Diabetologia (2006) 49:1989–1995
DOI 10.1007/s00125-006-0348-7

EDITORIAL

Declassifying diabetes

E. A. M. Gale

Published online: 5 July 2006

He who first gave names, gave them according to his conception of the things which they signified, and if his conception was erroneous, shall we not be deceived by him?

Socrates

Why classify?

Biologists devote a lot of time and energy to classification, and consider this a necessary prelude to more detailed investigation. Physicians adopt a more relaxed attitude, and a glance through a medical textbook soon reveals that we lack a rational, or indeed uniform, basis for classifying disease, or for sub-classifying a given disorder. Our schemes are pragmatic and based around the tools at our disposal, the assumptions we make when we employ them, and the use we intend to make of the definition.

Some common features do, however, emerge from the various descriptions. A disease classification incorporates (or assumes) a clinical description and relates this to the natural history of the disorder and the spectrum of associated disability. Clinical or laboratory measures are used to draw a boundary between disease and non-disease. Where possible, the condition is described in terms of the causal agent or mechanism; when this is not possible, the disorder is defined in terms of its footprints (e.g. histological change or functional disorder) or its prognostic implications. The condition may then be staged in terms of severity or progression, often with differing therapeutic implications for each stage.

A classification, in other words, is a construct—or paradigm—that encapsulates current scientific understanding of a disease, and offers guidance as to how this might translate into clinical practice. Based as it is upon incomplete knowledge and understanding, any such formulation can only be provisional, and this will apply with particular force to a condition such as diabetes whose causes are largely unknown. If the fit between theory and practice is good, a creative dialogue is established whereby science informs clinical care, and clinical outcomes guide basic research. Figure 1 shows how such a dialogue might evolve with respect to diabetes. The central path, followed by clinicians, provides the clinical guidelines that direct them from evaluation to therapy. The two flanking paths, taken by investigators, follow the aetiological route to prevention and the functional route to intervention—although in practice these pathways frequently overlap.

A constructive dialogue is established when the fit between clinical observation and basic science is good. But suppose the fit is not a good one. Constructive opportunities will be missed, and much time wasted in the exploration of blind alleys. As Socrates pointed out, a name implies a concept, a concept implies a set of invisible assumptions, and each assumption represents a question that someone has forgotten to ask. A classification may thus come to embody outworn concepts that prevent us from seeking or applying the new information we need. This, as I

E. A. M. Gale (✉)
Diabetes and Metabolism, Division of Medicine,
University of Bristol,
Bristol, UK
e-mail: Edwin.Gale@bristol.ac.uk

E. A. M. Gale
Diabetes and Metabolism,
Medical School Unit Southmead Hospital,
BS10 5NB,
Bristol, UK

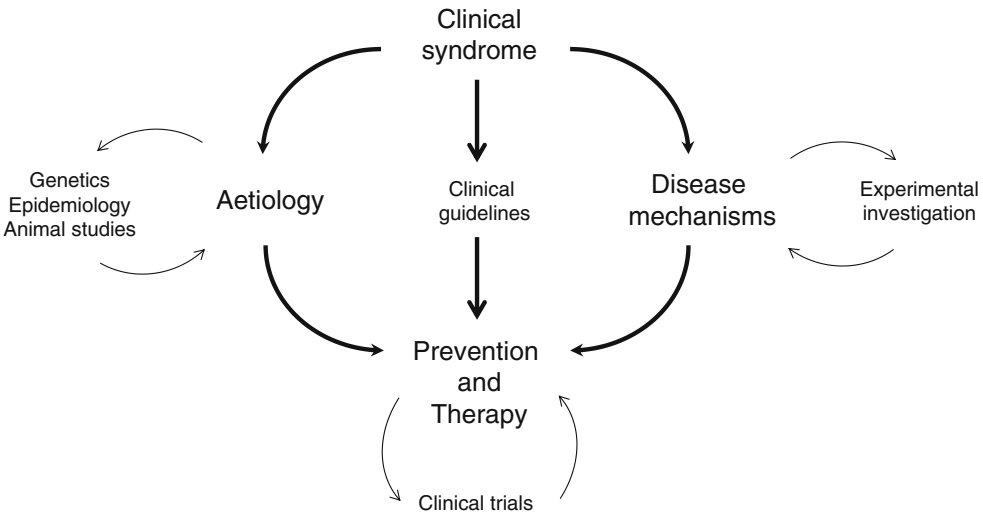

Fig. 1 The classification of diabetes begins with the identification of clinical syndromes. A constructive dialogue is established when the fit between clinical observation and investigation of aetiology and mechanism is good, but a poorly conceived classification can act as a brake on further understanding

