



Design Considerations for Artificial Pancreas Pivotal Studies

Steven J. Russell¹ and Roy W. Beck²

Diabetes Care 2016;39:1161–1167 | DOI: 10.2337/dc15-2449

The development of artificial pancreas systems has evolved to the point that pivotal studies designed to assess efficacy and safety are in progress or soon to be initiated. These pivotal studies are intended to provide the necessary data to gain clearance from the U.S. Food and Drug Administration, coverage by payers, and adoption by patients and clinicians. Although there will not be one design that is appropriate for every system, there are certain aspects of protocol design that will be considerations in all pivotal studies designed to assess efficacy and safety. One key aspect of study design is the intervention to be used by the control group. A case can be made that the control group should use the currently available best technology, which is sensor-augmented pump therapy. However, an equally, if not more, compelling case can be made that the control intervention should be usual care. In this Perspective, we elaborate on this issue and provide a pragmatic approach to the design of clinical trials of artificial pancreas systems.

A device that utilizes a continuous glucose monitor (CGM) and computer algorithms to automate, to various degrees, the calculation of insulin doses and the delivery of insulin is known as an artificial pancreas (AP) or closed-loop system. Some of these systems also automate the delivery of glucagon for the prevention and treatment of hypoglycemia. AP studies have evolved from small series of patients tested in highly controlled clinical research center environments for the assessment of algorithm performance, to studies done in hotels and camps, to short-term and medium-term outpatient studies with varying degrees of monitoring. This progression has set the stage for the conduct of “pivotal” studies designed to assess the efficacy and safety of AP systems using hardware planned for commercialization in patient populations for whom the commercialized product is intended. These studies will need to provide compelling data demonstrating efficacy and safety to gain not only U.S. Food and Drug Administration (FDA) approval but also coverage by payers and adoption by patients and clinicians. The study design will have an important influence on the labeling and coverage of these devices. The FDA has issued a guidance regarding applications for AP systems (1).

Although there will not be a single design that is appropriate for the demonstration of safety and efficacy of every system, there are certain aspects of protocol design that must be considered in all such studies. Here, we provide our pragmatic recommendations and views on certain key aspects of pivotal studies (summarized in Table 1, with an example study in Fig. 1) designed to demonstrate efficacy and safety of AP systems that both reduce insulin delivery to minimize hypoglycemia and increase insulin delivery to limit hyperglycemia, with applicability to both insulin-only systems and bihormonal systems that administer glucagon or other hormones, such as pramlintide, as well as insulin.

POSSIBLE STUDY DESIGNS

A pivotal study designed to assess efficacy and safety could be conducted as a randomized crossover trial or a parallel-group randomized clinical trial (RCT). A

¹Diabetes Unit, Massachusetts General Hospital, Boston, MA

²Jaeb Center for Health Research, Tampa, FL

Corresponding author: Roy W. Beck, MD, rbeck@jaeb.org.

Received 11 November 2015 and accepted 19 December 2015.

© 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, 1168, 1175, and 1180.

Table 1—Recommendations for design of pivotal trials designed to show long-term efficacy and safety of AP systems			
Design consideration	Recommendation	Alternative	Comment
RCT type	Parallel	Crossover	Crossover design requires long washout for HbA _{1c} outcome
Study population	Representative of population, patients who use multiple daily injections and continuous subcutaneous insulin infusion, with few exclusions	Adults only, high-risk patients excluded	Given the potential for off-label use, the FDA may not approve if the device is not demonstrated to be safe in a broad population and payers may limit coverage to only the population that was studied
Randomization (AP:control)	2:1	1:1	2:1 randomization provides greater exposure to AP; 1:1 randomization will require a smaller sample size or give greater power for same sample size if equal variance
Control group	Usual care	SAP	Both scientifically valid, usual care has numerous pragmatic advantages (see text)
Superiority vs. noninferiority	Superiority	Noninferiority	Noninferiority may be sufficient for approval but is not likely to drive reimbursement and adoption
Run-in period	Blinded CGM	Unblinded CGM (SAP training)	Unblinded run-in must be sufficient to achieve competency for SAP trial enrolling non-SAP users
Duration	6–12 months	3 months	3 months minimum for HbA _{1c} , longer duration shows continuation of use and durability of effect
Primary outcome(s)	HbA _{1c} , time <60 mg/dL	HbA _{1c} only	HbA _{1c} does not capture hypoglycemia; CGM more reliable and quantitative than participant recall

single-arm trial in which the performance of the AP system is compared with baseline data or historical data from a similar patient cohort is useful to collect data on the feasibility and functioning of the system and preliminary outcome data and may be sufficient for a functional labeling claim by the FDA.

