TYPE 1 DIABETES

Evaluating Preclinical Efficacy

Mark A. Atkinson

Recent disappointing results of clinical trials seeking type 1 diabetes (T1D) reversal suggest the need for a reevaluation of our translational efforts. This Commentary explores the need for standards in evaluating therapeutic efficacy in preclinical models of T1D.

The availability of nonobese diabetic (NOD) mice, a spontaneous animal model for type 1 diabetes (T1D), should in theory accelerate efforts to prevent or cure the disease. However, despite the discovery of multiple agents that demonstrate positive therapeutic outcomes in NOD mice, reversal-based immunotherapies have not to a large extent achieved their promise, including designated end points in recent phase III clinical trials. In this Commentary, I suggest that our limited translational success in the T1D field, and perhaps other disciplines, results in part from a lack of appropriate standards for evaluating therapeutic efficacy in preclinical models, leading to suboptimal trials in human patients. Implementation of such standards may yield meaningful improvements.

In the early 1980s, the discovery of two spontaneous animal models (NOD mice and BB rats) for investigating the pathogenesis of T1D (*1*, *2*), the advent of molecular biology, and the development of a variety of molecular reagents—new immunosuppressive drugs and monoclonal antibodies (mAbs) that sleuth out disease-causing immune cells—energized the scientific community. Many believed that this rich research environment would foster rapid answers to multiple questions about T1D pathogenesis and specify means by which the disease could be cured.

The ensuing decade saw a number of remarkable advances in identifying factors that contribute to T1D, including multiple disease susceptibility loci, autoantigen discovery, and the characterization of disease-related immune cells in both human patients and NOD mice, which rapidly became the preferred animal model for studies of T1D. One key hope remained unfulfilled, but only in humans: identifying a means to prevent the disease.

In contrast, in NOD mice the number of potential therapies that had the capacity for disease prevention grew rapidly: A mere five years after an early publication noted an initial handful of such efficacious therapies (*3*), more than 125 successful therapeutic methods were reported (*4*). This imbalance in therapeutic successes (T1D prevention) in NOD mice compared with human patients suggested the need to reassess the utility of this animal model to address key issues related to bench-to-bedside translation. Foremost among the issues requiring attention was the need to define and adopt standards for defining disease prevention in NOD mice, and to consider the practical implications of such approaches when applied to humans (such as age at treatment onset, safety, practicality, and the appropriate duration for clinical studies) (*5*). In addition, even the definition of "diabetes" was subject to profound variance amongst published works on T1D prevention in NOD mice (*6*). However, years later, the adoption of any real form of experimental standards for efficacy involving disease prevention that uses NOD mice has yet to occur.

BENCH TO BEDSIDE AND BACK

In terms of translation, the impact of this void in standards may have resulted in deleterious downstream effects. Essentially, all attempts to prevent T1D in humans—including many trials that tested therapies that showed efficacy in NOD mice (for example, oral nicotinamide and subcutaneous insulin)—either failed to achieve their primary trial end point or continue to require additional investigation to validate potential efficacy before movement beyond a research setting (*7*). Indeed, so disappointing were the early T1D prevention-based efforts in humans that, starting in the mid-2000s, many clinical investigators moved their efforts away from seeking the goal of disease prevention toward what is now often referred to as disease intervention [that is, reversal of overt T1D (hyperglycemia) once diagnosed].

This scenario formed an interesting setting for preclinical efforts in that the lack of success in the translation of prevention studies from NOD mice to humans swayed therapeutic research in a unique way; rather than progressing from mice to humans as normally occurs, the field experienced an influence of human study results on the design of experiments in mice. One needs only to view the trend in publications, now nearly 50 in number, on attempts to reverse T1D in NOD mice between 1990 and 2010 to see that a paradigm shift has occurred (table S1). Support for this change in research direction quickly materialized, with the therapeutic success of the immunomodulating agent antibody to CD3 (a mAb to the T cell protein CD3) being a prime example (*8*, *9*). This immunotherapy approach to disease reversal in NOD mice provided strong preliminary evidence for a potentially successful movement from bench to T1D bedside (*10*).

