Craig A. Beam,[1] Stephen E. Gitelman,[2] Jerry P. Palmer,[3,4] and the Type 1 Diabetes TrialNet Study Group

Recommendations for the Definition of Clinical Responder in Insulin Preservation Studies

Clinical responder studies should contribute to the translation of effective treatments and interventions to the clinic. Since ultimately this translation will involve regulatory approval, we recommend that clinical trials prespecify a responder definition that can be assessed against the requirements and suggestions of regulatory agencies. In this article, we propose a clinical responder definition to specifically assist researchers and regulatory agencies in interpreting the clinical importance of statistically significant findings for studies of interventions intended to preserve β-cell function in newly diagnosed type 1 diabetes. We focus on studies of 6-month β-cell preservation in type 1 diabetes as measured by 2-h–stimulated C-peptide. We introduce criteria (bias, reliability, and external validity) for the assessment of responder definitions to ensure they meet U.S. Food and Drug Administration and European Medicines Agency guidelines. Using data from several published TrialNet studies, we evaluate our definition (no decrease in C-peptide) against published alternatives and determine that our definition has minimum bias with external validity. We observe that reliability could be improved by using changes in C-peptide later than 6 months beyond baseline. In sum, to support efficacy claims of β-cell preservation therapies in type 1 diabetes submitted to U.S. and European regulatory agencies, we recommend use of our definition.

The European Medicines Agency (EMA) recently recommended that "responder analyses" be included as either primary or secondary analyses in applications submitted for the approval of medicinal products for diabetes treatment and/or prevention (1). For example, the EMA guidelines require that studies of glucose reduction in type 2 diabetes (T2D) "… should justify the clinical relevance of the observed effect by presenting responder analyses comparing the proportion of patients who reached an absolute (HbA$_{1c}$) value of ≤7 and/or 6.5% (≤53 and/or 48 mmol/mol) across the different treatment groups" (1). For treatment studies based on β-cell preservation in newly diagnosed type 1 diabetic (T1D) patients, the EMA recommends using several coprimary end points, one of which could be, if appropriately justified, "… the percentage of patients with (stimulated) C-peptide increases above a clinically meaningful threshold" (1).

Although, as of the time of the writing of this article, the U.S. Food and Drug Administration (FDA) has not required responder analyses in their draft guidance document for industry related to diabetes (2), the agency did endorse the practice as part of the International Conference on Harmonisation guidelines for statistical principles for clinical trials (3). Since the agency does advise sponsors of therapeutic interventions for preservation of endogenous β-cell function in patients with newly diagnosed T1D to demonstrate the maintenance of C-peptide from baseline, it is reasonable to expect that the agency would be interested in responder analyses defining a clinical responder as someone who maintains C-peptide.

The term "responder analysis" refers to the dichotomization of a continuous primary efficacy measure into

[1]Division of Epidemiology and Biostatistics, Department of Pediatrics, University of South Florida, Tampa, FL
[2]Department of Pediatrics, University of California, San Francisco, CA
[3]Department of Medicine, VA Puget Sound Health Care System, Seattle, WA
[4]University of Washington, Seattle, WA

Corresponding author: Craig A. Beam, craig.beam@med.wmich.edu.

Received 21 January 2014 and accepted 7 April 2014.

C.A.B. is currently affiliated with the Division of Epidemiology and Biostatistics, Department of Biomedical Sciences, Western Michigan University Homer Stryker M.D. School of Medicine, Kalamazoo, MI.

"responder" and "nonresponder" categories (4). Such responder classifications help us interpret data clinically and speak directly to the question of fundamental interest in clinical science and practice: "Is this therapy benefitting the patient?" Clinical responder classifications have been used as end points in clinical trials to compare treatments and to define subgroups that are then compared for differences in immunologic, metabolic, or mechanistic characteristics. Examples from the diabetes literature are numerous: Chiasson and Naditch (5) compared the efficacy of miglitol and metformin on glycemic control in T2D and defined a "clinically significant response" to be either a 15% or greater reduction from baseline in HbA_{1c} or an HbA_{1c} level <7.0%; Luque Otero and Martell Claros (6) compared hypertension control in T2D patients given either manidipine or enalapril monotherapy, defining a responder to be a subject having at the end of the study either a diastolic blood pressure <90 mmHg or a reduction in diastolic blood pressure of at least 10 mmHg; and Shaibani et al. (7) defined a responder to be a subject achieving a 30% reduction in mean daily pain score in patients with painful diabetic neuropathy.

