BMJ

BMJ 2011;343:d7995 doi: 10.1136/bmj.d7995 (Published 28 December 2011)

ANALYSIS

The idolatry of the surrogate

Easier to measure surrogate outcomes are often used instead of patient important outcomes such as death, quality of life, or functional capacity when assessing treatments. **John Yudkin**, **Kasia Lipska**, **and Victor Montori** argue that our obsession with surrogates is damaging patient care

John S Yudkin *emeritus professor of medicine*[1], Kasia J Lipska *Robert Wood Johnson clinical scholar*[2], Victor M Montori *professor*[3]

[1]University College London, UK; [2]Department of Medicine, Yale University School of Medicine, New Haven, CT, USA; [3]Knowledge and Evaluation Research Unit, Division of Endocrinology and Diabetes, Departments of Medicine and Health Sciences Research, Mayo Clinic, Rochester, MN, USA

Diabetes care is largely driven by surrogates. The US Institute of Medicine defines surrogates as "biomarker[s] intended to substitute for a clinical endpoint [and] expected to predict clinical benefit (or harm . . .) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence."[1] In diabetes, concentrations of glycated haemoglobin (HbA_{1c}) are used as a surrogate marker for outcomes that are important to patients, such as blindness or amputation. Other surrogates such as blood pressure, lipids, albumin excretion rates, and C reactive protein have been used to predict outcomes of cardiovascular disease and to guide clinical practice in people with or without diabetes. Much of the evidence for clinical interventions is based on their effect on surrogate outcomes rather than those that matter to patients such as quality of life or avoidance of vision loss or renal failure. Moreover, because these "hard" end points generally show much smaller responses to interventions than surrogate markers, many of the widely accepted strategies for diabetes may be based on artificially inflated expectations.

Recent studies have challenged the assumption that reliance on surrogates can accurately predict the effect of treatment on hard outcomes. There are the oral hypoglycaemic drugs that reduce HbA_{1c} but increase the risk of cardiovascular events,[2] antihypertensive drugs that do not reduce the risk of stroke,[3] and drugs that improve cholesterol profiles but do not reduce cardiovascular events.[4] Explanations for such phenomena include unwanted effects of the drug or an incomplete understanding of the pathophysiology of the disease.[5] But why have these examples been regarded as exceptions rather than radically challenging the value of surrogates in clinical practice or drug registration?

The obsession with surrogate markers within medical practice goes even further. Not only are markers given more importance than is justified by the evidence but they also begin to take on an existence of their own as new disease entities. And despite being far from perfect surrogates for outcomes, glucose, lipid, and blood pressure thresholds are used to evaluate quality of healthcare and to influence reimbursements.[6] So clinicians spend time exploring ways of reducing the level of the surrogate, even when the only options are interventions that do not improve, or may even worsen, a patient's outlook.[7] In this article, we use the example of type 2 diabetes to show how these surrogates are idols with feet of clay, and so challenging good medical practice. We suggest some possible strategies on how to counter this.

The problem with surrogates

Surrogate markers take several forms. They may be true risk factors involved in the causal pathway for the outcome. They may represent preclinical manifestations of organ damage. Or they may be bystanders without an active role but nevertheless correlate with the clinical outcome and mark the response to therapy. But regardless of their place in the spectrum, overinterpretation of surrogates can lead to misinterpretation of the evidence, as we discuss below.

Causal factors

Risk factors such as low density lipoprotein (LDL) cholesterol and blood pressure are thought to lie in the causal pathway for the disease process. Robust relations between their levels and cardiovascular outcomes across a range of interventions make them particularly attractive candidates as surrogates. Glycaemia is heavily touted as a comparably important surrogate, but its epidemiological relation with cardiovascular disease is much weaker than that of LDL cholesterol and blood pressure, and intensified glucose lowering has a substantially smaller effect on the absolute risk of vascular events.[5] Moreover, glycaemia's reputation as a valid surrogate end point has been tarnished by studies showing that intensified glucose lowering does not reduce cardiovascular disease[8] and by the finding that glucose lowering drugs such as rosiglitazone actually increase the risk

