Outcome Measures for Artificial Pancreas Clinical Trials: A Consensus Report

Diabetes Care 2016;39:1175–1179 | DOI: 10.2337/dc15-2716

Research on and commercial development of the artificial pancreas (AP) continue to progress rapidly, and the AP promises to become a part of clinical care. In this report, members of the JDRF Artificial Pancreas Project Consortium in collaboration with the wider AP community 1) advocate for the use of continuous glucose monitoring glucose metrics as outcome measures in AP trials, in addition to HbA1c, and 2) identify a short set of basic, easily interpreted outcome measures to be reported in AP studies whenever feasible. Consensus on a broader range of measures remains challenging; therefore, reporting of additional metrics is encouraged as appropriate for individual AP studies or study groups. Greater consistency in reporting of basic outcome measures may facilitate the interpretation of study results by investigators, regulatory bodies, health care providers, payers, and patients themselves, thereby accelerating the widespread adoption of AP technology to improve the lives of people with type 1 diabetes.

Since the publication of the Diabetes Control and Complications Trial (DCCT) in 1993 (1), the main outcome measures for glycemic control in people with type 1 diabetes have been hemoglobin A1c (HbA1c) due to the clear link to the development of complications and episodes of severe hypoglycemia (SH) as it is an immediate life-threatening event. Advances in diabetes treatment and technology have since resulted in improved care, reflected in lower HbA1c and rates of SH for people with type 1 diabetes (2–6). However, many patients still struggle with glucose control and have large and erratic swings in glycemia (7). Research on and commercial development of the artificial pancreas (AP), either as automated insulin-only delivery or as multihormonal delivery, continue to progress rapidly, and the AP promises to become a part of clinical care (8). An AP system may benefit individual patients in unique ways that would not be reflected in HbA1c improvements alone; for example, a patient with a low HbA1c and frequent hypoglycemia may have an increase in HbA1c on an AP system while hypoglycemia and quality of life improve.

OBJECTIVE AND RATIONALE

In this report, members of the JDRF Artificial Pancreas Project Consortium in collaboration with the wider AP community 1) advocate for the use of continuous glucose monitoring (CGM) glucose metrics as glycemic outcome measures in AP trials, in addition to HbA1c, and 2) identify a short set of basic, easily interpreted outcome measures to be reported in AP studies whenever feasible. Currently, the U.S. Food and Drug Administration accepts the use of various CGM glucose metrics in AP trials (9), but investigators do not always use a consistent set of measures that enables comparison. Thus, one rationale for the current report is to enable basic comparison between different AP research studies and with other clinical studies on...
glycemic control in type 1 diabetes. We acknowledge there are methodological limitations with between-study comparison that require careful consideration of study design differences. However, the standardization of these simple metrics provides a starting point for regulators, payers, health care providers, and patients to interpret AP and other study data with interventions on glycemic control. This will be especially important as AP systems become part of the daily lives of people with type 1 diabetes. Standardization of these measures does not preclude the addition of other metrics specific for a particular AP approach or used by particular research groups. In this report, we specifically advocate for the use of a basic set of CGM glucose metrics in AP studies. We suggest that their use in general type 1 diabetes studies is broadly applicable and highly relevant given the increasing adoption of CGM in research and clinical care (7).

Improvements in and adoption of AP-related technology, particularly in the reliability and accuracy of CGM systems, have focused attention on determining the best metrics for assessing outcomes in studies with people with type 1 diabetes (10–16). Currently available CGM systems with glucose readings up to every 5 min, or 288 times daily, provide considerably more data than do the American Diabetes Association recommendation of checking blood glucose 6–10 times daily (17) or the 7-point blood glucose measurements performed quarterly for research purposes in the DCCT. Although 7-point blood glucose profiles do provide insights into glycemic excursions that are not apparent with HbA1c, the profiles are very dependent on patient motivation and the chosen day of performance and provide only limited information about glucose control compared with glucose values provided by CGM systems. From a patient perspective, a CGM glycemic profile is more meaningful in that it shows highs, lows, trends, and variability as well as the effect of behaviors on glucose levels (18–21). This contrasts with HbA1c as a metric of integrated glycemic exposure over time. In particular, HbA1c does not provide information on frequency and extent of hypo- or hyperglycemia, which is a crucial aspect to evaluating glucose control in people with type 1 diabetes.

**BASIC OUTCOME MEASURES**

The recommended basic set of outcome measures presented in Table 1 includes CGM glucose metrics to define time spent in desired ranges as well as time in hypo- and hyperglycemia, measures of CGM glucose variability, safety measures such as SH and diabetic ketoacidosis, and technical metrics to evaluate AP system performance. It is intended that these measures be applicable across a wide range of AP study designs, including both short-term pilot studies and longer-term in-home or pivotal studies. Many of the glucose cut points and ranges are based on convention in AP research but were chosen to allow for comparison between studies.