Other efforts in NOD mice disease reversal that have joined the ranks in translation include therapies that target B lymphocytes, vaccines directed against β-cell proteins such as glutamic acid decarboxylase (GAD), or agents such as rapamycin and interleukin-2 (IL-2) that modify immune responses (*11*). Although some positive outcomes have been noted in human clinical trials, recently reported trial results from a variety of venues indicate that these mAb, β-cell proteins, and other agents have not fulfilled their promise for therapeutic efficacy; this is especially the case if one considers the important facets of long-term durability in preserving β-cell function and in inducing true immunological tolerance. With this in mind, a major question is but a simple one in terms of wording: Why?

A NOD TO STANDARDS

The history surrounding T1D prevention efforts provides a number of lessons that the T1D research community must consider in order to improve prospects for its future. First, as was the case for disease prevention, investigators who are attempting T1D reversal in NOD mice would benefit greatly from adopting a series of experimental standards for efficacy (Table 1). For example, there remains no universal definition of T1D onset in NOD mice, either in terms of a screening method (such as glycosuria or glycemia) or the degree of glycemia that defines disease. Indeed, the criteria for diagnosis of T1D in disease-reversal studies in NOD mice range from two blood glucose determinations of >162 mg/dl to multiple measurements of values >500 mg/dl (table S1). Yet, despite having no uniform quantitative value for defining

Department of Pathology, University of Florida, College of Medicine, Gainesville, FL 32610, USA.
E-mail: atkinson@ufl.edu

Table 1. Crossing the translational divide. Shown below are steps for improving the translational potential for therapeutic intervention studies of disease reversal in NOD mice. Initial recommendations for consideration are in italics.

Establish parameters for experimental design and reporting.
• Minimum group sample size for analysis.
o *Minimum of 12 animals per group for calculation of disease frequency (that is, remissions).*
• Definition for control groups (use of placebo controls versus historical controls).
o *Must use contemporaneous controls subject to similar handling and procedures as treatment groups.*
• Age limits (earliest and latest age for mice included in study).
o *Only use animals with disease onsets ≥10 weeks and ≤24 weeks. In addition, mice should be randomized to treatment groups to assure no age bias exists amongst treatment groups (ages, with range and SE, should be reported for all treatment groups).*
• Reporting of NOD colony frequency for type 1 diabetes.
o *Background (unmanipulated) frequency of NOD colony within contemporaneous time period (≤ 2 years) should be reported.*
• Standardize time from hyperglycemic onset to initiation of therapeutic intervention.
o *0 to no more than 48 hours after diabetes onset for studies of T1D reversal.*
Impart standards for diagnosis and treatment.
• Method for defining "diabetes" (hyperglycemia in blood versus glycosuria).
o *Hyperglycemia using blood obtained from IACUC approved method (such as tail vein).*
• Definition of diabetes onset using blood glucose (for example, specific mg/dl value and number of occasions in hyperglycemic range).
o *≥250mg/dl, on two occasions, separated by 16 to 24 hours, and no more than 48 hours.*
• Agreement on insulin therapy post-diagnosis (address question of use, route of administration, type of insulin, and duration of administration).
o *Must provide insulin treatment by IACUC approved means (such as insulin injections, insulin pellet, or pump) capable of imparting diabetes management (normoglycemia/mild hyperglycemia, acceptable with local IACUC) for a period until disease reversal occurs or intervention is considered a failure.*
Define principles to assess therapeutic efficacy.
• Blood glucose at time of treatment onset should be reported for all study groups.
o *Efficacy data must compare starting blood glucose values for all treatment groups to assure there is no bias.*
• Adopt universal definition for "reversal" (what level of dysglycemia, if any, would be considered a success as well as a failure).
o *Achieve three blood glucose levels of <250 mg/dl, separated by 16 to 24 hours, within a 3-day period, taking into account the potential influence of insulin therapy (daily insulin treatment versus the use of insulin pellets will potentially require different means of analysis).*
• Standard duration for follow-up (number of days animals must be monitored to assure reversal).
o *Monitor animals for at least 90 days, with blood glucose assessments at least once per week.*

normal blood glucose, most of these studies are reported as "reversals."

