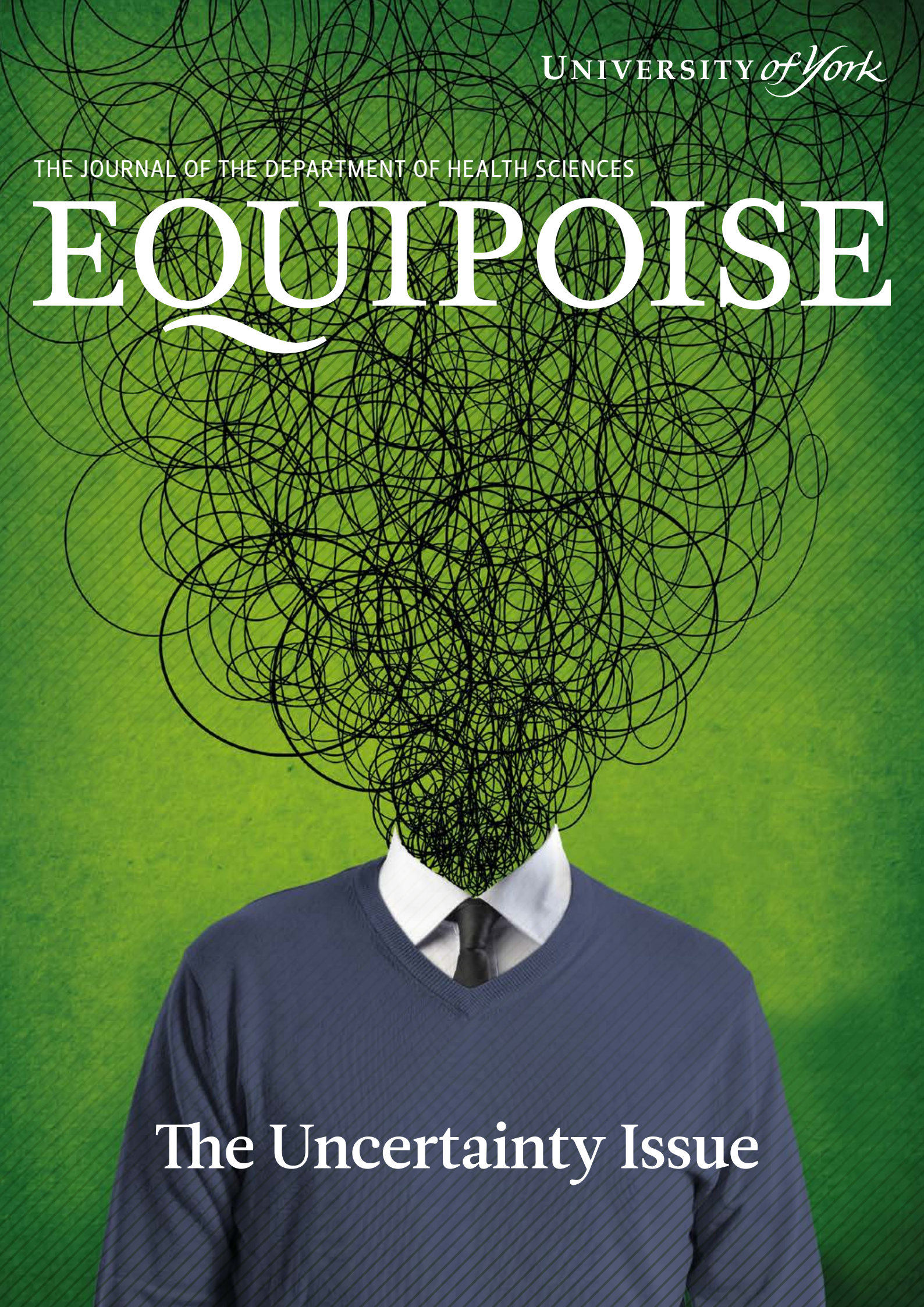
University of York

THE JOURNAL OF THE DEPARTMENT OF HEALTH SCIENCES

EQUIPOISE



The Uncertainty Issue

Funded Studentship for the **MSc Applied Health Research**

Applications are invited for a full-time one year studentship for the **MSc Applied Health Research** hosted within the Department of Health Sciences at the University of York, commencing **October 2014**.