A recent position paper by the Pharmaceutical Research and Manufacturers of America (PhRMA) organization points out that the role of responder analysis in regulatory guidance documents is to help the regulatory agency interpret the clinical relevance of statistically significant results and put forth the following requirements for responder analyses: "… *1*) the criteria for 'responder' should be generally accepted, fully validated; *2*) the responder analysis should be reliable, robust, sensitive to change of disease and also be able to discriminate an experimental treatment compared to a control in a clinical trial; *3*) the cut point for dichotomization should be prespecified in the study protocol and agreed up front with regulatory agencies" (8).

A series of new-onset T1D clinical trials have been conducted in the past decade to identify safe and effective means to preserve β-cell function. Interest stems from prior observations in the Diabetes Control and Complications Trial, showing that even modest preservation of endogenous insulin secretion confers clinical benefit, with better overall glycemic control and lower risk for long-term complications (9–11). Earlier studies usually reported aggregate analyses and usually did not identify "responder" subgroups with more (or less) marked benefit from the therapy being evaluated. However, several recently published new-onset T1D clinical trials have presented responder analyses, but apparently without consideration of regulatory guidelines or the recommendations discussed above, and these have not been prespecified but rather post hoc analyses (12–15). As the ultimate goal in diabetes research is to bring effective treatments and interventions to the clinic, and as this goal inevitably will require regulatory approval, clinical researchers in diabetes should consider responder analyses as part of their prespecified clinical research strategies and must ensure as well that regulatory requirements or recommendations are met by adhering to the standards stated by the PhRMA. In this article, we propose a clinical responder definition that would assist researchers and regulatory agencies in interpreting the clinical importance of statistically significant findings for studies of interventions to preserve β-cell function in patients with newly diagnosed T1D. We then compare this definition with two previously published alternatives (13,14). In addition, we evaluate our definition against the PhRMA requirements stated above.

RESEARCH DESIGN AND METHODS

Concepts and Definitions
Table 1 summarizes the key concepts used in this article. We describe below our implementation of these concepts.

Responder Criterion
A primary clinical objective in the treatment of T1D is the preservation of β-cell function with resultant preservation of endogenous insulin production. An objectively definable criterion of clinical response is, therefore, "preservation or increase from baseline in insulin secretion." Mixed-meal tolerance testing is often used to assess

Table 1—Concepts and assessment criteria for responder definitions

Responder criterion	For a continuous response variable (e.g., C-peptide), the threshold demarcating clinically favorable values (e.g., preservation)
Intrasubject variability	Sometimes referred to as "biologic variability," this term refers to the phenomenon that repeating an assay on material taken from the same subject but at different times can lead to different assay values
Bias	The amount by which a responder definition will, on average, over- or underestimate the responder percentage in a patient population
False responder rate	The percentage of subjects that are classified as responders by the responder definition but who actually are nonresponders
False nonresponder rate	The percentage of subjects that are classified as nonresponders by the responder definition but who actually are responders
Reliability	Refers to the false responder and false nonresponder rates collectively
External validity	In general, concordance with the findings of an external study that found statistically significant differences between groups (typically defined by therapy vs. control or placebo) in the underlying continuous variable

insulin secretion by measurement of the endogenous insulin metabolite C-peptide over a 2–4-h period. Responder definitions are then typically based on changes in C-peptide exceeding some threshold (referred in this article as the "responder criterion").

Since C-peptide preservation or improvement is the goal, a logical criterion to define a responder would be a postbaseline value that is 100% or more of the baseline value of C-peptide. However, two modifications of this responder criterion have been offered, and used, in the diabetes literature to address the problems that arise from intrasubject variability (test-retest variability) (13–15). Each definition effectively lowers the 100% threshold in an attempt to counteract the fact that a true responder might occasionally have an observed change slightly less than 100% due to intrasubject variation. One definition (13) lowers the threshold to 92.5%, an amount equal to one-half the interassay variation of the C-peptide assay. The other definition (14) essentially lowers the threshold to 87.2%, which is an amount equal to the median subject-level coefficient of variation found in a test-retest study of C-peptide. In this article, we compare all three criteria, namely 100, 92.5, and 87.2%.