There is also a lack of accounting for the marked diurnal variation that occurs in mice; over the course of a day, blood glucose values vary dramatically in mice near diabetes onset or thereafter (*12*). Second, the duration of disease reversal among reported studies varies greatly; outcome measurements range from as little as 12 days (that is, treatment provided continuously for a short period of time, and once the administration ended, the results were tallied) to 200 days (table S1). Yet here again, all of these results are described as successful disease reversals, leaving the question of the potential therapeutic impact of these great differences in treatment durability unanswered. Third, and certainly not least, is the notion of insulin treatment. Some studies test new therapies in conjunction with a variety of insulin therapies, whereas some use no insulin therapy—a practice that would seemingly have no comparator in human T1D patients with respect to clinical practice guidelines. Clearly, the influence of glycemic control in such efforts (that is, whether or not insulin therapy is provided) is too loosely defined to determine its impact on disease reversal and should be standardized. Differences in sample sizes and the definition of what one considers an adequate number of control animals for a given therapy are also causes for concern. These issues do not even address a less subtle, but perhaps equally meaningful, panel of parameters [such as age of disease onset, the colony frequency of T1D in the absence of any intervention (high or low), and animal housing conditions]. Until such standardization occurs, researchers are left with the notion that "all reversal therapies are created equal," which blocks their ability to draw meaningful conclusions from mountains of data. The recommendations in Table 1 hopefully will serve as initial guidelines for deliberation by an organized group of experts in the field; the resulting recommendations would be subject to adoption by the research community.

Today, a new generation of investigators and funding agencies appear to be in a position to recognize these issues and adopt meaningful efficacy standards for both human trials and mouse model studies. Following the initial lead of the National Institutes of Health's (NIH's) National Institute of Diabetes and Digestive and Kidney Disease's formation of a unit dedicated to preclinical trials performed by an independent research entity, other groups such as the NIH Immune Tolerance Network, the Juvenile Diabetes Research Foundation (JDRF) Autoimmunity Centers, NIH TrialNet, and others will hopefully develop programs for both standardization and confirmation of therapeutic efficacy for T1D translation efforts in studies of mice and humans. Two of these organizations—the Immune Tolerance Network and JDRF—recently developed a series of recommendations for combination therapies in NOD mice that should be subject to translational efforts in T1D with, hopefully, some degree of the aforementioned standards (*13*).

Studies of T1D reversal in humans are expensive to perform, exhausting for investigators to organize, and difficult for participating patients and their families (*14*). Clearly, the T1D community—patients, clinicians, researchers, foundations, and the pharmaceutical industry—has a vested interest in outcomes of preclinical research and would benefit over time by an adoption of such efficacy standards in mice so that the best trials, which have the highest likelihood of success, will be performed in humans.

That said, mice are mice and humans are humans. Indeed, a number of important

etiopathological differences exist between NOD mice and human subjects with T1D, which must be taken into account in therapeutic translation efforts (*15*). An abbreviated list includes contrasts such as gender bias (NOD mice show a propensity for female bias, whereas humans with T1D show equivalent disease frequencies as a function of gender), environmental exposures (NOD mice live a highly controlled and specific pathogen–free life-style, whereas humans encounter environmental variables), and the diverse number of immunological variants among humans versus a single variant in mice (*16*). Lastly, NOD mice are inbred and may merely represent one putative subform of human T1D.

Despite these variances, a number of seemingly important communalities exist between the two species, some of which would appear not to result from serendipity: the presence of self-directed autoantibodies against common autoantigens (such as insulin) and pancreatic islet cell inflammation, as well as overlap within genes that specify disease susceptibility, especially those that involve the major histocompatibility complex (*17*, *18*). With this set of commonalities and variances, it is logical to ask whether improvements could be made in the form of a new T1D animal model that is more representative of the human disease.