This MSc provides ideal training for anyone wanting to become a professional researcher in health, or for anyone wishing to develop a career in the health sector, where the ability to take a critical, evidence-based approach to health problems and solutions is increasingly valued. The course offers students an excellent grounding in applied health research methods, including systematic reviews, RCTs, epidemiology, applied statistics, qualitative methods and health economics. Teaching is led by senior staff in the Department of Health Sciences who are acknowledged experts in their fields, and the large majority of teaching sessions take place in small groups.

The MSc can be undertaken over one year (full-time) or two years (part-time). You will focus on the production, critical appraisal and use of scientifically rigorous research evidence, applied to a range of health-related areas.

This programme is particularly relevant to science and social science graduates and healthcare professionals who wish to develop their health-related research and evaluation skills, and to people from any health-related background interested in a career in health services research. A number of our students on this pathway have upgraded to MPhil/PhD study.

UNIVERSITY of York

www.york.ac.uk/healthsciences/gradschool/masters/health-serv-res/

CONTENTS

EQUIPOISE

Written and produced by
Department of Health Sciences, University of York, York YO10 5DD

Design by
Ball Design Consultancy

EDITORIAL

Sisyphus was the subject of a Greek myth, a man condemned by the gods to roll a huge boulder up a steep hill, only to find that each time he thought he was almost at the top, the boulder rolled back down again. At face value this allegorical story focuses on Sisyphus' fate as a punishment, but Albert Camus[1] sees things slightly differently: the chance that the stone *might* get to the top gives Sisyphus some hope, at least, a reason to continue in his task and even some pleasure in the task itself. For Sisyphus, uncertainty was something to be celebrated (through his efforts) not a reason for despair, despondency or inertia.

Uncertainty – and the challenges involved in reducing it – lie at the heart of research and scholarly endeavour. Often uncertainty is portrayed negatively (the fear of cancer) and whole belief systems have been developed to combat the dread fear of the uncertain (astrology is a prime example). Our beliefs and how we use them to respond to uncertainty, can have profound effects on our behaviours and the world around us. President Ronald Reagan famously drew on the advice of an astrologist before making decisions (exactly which decisions is disputed), but human beings are prone to handling uncertainty in ways that defy logic, rationality or reliable ways of thinking. Jon Ronson[2] devotes a chapter of his book "Lost at Sea" to stories of ordinary people who interpret chance events in which they had some kind of input as if they could actively influence the outcomes through various complex and (not so complex) schemas and belief systems. Psychologists have long recognised that spurious belief systems and awry responses in the face of uncertainty are a powerful driver for our behaviour and decisions[3].

As Camus recognised, uncertainty is not always a bad thing. As a thought experiment, imagine a world with NO uncertainty. No uncertainty would eliminate many aspects of our lives from which we derive pleasure, hope and a sense of achievement. Imagine knowing the punchline of every joke before it was delivered to you; knowing the date and year on which you would perish; or – as psychologists Reid Hastie and Robyn Dawes highlight – being told at age 18 that you possess the gene that will definitely lead to early onset Alzheimer's disease[3]. Uncertainty is part of our existence and even without tools to handle it more efficiently or effectively, human beings are remarkably adept at adapting to it. Hastie and Dawes use the example of a study in which people were "scored" according to how they thought a diagnosis of HIV would affect them. Most, unsurprisingly, anticipated negative (but occasionally positive) impacts. The reality was that, five weeks after diagnosis, they were actually far more "neutral" than they had anticipated – they adapted.[3]

Alongside our evolutionary adaptive capabilities, we do have the "tools" available to help us manage uncertainty in our thinking, reasoning, and behaviour. This issue of Equipoise examines some of the tools that health researchers have at our disposal: Professor Martin Bland outlines the role of statistics (a dread fear to rival uncertainty for some people) in putting chance in its place; Ryan Pulleyblank, an economist, raises the possibility that our confidence in things that seem intuitively appealing (who would argue with quality improvement in the NHS?) might be far less certain when we think differently about costs and consequences; Dr Peter Knapp illustrates how the successful communication of risk and uncertainty is dependent in part on our understanding of people's flaws and failings in understanding; Professor David Torgerson shows how chance itself can be harnessed to deliver more reliable research results; Professor Tim Croudace highlights a novel way of categorising things in a bid to reduce uncertainty.