Bias and Reliability

However, it stands to reason that lowering the threshold from 100% will lead to overestimates of the percentage of subjects maintaining or improving C-peptide. In addition, the presence of intrasubject variability can also result in the overestimation, or even underestimation, of responder percent (16,17). The net amount of this over- or underestimation, referred to in this article as "bias," is evaluated for each of the responder definitions using the statistical method described below; classifying individual subjects and including intrasubject variability of the C-peptide assay can also lead to the misclassification of true responders as nonresponders and vice versa. The rates at which these misclassifications occur are important characteristics of a responder definition. We refer to the percent of responders who are wrongly classified as nonresponders as the "false nonresponder rate." Similarly, we refer to the percent of nonresponders wrongly classified as responders as the "false responder rate." We refer to these two rates collectively as the "reliability" of the responder classification.

Evaluation Populations

We used two evaluation populations, one to assess bias and reliability and the other to determine external validity. For the former, we used parameters estimated from Greenbaum et al. (18), who combined data from three TrialNet onset treatment studies to represent the decline in C-peptide to be expected in newly diagnosed untreated subjects (Table 2). This evaluation population is largely white (93%) and male (61%), with mean (SD) age of 18.1 years (8.8). The mean (SD) C-peptide area under the curve (AUC) at baseline was 0.71 pmol/mL (0.34). As can be seen from

this table, subsequent AUC values decline across time. In addition, a TrialNet "test-retest" study (19) provided an estimate of intrasubject variability in C-peptide AUC of 0.4167, which is the variance in C-peptide AUC measured on the same individual on two occasions within 10 days of each other. The estimate was derived from a variance components analysis in which there were random between- and within-subject effects. Intrasubject variability was then taken to be the within-subject variance. This estimate was required for the simulations described below.

We used data from the rituximab study to evaluate external validity. This study, in fact, was included with the evaluation population collated by Greenbaum et al. (18). Yet, since the evaluation of bias and reliability within the placebo-treated population does not affect, nor is affected by, the evaluation of treatment effect (external validity), there are no issues of possible confounding results between the two analyses.

External Validity

As stated previously, one of the requirements suggested by the PhRMA is that the responder definition should be "reliable…and also to be able to discriminate an experimental treatment compared to a control in a clinical trial" (8). Our evaluation of reliability has been described above. We will refer to the ability to discriminate treatment from control as "external validity" of the responder definition, which will need to be evidenced by application in a real-world setting. As described above, in our evaluation, we used data from the TrialNet rituximab study (15) to test the external validity of the responder definition by comparing the difference in percent responders between the two treatment groups. An externally valid responder definition will also detect a difference between treatment groups.

Statistical Methods

In this section, we provide a brief overview of our statistical methods. The Supplementary Data provides greater detail.

We used Monte Carlo simulation to generate samples from the Greenbaum evaluation population. We first developed a multivariate normal measurement error model with parameters from the Greenbaum et al. (18) data and the intrasubject variability in C-peptide that was estimated in the test-retest study (19). The multivariate vector consisted of four measurements of C-peptide: two at baseline and two at follow-up. At each of these time points, there are two measurements: one is the actual C-peptide value and the other is the apparent value, which is the actual value plus measurement error. Inspection of the raw data led us to use the log-transformed C-peptide data as this gave approximate normality in the univariate distributions. We then used Monte Carlo simulation to generate samples from the joint density function in order to estimate responder proportions, bias, and reliability.

