All Insulin Pumps Are Not Equivalent: 
A Bench Test Assessment for Several Basal Rates

Sylvain Girardot, PhD,1–3 Pauline Jacquemier, MS,1–3 Flavien Mousin, MS,1 Carine Rendekeu, MS,1 Sébastien Hardy, MS,1 and Jean-Pierre Riveline, MD, PhD2–4

Abstract

Background and Aims: Continuous subcutaneous insulin infusion (CSII) is a widely adopted treatment for type 1 diabetes and is a component of an artificial pancreas. CSII accuracy is essential for glycemic control, however, this metric has not been given sufficient study, especially at the range of the lowest basal rates (BRs), which are commonly used in a pediatric population and in closed-loop systems (CLSs). Our study presents accuracy results of four off-the-shelf CSII systems using a new accurate method for CSII system evaluation.

Materials and Methods: The accuracy of four off-the-shelf CSII systems was assessed: Medtronic MiniMed 640G®, Ypsomed YpsoPump®, Insulet Omnipod®, and Tandem t:slim X2®. The assessment was performed using a double-measurement approach through a direct mass flow meter and a time-stamped microgravimetric test bench combined with a Kalman mathematical filter. CSII accuracy was evaluated using mean of dose error. Mean absolute relative difference (MARD) of error was calculated at different observation windows over the whole series of tests. Peakwise insulin deliverance was assessed regarding stroke regularity in terms of frequency and volume.

Results: Mean error values indicate a general tendency to underdeliver with up to −16%. MARD of error shows very wide results for each pump and each BR from 7.4% (2 UI/h) to 61.3% (0.1 UI/h). Peakwise analysis shows several choices for BR adaptation (frequency for Omnipod, volume for Tandem, both for YpsoPump and MiniMed 640G). Precision in interstroke time appears to be better (standard deviation [SD] at 0.1 UI/h: 4.6%–12.9%) than stroke volume precision (SD at 0.1 UI/h: 38.3%–46.4%).

Conclusions: The accuracy of four off-the-shelf CSII systems is model and BR dependent. CSII imprecision could be due to a variability in volume and/or frequency of strokes for every pump. Some models appear better adapted for the smallest insulin needs, or for inclusion in a CLS. The clinical implications of these delivery errors on glucose instability must be evaluated.

Keywords: Insulin pump, Artificial pancreas, Precision, Type 1 diabetes, Performance, Test bench.

Introduction

Continuous subcutaneous insulin infusion (CSII) is one of the gold standards for achieving glucose control in patients with type 1 diabetes (T1D).1–3 CSII is also a key component of closed-loop systems (CLSs).4 CSII allows insulin administration to be automatically adapted each 5 to 15 min according to several variables such as continuous glucose measurement (CGM), insulin-on-board, and physiologic parameters,5,6 which is done to optimize time-in-range (TIR). However, glucose control in patients treated with traditional CSII remains imperfect7,8 and normoglycemia in patients using CLS (i.e., TIR [70,180] >70%) is not always achieved.6,9

This statement has many explanations, such as continuous glucose measurement (CGM) imprecision, erratic insulin subcutaneous absorption, unstable lifestyle, and so on.10,11 CSII imprecision could also be a factor, and to-date, very few
independent studies have evaluated the accuracy of insulin administration system.

There are several pumps on the market, and all are able to deliver insulin at various basal rates (BRs). Despite consistent improvements over the years, the international gold standard for the assessment of CSII, recognized by many authorities along with introduction to various markets, do have some limitations. In brief, this standard is quite detached from daily life constraints, and tests rely on an indirect means of measurement with a low frequency of data acquisition. Moreover, the standard does not include compulsory testing for the lowest BR, although CSII are commonly used in both the pediatric population and CLS.

In this context, as previously suggested by Heinemann et al., the available evidence on the safety and efficacy of CSII remains limited. Due to this lack of evidence, the accuracy of insulin administration of pumps available on the market requires more in-depth evaluation, especially at different BRs. Moreover, CSII misprecision could be a potential factor for the persistent glycemic instability observed in T1D patients treated by CSII.

We recently introduced a new bench test method that is able to reach high-precision measurement of insulin administration at several BRs, including the lowest levels. This new method provides both continuous measurement of insulin flow rate and insulin volume with good uncertainties. In the present study, four off-the-shelf insulin pumps were tested at several BRs (2, 1, 0.5, and 0.1 UI/h) using the new method. The pumps were a Medtronic Minimed 640G, a Ypsomed YpsoPump, an Insulet Omnipod, and a Tandem t:slim X2.