We hope that you enjoy this third issue of Equipoise (which itself means uncertainty) and that, if it has made you even more certain that you want to undertake research, training or further/higher/continuing education, you will get in touch.

Professor Carl Thompson
Professor Kate Pickett

“The struggle itself towards the heights is enough to fill a man’s heart.”[1]

References

1. Camus, A., The Myth of Sisyphus. 1955, New York: Alfred A. Knopf, Inc.
2. Ronson, J., Lost at Sea: the Jon Ronson Mysteries. 2012, London: Penguin.
3. Hastie, R. and R.M. Dawes, Rational choice in an uncertain world. 2001, London: Sage.

UNCERTAINTY IS IN THE BLOOD

Blood is wonderful stuff. It flows around the body carrying fuel and building materials to cells and taking waste away. Its efficient working is essential to health and life. A key indicator of the blood system is blood pressure. It is quite easy to measure, at least in principle. We use an inflated cuff to compress the arm where there is an artery close to the surface, stopping the blood flow. Then we relax the pressure, listening to the artery beyond the pressure point, until we hear the thump which means that some blood has been forced through the constriction. This is the systolic pressure, the blood pressure at the moment when the heart pumps out, forcing blood around the system. We continue to relax the pressure listening to the sounds, until the thumping stops, the moment that blood starts to flow continuously. This is the diastolic pressure, the minimum pressure, when the heart draws in blood before pumping it out again. The pressure is measured by a manometer connected to the cuff. Not so long ago, this would be the height of a column of mercury which the pressure could support against gravity. We now use other types of manometer to avoid the dangers of mercury poisoning, but we still record the pressure in millimetres of mercury, or mm Hg. The whole device is called a sphygmomanometer and the listening may be by a human using a stethoscope or by a microphone connected to an automated system.

Before writing this, I measured my blood pressure as systolic 129 mm Hg and diastolic 81 mm Hg, which we usually write as 129/81. During this process, my blood pressure varied continuously between these two values. With each beat of my heart it rose to the maximum systolic pressure and fell to the minimum diastolic. These were not necessarily the same for each heartbeat. After making measurement, I waited a few minutes and then did it again, getting 119/80. A few more minutes and I got 125/80.

Blood pressure varies continuously, so what do I mean when I say my blood pressure is 125/80? I could mean that this was the last measurement I made, or that this was the average of several measurements. My actual blood pressure right now is uncertain. Not only does it change from heartbeat to heartbeat, but it changes systematically over time. It is highest during the day when I am active, lowest at night when I am asleep. It also changes over the year, tending to be higher in the winter than the summer. If a stethoscope is used, it may depend on the listener, on how they interpret what they hear and how fast they react to it. Blood pressure may be increased if you are being very active or excited. The observer and the person being measured may interact to change the pressure. A friend was a statistician for a large study of risk factors for heart disease among middle-aged men. At the training day for the research nurses, he had no problem in correctly predicting which nurse would measure the highest average blood pressure in the study. Because of all this uncertainty in measurement of blood pressure, the British Hypertension Society set up a working party to recommend a protocol for evaluating new automated sphygmomanometers. Some of the members were blood pressure specialist clinicians, but two of us were statisticians.[1]

Because of all this uncertainty in measurement of blood pressure, the British Hypertension Society set up a working party to recommend a protocol for evaluating new automated sphygmomanometers.

If your GP measures your blood pressure as being too high, they are unlikely to put you on treatment immediately. They will ask you to come back in three weeks or so for another measurement. Because of the uncertain nature of a single measurement they will wait to see if it is high for a second time. For the same reason, they will also make more than one measurement each time or use a 24-hour monitor. If they do decide to treat, they will also be uncertain as to what drug would be best. They would probably start an ACE-inhibitor. But you may not stay on it for long. Some people react to ACE-inhibitors with a persistent cough and switch to another drug. I was one of them. For some people one drug is not enough, they need two or three, acting on different mechanisms to control the blood pressure.