shall argue, is now the case with our conventional distinction between type 1 and type 2 diabetes. We need to consider whether this distinction is always useful. Does it act as a stimulus for thought, or as a substitute for it? Does our problem lie in the pathology of disease, or in the way we think about it?

Orphan observations

Although the existence of two main forms of diabetes could be considered the central dogma of research and clinical practice in diabetes, the diabetes community once believed, with equal conviction, that there was only one form of the condition. The transition from one set of beliefs to the other was abrupt, and the terms type 1 and type 2 diabetes, ignored when first proposed in 1951, were rapidly accepted when reintroduced in 1976 [1]. The dual nature of diabetes was endorsed by US National Diabetes Data Group (NDDG) in 1979, by the World Health Organization (WHO) expert committee in its 1980 classification [2, 3], and has been axiomatic ever since.

The earlier belief in the unity of diabetes persisted in the face of a mounting series of orphan observations. An orphan observation is a fact that does not fit *and is therefore ignored*. This situation persists until the intrusive observation is reconciled with previous knowledge, at which point the paradigm shifts, and the orphan suddenly finds that many parents have come forward to claim it.

Early orphan observations relating to the duality of diabetes included the obvious difference between young thin patients whose lives depended on insulin, and overweight older patients who could survive without it. To this was added Himsworth's observation of insulin-sensitive and insulin-insensitive forms of diabetes in 1936 [4], followed in 1951 by the first use of the terms type 1 and type 2 diabetes [5]. The latter paper linked central obesity to insulin resistance, hypertension and arterial disease in its description of type 2 diabetes. None of these observations seriously dented the prevailing belief in the unity of diabetes. Instead, the heterogeneity of diabetes only came to be accepted when the concept of autoimmunity had been developed, islet cell antibodies had been identified, and the HLA associations of diabetes in the young were described. The prevailing assumption that diabetes was a single disease then disappeared like a puff of smoke. No one paused to defend what everyone had previously believed [1].

We may wonder why it took so long for the message to get through, but it would be wrong to dismiss our predecessors as doctrinaire or obtuse—or to assume that we ourselves are exempt from these characteristics [6]. Elliott Joslin, the leading advocate of the unity of diabetes, was well aware of Himsworth's work on insulin sensitivity and insensitivity, and went so far as to repeat Himsworth's test, although he did not publish the study. He concluded that patients could indeed be divided according to insulin sensitivity, but that the overlap between clinical phenotypes was so great that the test added little useful information [7]. The difference, as we can now see, was one of interpretation. Himsworth focused on the extremes, and saw two distinct conditions, whereas Joslin focused on the variation in between, and saw continuity.

Another way of expressing the two points of view is to say that phenotypic variation within the diabetes spectrum could be considered either as a *categorical* variable or as a *dimensional* variable [8]. The Himsworth school of thought is categorical, and assigns each case of diabetes to one of two distinct clinical entities. The Joslin school is dimensional, and sees two clinical extremes, linked by a range of overlapping phenotypes, with no

clear dividing line down the middle. From this perspective, it could reasonably be argued that *both* Himsworth and Joslin were correct. Comparison of extremes will inevitably suggest that there are two distinct forms of diabetes, whereas consideration of the whole range of phenotypic variation will give the impression of a continuum.