**Materials and Methods**

The methods have been precisely described in reference. The direct mass flow meter BL100 and weighing scale XPE56

The measurement devices included a Mettler Toledo XPE56 weighing scale and a BL100 direct mass flowmeter by Bronkhorst according to the setup designed and described in a previously published article. In brief, the CSII under study is connected to the flowmeter by standard catheter tubing as shown in Figure 1. The flowmeter’s exit is connected to similar tubing that plugs hermetically to an 18G, 1.2 mm, ISO 7864 needle immersed in a test tube on the weighing scale. A thin oil layer prevents any evaporation from occurring inside the tube during the several hours of the test. As described in a previous article, numerous microfluidic phenomena as well as environmental constraints relative to measurement were controlled before and during each test.

**Kalman filter-based assessment method**

A Kalman filtering method combining signals from both devices was implemented, as detailed in a previously published study. Here, the Kalman filter minimizes the mean square error of the two measured parameters. This step allows for the high accuracy of the weighing scale to compensate for the uncertainty of the flowmeter measurement for the very low flowrates under study. Similarly, the low acquisition frequency of the balance is compensated by the high frequency of the flowmeter acquisition. The output of the Kalman filter is composed of two signals, a corrected cumulated mass signal and a corrected flow rate signal.

As displayed in our previous article, no external validation protocol allowed to approach the imprecision of our measurement method. Nevertheless, we took care of using the Kalman estimator within a proper mathematical framework, and we assessed the estimation output with mathematical indicators to approach the validity of our results. More importantly, the general insulin pump precision results we provide in this study (Figs. 2 and 3) rely on the weighing scale for which uncertainties are totally mastered and provided in the method article.

**Design of experiment**

Three wired pumps, MiniMed 640G (Medtronic), YpsoPump (Ypsomed), and t:slim X2 (Tandem), and one patch pump, Omnipod (Insulet), were assessed. All four were tested under four BRs: 2, 1, 0.5, and 0.1 UI/h. Each test lasted for 8 h at a constant BR, and was reproduced four times each. The insulin used was Novorapid (Novo Nordisk). The overall results of sixty-four 8 h tests are presented hereunder.

**Numerical indicators**

**Mean error.** For all tests, the mean error in insulin delivery was computed over the whole 8 h test (Eq. 1). The difference between actual insulin dose (AID) and expected
insulin dose (EID) was considered for each $i$-th interval upon the total $n$ time intervals.

\[
\text{Mean (Error)} = \sum_{i=1}^{n} \left[ \frac{|\text{AID}_i - \text{EID}_i|}{\text{EID}_i} \cdot 100 \right] \cdot \frac{1}{n} \quad (Eq. 1)
\]

Mean absolute relative difference error. Also, the mean absolute relative difference (MARD) between AID and EID was computed (Eq. 2) for each $i$-th interval upon the total $n$ time intervals of the tests. This was done so that successive underdeliverance and overdeliverance would not compensate for each other, as they do for simple mean computations.

\[
\text{MARD (Error)} = \sum_{i=1}^{n} \left[ \frac{|\text{AID}_i - \text{EID}_i|}{\text{EID}_i} \cdot 100 \right] \cdot \frac{1}{n} \quad (Eq. 2)
\]

**FIG. 2.** Boxplot of mean error and MARD. CSII system accuracy results for each BR with mean (A) and 15-min observation windows MARD errors (B) grouped by pump model. BR, basal rate; CSII, continuous subcutaneous insulin infusion; MARD, mean absolute relative difference.

**FIG. 3.** Evolutive MARD. For each pump model, MiniMed 640G (A), YpsoPump (B), Omnipod (C), t:slim X2 (D), and each set BR, MARD error results is calculated for several time windows intervals along the four replicas. Evolutive MARD is the graphic representation of MARD for several observation windows in each situation.
MARD was also computed for different observation windows (30, 60, 120, 240 min) over the duration of the whole 8 h experiments. The resulting MARD was plotted according to the size of the considered time window for each pump and BR to evaluate if errors were being compensated for larger observation windows. Indeed, all sub-over deliveries comprised within the length of those time windows are then being compensated.