We ran five separate Monte Carlo simulations representing different levels of preservation. Each simulation used

Table 2—Evaluation populations

Purpose	Bias and reliability		External validity+	
Source	Test-retest study (*n* = 148)	Greenbaum study (*n* = 191)	Rituximab treated (*n* = 57)	Rituximab placebo (*n* = 30)
Age (years)*	16.2 ± 6	18.1 ± 8.8	19.0 ± 8.6	17.3 ± 7.8
Sex (% female)	39%	38%	37%	40%
Race (% white)	86%	93%	96%	97%
Ethnicity (% Hispanic)	7%	7%	5%	10%
C-peptide 2-h AUC (pmol/mL)*				
Baseline	0.49 ± 0.50	0.71 ± 0.34	0.77 ± 0.43	0.71 ± 0.37
Follow-up				
<10 days	0.46 ± 0.44			
6 months		0.56 ± 0.39	0.79 ± 0.57	0.59 ± 0.51
12 months		0.43 ± 0.34	0.62 ± 0.42	0.45 ± 0.48
24 months		0.36 ± 0.37	0.47 ± 0.46	0.35 ± 0.42

*Mean ± SD. +Both data sets are from the rituximab study. The C-peptide data reported here are only from subjects used in the intention-to-treat analysis. For baseline, mean and SD do not match the values reported in Table 1 of the rituximab study, which included all subjects.

a different assumed percent responder (i.e., the percent with preserved C-peptide) in the population and generated 100,000 observations. One simulation assumed a 20% responder percent, which is representative of the rate (21.2%) seen in the untreated population of Greenbaum et al. (18). The other simulations used increasing rates of responders, namely 30, 40, 50, and 60%, and are therefore reflective of increasing levels of preservation.

For each sample of 100,000, we computed responder percentages for each criterion and false-positive and false-negative percentages using simple proportions. This number of simulations enabled us to estimate each of these quantities within, at most, 0.3%, using a 95% CI.

Bias was computed as the difference between the apparent responder percentage and that assumed in the simulation. Reliability was measured by computing the false responder and false nonresponder proportions using the method described in the Supplementary Data. External validity was evaluated with the rituximab study using conventional methods for estimating and comparing (χ^2 test) two independent proportions of responders between the treatment groups.

RESULTS

Bias: Over- or Underestimating C-peptide Preservation

Table 3 compares bias among the three criteria at each of three times of follow-up after baseline. In every case, the 100% criterion is nearly unbiased, with overestimation of the assumed responder percent by no more than 0.4%. On the other hand, the other two criteria have greater bias ranging from 1.8 to 9.3%. In addition, comparison of bias across time suggests that less bias occurs in later time points.

Based on the comparison of bias, we conclude that lowering the criterion from 100% leads to substantial overestimation of percent responders in the population

and that the two alternative criteria have unacceptably large bias. We therefore focus the remainder of our evaluation on the 100% criterion for C-peptide preservation.

Reliability: Incorrect Classification of the Individual Subject

Figure 1 summarizes the rates of the incorrect classification of individual subjects for the responder criterion of 100% with different assumed values of percent responders in the population and at different lengths of follow-up. False responder rates ranged from 6 to 36%. False nonresponder rates ranged from 3 to 22%. These two rates change in opposite directions as the assumed

Table 3—Bias† in proposed vs. published responder criteria

	Assumed percent maintaining C-peptide at 6 months				
Criterion	20%	30%	40%	50%	60%
100%	+0.2	+0.0	−0.4	+0.0	−0.3
92.5%	+4.0	+4.9	+4.9	+5.3	0+4.9
87.2%	+7.4	+8.7	+9.1	+9.3	+8.6
	Assumed percent maintaining C-peptide at 1 year				
Criterion	20%	30%	40%	50%	60%
100%	+0.0	−0.3	+0.1	−0.2	−0.1
92.5%	+2.0	+2.0	+2.7	+2.5	+2.6
87.2%	+3.5	+3.9	+4.8	+4.5	+4.5
	Assumed percent maintaining C-peptide at 2 years				
Criterion	20%	20%	20%	20%	20%
100%	−0.1	−0.3	−0.1	+0.0	−0.0
92.5%	+1.8	+2.0	+2.5	+2.8	+2.6
87.2%	+3.4	+4.0	+4.5	+4.9	+4.5