However, such a design is not sufficient to provide strong evidence for efficacy and safety because any improvements could represent a study effect. Having a participant serve as his/her own control in a two-period crossover design is efficient, but the study duration will necessarily be at least twice as long as a

parallel-group design with the same duration of exposure to the AP system and will require a sufficient washout period between the two study periods, which will be long for studies using glycated hemoglobin (HbA_{1c}) as a primary outcome. Therefore, to provide sufficient data over long enough periods of AP use to assess long-term performance and tolerability, a parallel-group RCT comparing an AP group with a concurrent control group has the strongest rationale. The remainder of this Perspective will assume a parallel-group RCT (Fig. 1) for an AP system that is designed to decrease and actively increase insulin dosing to reduce hyperglycemia and hypoglycemia.

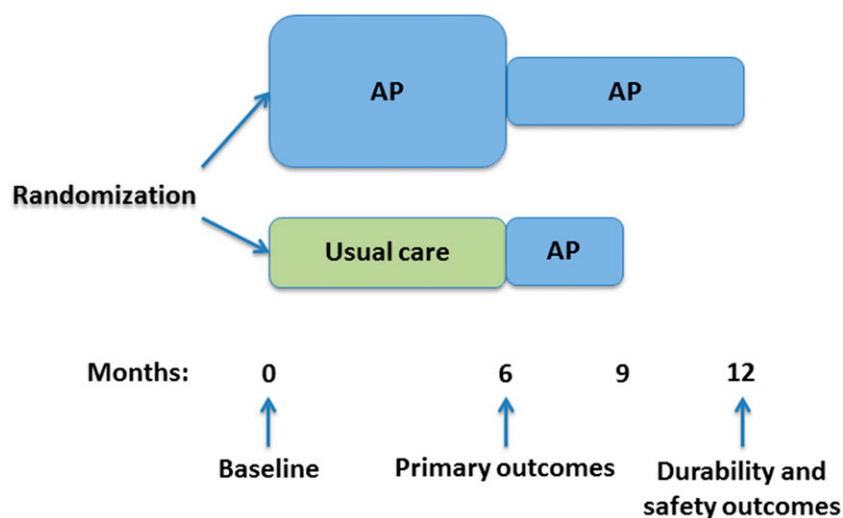


Figure 1—An example of an AP parallel-group RCT. In this example, the control group is usual care with 2:1 randomization, which is used to increase exposure to the AP system. The coprimary outcomes (change in HbA_{1c} from baseline and change in time <60 mg/dL from baseline) are computed at 6 months. A randomly selected group of participants in the AP group are asked to continue in the study for an additional 6 months, with an assessment of durability of outcomes and long-term safety at 12 months. Participants randomized to usual care may crossover to AP after 6 months to increase recruitment and encourage compliance with the research protocol in the usual care group. The height of the boxes is proportional to the number of participants and the width of the boxes is proportional to the duration of the period.

PATIENT POPULATION

The patient population should be as broad as possible. At present, many patients with type 1 diabetes do not use insulin pumps (~30–60% are pump users) (2,3) and a substantially smaller proportion use CGM technology (~11%) (2). However, the promise of better glycemic control and the reduced patient effort associated with some AP designs may entice a larger proportion of patients who are currently managing their diabetes with insulin injections without CGM. Therefore, it is important for trials to enroll a sufficient number of pump- and CGM-naïve patients so that safety and

efficacy analyses can be performed in these subgroups. Preadolescent children present different challenges from adolescents and young adults, who are likewise distinct from mature adults and the elderly; thus, trials should test efficacy and safety in each of these groups. Patients with hypoglycemia unawareness and complications of diabetes should not be excluded unless absolutely necessary for safety reasons. Given the potential for off-label use, the FDA has signaled that label restrictions may not be sufficient and that it may not approve an AP device unless it has been shown to be safe in a population representative of patients who may use the device. A study design could include quotas for minimum numbers of participants using pumps and injections, of patients with different HbA_{1c} levels, and of age-groups to assure that the study population is reasonably representative of the intended-use population and to provide sufficient numbers for subgroup analyses.