The two faces of diabetes

THE NDDG and WHO reports, submitted in 1979 and 1980, respectively, endorsed the duality of diabetes [2, 3]. Their major achievement was, however, to provide standardisation of glucose testing and an estimate of vascular outcomes based on prospective analysis of populations studied with the glucose tolerance test. This analysis identified two levels of vascular risk: diabetes (retinopathy and arterial risk) and impaired glucose tolerance (arterial risk but no retinopathy). Each required a different approach to therapy: non-specific reduction in vascular risk for those with impaired glucose tolerance, and specific glucose-lowering measures for those with a diagnosis of diabetes.

Diabetes was then subdivided according to severity, as judged by a perceived need for insulin to maintain a healthy life, while those who did not need insulin were further subdivided into the obese and the non-obese. The approach was unashamedly pragmatic, and fed directly into treatment guidelines ranging from lifestyle change to insulin. Subdivision of the non-insulin-requiring by weight also had therapeutic implications, since the obese would be treated with diet followed by metformin (except in the USA), whereas the lean would get sulfonylureas. Each step in classification thus implied a corresponding therapy (Fig. 2).

In 1997, the American Diabetes Association (ADA) expert committee [9] argued that a classification based on therapy was unsatisfactory. There were good reasons for this, not least the widespread clinical trend towards earlier elective use of insulin in type 2 diabetes. The new classification was therefore based, not on a clinical decision concerning the use of insulin, but rather on the functional basis of such a decision, i.e. the presence or absence of useful residual insulin secretion. This allowed type 1 diabetes to be defined in terms of insulin deficiency, and then subdivided on aetiological grounds into type 1A (immune-mediated) and type 1B (non-immune-mediated). Type 2 diabetes, in contrast, remained a purely functional category, characterised by ongoing insulin secretion, exclusion of other known causes of diabetes, and absence of the aetiological hallmarks of type 1A diabetes.

There are some obvious limitations to this formulation, and some that are less obvious. An obvious limitation is that the classification describes differences that it fails to define. It does not explain how residual insulin secretion should be measured or where the cut-off should be drawn. Nor does it specify how type 1A diabetes should be diagnosed. A further limitation is that a relatively distinct clinical entity—type 1A diabetes in the young—is set against a conglomeration of undefined pathologies summarised under the heading of type 2 diabetes. Less obvious, but equally important, the formulation incorporates several errors of reasoning. Before we explore these in more detail, we will first consider the one form of diabetes that does appear to have a secure aetiological basis and question how secure this basis really is.

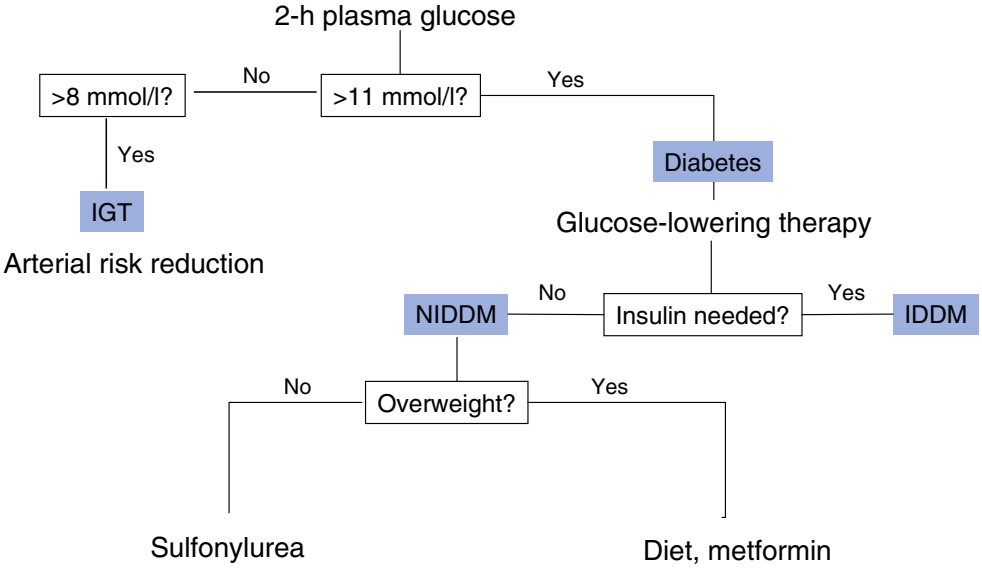

Fig. 2 The WHO classification of diabetes related glucose levels to vascular risk, and used this to derive a therapeutic pathway although specific oral agents are not mentioned. Changes in insulin use soon undermined this approach. *IDDM* Insulin-dependent diabetes mellitus, *NIDDM* non-insulin-dependent diabetes mellitus