Values denoted by + (or −) "0.0" were such that they were slightly above (below) zero and rounded to zero. †Bias is defined to be proportion in excess (+) or in deficit (−) of the assumed percent retaining C-peptide in the population.
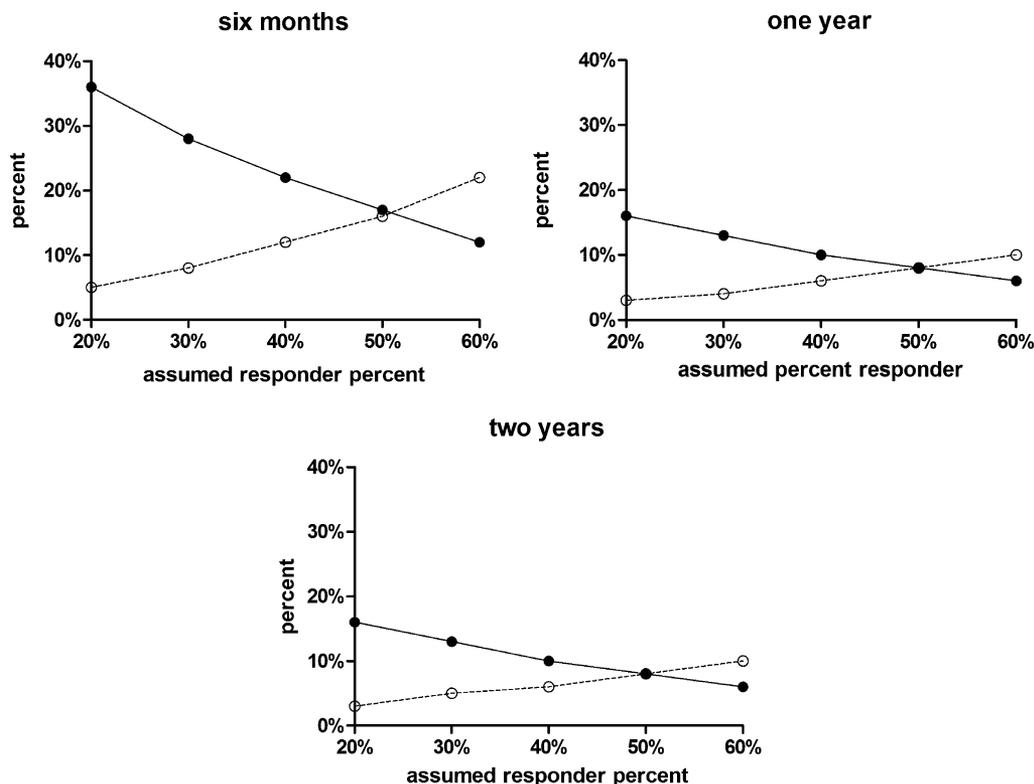
Figure 1—False nonresponder (○) and false responder (●) percentages for the 100% criterion as a function of the percent responders assumed for the population and the time of the C-peptide end point (6 months and 1 and 2 years).

percent responders increases from 20 to 60%. Rates of false responders decrease and rates of false nonresponders increase as the percent of true responders in the population increases. Although the greatest rates of incorrect classification occur at 6 months of follow-up, the rates at 1- and 2-year follow-up were nearly identical.

External Validity: Detection of a Treatment Difference When One Exists

The rituximab study (15) found statistically significant higher mean C-peptide AUC in the treated subjects at 6 months, 1 year, and 2 years of follow-up. Figure 2 presents responder percentages using the 100% responder criterion in the rituximab evaluation population. This responder definition also determined there to be a benefit from rituximab in that the preservation of C-peptide was greater in the treated than the untreated subjects at each follow-up period. Although the percentage responders were statistically different only at 6-month follow-up, we note that the differences were smaller at later points (although still substantive), and lack of statistical significance might be simply due to limited sample size. Thus, we conclude that the proposed responder criterion of 100% was able to detect a previously determined efficacious treatment and so is externally valid.