RANDOMIZATION

Randomization to the AP group versus the control group could be with an equal (1:1) or unequal (e.g., 2:1) allocation. There is a statistical advantage for 1:1 allocation in that the required sample size will be about 10–15% smaller than for 2:1 allocation, assuming equal variance between groups and the same statistical power. The main advantage of a 2:1 allocation, which often outweighs the need for a modestly larger sample size, is the ability to have greater exposure to the intervention for assessing adherence and safety. Although in some settings a 2:1 randomization can enhance recruitment, for AP trials, recruitment should not be difficult if the study participants in the control arm are given the opportunity to use the AP system in an extension study (Fig. 1). The randomization scheme should include stratification for factors that are most important for balance between treatment arms. Imbalances in predictive factors occurring despite randomization may be adjusted for in analysis.

CONTROL GROUP

There are two types of control groups that could be used in an AP parallel-group RCT: 1) sensor-augmented pump (SAP) therapy using the same pump and CGM that is part of the AP system, which also

could include a low-glucose suspend feature if available, and 2) usual care, which could include both pump and injection users as well as CGM users and nonusers. There are advantages and disadvantages for both approaches. The main rationale for an SAP control group is the belief that new technology needs to be demonstrated to be better than the best existing technology in efficacy, safety, or both. However, data from the T1D Exchange clinic registry indicate that only about 11% of patients with type 1 diabetes are using SAP to manage their diabetes (2). Even this low rate of adoption may be an overestimate as all of the registry participants are seen by an endocrinologist. Thus, SAP is not usual care for the vast majority of patients with type 1 diabetes. Furthermore, T1D Exchange data on those who used CGM and discontinued it showed that lack of insurance coverage was not a major reason for discontinuing. Rather, it was annoyances and/or added burden that outweighed the perceived benefit and led to discontinuation (4). Most trial designs using SAP as a control include a run-in period for several weeks during which participants must demonstrate their willingness to wear a sensor and ability to execute SAP therapy effectively. Excluding unsuccessful participants selects for more technology-savvy individuals and those more tolerant of burdens and annoyances associated with technological solutions and thereby reduces the applicability of the results to the broader population of patients with type 1 diabetes. If we want to know what impact the adoption of AP could have on the control of the broader population of patients with type 1 diabetes, it makes most sense not to exclude patients who have rejected or who are otherwise not good candidates for SAP therapy.

There are several important rationales for a usual care control group. First, a new intervention needs to be demonstrated to be better than what the majority of patients currently are doing to manage diabetes. As noted, approximately 90% of patients use something other than SAP therapy. Second, for health economic analyses, a comparison with a population representative of the population with type 1 diabetes at large will provide the best information on how much benefit may be gained from the use of the AP device if it is widely adopted, which is an important consideration for payers

and policy makers. In addition to the low overall adoption of SAP in the U.S., there is currently no coverage for SAP therapy in some patient populations (e.g., Medicare) and in some other countries (e.g., Australia); thus, comparison with an SAP control group will be particularly unhelpful in these settings. Third, for health-related quality-of-life analyses, a comparison with a control group representing usual care will be far more informative than a comparison with a control group using an intervention used by only a small minority of patients with type 1 diabetes. Fourth, it is important to show that the device can be used effectively and safely by technology-naïve patients without prohibitive amounts of training and preparation and that the patients do not discontinue use of the device at high rates. Fifth, for the populations in which SAP has been shown to have benefit, the sample size required for a trial with a usual care control arm will be less than a trial with SAP for the control arm due to a larger treatment effect of AP compared with usual care than with SAP. Even relatively small differences in the treatment effect can result in large differences in the sample size required to demonstrate efficacy. For instance, for a mean change in HbA_{1c} outcome, the sample size to detect a treatment group difference if the true treatment group difference is 0.75% will be about half the sample size for a 0.5% difference. This may be particularly important for showing efficacy in subgroups if the trial is so powered. Sixth, overly restrictive inclusion criteria (including the ability to complete a run-in period) may negatively impact the label as well as coverage by payers of any approved device. If an AP system is only indicated (or reimbursed) for patients who have demonstrated the ability to use SAP, then only a small portion of the population will be served. If only such patients have been included in trials, then the AP devices may not be safe for the broader population for which they may be prescribed. Also, inclusion of poor SAP candidates in an SAP comparator group may exaggerate the benefit of the AP system. If an AP system is able to increase benefit with the same or reduced burden as the current SAP, then it can be expected that the proportion of patients using an AP system will eventually greatly exceed the proportion currently using SAP. Seventh, with this in

mind, SAP without any form of automation of insulin delivery is unlikely to exist in a few years, yet the importance of understanding the magnitude of benefit of AP compared with usual care will remain. With these considerations in mind, a usual care control group would be reasonable, if not preferred, in an AP RCT.