**Insulin stroke analysis**

To decipher more precisely the mechanism of insulin pump accuracy, we evaluated insulin stroke frequency and amplitude reproducibility. From the Kalman-filtered flowrate signal, local maxima of insulin flowrate was spotted, allowing the identification of stroke positions, which was done using an algorithm detailed in the Supplementary Data. This allowed us to compute for each stroke of the injected volume and the interstroke time to evaluate the intrinsic variability of insulin administration.

**Results**

**Global error assessment**

Mean error and MARD were different according to BR and according to the pump as shown by boxplots in Figure 2.

The values of mean error percentage shown in Figure 2A are mostly negative and suggest a tendency to under deliver.

Values for the 15 min observation windows for MARD presented in Figure 2B highly increase as BR drops, exhibiting a very high difference between the set point and actual injected volume. At 2 UI/h BR, mean MARD of Ypsopump and Tandem were close, 7.9% and 7.4%, respectively, while MiniMed 640G had the highest value, although with a mean MARD of 11.8%. Some important differences appear between pumps as BR decreases. For 15 min observation windows, Omnipod mean MARD reached 61.3% at 0.1 UI/h, while t:slim X2 had the smallest mean MARD results with a mean MARD at 22.7% (Fig. 2B). A table of these values can be found in the Supplementary Data.

The evolution of MARD for different observation windows (15, 30, 60, 120, and 240 min), presented in Figure 3, differs between pump models and BR values. Overall, we observed two distinct behaviors. First, when positive errors immediately compensate negative errors, the resulting MARD decreases when observations windows become wider.

This explains why some MARD evolution plots quickly plummet. For instance, the error of Omnipod at 0.1 UI/h was a 68% error with 15 min observation windows versus 19% at a 60 min time interval. Later in our discussion, we will refer to this phenomenon as short-term inaccuracy.

Second, some errors, either positive or negative, are maintained over longer time windows, sometimes even over the whole 8 h test. MARD does not evolve much when the observation window gets wider, making the MARD evolution plot nearly flat (Fig. 3: Omnipod 0.5 UI/h 23% error at 15 min interval windows against 21% error at a 60 min time interval). This behavior will be later referred to as long-term inaccuracy.

**Evaluation of stroke amplitude and frequency in each pump**

To decipher causes for the inaccuracies, we evaluated the volume and frequency of each stroke according to BR and pump. As observed in Figure 4, pump models adopt different strategies to adjust BR. Omnipod made the choice to maintain stroke volume and adapt interstroke time. Oppositely, t:slim X2 adjusts stroke volume while keeping the same interstroke value for each set BR. The other models, MiniMed 640G and YpsoPump, adopted both strategies changing either interstroke time value or insulin stroke volume depending on BR range. Also, Figure 4 highlights the overall imprecision of different mechanisms: every pump model shows approximately the same intrinsic stroke volume variability except for MiniMed 640G at 2 UI/h. However, this model had unclear delimitied strokes, allowing a less precise discrimination between strokes, whereas the segmentation was performed by the human eye or by use of our algorithm (Supplementary Data). A good reproducibility for interstroke time at all BR was observed except for the YpsoPump and MiniMed 640G at 0.1 UI/h BR. Stroke volume variability was overall wider. As an example, standard deviations of observed stroke volumes at 0.1 UI/h reach 38.3% (MiniMed 640G), 51.1% (Omnipod), 36.6% (YpsoPump), and 46.4% (t:slim X2) of the mean stroke volume value while interstroke time variability at 0.1 UI/h is 12.9% (MiniMed 640G), 4.6% (Omnipod), 5.7% (YpsoPump), and 7.5% (t:slim X2) of the mean interstroke time value. Insulin stroke features can be found in Table 1.

**Discussion**

This study evaluates for the first time, off-the-shelf CSII accuracy for results at several BRs, from 0.1 to 2 UI/h, using a high precision and reactive method. Overall, our results showed a general tendency to underdeliver with up to −16% of the announced insulin volume delivered over 8 h. The three manufacturers claim error ranges within ±5% in basal delivery for all BRs to be delivered. Only YpsoMed indicates an error of ±30% (at 23°C±2°C) for 0.1 UI/h BR and below, and this statement matches with our observations. We also demonstrated a huge heterogeneity of CSII accuracy according to BR and CSII model. Using usual BRs (1 to 2 UI/h), MARD goes from 8% (Ypsopump and t:slim X2) to 21% (Omnipod). At lowest BRs, MARD errors become commonly higher, from 22% for t:slim X2 to 61.3% for Omnipod at 0.1 UI/h. We also generally observed an overall less reproducible error for lower BRs. Even a few cases of overdelivery were observed.