More problems of uncertainty arise when we want to test a new drug for treating high blood pressure. We cannot simply find someone newly diagnosed, give them the drug, and see whether their blood pressure falls. The variability of blood pressure itself and the variability of reactions to drugs mean that we need a lot of people, not one. Nor is it enough to treat a group of the newly diagnosed; uncertainty has a very sneaky trap for the unwary. Those people have just been chosen because their blood pressure is particularly high. It is likely to be higher than their long-term average and so likely to fall, with or without treatment, a process termed "regression towards the mean". We need another group of people with high blood pressure,

a control group, who may be untreated for a while or may get an established, standard treatment. Statisticians have shown that the best way to select them is to recruit our trial participants and let chance decide; we allocate them randomly, as by a flip of a coin.

Once we have run our trial and collected our final blood pressure measurements, uncertainty has a further part to play. We may find that those on our new treatment do better, but is the difference big enough for us to conclude that the new treatment really is better? After all, if the treatments were exactly the same, half the possible ways we could have put our participants into two groups would have resulted in the new treatment appearing better, half worse. Statisticians solved this one with something we call a "significance test", which says that if the difference we see in our participants would be seen in only a very small proportion of those possible allocations, then we have strong evidence that the treatments really do differ. The difference which we see is not necessarily the actual difference between the treatments either: some samples will give us a difference smaller than we would get if we could treat all people with high blood pressure, some will give us a difference which is larger. So we produce a range of possible values which we think are compatible with what we found in the trial, called a "confidence interval".

Even after all that, uncertainty still plays a few more games with us. Not all statisticians agree that these methods are the best, or even that they are correct, and propose entirely different solutions to the problem. (Fortunately, they give us much the same answers.) In the past, statisticians fought like pitbulls over whose way was right. At least we don't do that anymore. A good thing – and of that I am certain.

Martin Bland is a Professor of Medical Statistics in the Department of Health Sciences.

References

1. O'Brien E, Petrie J, Littler W, de Sweit M, Padfield P, O'Malley K, Jamieson M, Altman D, Bland M, Atkins N. The British Hypertension Society protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special references to ambulatory systems. *Journal of Hypertension* 1990; 8, 607–619.

THE UNCERTAIN VALUE OF QUALITY IMPROVEMENT IN THE NHS

When organisations such as the National Institute for Health and Care Excellence (NICE) make recommendations on what the NHS should and should not be doing, it takes time for the message to spread and become widely adopted. Quality Improvement (QI) interventions are often developed and deployed to try and increase the health of the public by increasing adoption rates for the recommended treatment practices. The NHS spends a significant amount of money on improving the quality of treatment delivered to patients. The exact amount is unknown but much of this activity is delivered by specialised teams who develop and deliver interventions to encourage clinicians to adopt guideline recommendations.

Who could argue that improving the quality of patients' treatment was anything but a good thing? Well, it's not that straightforward: not all of the techniques employed by what might be termed the NHS "Quality Industry" are effective all of the time for all of the service. Because of this, it is justifiable to ask, "is NHS QI activity cost effective when we consider how much it costs?" With finite resources available to fund healthcare, it is important that policy makers consider the efficiency of all decisions that commit resources; this includes considering the likely costs, as well as the impacts of quality improvement. This is easier said than achieved.

The best guide to the impact of QI needing funds from the cash-limited public purse is to evaluate evidence from QI that has *already* happened. However, there are some basic measurement problems that must be overcome before we can decide whether QI is "worthwhile". For the patient in a service that is being "quality improved" the choice of whether to follow a treatment recommendation is not guided solely by the health outcomes that may result. Rather, the true impact is the difference in health outcomes that result and what *would have been the case* if the alternative treatment had been followed. New and exciting ways of doing things in the NHS may have greater health benefits, but the magnitude of the impacts that arise because of following a recommended treatment, rather than the alternative, is often unknown. The problem for the policy maker wishing to spend QI funds wisely is that where the expected health benefit associated with an average patient is uncertain, then the benefits for whole groups of patients in receipt of a particular health service will also be uncertain. Quantifying the impact of QI on a service requires summing the health impacts on patients who followed the recommended treatment as a *result* of the QI intervention. This picture is often missing.