DISCUSSION

Responder studies should contribute to the translation of effective treatments and interventions to the clinic. Since ultimately this translation will involve regulatory approval, we recommend that every clinical trial prespecify a responder analysis as part of the statistical analysis plan, and that the responder definition reflect the requirements and suggestions of regulatory agencies. Consequently, we suggest that responder definitions be viewed as being akin to any assay or test and be required, therefore, to be evaluated for bias, reliability, and external validity prior to its use in clinical research. Further, in any published study reporting the results of the analysis of responders, we recommend that such evaluations need to either be incorporated with the introduction of a new responder definition or, otherwise, a citation to the source of such evaluation.

In this article, we have introduced a clinical responder definition specifically targeted at studies of the 6-month preservation of β-cell mass and/or function in T1D as measured by 2-h–stimulated C-peptide AUC determined from mixed-meal tolerance testing. Our focus on 6-month change is relevant to studies needing an early end point to show efficacy. We show that this definition meets criteria recently recommended by the PhRMA and is less biased than published alternatives. In fact, it can be shown mathematically that this definition has the least bias among alternative responder definitions that are based on changing the threshold from 100% of baseline out to 6 months. We therefore recommend its use in studies submitted for regulatory approval in both the U.S. and Europe.
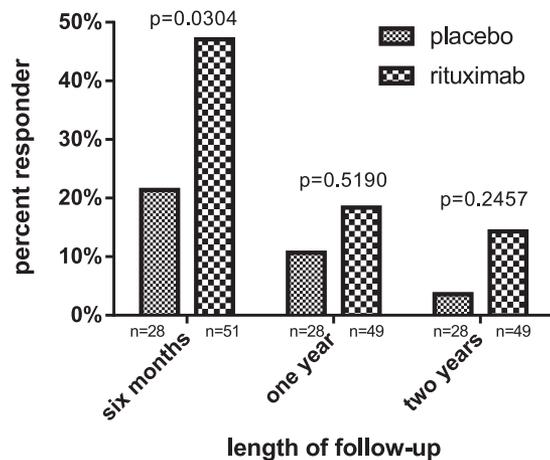
Figure 2—Responder percentages based on the 100% criterion in the rituximab study.

However, a fundamental question faces every study aiming to preserve β-cell mass/function: whether or not it is realistic to expect no decline after diagnosis, particularly as one pushes the primary end point out further from diagnosis, often to 12 or 24 months. It might be more realistic to consider "preservation" to be the slowing of decline. Responder definitions based on this definition of preservation would then require specification of a level of decline that would still be considered a positive response to treatment. Without clinical studies tying C-peptide to a definitive clinical end point, however, the selection will be arbitrary. This challenge will be faced if we consider other possible end points as well. For example, one might consider using the time-to-peak C-peptide falling below 0.2 nmol as an end point and then define a responder as someone whose time exceeds some threshold. Here, again, selection of the threshold is arbitrary until the extension of time can be directly linked to clinical outcome. Similar challenges face the use of composite end points, e.g., using both HbA_{1c} and insulin use, to define the clinical responder (20,21).

Yet, the use of arbitrary thresholds to define a responder is found in other medical settings. Responder criterion selection is often arbitrary because typically there exists no objectively definable single point in the measurement of therapeutic response that unequivocally demarcates favorable response from unfavorable response. For example, the RECIST (Response Evaluation Criteria In Solid Tumors) threshold of partial response (30% reduction in tumor volume) was established by expert consensus and, according to Therasse et al. (22) is an "arbitrary convention." We recommend that the field of T1D similarly establishes, by consensus, a definition of responder related to the slowing of β-cell decline. (In addition, a definition of "deleterious effect" of treatment should be established as well.) Of course, such a definition will need to also be assessed to ensure that it meets the PhRMA

recommendations if it is to be used to support efficacy claims in insulin preservation studies submitted to U.S. and European regulatory agencies. We recommend the methodology used in this article be considered for that purpose.

The "placebo responder" presents a dilemma when interpreting responder analyses. In our analysis of the rituximab data, we observed a 32% responder rate in the subjects treated with placebo. But, how can a placebo-treated subject be declared a "responder"—what are they responding to? We recommend that a complete responder analysis include a clinical interpretation of the placebo responder (e.g., the responses reflect the effectiveness of insulin therapy) and assess as well the clinical significance of the increase in percent responders above the placebo responder rate that is due to treatment. For example, recalling our validation data, was the 10% increase in the percentage of subjects with C-peptide preservation at 6 months by rituximab above the percentage expected to preserve without treatment (as seen in the placebo group) large enough to clinically justify the adoption of rituximab as first-line therapy?