The aetiological dimension

There is good evidence that one variant of diabetes, type 1A, is mediated by immune processes. This is an important first step towards an aetiological basis for classification—although cynics might point out that we have merely explained one process of unknown cause in terms of another. The conventional description of type 1 diabetes as a childhood disorder, fatal unless treated with insulin, is highly distinctive. Not surprisingly, almost all our incidence data are based on children diagnosed under the age of 15 years, and almost all studies of its natural history are based on individuals diagnosed in childhood or youth. Such cases form the basis of our understanding of the genetic basis of the condition and of its immune pathogenesis.

The limitation of this approach is that it focuses exclusively upon one end of a clinical spectrum. Most cases of type 1 diabetes, or so it is believed, present in adult life. How many are there? No one knows. Our ignorance of the overall contribution of immune processes to the lifetime burden of diabetes remains an embarrassing gap in our understanding of the disorder. This is because the diagnostic criteria that work so well in childhood falter with increasing age at onset, the immunogenetic criteria blur, and progression to insulin deficiency is less rapid. The final frontier, latent autoimmune diabetes in adults (LADA) can be identified only by the softest of all endpoints, the selected cut-off level of an antibody assay. Our habit of referring to such patients as antibody 'positive' or 'negative' satisfies our thirst for converting continuously distributed variables into categories, but implies certainty where none exists [10]. Even the most distinctive feature of childhood-onset type 1 diabetes—the fact that it presents in childhood—has been eroded by the rise of obesity-related diabetes in the same age group. In other words, our starting assumption that type 1 diabetes is a distinct category has prevented us from seeing that it actually behaves very much like a dimension.

The purpose of this discussion is not to argue type 1 diabetes out of existence. Focus on its most distinctive expression has opened up possibilities of disease prediction and prevention that are unique to this form of diabetes. This advance in knowledge has, however, come at a cost. The cost is that we buy into a whole set of invisible assumptions. The mental circuit operates in the following way: focus on the extreme (as against the more prevalent) form of the disorder emphasises its distinctive nature. This encourages us to view it as a separate disease category. A categorical view of type 1 diabetes encourages us to see it as an all-or-none disorder, which an individual either acquires in childhood or never does. It is a short step from here to the assumption that external factors determine this outcome, and the result is the largely fruitless hunt for 'trigger factors'. As I have pointed out elsewhere, a dimensional view of immune-mediated diabetes opens up a whole new perspective on the nature and possible prevention of the disorder [11].

Meanwhile, our understanding of the involvement of immune processes in the causation of diabetes remains limited. We do not know whether disordered immune function is the cause, precondition or consequence of beta cell damage. Our notion of type 1 diabetes is based around a pathological lesion that cannot be measured in life, and a sequence of immunopathological events described mainly in the mouse. The concept of autoimmunity, in other words, is valuable in certain contexts but does not provide a secure, easily measurable, or consistently useful means of dichotomising diabetes. A further limitation is that this mindset traps us into the assumption that immune processes are either totally responsible for the development of diabetes or not at all, excluding the more realistic possibility that both immune-mediated and non-immune-mediated processes might act in synergy, especially in later onset cases.

The functional dimension

We have seen that the 1979 NDDG classification subdivided diabetes according to a perceived need for insulin therapy, and that the 1997 classification sought to replace this with a functional measure based on insulin deficiency. The rationale is clear: type 1 diabetes is the consequence of beta cell destruction, and the resulting state of insulin deficiency should therefore be measurable in terms of circulating C-peptide. Low C-peptide levels at diagnosis will indeed identify insulin-sensitive individuals with advanced beta cell loss; however, it is worth pointing out that this situation is generally obvious on clinical grounds.