HbA1c remains the best currently available measure to assess long-term glycemic control and should be assessed in any AP study of 3 months or longer. However, it is clear that HbA1c only captures average glycemia and does not provide information on the frequency or severity of hypoglycemic events. Although HbA1c is currently the most accepted metric for risk stratification of long-term complications of diabetes, that these measures be applicable across a wide range of AP study designs, including both short-term pilot studies and longer-term in-home or pivotal studies. Many of the glucose cut points and ranges are based on convention in AP research but were chosen to allow for comparison between studies.

Table 1—Recommended basic outcome measures to be reported for AP clinical trials

<table>
<thead>
<tr>
<th>Glycemic metrics*†</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbA1c</td>
<td>If intervention period ≥3 months</td>
</tr>
<tr>
<td>Mean CGM glucose</td>
<td></td>
</tr>
<tr>
<td>% CGM time &lt;50 mg/dL (&lt;2.8 mmol/L)</td>
<td></td>
</tr>
<tr>
<td>% CGM time &lt;60 mg/dL (&lt;3.3 mmol/L)</td>
<td></td>
</tr>
<tr>
<td>% CGM time &lt;70 mg/dL (&lt;3.9 mmol/L)</td>
<td></td>
</tr>
<tr>
<td>% CGM time 70–140 mg/dL (3.9–7.8 mmol/L)</td>
<td></td>
</tr>
<tr>
<td>% CGM time 70–180 mg/dL (3.9–10.0 mmol/L)</td>
<td></td>
</tr>
<tr>
<td>% CGM time &gt;180 mg/dL (&gt;10.0 mmol/L)</td>
<td></td>
</tr>
<tr>
<td>% CGM time &gt;250 mg/dL (&gt;13.9 mmol/L)</td>
<td></td>
</tr>
<tr>
<td>% CGM time &gt;300 mg/dL (&gt;16.7 mmol/L)</td>
<td></td>
</tr>
<tr>
<td>SD and coefficient of variation of CGM values</td>
<td>SD is much more dependent on the mean than coefficient of variation</td>
</tr>
<tr>
<td>Fasting blood glucose, mg/dL (mmol/L)</td>
<td>If available, depending on study design; CGM glucose at 06:00 can be taken as proxy</td>
</tr>
<tr>
<td>Safety metrics</td>
<td></td>
</tr>
<tr>
<td>SH events</td>
<td>As defined by ADA (adults) (32) and ISPAD (children and adolescents) (31)</td>
</tr>
<tr>
<td>Diabetic ketoacidosis events</td>
<td>Per ADA definition (41)</td>
</tr>
<tr>
<td>Technical performance metrics*</td>
<td></td>
</tr>
<tr>
<td>% Time closed-loop active</td>
<td></td>
</tr>
<tr>
<td>Total daily dose of insulin</td>
<td></td>
</tr>
<tr>
<td>Total daily dose of glucagon or other hormones</td>
<td>If applicable</td>
</tr>
</tbody>
</table>

ADA, American Diabetes Association; ISPAD, International Society for Pediatric and Adolescent Diabetes. *Metrics may have a skewed distribution. Report median (quartiles) instead of mean if not normally distributed. †All CGM measures should be reported for the overall 24-h period (if applicable) and also stratified by daytime and nighttime periods. The time period 00:00 to 06:00 is proposed as a definition of the nighttime period to exclude postprandial data as much as possible for a typical study population, though this definition may not be appropriate for all studies.

---

20Department of Endocrinology, Diabetes, and Nutrition, Montpellier University Hospital, INSERM Clinical Investigation Centre 1411, Institute of Functional Genomics, CNRS UMR 5203, INSERM U1191, University of Montpellier, Montpellier, France
21Diabetes Unit and Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA
22Yale University School of Medicine, New Haven, CT

© 2016 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered.

See accompanying articles, pp. 1123, 1127, 1135, 1143, 1151, 1161, 1168, and 1180.
the proposed metrics more comprehensively describe glycemia.

ADDITIONAL RECOMMENDATIONS AND LIMITATIONS

Graphical presentation of outcome data is also important, although standardization is less straightforward than with tabular data. A common figure for visualization of pooled-subject AP performance is the modal day glycemic control plot (or analogous insulin delivery plot) with median line and interquartile range bands. Inclusion of a cumulative histogram of CGM data would support the extraction of arbitrary glycemic ranges for comparison purposes (22). Numerous other graphical representations of data have been developed, and the choice of figures should be individualized for the data and the target audience (23).

The number of symptomatic hypoglycemia events per week may also be valuable as a meaningful clinical index of diabetes burden to the patient (24–27). Indeed, time spent below targeted glucose range according to CGM data may not fully capture the patient’s experience with debilitating glucose-related events, which might better illustrate diabetes burden. Because reliable capture of symptomatic hypoglycemia events requiring treatment may be challenging in longer-term AP studies, biochemical hypoglycemia event rate as measured by CGM could be reported as a proxy. For example, the rate of CGM excursions below 70 or 55 mg/dL (3.9 or 3.0 mmol/dL) for at least 10 or 30 min or longer time periods could be reported, as could other metrics including area under the curve (28–32). Many other novel measures of AP performance and algorithms have been developed (33–36), and this is an area of active research.