Who could argue that improving the quality of patients' treatment was anything but a good thing? Well, it's not that straightforward: not all of the techniques employed by what might be termed the NHS "Quality Industry" are effective all of the time for all of the service.

Directly measuring health outcomes (other than did someone die or survive) associated with QI may not be feasible. Often, it is only possible to measure activity in a service (e.g. which drugs have been prescribed or treatments delivered). This may not be a problem: where such activity is well defined and clearly measurable and where activity is demonstrably linked to better health. In these cases, looking at activity (process) rather than outcome is fair and justified. Some examples include aspirin prescription for ischaemic heart disease, or regular examination of the feet in diabetes care to prevent complications of diabetes. Often though, treatment recommendations are not always well defined (for example in mental healthcare) and clear measures of such activity may not be available. Consequently, the impacts on health outcomes which can be implied through the available measures remains uncertain. Without accurate measures of activity, the extent that current practice deviates from guideline recommendations is also unknown. Without this picture the scope for improvement via QI will always be down to guesswork.

We can see then that the impact of QI is reflected by looking at treatment patterns following QI and the pattern of treatments had QI not been undertaken. A further problem here is that it will always be impossible to measure *both* treatment delivered with and without delivering a QI intervention. One solution is to randomly allocate services to QI or no QI arms of a clinical trial (expensive and not always feasible). Another, partial – but arguably more feasible – work around is to measure activity repeatedly over time. Where such a series of measurements includes the time before and after a QI intervention then the analysis can go some way to illuminating the impact of QI. Because we have the "before" data, we can look at the difference between measured trends after the intervention, and projections of trends that may have continued had no QI intervention been delivered. Often though such time series analyses are too short and so the persistence of any impact remains unknown. Nevertheless, the greater the magnitude and persistence of a QI impact, the more likely it is that QI will be good value.

Whether QI is "worthwhile" depends on the relationship between its impact on health outcomes and treatment costs, *as well as* the cost of delivering the QI. With these three ingredients we can work out the overall *additional* cost to society of obtaining *additional* quantity of health benefits. So, what is the cost of delivering a QI intervention? To be sure, it is much more than the cost of renting a conference room in a local hotel for a workshop and serving sandwiches to local clinicians and managers – inputs for which there are uniquely identifiable receipts. Everything which contributes to the production of QI comes at a cost, and should be considered. The largest cost of delivering QI often comes in the labour contribution of the participants. Proper accounting for a QI project requires including all of the labour hours involved in developing and delivering the QI. This will include the time that clinicians allocate to engaging with a QI project and often there is no accurate tracking of this. When people have multiple roles, it can be challenging to reliably identify the amount of time allocated to particular responsibilities (i.e. working on a particular QI project). Furthermore, accurate unit pricing for contributors' time may be unavailable, and there may be substantial uncertainty around the accuracy of any proxies which are available.

No patient ever complained of receiving too high a quality of treatment. However, this should not be taken as evidence that every effort to improve the quality of treatment delivered reflects sufficient benefit to justify the costs. Of course, some QI represents good value and other QI poor value. The point at the current time is we don't know which! Generating an honest picture for the public and policy makers about the costs and consequences of QI in the NHS will require NHS and academia to work together to estimate the true costs and benefits of delivering QI.

Ryan Pulleyblank is a Research Fellow and Health Economist on the TRiPLaB project in the Department of Health Sciences and the NIHR CLAHRC for Leeds York and Bradford. See www.trip-lab.com for more details. He also teaches NHS professionals and managers about "thinking like a health economist" when faced with quality improvement challenges (see www.york.ac.uk/healthsciences/gradschool/ri-kt) for more details.

COMMUNICATING UNCERTAINTY

In October 2012 six Italian seismologists and a government official were convicted of manslaughter and jailed for six years. Their crime: failing to predict the devastating 2009 earthquake in L'Aquila that led to 300 deaths. Whilst the scientists *had* identified a pattern of small tremors they stated that they "could not be confident that an earthquake would occur". The public largely interpreted this as "the scientists were confident that an earthquake would *not* occur". The residents of L'Aquila stayed overnight in their homes, many of which were destroyed, rather than seeking refuge elsewhere as they had on occasions when the risk was thought to be high.[1] At the heart of this confusion was the difficult task of communicating risk.