The problem, as always, lies in how to interpret the test and where to draw the line. Residual insulin secretion, as measured at the onset of hyperglycaemia, reflects the existing balance between insulin supply and insulin demand. Insulin-sensitive individuals present with advanced beta cell failure because they are insulin sensitive, and insulin-insensitive individuals retain insulin secretory reserve at the time of diagnosis because they are insulin-insensitive. C-peptide levels at diagnosis therefore tell us more about insulin sensitivity than the causes of beta cell failure, or the rate at which this has progressed. The rationale for discriminating between aetiological types of diabetes by estimation of residual insulin secretion at the onset of diabetes therefore risks a logical loop from which there is no exit.

Self-confirming circularity

It is time to apply the litmus test of rationality to the conventional division of diabetes. The logical pitfalls are not far to seek, and circular reasoning takes pride of place among them. A fable may help to clarify this point. Let us assume that I want to study the characteristics of tall and short people. I select people who are obviously tall, and people who are obviously short, and soon notice that that the groups differ in other ways. There are genetic differences (tall people have taller parents; there are genetic syndromes associated with short and tall stature), ethnic differences, developmental differences (short people may have been less well nourished in the uterus), and environmental differences (taller people may have been better nourished following birth). Easily verifiable though these differences are, I would be considered foolish if I went on to propose that there were two distinct variants of our species, known as type 1 humans and type 2 humans. But suppose (this is a fable) that others believed me, and accepted the classification. All future research studies would focus on the comparison between short and tall people, and humans of intermediate stature would be excluded from their investigations. The distinction between type 1 and type 2 humans would be repeatedly confirmed, and future expert committees would be called upon to debate whether our species should be divided at 168 or 170 cm (Fig. 3).

The fallacy of such a distinction is obvious. We know that height is a continuously distributed variable, and we know that the comparison between short and tall people merely demonstrates that people selected for certain characteristics will differ in associated ways, and is therefore circular. But is it not equally circular to compare the characteristics of diabetes in lean children and obese adults? Have we advanced knowledge by establishing that children are leaner, more sensitive to insulin, and less prone to atherosclerosis, hypertension and hyperlipidaemia? The real test of the existence of two forms of diabetes would be to demonstrate that they could be distinguished from one another in individuals matched for age and BMI. And this could scarcely be contemplated, because our clinical classification is based upon *the context in which diabetes develops*, rather than the nature of the disease process itself.

Circular reasoning apart, two other logical errors enter our scheme of classification. The first is inappropriate lumping of clinical features that have no necessary relationship with one another. Youth, leanness, autoimmunity and insulin sensitivity cluster at one end of our clinical scale, and middle age, obesity, lack of evidence of autoimmunity and insulin insensitivity at the other, but there is no necessary interconnection between these associated features. Obesity does indeed predispose to diabetes, much as cigarette smoking does to lung cancer, but lung cancer is not defined in terms of cigarettes, and obesity does not define diabetes. We thus confuse the contributory factors that determine *how* diabetes will present, with the more fundamental mechanisms that determine *whether* it will develop at all, and fling them into the same pot for analysis. Progress in biology is made when complex phenomena are analysed into their component elements, each individually capable of solution. Progress is unlikely to be made when loosely connected phenomena are lumped together for consideration, yet this is precisely what we do when classifying diabetes.

A third logical error is known as the *essentialist* or *ontological* fallacy [12, 13]. We fall into this when we invent categories for our own convenience and then treat them as if they had a real existence. Type 2 diabetes is a catch-all category that includes most forms of non-immune diabetes, and has no single defining characteristic. So far, so good—but we then go on to talk as if this loose constellation of clinical and pathological features has a real existence. We write about its aetiology, its therapy, and—the ultimate delusion—type 2 diabetes in the mouse.