CSII accuracy has been previously evaluated by dependent and independent teams testing several CSII models with numerous methodologies. Their conclusions generally agreed on a CSII with lower precision for the smallest BRs and on the fact that accuracy is model dependent. However, some major differences between results exist in the literature. For example, Omnipod overall precision was found to be highly accurate by Zisser et al., while others found it inaccurate. Our results are unique to the best of our knowledge for several reasons: we addressed some CSII models that have not been yet assessed, such as t:slim X2 and YpsoPump. In addition, accuracy using a BR below 0.5 UI/h had never been tested. Most importantly, we applied a leading-edge measurement method that is able to measure insulin flow precisely and continuously.

The smallest flowrates are intuitively the hardest to deliver correctly for CSII-based microfluidic systems. Therefore,
testing CSII especially at those BR is necessary. It is clear that CSII is known to improve glycemic control\textsuperscript{23,24} even in a pediatric population, even under low BR settings.\textsuperscript{25} However, it appears that some glycemic variabilities (GVs) do persist, and hypoglycemia are still being observed even when the amount of severe hypoglycemia is being reduced by the use of CSII.\textsuperscript{7,23,26,27} More surprisingly, a 2009 meta-analysis shows a tendency to higher rates of hypoglycemia in children under CSII compared to multi-daily injections.\textsuperscript{28} This appears to be confirmed by continuous glucose measurement (CGM) clinical trials who still observe a GV and hypoglycemic episodes in patients treated by CSII, although the use of the CGM also contributes to reducing GV in those patients.\textsuperscript{29,30}

**FIG. 4.** Insulin stroke regularity. The stroke unit volume is displayed according to its corresponding interstroke time, for all observed strokes. The uniaxial distributions of each test situation are represented with gaussians.

<table>
<thead>
<tr>
<th>BR (UI/h)</th>
<th>Minimed 640 G</th>
<th>Omnipod</th>
<th>Ypsopump</th>
<th>t:slim X2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Volume per stroke (UI) mean (SD)</td>
<td>Stroke period (min) mean (SD)</td>
<td>Volume per stroke (UI) mean (SD)</td>
<td>Stroke period (min) mean (SD)</td>
</tr>
<tr>
<td>2 UI/h</td>
<td>0.28 (0.086)</td>
<td>8.54 (1.67)</td>
<td>0.049 (0.011)</td>
<td>1.50 (0.079)</td>
</tr>
<tr>
<td>1 UI/h</td>
<td>0.046 (0.014)</td>
<td>2.82 (0.50)</td>
<td>0.050 (0.014)</td>
<td>2.99 (0.15)</td>
</tr>
<tr>
<td>0.5 UI/h</td>
<td>0.024 (0.017)</td>
<td>3.05 (0.47)</td>
<td>0.050 (0.040)</td>
<td>5.98 (0.33)</td>
</tr>
<tr>
<td>0.1 UI/h</td>
<td>0.024 (0.009)</td>
<td>14.6 (1.89)</td>
<td>0.045 (0.023)</td>
<td>29.7 (1.38)</td>
</tr>
</tbody>
</table>

Insulin stroke regularity is calculated for each pump model at each BR regarding interstroke time and stroke volume regularity. BR, basal rate; SD, standard deviation.
Today, a few studies have discussed clinical consequences of CSII errors: Heinemann et al. showed that “in adult with normal insulin requirement (0.12 UI/kg), changing BR from 1 to 2 UI/h had an effect on glucose infusion rate after 30–60 min. For the same patient type, BR from 0.1 to 0.5 UI/h did not reveal any particular systemic consequences.”

However, many studies raise the question of the impact of CSII accuracy for patients with low insulin needs using small BR ranges. A more recent simulation study also demonstrated that a variation of 0.1 UI/h in BR or 0.3 UI in bolus within 4 h time span has also a systemic incidence. Hence, more and more insights suggest to reinforce knowledge on CSII delivery behavior, especially for patients using low BR, which has been our goal in this study.

The origins of the inaccuracies we observed remain unknown. The analysis and the comparison of insulin stroke deliverance parameters, such as frequencies and stroke volumes, for each CSII models at several BR ranges, could bring a piece of the answer. Our method provides for the first time a continuous measurement method that is able to study peakwise CSII accuracy.