Trials conducted solely to provide evidence to support regulatory approval of an AP system have no incentive to include more than one control arm. However, the ideal design of a trial on a commercially available AP system might be three arms to compare the AP system to SAP alone and to usual care, and the additional data may be useful to support reimbursement of AP systems.

RUN-IN PERIOD

In all study designs, obtaining baseline CGM glucose data that are representative of the participant's usual glycemic state will be useful. Wearing a blinded CGM for 7–14 days can provide a good representation of glycemic control (5). To avoid bias, the same CGM used for calculation of outcomes during the trial should be used for baseline data collection.

Depending on the control group intervention that is selected, a run-in period with an unblinded CGM device also may be needed, particularly for sensor-naïve participants. When the control group will be using SAP, the purpose of an unblinded run-in is twofold. First, it helps to identify which participants are less likely to use the AP system or SAP regularly in an RCT and thus can be dropped prior to randomization (although this will reduce the generalizability of the results). Second, it provides an opportunity to train the participants on CGM use and have CGM-related improvements in glycemic control occur prior to the start of the RCT. However, a long run-in period may affect AP labeling, could lead to a requirement that patients have a successful supervised SAP period before payers will provide coverage, and could increase the provider workload associated with initiating AP therapy. When control group participants will be following their usual care, training on CGM use prior to randomization is not practical, and in designing the study, the sample size should be increased to account for a proportion of the participants in the AP group discontinuing use after randomization.

SUPERIORITY VERSUS NONINFERIORITY

A study can be designed to demonstrate that the intervention is superior to the control or can be designed to show that the intervention is “at least as good as” the control. The latter requires a definition of “at least as good as,” which in statistical terms is called the noninferiority limit. The control intervention (usual care or SAP) is an important consideration in deciding between a superiority and noninferiority approach. Designing the study to demonstrate superiority will be meaningful irrespective of the control intervention. However, designing the study to demonstrate noninferiority only seems defensible when SAP is the control and, importantly, only in populations where SAP already has been demonstrated to be better than usual care without SAP. When choosing an end point, it will be important to consider not only the requirements for approval but also the effect such end points will have on labeling and on coverage by payers. Although it may be sufficient for regulatory approval, noninferiority of AP systems that are more expensive than current standards of care is unlikely to be compelling to payers.

TRIAL CONDUCT AND DURATION OF FOLLOW-UP

In designing a trial, it is important that the AP group and control group have similar levels of contact with study staff and similar degrees of general diabetes education. AP systems that are easy to use and reduce burden should not require a great deal of contact and the initial training period should be brief. Systems that require more of the user may require more contact. Increased contact limited to the early months of the study is less likely to affect outcome comparisons at the end of the trial. Therefore, amount and timing of contact with the AP versus comparator groups should be reported.

The efficacy of an AP system in reducing mean glucose, increasing time in range, and reducing hypoglycemia and hyperglycemia can be demonstrated in short-term studies of less than a week in duration. However, these data cannot be extrapolated to know that there will be long-term benefit, and the time period is too short to assess safety or changes in HbA_{1c}. It is likely that AP systems can

replicate short-term results indefinitely if they are used properly, but it is unknown whether patients will continue to use a particular AP system properly over extended periods of time. This should be evaluated with a study of at least 3 months, and preferably 6–12 months. In the JDRF CGM RCT as well as other studies, compliance with the use of CGM dropped off after the first few weeks, even with the encouragement that was given as part of the study (6). The data on continuation of use and durability of effect are likely to be particularly important to payers when they are making coverage decisions.

OUTCOMES AND ANALYSES

The primary analysis should follow the intent-to-treat principle. In order to minimize bias, a high proportion of study participants must remain in the study through the primary outcome time point. Participants who discontinue or poorly comply with the intervention protocol should be strongly encouraged to remain in the study and return at least for the primary outcome exam, unless there is a safety concern with continued participation.

Efficacy, safety, and quality of life are all important outcomes for an AP RCT. Efficacy equates with improved glycemic control, a reduction in mean glucose, hypoglycemia, and/or hyperglycemia, which can be measured with HbA_{1c} and with CGM glucose metrics. The main safety outcomes are severe hypoglycemia and diabetic ketoacidosis. Quality of life in an AP RCT relates to measurement of diabetes-specific issues, such as fear of hypoglycemia and burden of diabetes management, as well as more general well-being. Aside from insurance coverage issues, the degree to which the use of AP systems are adopted by individuals with type 1 diabetes will depend on the perception of burden relative to benefit. Thus, assessment of quality of life is an important outcome measure, even though it is unlikely to be the primary outcome measure in a pivotal RCT.

