors, defined as [blood glucose – device glucose result]. The percentage of data points that fell within ± 15 mg/dL of the blood glucose results for blood glucose results < 75 mg/dL and within 20% of the blood glucose results for blood glucose results ≥ 75 mg/dL were reported according to the current International Standards Organization standard (ISO15197) [32]. We also used Clarke error grid analyses to show the percentage of paired data values falling within each zone of the Clarke error grid [30], and Bland–Altman plot [31]. The Clarke error grid is divided in 5 paired ‘zones’: zones A (measurement within 20% of the reference or glucose levels < 70 mg/dL); zones B (measurement more than 20% different from the reference but still clinically acceptable as they would not change the rate of insulin infusion); zones C (measurement that would lead to unnecessary changes in insulin infusion, i.e., overcorrecting acceptable glucose levels); zones D (potentially dangerous hypo- or hyperglycemic events are missed); and zones E (levels that would lead to a decision opposite to that required, i.e., treatment for hypoglycemia instead of hyperglycemia). General consensus is that 95% of the values should be in zones A and 5% in zones B.
The Bland–Altman plot is presented with bias (mean difference between the device glucose results and blood glucose results) and limits of agreement (bias ± 1.96 x standard deviation of the bias) to analyze the agreement between the device glucose results and blood glucose results. In a post–hoc analysis we also determined point accuracy according to the recently published consensus recommendations [1]. For this, the percentage of data points that fell within 12.5% of the blood glucose results, or within 10 mg/dL for readings < 99 mg/dL were reported. In a round the table meeting of ICU experts it was recommended to report the mean absolute relative difference and values should be <14%; values >18% were considered to represent poor accuracy [32]. Furthermore, we analyzed the accuracy following the recently published the surveillance error grid [33].

**Definitions of metrics for device reliability**

The following metrics and definitions were used to assess device reliability, including those suggested by recent consensus recommendations [1]:

- Connection time – time between first device glucose results and last glucose result
- Start–up time – time between the start of initialization of sensor and first device glucose result after calibration, including blood glucose measurement time and time for nurse to enter value into the device.
- Initialization time – time between initialization of sensor and ready for calibration
- Real-time data – time when device glucose results were available
- Percentage of real-time data – percentage of time device glucose results were available divided by total connection time
- Skips in data acquisition all causes – total time when the monitor gave no results
- Percentage of skips in data acquisition all causes – percentage of time when the monitor gave no results divided by the connection time
- Skips in data acquisition poor sensor signal – percentage of time when the monitor gave no results caused by poor sensor signal
- Percentage of skips in data acquisition poor sensor signal – percentage of time of skips in data acquisition caused by poor sensor signal device divided by the connection time minus the time of skips in data acquisition caused by other reasons
- Skips in data acquisition other reasons – time of skips in data acquisition caused by other reasons than poor sensor signal
- Percentage of skips in data acquisition other reasons – percentage of time of skips in data acquisition caused by other reasons divided by the connection time minus time of skips in data acquisition caused by poor
• ‘Poor Sensor Signal’ – a device alert indicating that the sensor may be experiencing decreased performance. This alert removes the real time sensor glucose value display until a requested reference calibration value is entered to recover sensor performance.

**FACTORS THAT AFFECT POINT ACCURACY**

**Background**
The aim of the primary study was to test the point accuracy and reliability of an interstitial CGM device in a mixed medical–surgical ICU. We found a low point accuracy of an interstitial CGM device in a mixed medical–surgical ICU. We were interested if this was dependent on particular variables. Therefore, we performed a post-hoc analysis to determine which variables influence the accuracy of the device.

**Methods**
We used a linear mixed model to determine which variables influence the accuracy of the device. For this, patient and sensor were used as random intercepts to account for repeated measurements. The absolute difference between the arterial blood glucose level and device glucose level was the dependent variable. The absolute difference was logarithmically transformed (using the natural logarithm) to obtain a normal distribution. The following variables were chosen based on clinical relevance and previous trials testing other CGM devices [18, 19, 28]: demographic variables including gender, age, body mass index and history of diabetes; disease severity variables including the APACHE II score and the circulation score of the Sequential Organ Failure Assessment (SOFA) Score on the day of measurements; in addition, we added time between calibrations (as shorter time between calibrations could improve accuracy) and the rank order of the paired glucose results (as more calibrations could improve accuracy) [18]. All variables were added to the model without considering further model reduction strategies. Visual inspection of residuals was done. Correlation between covariates was assessed to investigate collinearity. The effect of covariates on the absolute difference was reported as the percentage of change in the absolute difference with the standard error.