CGM and pump make and model and the kind of device running the control algorithm (e.g., laptop, smartphone) should be specified, including any relevant CGM signal conditioning algorithm details. We note that bias can occur when the same CGM that informs the AP controller is also used to assess glycemic outcomes (37), but there often is no practical alternative to this approach. Any special system- or protocol-related design elements should be disclosed if they are intended to improve safety or impact glycemic control or if they place additional burden on the user. The amount and timing of contact between study staff and participants in both the AP and comparator arms of the studies should be reported for in-home studies. In addition, CGM calibration logistics should be disclosed, along with a description of how conventional capillary blood glucose measurements are performed. Determination of median (or mean) absolute relative difference (MARD) for CGM versus capillary blood glucose is often used to characterize CGM accuracy in AP studies, though blood glucose sampling bias may limit the generalizability of these results (38).

Depending on the study design, the outcomes described could be reported for the entire cohort of a study or the study participants could be stratified into relevant subgroups with outcomes reported separately. For example, improved HbA1c without increased risk for hypoglycemia could be reported for those who were poorly controlled at the baseline (e.g., baseline HbA1c >8% [>64 mmol/mol]), whereas reduced incidence of hypoglycemia without deterioration in HbA1c could be reported for those with well-controlled average glycemia at the baseline. The analysis of the primary and other important outcomes should be performed on an intention-to-treat basis.

Future areas of need for AP technology include expanded standardized metrics to evaluate the technical performance of AP systems (22) and to assess patient/caregiver usability, including psychosocial metrics such as quality of life and other measures of reduction of burden, which need to be developed, including stress, anxiety, depression, and use during exercise (39). In addition, the development of standard measures to assess patient preference may be used to support regulatory approval and to serve to inform health care providers and patients of the potential impact of the use of AP systems (40). Compelling health economic measures comparing AP system costs with the potential short- and long-term economic benefits are required to establish the financial viability of these systems and to drive acceptance by health care providers and people with diabetes (8).

Multiple large and longer in-home clinical trials will soon begin with different AP systems supported by the National Institute of Diabetes and Digestive and Kidney Diseases, JDRF, the Helmsley Charitable Trust, and other funders, as well as those being supported by the industry. Some of these have been designed as pivotal trials to provide data about relevant end points to be presented to U.S. Food and Drug Administration and other regulatory authorities for approval of AP products for clinical use and to support reimbursement. This report emphasizes the need for a set of basic, uniform, standardized, and comparable outcome measures with different AP systems. As AP technologies become available for clinical use, common data reports that compare systems will be desired for health care providers, payers, and people with type 1 diabetes and their families.

In summary, members of the JDRF Artificial Pancreas Project Consortium and the larger AP community advocate for the adoption of a set of basic outcome metrics that will allow for comparison between studies with different AP systems. This, in turn, will facilitate the interpretation of the information from trials and contribute to the ultimate goal of widespread adoption of AP technology to improve the life of people with type 1 diabetes.

Funding and Duality of Interest. D.M.M. reports personal fees from Abbott and Insulet and research support from Medtronic, Dexcom, and Roche Diagnostics. B.A.B. reports grant support from JDRF, Medtronic, and Dexcom; personal fees from Medtronic, Sanofi, Tandem Diabetes Care, and Novo Nordisk; and other support from Roche Diagnostics, Dexcom, and Abbott. J.C.R. has a significant financial interest in Pacific Diabetes Technologies, Inc., a company that may have a commercial interest in the results of this research and technology. This potential conflict of interest has been reviewed and managed by Oregon Health & Science University. A.C. reports personal fees from Medtronic and a patent “Automatic insulin pumps using recursive multivariable models and adaptive control algorithms” (#US8690820) issued. E.R.D. reports a patent “Fully automated control system for type 1 diabetes” (#US7806854) issued, grants from JDRF and the National Institutes of Health (NIH), and personal fees from Automated Glucose Control LLC and Animas Corp. J.H.D. reports grants from Medtronic and personal fees from Johnson & Johnson and Sanofi. F.J.D. reports grants from JDRF and NIH and personal fees from Automated Glucose Control LLC and Animas Corp. A.H. owns intellectual property in the field of artificial pancreas (dosing algorithm). A.H. has received consultant and speaker honoraria from SNELL Medical Communication and the Diabetic Children’s Foundation and research support from Medtronic. R.H. reports personal fees from Medtronic, Eli Lilly,
Diabetes Care. H.Z. reports grants from NIH and JDRF. No other potential conflicts of interest relevant to this article were reported.

References