HbA_{1c}, which conventionally has been the gold-standard outcome measure to assess glucose control in clinical trials, has a number of properties that make it a good outcome measure for an RCT. It can be measured with a high degree of precision in a central laboratory, is not dependent on the use of a device such

as CGM or blood glucose meter at home, and is understood by most patients. Perhaps most compellingly, lower HbA_{1c} levels are associated with lower risk of chronic diabetes complications as shown in the Diabetes Control and Complications Trial (DCCT) and other studies (7,8). However, there are certain drawbacks to using HbA_{1c} as the primary outcome measure. First, the goal of therapy may not be to lower HbA_{1c} in all patients, such as those who already have a normal or near-normal HbA_{1c} level but experience hypoglycemia or the elderly, in whom hypoglycemia may be more of a concern than hyperglycemia. Hypoglycemia, if frequent, can lower the HbA_{1c}, so reduction in hypoglycemia can appear to have a negative effect on glycemic regulation as measured by the HbA_{1c}. Second, the relationship between mean glucose and glycation rates vary among individuals (9–11). Differing glycation rates are less of an issue for an RCT in which the data are pooled across the treatment group for analysis (and presumably balanced through randomization) than it is for assessing the level of glycemic control for an individual patient. A small percentage of patients may have a hemoglobinopathy that affects the interpretation of the HbA_{1c} value, but these patients are relatively rare and could be excluded at screening. Third, there is a ceiling effect on the amount of improvement in HbA_{1c} that can occur when baseline HbA_{1c} approaches 6% (42 mmol/mol), although the reduction in complications associated with further lowering of HbA_{1c} below 7% (53 mmol/mol) is low in absolute terms (12,13) and consistent control at a higher level of HbA_{1c} may be sufficient to prevent microvascular complications (14). Fructosamine and glycated albumin are other measures of glycemic control that could avoid some of these issues, but they have not been widely used and are not validated as outcomes measures for a clinical trial. Furthermore, these assays are not standardized in the same way as HbA_{1c} and the correlation of glycated albumin with HbA_{1c} depends on the specific assay used (15,16).

CGM provides the opportunity to measure actual glucose levels during day-to-day living and provides assessments of both hyperglycemia and hypoglycemia. As such, it could be considered the optimal method for assessing outcomes in an

AP RCT. The value of CGM as a primary outcome measure has been demonstrated in long-term randomized trials assessing CGM as an intervention in patients with normal or near-normal HbA_{1c} levels (17,18). CGM can be used to separately analyze glycemic regulation during the daytime and overnight. However, there are several considerations when CGM is used as an outcome measure. There will be a certain amount of sensor inaccuracy, but this can be addressed in the study design by increasing sample size to account for greater variance of continuous outcome variables and to account for misclassification of binary outcome variables. A more difficult problem is that the outcome data will not be available for study participants who discontinue use of AP or SAP, leading to missing data and potential bias. This could be mitigated by having those who discontinue use of AP or SAP wear a blinded sensor, similar to an untreated control group, to provide outcome data for analysis. Consideration should be given to having both the AP group and the control group wear the same blinded sensor during the run-in period (to provide baseline data) and for 7–14 days at intervals (e.g., every 1–3 months) during the trial to have a direct comparison between the groups. Using a different sensor than the one used as part of the AP system has the advantage of reducing, although not eliminating, the potential bias associated with using the CGM data that drive the AP to determine time in range or above or below a threshold. This bias will tend to overestimate the proportion of true glucose values in the target range but only when a true benefit exists; however, it will not affect the mean glucose (19). A stochastic adjustment of the data has been proposed to offset the bias, but simulations have shown that in many scenarios this adjustment does not reduce the bias and may actually increase it (19). Retrospective recalibration of the CGM glucose values using blood glucose meter measurements might improve the accuracy of the CGM glucose values, but it is uncertain whether this will have a meaningful effect on the analysis of RCT data. Using a blinded CGM that does not require calibration can avoid bias that may occur if participants in the control group do not calibrate the CGM as frequently as those in the AP group.

CGM metrics that could be used for AP RCT outcomes include ones that provide a measure of overall control (mean glucose, time within a target range), hyperglycemia, hypoglycemia, and glycemic variability. Mean glucose is useful as an overall measure and is well correlated with HbA_{1c}, but similar to HbA_{1c} it can be misleading without data on hypoglycemia. Time within a target range, typically 70–180 mg/dL, is a metric understood by clinicians and patients. Because an AP system overnight should produce better glycemic control during the day, some investigators have used a target of 70–140 mg/dL overnight and 70–180 mg/dL during the daytime. However, there are little data relating time in range to the risk of complications.