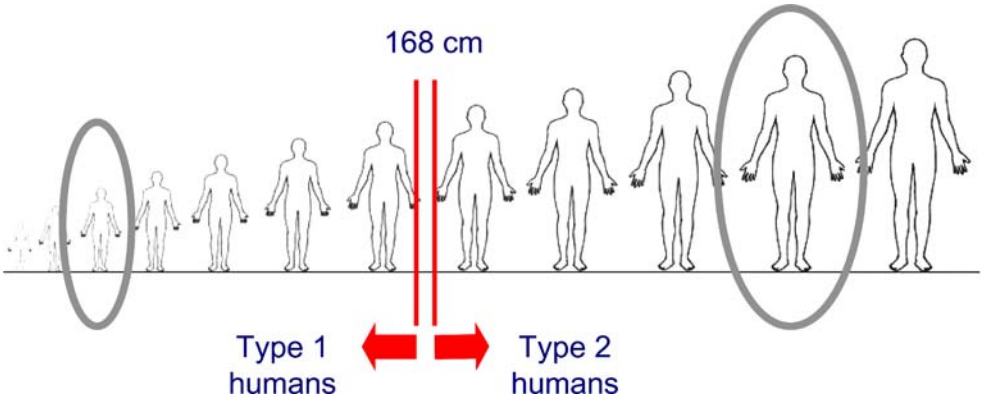

Fig. 3 Height is a continuously distributed variable. Comparison of extremes might, however, suggest that humans fell into two distinct categories. This belief would be reinforced if all subsequent studies were based on comparisons between the very tall and the very short

Utility or futility?

The current classification of diabetes is based upon the assumption of two categories of diabetes, but offers no rational basis for this assumption, and does not explain how the difference might be defined. Clinical papers routinely cite detailed WHO or ADA criteria for diagnosing diabetes; the lack of equivalent criteria for distinguishing one type of diabetes from another is equally evident. In truth, as we have seen, the distinction only works well when the two extremes are compared. This makes little difference to clinicians, who are generally quite happy to carry on assigning patients to treatment on the basis of age, ketonuria, adiposity and perceived need for insulin, just as they have been doing since the 1920s. Clinicians do not need a committee to tell them this.

Clinical trials require a more formal basis for assigning patients to one category or another, and triallists have filled this diagnostic vacuum by devising their own criteria. Take, for example, the UKPDS, the iconic study of type 2 diabetes. Patients were assigned this diagnosis if they were over 25 years of age at diagnosis, had a fasting plasma glucose below 15 mmol/l, and no ketonuria [14]. This is about as broad a definition of type 2 diabetes as one could possibly imagine, and a fair proportion of the younger patients must also have met many of the standard criteria for type 1 diabetes. This diagnosis was, however, excluded by the entry criteria, so patients subsequently identified by autoantibody testing were lumped together under the designation of LADA. It so happened that our own laboratory retested samples from some of these patients, using a later generation assay, and found (as is inevitable), that a proportion could not now be said to have LADA. So, who had type 1 diabetes, who had type 2 diabetes, and who had LADA?

More important, did any of this matter? The largely unconsidered message of UKPDS—a study that was more interested in treatment schedule than classification of diabetes—was that autoantibody status made little difference to treatment or outcome, other than the fact that those with autoantibodies were more likely to require insulin. This apart, patients retrospectively assigned a diagnosis of LADA had the same prognosis as those without, and initial randomisation to insulin as against diet or tablets made no difference to the outcome at 10 years [15]. This large study therefore suggests that patients who are aged 25 or over at onset of diabetes, and who do not need immediate insulin treatment, can safely be offered empirical management along the same treatment pathway. The way we classify their diabetes need not influence their management, and there is no evidence—nor indeed any logical argument—to support the common assumption that all forms of autoimmune diabetes need early insulin therapy.

Conclusions

Harry Himsworth summarised our present situation when he said that ‘what we know already is a great obstacle to finding out those things that we do not know’. The same point was made by the man who said that ‘it’s not what people don’t know that’s the problem. It’s the things they do know that ain’t so’. The central point of this and previous editorials [10, 16] is not whether entities such as the metabolic syndrome, LADA, or type 1 and type 2 diabetes really exist. The issue is whether the mental constructs we create around these concepts are a help or a hindrance to further understanding. I believe that such constructs should be considered as tools, helpful for some tasks but not for others, and that we should use them, consciously, as a guide to thought, rather than a substitute for it. Above all, we need to avoid the humiliation of being taken prisoner by constructs invented for our own convenience. The main drawback to this more flexible approach is that we might need to fall back on that most unconsidered of all research tools, the human brain.