Correspondence to: J S Yudkin, London N7 0AG, UK j.yudkin@ucl.ac.uk

of cardiovascular events.[2] Even the validity of LDL cholesterol and blood pressure as surrogates has been challenged after studies of drugs targeting them have had negative results.[3 4] Furthermore, the Institute of Medicine report on biomarkers raises similar questions in other areas such as cardiology, oncology, and HIV medicine.[1]

Subclinical indicators

Surrogates can be subclinical markers of disease or of treatment response, such as carotid artery intima-media thickness or retinal morphology. A three step progression of retinal morphology using the Early Treatment Diabetic Retinopathy Study (ETDRS) classification has been used in numerous studies as a surrogate of response to treatment intended to prevent severe vision loss. This surrogate was developed over two decades ago using data from the ETDRS, in which one eye of each patient was assigned to early photocoagulation. Observation of the natural course of retinopathy over 10 years in the initially untreated eye allowed for definition of abnormalities predicting progression to proliferative retinopathy.[9] However, recent findings show that increasing severity of retinopathy below step 9 of the 15 point ETDRS scale had little effect on visual handicap or vision related quality of life. Above this threshold, functional visual decline occurs rapidly, especially with bilateral eye disease.[10] Thus the baseline level of retinopathy seems crucial to interpretation of such progression. This may help explain why the ACCORD Eye Study found that three step progression on the ETDRS scale substantially overestimated the incidence of moderate vision loss in response to both intensified glucose lowering (33% *v* 12%) and fenofibrate (40% *v* 5%).[11] A meta-analysis of the four major studies of intensified glucose lowering found that a 1% reduction in HbA_{1c} levels was associated with a reduction in blindness or severe vision loss of 6% (95% confidence interval −10% to 20%) over 4.1 years,[5] substantially less than expected from studies like the Diabetes Control and Complications Trial,[12] which used progression on the ETDRS scale as an end point.

Correlated factors

Raised albumin excretion rates and C reactive protein concentrations are associated with current and future cardiovascular disease and have been used for risk stratification, even though there is no evidence of direct involvement in the pathology.[13 14] Microalbuminuria has also been used as a putative marker of the renoprotective response of interventions for diabetes or hypertension in studies of both type 1 and type 2 diabetes, with the assumption that the reduction in risk of microalbuminuria is roughly equivalent to the reduction in risk of end stage renal failure.[15] However, the ACCORD microvascular study found that although intensified glycaemic control reduced the incidence of microalbuminuria by 21%, end stage renal disease was reduced by only 5%.[16] In a meta-analysis of the four major studies of intensified glucose lowering, 1% lower levels of HbA_{1c} were associated with a 12% reduction (−11% to 30%) in the incidence of renal replacement therapy, renal failure, or renal death over 4.4 years,[5] and a Cochrane review found a reduction in end stage renal failure of 13% (−6% to 29%) over 10 years.[17] Thus, despite studies of over 27 000 patients treated for more than four years, it is unclear whether intensified glucose control prevents clinical renal disease. As with retinal morphology, the effect on a surrogate has given false hope for patients.

Despite their shortcomings, these markers of risk are being invested with new clinical importance. Microalbuminuria became promoted as warranting treatment in its own right[18] and, more recently, has become a target for prevention. The ROADMAP Study[19] randomised normoalbuminuric patients with type 2 diabetes to olmesartan or placebo to prevent microalbuminuria. The study showed a reduction in the incidence of microalbuminuria from 9.8% to 8.2% with olmesartan. The follow-up measurement, however, was taken while the patients were still taking the drug, and it is thus unclear whether the reduction resulted from an effect on the pathological processes responsible for microalbuminuria, or on those linking it with renal failure or cardiovascular disease. The data suggest that treating 1000 people with olmesartan for 3.2 years would result in 16 fewer people developing microalbuminuria, so by implication obviating their twofold increase in cardiovascular risk.[20] However, the number of deaths prevented by this is likely to be less than the 5.4 excess deaths per 1000 observed in the olmesartan group.[19 21]

A further problem arises when hard end points are combined with surrogates in a composite end point to improve a study's statistical power. In such instances, surrogates generally accumulate the largest number of events and show the largest intervention effect, while the more important patient relevant outcomes accumulate few events with much smaller effects of treatment.[22] Thus, although the surrogate enriched composite end point permits a smaller and faster trial, it misleads by reflecting the effect of therapy mostly on the surrogate rather than on the important outcome.