There are several highly correlated metrics for assessing hypoglycemia, but time below a threshold (e.g., 70 mg/dL, 60 mg/dL, or 50 mg/dL) is the most understandable and, as a result, should be considered as the main metric for biochemical hypoglycemia. One study showed that time below 70 mg/dL correlated 0.97 with area under the curve and 0.98 with the low blood glucose index (20). Use of a threshold lower than 70 mg/dL may be preferred because individuals without diabetes can have glucose concentrations 60 to <70 mg/dL without symptoms and glucose concentrations <60 mg/dL in individuals without diabetes are rare (21). The higher threshold of 70 mg/dL has been used clinically partly due to the historical inaccuracy of meters and to the possibility that a glucose concentration below 70 mg/dL is a harbinger of more severe hypoglycemia. As the accuracy of glucose meters and CGM devices has improved, the 60 mg/dL threshold is likely to be a more meaningful measure of clinically important hypoglycemia.

Biochemical hypoglycemia also can be evaluated as events. However, this sort of analysis tends to discount the difference between short, transient episodes of hypoglycemia and longer, potentially more severe episodes and may be insensitive to hypoglycemic events of short duration, depending on the definition. Capturing symptomatic hypoglycemia reliably with a journal or participant recall may be difficult in long-term trials. Therefore, duration of time in hypoglycemia using CGM data appears to be the best single measure. Table 2 provides an operational definition for an event analysis.

Table 2—Analytic definition of a CGM-determined hypoglycemic event

1. A hypoglycemic event is defined as 15 consecutive minutes with a sensor glucose value below a threshold (such as <70 mg/dL, which is continued through the example but could be replaced by 60 or 50 mg/dL).
 - a. At least two sensor values <70 mg/dL that are 15 or more minutes apart plus no intervening values ≥ 70 mg/dL are required to define an event.
 - a.i. This accounts for a CGM device such as the Abbot Libre only having measurements every 15 minutes.
 - a.ii. This accounts for potential missing sensor values for devices that record a glucose concentration every 5 minutes.
2. The end of the hypoglycemic event is defined as a minimum of 15 consecutive minutes with a sensor glucose concentration ≥ 70 mg/dL and ≥ 10 mg/dL above the nadir of the event (note: the latter requirement only comes into play if the nadir is 61–69 mg/dL).
 - a. At least two sensor values ≥ 70 mg/dL that are 15 or more minutes apart with no intervening values <70 mg/dL are required to define the end of an event.
 - a.i. This accounts for a CGM device such as the Abbot Libre only having measurements every 15 minutes.
 - a.ii. This accounts for potential missing sensor values for devices that record a glucose concentration every 5 minutes.
3. When a hypoglycemic event ends, the study participant becomes eligible for a new event.

Reduction in biochemical hypoglycemia is a worthy goal because there is an association between biochemical hypoglycemia and subsequent severe clinical hypoglycemic events (although the positive predictive value is low) (22), because hypoglycemia can increase the risk of cardiovascular events as well as falls and other accidents (23,24), and because symptomatic hypoglycemia negatively affects patient quality of life, function, and productivity (25–27). In addition, severe hypoglycemia must be a safety outcome in an AP RCT to assure that the system is not increasing the risk of such events. Severe hypoglycemia also can be considered an efficacy outcome as one objective of an AP system is to minimize the frequency of such events. The difficulty in designing an RCT to demonstrate a reduction in severe hypoglycemic events is that the required sample size is very large due to the low event rate, and thus severe hypoglycemia as a primary outcome may only be feasible in a study limited to patients at very high risk based on frequent prior severe hypoglycemic events. Assuming a control group severe hypoglycemia rate of 15 per 100 person-years in a 6-month RCT, the sample size would need to be >1,000 for a 50% reduction in the rate using an AP system. If eligibility is restricted to those with frequent severe hypoglycemia who have a rate of 45 events per 100 person-years, the sample size is reduced by more than half to ~500, but this is still a formidable number considering the restrictive eligibility criterion (28).

There are a number of metrics for hyperglycemia that are all highly correlated with each other, and the simplest, time above a threshold such as 180 mg/dL or 250 mg/dL, may be preferred for an RCT as it is more understandable than area under the curve, high blood glucose index, or other metrics that account for the time and magnitude of hyperglycemia. Both time in range and mean glucose correlate highly with hyperglycemia, which in virtually all patients is much more frequent than hypoglycemia.