Insulin stroke analysis (Fig. 4) shows that CSII imprecisions appear to come from variability in insulin strokes frequency. This phenomenon is accentuated when BR is reduced and is of various significance depending on pump model. Indeed, interstroke time is observed as having a larger spread for smaller BRs, especially for MiniMed 640G and YpsoPump at 0.1 UI/h. Oppositely, t:slim X2 showed the lowest variability in interstroke time for the same BR. Variabilities in stroke volumes appear to be similarly spread whatever the BR. In the specific case of MiniMed 640G at 2 UI/h, the stroke volume appears to be exceptionally spread. At 2 UI/h, this model showed unclear delimited strokes (Supplementary Data), and its flowrate oscillates around an average value. However, it presents no extensive delivery error, looking at both the 15 min MARD and the mean error (Fig. 2). Therefore, we have no element to establish the clinical relevance of this delivery.

Figure 4 also suggests that different strategies were chosen by the various manufacturers regarding the key parameter used to modulate BR. For instance, Omnipod clearly seems to use a variation in stroke frequency with constant stroke volume to modify BR, whereas t:slim X2 apparently chose to keep a constant stroke frequency and to change the stroke volume to achieve this. YpsoPump and MiniMed 640G seem to use a combination of both strategies, using stroke volume adaptation for the highest range of BRs and interstroke adaptation for the lowest. These observations are made clear by analyzing the relative positions of each cloud of points for each pump model in Figure 4.

Manufacturer choices and CSII accuracy might have clinical consequences. Reducing BR by acting on stroke frequency implicates that in this situation, strokes as rare as one every 30 min could be observed.

In the results section, we introduced two behaviors with regard to how MARD plots evolve according to time, namely short-term instability (YpsoPump and Omnipod at 0.1 UI/h) and long-term instability (MiniMed 640G at 0.5 and 0.1 UI/h and Omnipod at 0.5 UI/h). To date, no evidence has established a causality between insulin administration errors and glycemic instability, whether short-term or long-term. One might indeed argue that successive over- and underdeliveries lasting less than 15 min (short-term instability) could physiologically have no clinical impact at all. Conversely, long-term instability could foster overall glycemic instability, as it implies long exposures to either higher doses or lower doses of insulin, and naturally lead to a succession of hypoglycemia and hyperglycemia.

CSII short-term instability would appear acceptable compared to long-term as long as the global mean error remains low and the plateau shape is reached as fast as possible. In the opposite case, long-term instability could be acceptable since a constant error makes it easier for the patient to adapt his treatment, provided of course that no major changes from under- to overdelivery occur, but rather there is just a simple and stable long-lasting error. In the context of artificial pancreas (AP) systems, this last kind of predictable error could be integrated in an AP algorithm to correct these inaccuracies. In contrast, short-term instability, which corresponds to a high 15 min time interval MARD, could be seen as problematic since control algorithms frequently adapt the dose and thus could base their calculations on frequent errors.

Conclusions

In conclusion, we demonstrated that a more precise evaluation of CSII shows a slight global tendency to underdeliver, important errors in delivery over 15 min time intervals, and errors that are enhanced with the diminution of BR. Defects in reproducibility of CSII delivery were also highlighted by identifying strokes characteristics. It appears necessary to qualify the characteristics of inaccuracies according to CSII models to develop closed loop algorithms that take these specific inaccuracies into account. Further investigation is required to specifically assess the clinical effect of the observed short-time misdelivery on patients.

Acknowledgments

Mr. Sylvain GIRARDOT is a recipient of a doctoral fellowship (CIFRE: Conventions Industrielles de Formation par la Recherche) from the Association Nationale de la Recherche et de la Technologie (ANRT, Paris, France).

Author Disclosure Statement

J.-P. Riveline discloses financial support for clinical research from Insulet and Medtronic.

Funding Information

Sylvain Girardot’s work was funded by Association Nationale de la Recherche et de la Technologie (ANRT). The material was funded by Air Liquide Healthcare S.A.

Supplementary Material

Supplementary Data
Supplementary Table S1
Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3

References


Address correspondence to:
Jean-Pierre Riveline, MD, PhD
Department of Diabetes and Endocrinology
Lariboisière Hospital
APHP
Paris 75010
France

E-mail: jeanpierre.riveline@aphp.fr