One potential advance that is under way is the development of humanized mouse models (*19*). The identification of immunodeficient mice that lack a functional IL-2 receptor γ chain dramatically improved the ability to observe the functional activity of these animals when engrafted (that is, receiving hematopoietic transplants) with cells from the human immune system. Natively immunodeficient mice experimentally manipulated to possess functional human immune systems allow for assessment of a variety of physiological activities, including the ability to produce autoantibodies, reject tissue allografts, mount antiviral responses, and more. Observing such outcomes provides much promise for improved investigations of disease pathogenesis and therapeutic efficacy far beyond T1D (for example, in infectious disease, oncology, and transplantation translational research). This promise does, however, remain somewhat unfulfilled because the requirement to genetically introduce (by traditional means) the plethora of presumed "necessary" human elements (growth factors, cytokines, and histocompatibility molecules) requires time. But recent reports have provided reason for optimism (*20*–*22*). Beyond rodent models, the potential also exists that nonhuman primate models for T1D could one day be generated, allowing for an improvement in our means of analyzing both therapeutic and pathogenic aspects of the disease. However, thus far no spontaneous nonhuman primate models have been uncovered, and investigations in which induced models of the disease (such as with streptozotocin treatment) might provide therapeutic insights have been hampered by fiscal and ethical limitations.

It is also possible that our collective experimental designs and interpretations have relied too much on one animal model—the NOD mouse. Perhaps observing therapeutic efficacy in a second animal model (for example, BB rats, rat models rendered insulin-dependent by poly I:C or viral treatment, or transgenic mouse models) would enhance the likelihood of translational success in humans. Researchers must not repeat the mistakes of our past but move forward using NOD mice in an optimal and proper fashion to prevent and cure T1D. Indeed, this important lesson applies to a variety of disorders for which animal models are used to guide therapeutic decisions for human clinical trials.