Glycemic variability also is a popular CGM metric, although the importance of glycemic variability as a predictor of diabetes complications is uncertain. There are numerous metrics to assess glycemic variability including SD, coefficient of variation, interquartile range, mean amplitude of glycemic excursion (MAGE), mean of daily differences (MODD), continuous overall net glycemic action (CONGA), and others (29). One limitation is that SD and MAGE typically increase with mean glucose, making it difficult to separate the effect of the intervention on glycemic variability from the effect on the mean glucose. Thus, when comparing treatment groups, SD may not be as good an indicator of glycemic variability as is the coefficient of variation, which is the SD divided by the mean glucose. The coefficient of variation is approximately independent of mean glucose and therefore can be used to determine whether the difference in glycemic variability between two interventions differs

more than expected given the difference in mean glucose.

For an AP system designed to reduce both hypoglycemia and hyperglycemia, a combined outcome has a compelling logic. An appealing combined outcome is one in which success is defined as achieving both a reduction in HbA_{1c} (or alternatively, CGM-measured mean glucose) and a reduction in biochemical hypoglycemia. However, it would be reasonable to consider an intervention to be a success if there was either 1) a reduction in HbA_{1c} (or alternatively, CGM-measured mean glucose or time in range) with no increase in CGM-measured hypoglycemia or 2) a reduction in CGM-measured hypoglycemia without an increase in HbA_{1c} (or alternatively, CGM-measured mean glucose or time in range) as was done in the well-controlled cohort in the JDRF CGM RCT (18).

CONCLUSIONS

This is an exciting time in the evolution of the development of an AP. Progress is being made rapidly and a number of systems are ready to be tested in rigorous long-term RCTs. There are numerous issues to consider to assure that these trials are sound in design and efficient and will provide the efficacy and safety data needed to gain FDA approval, coverage by payers, and adoption by patients and clinicians.

Duality of Interest. S.J.R. reports pending patent applications for a blood glucose control system assigned to Partners HealthCare and Massachusetts General Hospital; reports receiving loaned equipment, support in kind, and/or technical assistance from Dexcom, Tandem Diabetes, and Eli Lilly; reports honoraria for lectures from Tandem Diabetes, Dexcom, Sanofi, and Eli Lilly; reports consulting fees from Sanofi; and reports serving on the scientific advisory boards for Tandem Diabetes and Companion Medical. R.W.B. reports that his nonprofit employer has received consultant payments on his behalf from Animas, Tandem, and Bigfoot Biomedical with no personal compensation to him and grant funds from Dexcom. No other potential conflicts of interest relevant to this article were reported.

References

1. U.S. Food and Drug Administration Center for Devices and Radiological Health. *Guidance for Industry and Food and Drug Administration Staff: The Content of Investigational Device Exemption (IDE) and Premarket Approval (PMA) Applications for Artificial Pancreas Device Systems*. Rockville, MD, U.S. Department of Health and Human Services, 2012