Understanding and communicating uncertainty (often called *risk*) is a core part of being a healthcare professional.

Understanding and communicating uncertainty (often called *risk*) is a core part of being a healthcare professional: diagnosing disease, considering prognosis, judging potential benefit of the available options, are all common problems that are mired in uncertainty but which necessitate communication with patients, families and colleagues. Such risk communication has been called "the main work of doctors".[2]

Researchers must also communicate uncertainty revealed in their data analysis. They may start out thinking this is easy, "Construct a table, work out an average or two, talk about p-values and confidence intervals and chuck in a couple of graphs." Many researchers work out pretty quickly that this strategy is unlikely to lead to widespread understanding and effective communication of uncertain risks, harms and benefits – risk communication is complex.

The complexity starts with what it is that's being communicated. Are you trying to convey the most probable outcome – perhaps the thing most likely to happen given a particular patient's situation? Or is it the range or scope of possible outcomes, perhaps with estimates of likelihood attached to each of them? Alternatively are you trying to convey the fact of uncertainty surrounding a finding or problem (something that is often overlooked by the mass media). For example, that we simply don't know whether this *particular* analgesic will work with this *particular* patient.[3]

I have been working with colleagues to research uncertainty for the past decade e.g.[4]; and I am not alone: entire journals and research groups are dedicated to researching risk and how we communicate it. Our collective efforts are paying dividends – we are now much more informed about uncertainty and how we think about, calculate and communicate it.

There's an important distinction between communicating the level of risk (let's think of this as a *technical* problem: such as whether to use a graph, a percentage or a proportion to aid communication) and the presence of uncertainty itself (let's think of this as a more *conceptual* issue: what might be the origins of any uncertainty and what might be the consequences?). Research has mostly concentrated on the former[5] examining questions such as:

- how do people understand words such as *common* or *rare*? (answer: many people use these terms, they are often ill-defined, and so there is little consensus about what they mean accordingly, they are best avoided for conveying risk);
- how are percentages and proportions interpreted? (answer: they work well for most people but some struggle, particularly when values are smaller than 1%);
- is it better to exploit the potential of *natural frequencies* such as "for every 100 people who take this medicine, 12 will get an upset stomach?" (answer: frequencies work well but become awkward when there are lots of data to communicate);
- how should I use graphs? (answer: it's complex! Graphs seem to vary in their effectiveness according to context and the crucial point is to test that they are understood as intended).[6]

A lack of understanding of these conceptual aspects of uncertainty artificially limits our understanding of risk and has wider implications for a goal of a more informed society.

Not surprisingly, it can be hard for people to make sense of so many mixed messages when it comes to communicating risk. Perhaps the answer lies in trying to work with the information people actually use and understand rather than trying to get them to think differently. Information is often fuzzy and impressionistic rather than detailed; people may understand the "gist" of what is being communicated rather than the detail. Some people have a strong preference for this kind of gist information and, in some circumstances, it may be more useful for patient and clinician decision making.[7] The problem is that we don't really know when patients prefer this kind of gist approach to communication and how we might use it in ways that can be trusted to inform patient choices.

What about the (more neglected) conceptual aspects of uncertainty? A lack of conceptual understanding might help explain why some patients struggle with quick information-communication solutions such as the Number Needed to Treat (NNT). An example of benefit communicated this way is, "20 patients would have to take this medicine for five years for one additional patient to be prevented from having a heart attack". In order to properly understand a NNT, you have to know that it describes an incremental benefit (or harm). Also that the data comes from a clinical trial (or a set of trials), in which one group of patients received a treatment and another group received something else. Consequently the reported benefits come from a group of patients and it's not possible to know precisely who will benefit (or who will not). Similarly, some patients *taking* the treatment will still have a heart attack and some of those *not taking* it will avoid one. Applying effects seen in a group to an individual patient is an uncertain affair.[8]

A lack of understanding of these conceptual aspects of uncertainty artificially limits our understanding of risk and has wider implications for a goal of a more informed society. A public who understand *how* we reduce uncertainty (let's call it "science") might go some way to more effective shared decision making between professionals and patients – assuming of course that we want shared decision-making to be "the norm".[9]

It is a challenging prospect to raise the veil on the fact that much of what we know in healthcare is uncertain. It is important though that the communicators of this message avoid the fate of the Italian seismologists: sometimes in healthcare it is better (and always more honest) to say clearly when we don't know than to use ways of communicating that leave open the potential for misinterpretation.