SUPPORTING ONLINE MATERIAL

www.sciencetranslationalmedicine.org/cgi/content/full/3/96/96cm22/DC1
Table S1

REFERENCES AND NOTES

1. S. Makino, K. Kunimoto, Y. Muraoka, Y. Mizushima, K. Katagiri, Y. Tochino, Breeding of a non-obese, diabetic strain of mice. *Jikken Dobutsu* **29**, 1–13 (1980).
2. A. A. Like, L. Butler, R. M. Williams, M. C. Appel, E. J. Weringer, A. A. Rossini, Spontaneous autoimmune diabetes mellitus in the BB rat. *Diabetes* **31**, 7–13 (1982).
3. M. A. Bowman, E. H. Leiter, M. A. Atkinson, Prevention of diabetes in the NOD mouse: Implications for therapeutic intervention in human disease. *Immunol. Today* **15**, 115–120 (1994).
4. M. A. Atkinson, E. H. Leiter, The NOD mouse model of type 1 diabetes: As good as it gets? *Nat. Med.* **5**, 601–604 (1999).
5. P. E. Beales, T. L. Delovitch, A. Signore, P. Pozzilli, Standardizing experiments with NOD mice. *Autoimmunity* **24**, 127–129 (1996).
6. L. K. Shoda, D. L. Young, S. Ramanujan, C. C. Whiting, M. A. Atkinson, J. A. Bluestone, G. S. Eisenbarth, D. Mathis, A. A. Rossini, S. E. Campbell, R. Kahn, H. T. Kreuwel, A comprehensive review of interventions in the NOD mouse and implications for translation. *Immunity* **23**, 115–126 (2005).
7. J. S. Skyler, C. Ricordi, Stopping type 1 diabetes: Attempts to prevent or cure type 1 diabetes in man. *Diabetes* **60**, 1–8 (2011).
8. L. Chatenoud, E. Thervet, J. Primo, J. F. Bach, Anti-CD3 antibody induces long-term remission of overt autoimmunity in nonobese diabetic mice. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 123–127 (1994).
9. D. Bresson, L. Togher, E. Rodrigo, Y. Chen, J. A. Bluestone, K. C. Herold, M. von Herrath, Anti-CD3 and nasal proinsulin combination therapy enhances remission from recent-onset autoimmune diabetes by inducing Tregs. *J. Clin. Invest.* **116**, 1371–1381 (2006).
10. L. Chatenoud, Immune therapy for type 1 diabetes mellitus—What is unique about anti-CD3 antibodies? *Nat. Rev. Endocrinol.* **6**, 149–157 (2010).
11. C. Greenbaum, M. A. Atkinson, Persistence is the twin sister of excellence: An important lesson for attempts to prevent and reverse type 1 diabetes. *Diabetes* **60**, 693–694 (2011).
12. U. Klueh, Z. Liu, B. Cho, T. Ouyang, B. Feldman, T. P. Henning, M. Kaur, D. Kreutzer, Continuous glucose monitoring in normal mice and mice with prediabetes and diabetes. *Diabetes Technol. Ther.* **8**, 402–412 (2006).
13. J. B. Matthews, T. P. Staeva, P. L. Bernstein, M. Peakman, M. von Herrath; ITN-JDRF Type 1 Diabetes Combination Therapy Assessment Group, Developing combination immunotherapies for type 1 diabetes: Recommendations from the ITN-JDRF Type 1 Diabetes Combination Therapy Assessment Group. *Clin. Exp. Immunol.* **160**, 176–184 (2010).
14. M. A. Atkinson, It's time to consider changing the rules: The rationale for rethinking control groups in clinical trials aimed at reversing type 1 diabetes. *Diabetes* **60**, 361–363 (2011).
15. B. O. Roep, M. Atkinson, M. von Herrath, Satisfaction (not) guaranteed: Re-evaluating the use of animal models of type 1 diabetes. *Nat. Rev. Immunol.* **4**, 989–997 (2004).
16. B. O. Roep, M. S. Atkinson, Animal models have little to teach us about type 1 diabetes: 1. In support of this proposal. *Diabetologia* **47**, 1650–1656 (2004).
17. E. H. Leiter, M. von Herrath, Animal models have little to teach us about type 1 diabetes: 2. In opposition to this proposal. *Diabetologia* **47**, 1657–1660 (2004).
18. J. A. Bluestone, K. Herold, G. Eisenbarth, Genetics, pathogenesis and clinical interventions in type 1 diabetes. *Nature* **464**, 1293–1300 (2010).
19. L. D. Shultz, F. Ishikawa, D. L. Greiner, Humanized mice in translational biomedical research. *Nat. Rev. Immunol.* **7**, 118–130 (2007).
20. S. Jaiswal, T. Pearson, H. Friberg, L. D. Shultz, D. L. Greiner, A. L. Rothman, A. Mathew, Dengue virus infection and virus-specific HLA-A2 restricted immune responses in humanized NOD-scid IL2rgammanull mice. *PLoS ONE* **4**, e7251 (2009).
21. S. J. Libby, M. A. Brehm, D. L. Greiner, L. D. Shultz, M. McClelland, K. D. Smith, B. T. Cookson, J. E. Karlinsey, T. L. Kinkel, S. Porwollik, R. Canals, L. A. Cummings, F. C. Fang, Humanized nonobese diabetic-*scid IL2rγ*^{null} mice are susceptible to lethal *Salmonella* Typhi infection. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15589–15594 (2010).
22. T. Willinger, A. Rongvaux, H. Takizawa, G. D. Yancopoulos, D. M. Valenzuela, A. J. Murphy, W. Auerbach, E. E. Eynon, S. Stevens, M. G. Manz, R. A. Flavell, Human IL-3/GM-CSF knock-in mice support human alveolar macrophage development and human immune responses in the lung. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2390–2395 (2011).

23. **Competing interests:** The author serves in an advisory role for the following companies: Genzyme, Diamyd, Amylin, GSK, Tolerx, and Diakine. **Acknowledgments:** The author thanks M. Parker for his remarkable effort in gathering the material in Table 1 and D. Schatz, C. Wasserfall, T. Brusko, and M. Haller for critical discussions leading to this manuscript. **Funding:** This views expressed here were developed, over time, with funding from NIH, JDRF, American Diabetes Association, Brehm Coalition for Type 1 Diabetes Research, and Jeffrey Keene Family Professorship.

10.1126/scitranslmed.3002757

Citation: M. A. Atkinson, Evaluating preclinical efficacy. *Sci. Transl. Med.* **3**, 96cm22 (2011).

Science Translational Medicine

Evaluating Preclinical Efficacy

Mark A. Atkinson

Sci Transl Med **3**, 96cm2296cm22.
DOI: 10.1126/scitranslmed.3002757