**Results** (table S1)
We performed a linear mixed-effects model with a fit by maximum likelihood. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. Pearson correlation coefficients were all under 0.5 showing no collinearity.
In the linear mixed model only history of diabetes \((P = 0.02)\) and number of calibrations per sensor \((P = 0.04)\) affected the absolute difference between blood glucose and device result. Per each new calibration the absolute difference decreased with 1.4\% (standard error of 0.006\%), meaning that the sensor performance increased. The effect of a history of diabetes was bigger, though, since diabetes increased the absolute difference with 34.3\% (standard error of 13.0\%). Therefore we stratified our accuracy metrics by diabetic status (see figure S1 and table S2).

The formula for the final mixed model was:
\[
\log(\text{Absolute difference}) = 2.419 + \text{random intercept per patient} + \text{random intercept per sensor} + 0.295 \times \text{Diabetes} - 0.014 \times \text{rank order of measurement} + 0.011 \times \text{Sofa Circulation Score} + 0.025 \times \text{BMI} - 0.073 \times \text{Gender} - 0.003 \times \text{Age} - 0.009 \times \text{APACHE II} + 0.0001 \times \text{time between calibration}
\]

**RELIABILITY ANALYSIS**

**Background**
In the present study we found that more than half of the sensors had to be removed before 72 hours. We wanted to know reasons for disconnection and when this happened. Therefore we did a post–hoc analysis to investigate reasons for early disconnection.

**Methods**
Early disconnection was defined as the removal of a sensor before 72 hours, which could be caused by:

- Poor sensor signal – sensor performance issue, in which the system requests additional calibrations to solve. Nurses were able to remove the sensor when the monitor gave a poor sensor signal alarm without attempt to solve.
- Accidental removal of the sensor
- Device error – the device monitor had a technical failure

Furthermore the connection time was calculated (time between first device glucose results and last glucose result) for sensors, which were removed before 72 hours. The time between calibrations using an incorrect glucose value entry and the next calibration was extracted from the total connection time of the device.
Figure S1.
Bland–Altman plot with bias and limits of agreement (bias ± 1.96 standard deviation of the bias), glucose prediction errors, and Clarke error grid analyses stratified by diabetic status.
<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient ID</td>
<td>0.036</td>
<td>0.189</td>
</tr>
<tr>
<td>Sensor ID</td>
<td>0.047</td>
<td>0.217</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Value</th>
<th>Standard error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.419</td>
<td>0.489</td>
<td>0.000</td>
</tr>
<tr>
<td>History of Diabetes</td>
<td>0.295</td>
<td>2.408</td>
<td>0.021</td>
</tr>
<tr>
<td>Rank order</td>
<td>-0.014</td>
<td>0.007</td>
<td>0.037</td>
</tr>
<tr>
<td>Sofa circulation score</td>
<td>0.011</td>
<td>0.026</td>
<td>0.681</td>
</tr>
<tr>
<td>BMI</td>
<td>0.025</td>
<td>0.015</td>
<td>0.107</td>
</tr>
<tr>
<td>Male gender</td>
<td>-0.073</td>
<td>0.109</td>
<td>0.506</td>
</tr>
<tr>
<td>Age in years</td>
<td>-0.003</td>
<td>0.004</td>
<td>0.524</td>
</tr>
<tr>
<td>APACHE II score</td>
<td>-0.009</td>
<td>0.007</td>
<td>0.204</td>
</tr>
<tr>
<td>Time between calibration in minutes</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.629</td>
</tr>
</tbody>
</table>

Table S1.

<table>
<thead>
<tr>
<th></th>
<th>Diabetic</th>
<th>Non-diabetic</th>
<th>All patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of paired samples</td>
<td>337</td>
<td>592</td>
<td>929</td>
</tr>
<tr>
<td>Mean absolute relative difference</td>
<td>16.0</td>
<td>14.2</td>
<td>14.8</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>0.84</td>
<td>0.71</td>
<td>0.81</td>
</tr>
<tr>
<td>R2</td>
<td>0.70</td>
<td>0.50</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Consensus recommendations

- percentage of measurements within 12.5% blood glucose results (or within 10 mg/dL for results < 99 mg/dL)
  - Diabetic: 55
  - Non-diabetic: 59
  - All patients: 58

- percentage of measurements within 20% blood glucose results
  - Diabetic: 72
  - Non-diabetic: 77
  - All patients: 75

Table S2.