Surrogate markers are not intrinsically flawed. When interpreted appropriately, they can be helpful in risk stratification and in treatment. However, rather than a "one size fits all" treatment target, global measures of risk, based on a range of clinical features and risk factors,[23] are better suited to identifying high risk patients in whom intervention is most likely to yield benefit.

False idols

Why have doctors become so invested in surrogate markers? The main reason is that the evidence base is built from trials that focus on the effect on surrogates. Since they respond sooner than outcomes that are important to patients, surrogates are better suited as end points in clinical trials that need to be completed quickly and at low cost. Evidence that builds in this way shapes practice and policy. Consequently, clinicians see this evidence converted into guidelines, quality of care measures, and pay for performance targets. We could speculate that the short term goals of the drug industry contribute to the predominance of surrogates in clinical practice. But this is an oversimplistic analysis. A historical view points more broadly to an alliance of public health advocates, scientists and clinicians, professional societies, and test and treatment companies who see their interests coincide.[24]

Idolisation of the surrogate end point has turned doctors away from the focal point of patient centred therapy based on hard end points. Patients with diabetes may be asymptomatic but are treated to achieve levels of surrogates set as treatment targets by committees. When targets are not reached, patients are started on drugs that are licensed because they have been shown to affect surrogate end points rather than more relevant outcomes, with their promotion heavily dependent on these effects. It is only later that the excess risks of vascular disease or cancer become apparent, by which time the drugs may be off-patent with new wonder drugs promoted in their place. Meanwhile the patient continues treatment with new drugs to achieve target surrogate end points or perhaps to prevent the onset of new risk factors like microalbuminuria. These decisions are usually made in patients' best interest but often without their involvement.

The surrogate end point carries no information to which patients can relate, so removing the discussion further from the patient.

New approach

The growing trend has been for the focus on surrogates to dominate both research and clinical agendas on non-communicable diseases, with the connivance of public health, professional societies, and drug companies. We argue that the disconnect between surrogate and hard outcomes in terms of degree of benefit or harm, or even its direction, makes it important to review this. Changes are needed in both current criteria for registration of new drugs targeted at reducing risk of complications[25] and current formulas for measuring quality and reimbursing doctors.[26]

Such a refocus is beginning to emerge. In the wake of the rosiglitazone saga, the US Food and Drug Administration has instigated a requirement that new hypoglycaemic drugs must be shown to have no harmful effect on cardiovascular event rates (although they still don't have to show benefit).[25] These proposals mean that studies of hard end points have to be done during, rather than after, drug development, adding to both costs and duration, and drug companies may have to be compensated by extending patent life.[25] The risk of stifling innovation is often cited as a reason for expediency, yet the cost of false positive innovations for patients and society may exceed their value, especially when effective treatments are already available. Finally, studies of hard end points are necessary to practise truly patient centred medicine. In order to fully engage our patients in treatment decisions, we must understand how therapies affect outcomes that are important to them. Surrogate end points will not provide us with these answers.

Contributors and sources: JSY has researched cardiovascular risk in diabetes and did a meta-analysis of intensive glucose lowering in 2010. The discrepancy of the results with those predicted from previous literature relying on surrogate markers led him to explore this with KJL and VMM, two diabetes outcomes researchers. All authors contributed to the concepts and structure of this manuscript. JSY is the guarantor.