2. Miller KM, Foster NC, Beck RW, et al.; T1D Exchange Clinic Network. Current state of type 1 diabetes treatment in the U.S.: updated data from the T1D Exchange clinic registry. *Diabetes Care* 2015;38:971–978
3. Willi SM, Miller KM, DiMeglio LA, et al.; T1D Exchange Clinic Network. Racial-ethnic disparities in management and outcomes among children with type 1 diabetes. *Pediatrics* 2015;135:424–434
4. Wong JC, Foster NC, Maahs DM, et al.; T1D Exchange Clinic Network. Real-time continuous glucose monitoring among participants in the T1D Exchange clinic registry. *Diabetes Care* 2014;37:2702–2709
5. Xing D, Kollman C, Beck RW, et al.; Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. Optimal sampling intervals to assess long-term glycemic control using continuous glucose monitoring. *Diabetes Technol Ther* 2011;13:351–358
6. Tamborlane WV, Beck RW, Bode BW, et al.; Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. Continuous glucose monitoring and intensive treatment of type 1 diabetes. *N Engl J Med* 2008;359:1464–1476
7. Nathan DM; DCCT/EDIC Research Group. The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications study at 30 years: overview. *Diabetes Care* 2014;37:9–16
8. The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 1993;329:977–986
9. Nathan DM, Kuenen J, Borg R, Zheng H, Schoenfeld D, Heine RJ; A1c-Derived Average Glucose Study Group. Translating the A1C assay into estimated average glucose values. *Diabetes Care* 2008;31:1473–1478
10. Wilson DM, Xing D, Beck RW, et al.; Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. Hemoglobin A1c and mean glucose in patients with type 1 diabetes: analysis of data from the Juvenile Diabetes Research Foundation continuous glucose monitoring randomized trial. *Diabetes Care* 2011;34:540–544
11. Wilson DM, Xing D, Cheng J, et al.; Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. Persistence of individual variations in glycosylated hemoglobin: analysis of data from the Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Randomized Trial. *Diabetes Care* 2011;34:1315–1317
12. Lachin JM, Genuth S, Nathan DM, Zinman B, Rutledge BN; DCCT/EDIC Research Group. Effect of glycemic exposure on the risk of microvascular complications in the Diabetes Control and Complications Trial—revisited. *Diabetes* 2008;57:995–1001
13. The Diabetes Control and Complications Trial Research Group. The relationship of glycemic exposure (HbA1c) to the risk of development and progression of retinopathy in the Diabetes Control and Complications Trial. *Diabetes* 1995;44:968–983
14. Nordwall M, Abrahamsson M, Dhir M, Fredrikson M, Ludvigsson J, Arnqvist HJ. Impact of HbA1c, followed from onset of type 1 diabetes, on the development of severe retinopathy and nephropathy: the VISS Study (Vascular Diabetic Complications in Southeast Sweden). *Diabetes Care* 2015;38:308–315
15. Beck R, Steffes M, Xing D, et al.; Diabetes Research in Children Network (DirecNet) Study Group. The interrelationships of glycemic control measures: HbA1c, glycosylated albumin, fructosamine, 1,5-anhydroglucitol, and continuous glucose monitoring. *Pediatr Diabetes* 2011;12:690–695
16. Nathan DM, McGee P, Steffes MW, Lachin JM; DCCT/EDIC Research Group. Relationship of glycosylated albumin to blood glucose and HbA1c values and to retinopathy, nephropathy, and cardiovascular outcomes in the DCCT/EDIC study. *Diabetes* 2014;63:282–290
17. Battelino T, Phillip M, Bratina N, Nimri R, Oskarsson P, Bolinder J. Effect of continuous glucose monitoring on hypoglycemia in type 1 diabetes. *Diabetes Care* 2011;34:795–800
18. Beck RW, Hirsch IB, Laffel L, et al.; Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. The effect of continuous glucose monitoring in well-controlled type 1 diabetes. *Diabetes Care* 2009;32:1378–1383
19. Kollman C, Calhoun P, Lum J, Sauer W, Beck RW. Evaluation of stochastic adjustment for glucose sensor bias during closed-loop insulin delivery. *Diabetes Technol Ther* 2014;16:186–192
20. Beck RW, Calhoun P, Kollman C. Use of continuous glucose monitoring as an outcome measure in clinical trials. *Diabetes Technol Ther* 2012;14:877–882
21. Fox LA, Beck RW, Xing D; Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. Variation of interstitial glucose measurements assessed by continuous glucose monitors in healthy, nondiabetic individuals. *Diabetes Care* 2010;33:1297–1299
22. Fiallo-Scharer R, Cheng J, Beck RW, et al.; Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. Factors predictive of severe hypoglycemia in type 1 diabetes: analysis from the Juvenile Diabetes Research Foundation continuous glucose monitoring randomized control trial dataset. *Diabetes Care* 2011;34:586–590
23. Frier BM. Hypoglycaemia in diabetes mellitus: epidemiology and clinical implications. *Nat Rev Endocrinol* 2014;10:711–722
24. Sanon VP, Sanon S, Kanakia R, et al. Hypoglycemia from a cardiologist's perspective. *Clin Cardiol* 2014;37:499–504
25. Brod M, Christensen T, Bushnell DM. The impact of non-severe hypoglycemic events on daytime function and diabetes management among adults with type 1 and type 2 diabetes. *J Med Econ* 2012;15:869–877
26. Brod M, Christensen T, Thomsen TL, Bushnell DM. The impact of non-severe hypoglycemic events on work productivity and diabetes management. *Value Health* 2011;14:665–671
27. Brod M, Pohlman B, Wolden M, Christensen T. Non-severe nocturnal hypoglycemic events: experience and impacts on patient functioning and well-being. *Qual Life Res* 2013;22:997–1004
28. Beck RW, Kollman C, Xing D, Buckingham BA, Chase HP. Outcome measures for outpatient hypoglycemia prevention studies. *J Diabetes Sci Technol* 2011;5:999–1004
29. Rodbard D. Interpretation of continuous glucose monitoring data: glycemic variability and quality of glycemic control. *Diabetes Technol Ther* 2009;11(Suppl. 1):S55–S67