We have shown that the bias of the 100% C-peptide preservation criterion is, for practical purposes, negligible (0.2%) and that it is sensitive enough to detect an efficacious treatment. Therefore, to support efficacy claims of β-cell preservation therapies in T1D submitted to U.S. and European regulatory agencies, we recommend use of the 100% C-peptide preservation criterion. Individual classification is, however, unreliable and thus we agree with the PhRMA position paper, which also advises that responder analysis should be performed only after statistical significance of the original continuous variable (in our case, C-peptide change from baseline) has first been established and then the responder analysis should be considered as supportive analysis (8).

Caution must still be exercised even when a responder analysis is relegated to "supportive" or "secondary" status. It is well known in the statistical literature that the dichotomization of a continuous measurement reduces statistical power (23). Moreover, in the presence of "measurement error," as is the case with the C-peptide assay, the estimated treatment difference between responder percentages will be underestimated (17). Further issues arise from assay variation by the misclassification of true responders as nonresponders and true nonresponders as responders. Such misclassification leads to estimation bias and lowers the power of tests of association between categorical variables, such as the presence of a gene allele or sex of the subject, and responder status (24,25).

One limitation of our study is that we have not considered the use of responder analyses that attempt to identify factors, such as genotype or phenotype, that determine response to therapy, i.e., responder studies restricted to only treated subjects. However, such studies are typically exploratory and not intended to support efficacy claims. Although regulatory considerations are

not directly relevant, we do recommend that the responder definition used in such studies be at least assessed for bias and reliability so that a realistic interpretation of findings can be made. Another limitation is that our evaluations required computer simulations with assumed values. In the case of the estimate of intrasubject variability, we note that the parent study collected C-peptide measurements from subjects having a broad range of diabetes duration at baseline. It is conceivable that data from a study with more homogeneous diabetes experience might have led to different estimates of within-subject variability and altered some of our findings.

In summary, we recommend that the use of responder analysis in insulin preservation studies be considered circumspectly as a means to help interpret the clinical relevance of statistically significant results based on C-peptide measurement and not as a primary analytic end point. Every responder definition should be prespecified and evaluated for bias, reliability, and external validity prior to its use in clinical research. When insulin preservation is the focus, the logical threshold for defining a C-peptide responder is "no change or increase from baseline." There are no advantages to, or need for, alterations to this threshold to accommodate intrasubject variation. In addition, this "logical threshold" meets the recommendations of the PhRMA, FDA, and EMA for responder analysis in β-cell preservation studies. This logical threshold provides nearly unbiased interpretations that were shown to be sensitive to the presence of a treatment effect in the case of the TrialNet rituximab study. Finally, we recommend that a responder definition based on the notion of an "acceptable" reduction of β-cell decline be established via consensus by the T1D community.