Peter Knapp is a Senior Lecturer in Evidence-Based Decision Making in the Department of Health Sciences and the Hull York Medical School.

References

1. http://www.bbc.co.uk/news/world-europe-20025626.
2. Smith R. Communicating risk: the main work of doctors. *BMJ* 2003; 327: 7417.
3. Moore A, Derry S, Eccleston C, Kalso E. Expect analgesic failure; pursue analgesic success. *BMJ* 2013; 346: f2690 doi: 10.1136/bmj.f2690.
4. Knapp P, Gardner P H, Raynor D K, Woolf E, McMillan B. Perceived risk of tamoxifen side effects: A study of the use of absolute frequencies or frequency bands, with or without verbal descriptors. *Patient Education & Counseling* 2010; 79: 267–271.
5. Gigerenzer G. *Reckoning with risk*. Penguin Books, 2003.
6. Spiegelhalter D, Pearson M, Short I. Visualising uncertainty about the future. *Science* 2011; 333: 10.1126/science.1191181
7. Fagerlin A, Zikmund-Fisher B J, Ubel P A. Helping Patients Decide: Ten Steps to Better Risk Communication. *Journal of the National Cancer Institute* 2011; 103: 1436–1443. doi: 10.1093/jnci/djr318.
8. Han P K J. Conceptual, methodological, and ethical problems in communicating uncertainty in clinical evidence. *Medical Care Research and Review* 2013; doi: 10.1177/1077558712459361.
9. Department of Health. *Liberating the NHS* (White Paper). The Stationery Office 2010.

FROM MICRO-FINANCE TO BADGERS AND ASBOS

THE POWER OF RANDOMISATION IN SOCIAL POLICY

A randomised controlled trial (RCT) is the best method for testing the effectiveness of new treatments or policies. Since the first trials in the 1940s, many thousands have now been carried out. RCTs have revolutionised medicine by providing reliable answers to the questions that clinicians and policy makers ask, in a desire to provide higher quality, more effective and safer care.

Outside of medical care relatively few RCTs are undertaken – even at the level of health policy and decisions that might impact on millions of people. For example, we could have randomised GP practices to fund holding (or non fund holding) or hospitals to Foundation Trust (or centrally managed) status and evaluated the outcomes. In many ways, RCTs in social policy interventions are more important than in healthcare. For example, the consequences of a policy such as Anti Social Behavioural Orders (ASBOS)

impact on us all and that impact happens whether we like it or not. If ASBOs do not work or make criminal behaviour more likely then we have no choice. Of course, in healthcare we often have the option of refusing health treatments.

RCTs in the social realm are not new: the first RCTs were used to evaluate university educational interventions in the 1930s, when 'delinquent' undergraduates were randomised to receive 'mentorship' from more senior students. RCTs' expansion within the social sciences has been slower than it should have been, in part because of the [futile] paradigm wars associated with the method of testing ideas. Some researchers and academics (often from a position of relative ignorance) have viewed with suspicion any quantitative approach to evaluation. A special vehemence has been reserved for the randomised trial. Other methods of evaluation though have failed to answer the 'what works?' question.

The last decade has witnessed a resurgence in the use of RCTs in social policy. The World Bank, among others, encourages the use of RCTs wherever possible to evaluate interventions in developing countries. Indeed, large rigorously conducted trials have been undertaken in challenging circumstances but have generated useful answers: teaching children in ability groups improves outcomes for all ability groups in Kenya; micro-credit finance is not as good as many claim; international electoral monitoring increases female voter turnout; increasing female local political representation in India reduces negative attitudes to women.