Our current classification is a mental construct that assumes the existence of two forms of diabetes, and confirms this by circular reasoning. It lumps disease characteristics that relate to cause with others that relate to consequence, making logical analysis impossible. It treats dimensions as categories, and then leads us to believe that these correspond to objective entities. Since these so-called entities defy precise definition, and there is no agreed procedure for telling them apart, this belief is hard to justify. Our routine attempt to assign patients to one or other type of diabetes is therefore neither entirely rational, nor measurable, nor useful.

How might we do better? To begin with, there is no point in scrapping the existing classification: it works well enough for some purposes, and we do not yet know enough to devise a better one. But, as we look to the future, we should avoid categorical thinking, and should instead concentrate on disease mechanisms, regardless of the presumed aetiology of diabetes. We should be open to evidence that immune-mediated processes are relevant to the pathogenesis of type 2 diabetes, and that non-immune processes may also play a role in type 1 diabetes—which is not the same as assuming that the two conditions are necessarily the same. We should continue to seek for and define new genetic subtypes, an exercise that can sometimes make a great difference to the lives of the individuals concerned. We should also look for mechanisms and final common pathways that may be common to all forms of diabetes. In this way we can be sure that the vast area of *terra incognita* on our map of diabetes will slowly be eaten away—and will be less tempted to imitate the ancient map-

makers by decorating the blank spaces with mythical creatures.

Acknowledgements This editorial has given rise to many stimulating discussions, and I would particularly like to thank Peter Butler, Paul Dieppe, Andrew Hattersley, Roy Taylor and John Yudkin for their comments. Special thanks also to Jessica Hughes for her skilled and patient help with this and other editorials, and for knowing when to tell me to start again.

References

1. Gale EAM (2001) The discovery of type 1 diabetes. Diabetes 50:217–226
2. National Diabetes Data Group (1979) Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. Diabetes 28:1039–1057
3. WHO Expert Committee on Diabetes Mellitus (1980) Second Report. World Health Organization, Geneva
4. Himsworth HP (1936) Diabetes mellitus: its differentiation into insulin-sensitive and insulin-insensitive types. Lancet i:127–130
5. Lister J, Nash J, Ledingham U (1951) Constitution and insulin sensitivity in diabetes mellitus. BMJ i:376–379
6. Barber B (1961) Resistance by scientists to scientific discovery. Science 134:596–602
7. Joslin EP, Root HF, White P, Marble A, Bailey CC (1947) The treatment of diabetes mellitus. Eighth edition. Henry Kimpton, London, pp 144–146
8. Rutter M (2003) Categories, dimensions, and the mental health of children and adolescents. Ann N Y Acad Sci 1008:11–21
9. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus: Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 20:1183–1197
10. Gale EAM (2005) Latent autoimmune diabetes in adults: a guide for the perplexed. Diabetologia 48:2195–2199
11. Gale EAM (2005) Spring harvest? Reflections on the rise of type 1 diabetes. Diabetologia 48:2445–2450
12. Faber K (1923) Nosography in modern internal medicine. Oxford University Press, Oxford
13. Mayr E (1982) The growth of biological thought. The Belknap Press of Harvard University Press, Cambridge, Massachusetts
14. UK Prospective Diabetes Study (UKPDS) (1995) 13: Relative efficacy of randomly allocated diet, sulphonylurea, insulin or metformin in patients with newly diagnosed non-insulin dependent diabetes followed for three years. BMJ 310:83–88
15. Davis TME, Wright AD, Mehta ZM et al (2005) Islet autoantibodies in clinically diagnosed type 2 diabetes: prevalence and relationship with metabolic control (UKPDS 70). Diabetologia 48:695–702
16. Gale EAM (2005) The myth of the metabolic syndrome. Diabetologia 48:1679–1683