Competing interests: All authors have completed the ICJME unified disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; and no other relationships or activities that could appear to have influenced the submitted work.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Institute of Medicine. Evaluation of biomarkers and surrogate endpoints in chronic disease. 2010. www.iom.edu/Reports/2010/Evaluation-of-Biomarkers-and-Surrogate-Endpoints-in-Chronic-Disease.aspx.
2 Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes [correction in: *N Engl J Med* 2007;357:100]. *N Engl J Med* 2007;356:2457-71.
3 Law MR, Morris JK, Wald NJ. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. *BMJ* 2009;338:b1665.
4 Barter PJ, Caulfield M, Eriksson M, Grundy SM, Kastelein JJ, Komajda M, et al. Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med* 2007;357:2109-22.
5 Yudkin JS, Richter B, Gale EAM. Intensified glucose lowering in type 2 diabetes: time for a reappraisal. *Diabetologia* 2010;53:2079-85.
6 Calvert M, Shankar A, McManus RJ, Lester H, Freemantle N. Effect of the quality and outcomes framework on diabetes care in the United Kingdom: retrospective cohort study. *BMJ* 2009;338:b1870.
7 Shah ND, Montori VM, Krumholz HM, Tu K, Alexander GC, Jackevicius CA. Responding to an FDA warning—geographical variation in the use of rosiglitazone. *N Engl J Med* 2010;363:2081-4.
8 Action to Control Cardiovascular Risk in Diabetes (ACCORD) Study Group. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med* 2008;358:2545-59.
9 Early Retinopathy Research Group. Fundus photographic risk factors for progression of diabetic retinopathy. ETDRS report number 12. *Ophthalmology* 1991;98:823-33.
10 Mazhar K, Varma R, Choudhury F, McKean-Cowdin R, Shtir CJ, Azen SP, et al. Severity of diabetic retinopathy and health-related quality of life: the Los Angeles Latino Eye Study. *Ophthalmology* 2011;118:649-55.
11 ACCORD Study Group and ACCORD Eye Study Group. Effect of medical therapies on retinopathy progression in type 2 diabetes [correction in: *N Engl J Med* 2011;364:190]. *N Engl J Med* 2010;363:233-44.
12 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;329:977-86.
13 Yudkin JS, Chaturvedi N. Developing risk stratification charts for diabetic and nondiabetic subjects [correction in: *Diabet Med* 1999;16:972-3]. *Diabet Med* 1999;16:219-27.
14 Emerging Risk Factors Collaboration. C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality. *Lancet* 2010;375:132-40.
15 UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998;352:837-53.
16 Ismail-Beigi F, Craven T, Banerji MA for the ACCORD Trial Group. Effect of intensive treatment of hyperglycaemia on microvascular outcomes in type 2 diabetes: an analysis of the ACCORD randomised trial [correction in: *Lancet* 2010;376:1466]. *Lancet* 2010;376:419-30.
17 Hemmingsen B, Lund SS, Gluud C, Vaag A, Almdal T, Hemmingsen C, et al. Targeting intensive glycaemic control versus targeting conventional glycaemic control for type 2 diabetes mellitus. *Cochrane Database Syst Rev* 2011;2:CD008143.
18 Heart Outcomes Prevention Evaluation (HOPE) Study Investigators. Effects of ramipril on cardiovascular and microvascular outcomes in people with diabetes mellitus: results of the HOPE study and MICRO-HOPE substudy. *Lancet* 2000;355:253-9.
19 Haller H, Ito S, Izzo JL, Katayama S, Menne J, Mimran A, et al. Olmesartan for the delay or prevention of microalbuminuria in type 2 diabetes. *N Engl J Med* 2011;364:907-17.
20 Gerstein HC, Mann JF, Yi Q, Zinman B, Dinneen SF, Hoogwerf B, et al. for the HOPE Study Investigators. Albuminuria and risk of cardiovascular events, death, and heart failure in diabetic and nondiabetic individuals. *JAMA* 2001;286:421-6.
21 Yudkin JS, Lehman R. Olmesartan, microalbuminuria, and type 2 diabetes. *N Engl J Med* 2011;364:2260.
22 Ferreira-González I, Busse JW, Heels-Ansdell D, Bryant DM, Montori VM, Alonso-Coello P, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ* 2007;334:786.
23 Timbie JW, Hayward RA, Vijan S. Variation in the net benefit of aggressive cardiovascular risk factor control across the US population of patients with diabetes mellitus. *Arch Intern Med* 2010;170:1037-44.
24 Yudkin JS, Richter B, Gale EAM. Intensified glucose control in type 2 diabetes—whose agenda? *Lancet* 2011;377:1220-1.
25 Lehman R, Yudkin JS and Krumholz HM. Licensing drugs for diabetes. *BMJ* 2010;341:c4805.
26 O'Connor PJ, Bodkin NL, Fradkin J, Glasgow RE, Greenfield S, Gregg E, et al. Diabetes performance measures: current status and future directions. *Diabetes Care* 2011;34:1651-9.

Cite this as: *BMJ* 2011;343:d7995