References

1. Committee for Medicinal Products for Human Use. *Note for Guidance on the Clinical Investigation of Medicinal Products in the Treatment of Diabetes Mellitus.* London, European Medicines Agency Report, 2012 (CPMP/EWP/1080/00 Rev. 1)
2. Office of Training and Communications, Division of Drug Information, Center for Drug Evaluation and Research, U.S. Food and Drug Administration. *Draft Guidance for Industry: Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention.* 2008
3. Harmonisation ICo. Harmonized Tripartite Guideline Statistical Principles for Clinical Trials: E9. Vol 63. Fed Regist 1998;49583
4. Snapinn SM, Jiang Q. Responder analyses and the assessment of a clinically relevant treatment effect. Trials 2007;8:31
5. Chiasson JL, Naditch L; Miglitol Canadian University Investigator Group. The synergistic effect of miglitol plus metformin combination therapy in the treatment of type 2 diabetes. Diabetes Care 2001;24:989–994
6. Luque Otero M, Martell Claros N; Study Investigators Group. Manidipine versus enalapril monotherapy in patients with hypertension and type 2 diabetes mellitus: a multicenter, randomized, double-blind, 24-week study. Clin Ther 2005;27:166–173
7. Shaibani A, Fares S, Selam JL, et al. Lacosamide in painful diabetic neuropathy: an 18-week double-blind placebo-controlled trial. J Pain 2009;10:818–828
8. Uryniak T, Chan ISF, Fedorov VV, et al. Responder Analyses-A PhRMA Position Paper. Statistics in Biopharmaceutical Research 2011;3:476–487
9. Palmer JP, Fleming GA, Greenbaum CJ, et al. C-peptide is the appropriate outcome measure for type 1 diabetes clinical trials to preserve beta-cell function: report of an ADA workshop, 21-22 October 2001. Diabetes 2004;53:250–264
10. Steffes MW, Sibley S, Jackson M, Thomas W. Beta-cell function and the development of diabetes-related complications in the diabetes control and complications trial. Diabetes Care 2003;26:832–836
11. Lachin JM, McGee P, Palmer JP; DCCT/EDIC Research Group. Impact of C-peptide preservation on metabolic and clinical outcomes in the Diabetes Control and Complications Trial. Diabetes 2014;63:739–748
12. Herold KC, Gitelman SE, Willi SM, et al. Teplizumab treatment may improve C-peptide responses in participants with type 1 diabetes after the new-onset period: a randomised controlled trial. Diabetologia 2013;56:391–400
13. Herold KC, Pescovitz MD, McGee P, et al.; Type 1 Diabetes TrialNet Anti-CD20 Study Group. Increased T cell proliferative responses to islet antigens identify clinical responders to anti-CD20 monoclonal antibody (rituximab) therapy in type 1 diabetes. J Immunol 2011;187:1998–2005
14. McGee P, Kraus-Steinrauf H, Greenbaum C, et al.; the Type I Diabetes TrialNet Study Group. Classification of a clinical responder for evaluation of treatment efficacy in trials of new-onset type 1 diabetes. Abstract presented at the 71st Scientific Sessions of the American Diabetes Association, 24–28 June 2011, San Diego, CA
15. Pescovitz MD, Greenbaum CJ, Krause-Steinrauf H, et al.; Type 1 Diabetes TrialNet Anti-CD20 Study Group. Rituximab, B-lymphocyte depletion, and preservation of beta-cell function. N Engl J Med 2009;361:2143–2152
16. Chesher A. The effect of measurement error. Biometrika 1991;78:451–462
17. Oppenheimer L, Kher U. The impact of measurement error on the comparison of two treatments using a responder analysis. Stat Med 1999;18:2177–2188
18. Greenbaum CJ, Beam CA, Boulware D, et al.; Type 1 Diabetes TrialNet Study Group. Fall in C-peptide during first 2 years from diagnosis: evidence of at least two distinct phases from composite Type 1 Diabetes TrialNet data. Diabetes 2012;61:2066–2073
19. Greenbaum CJ, Mandrup-Poulsen T, McGee PF, et al.; Type 1 Diabetes Trial Net Research Group; European C-Peptide Trial Study Group. Mixed-meal tolerance test versus glucagon stimulation test for the assessment of beta-cell function in therapeutic trials in type 1 diabetes. Diabetes Care 2008;31:1966–1971

20. Sherry N, Hagopian W, Ludvigsson J, et al.; Protégé Trial Investigators. Teplizumab for treatment of type 1 diabetes (Protégé study): 1-year results from a randomised, placebo-controlled trial. Lancet 2011;378:487–497

21. Hagopian W, Ferry RJ Jr, Sherry N, et al.; Protégé Trial Investigators. Teplizumab preserves C-peptide in recent-onset type 1 diabetes: two-year results from the randomized, placebo-controlled Protégé trial. Diabetes 2013;62:3901–3908

22. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst 2000;92:205–216

23. Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. Pharm Stat 2009;8:50–61

24. Goldberg JD. The effects of misclassification on the bias in the difference between two proportions and the relative odds in the fourfold table. J Am Stat Assoc 1975;70:561–567

25. Mote VL, Anderson RL. An investigation of the effect of misclassification on the properties of chi-2-tests in the analysis of categorical data. Biometrika 1965; 52:95–109