If such trials can be done in developing countries there is no reason they cannot be done in the UK. For instance, the Mexican government allowed its flagship poverty reduction programme, PROGRESA, to use randomisation to assess its effectiveness. Because the trial showed significant benefits the new Mexican government, in opposition when PROGRESA was implemented, felt it could not dispute the trial's results and therefore retained the programme. Variants of this programme have been implemented across Latin America. In contrast, the previous Labour administration in this country refused all calls to use random allocation to evaluate one of its flagship poverty reduction programmes: Sure Start. The main objection against random allocation was that depriving the poorest areas of the 'benefits' of Sure Start was unethical. However, when Sure Start areas were compared against 'control' areas, the control areas had lower deprivation scores. This 'ethical' argument was spurious! Despite vast amounts of public funding we still don't know whether Sure Start represents good value for money. In contrast, the previous government embraced the findings of a large randomised controlled trial testing the effectiveness and cost effectiveness of culling badgers as a means of controlling bovine tuberculosis (TB). The £20 million trial showed clear evidence that 'reactive' culling (killing badgers around the area of a TB outbreak) made the situation worse, and proactive culling (killing badgers before any outbreak) led to only modest reductions in bovine TB and was not cost-effective. Hilary Benn, the then Environment secretary, accepted these results and did not implement badger culling. The current government has ignored the evidence and gone ahead with badger culling. Why the differences? Differential respect for the evidence or a desire to please constituents (farmers tend not to vote Labour).

RCTs in the social realm are not new: the first RCTs were used to evaluate university educational interventions in the 1930s, when 'delinquent' undergraduates were randomised to receive 'mentorship' from more senior students.

Despite ignoring the evidence on badgers, the current government is more favourably disposed towards the use of RCTs in social policy than the previous administration. The Educational Endowment Fund has been awarded £100 million over 10 years to commission RCTs of educational interventions. The York Trials Unit is leading on three of these trials and supporting a number of others. In addition, the 'Nudge Unit' in the Cabinet Office is supporting the use of RCTs across government. Members of the Trials and Statistics Group have been involved in several of the Nudge Unit's trials, including: RCTs of methods to encourage the prompt payment of fines; of paying tax; paying road tax; and encouraging people back to work. These new areas of work have provided us with opportunities to use novel designs – the multi-armed multi stage (MAMS) adaptive design and stepped wedge approaches – that we have not had the chance to employ in our healthcare trials. Our involvement in trials in the social world has thrown up some interesting results:

- Paying adults £5 per session to attend evening literacy classes (a policy rolled out without any actual evidence) actually *reduced* attendance.
- Year 7 pupils randomised to use a popular software programme used in many schools and designed to increase literacy and actually found that literacy levels *declined* after just a term's use.
- In contrast, providing simple guidance for probation officers reduced re-offending.

We at the York Trials Unit are the only registered UK trials unit routinely undertaking and supporting non-health trials. So what? You might ask. Working in both the health and social arenas means that we can apply and develop methodological advances in healthcare trials to social policy trials and vice versa. The upshot is fairer and higher quality tests of ideas that seemed to make so much sense on a desk in Whitehall but which fail when exposed to the real world of the general public and professionally delivered services.

Professor David Torgerson is Director of the York Trials Unit in the Department of Health Sciences.

References

Brooks, G., Burton, M., Coles, P., Miles, J., Torgerson, C., Torgerson, D. Randomised controlled trial of incentives to improve attendance at adult literacy classes *Oxford Review of Education* 2008;34:493–504.

Brooks G, Miles JNV, Torgerson CJ, Torgerson DJ. Is an intervention using computer software effective in literacy learning? A randomised controlled trial. *Educational Studies* 2006;32:133–43.

Haynes LC, Green DP, Gallagher R, John P, Torgerson DJ. Collection of delinquent fines: an adaptive randomised trial to assess the effectiveness of alternative text messages. *Journal of Policy and Management* 2013: In press.

Haynes L, Service O, Godacre B, Torgerson D. Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials. 2012 Cabinet Office Behavioural Insights Team.

Torgerson DJ, Torgerson CJ. *The design of randomised trials in health, education and the social sciences.* Palgrave Macmillan, Basingstoke, 2008.

BETTER THE QUESTION LEFT UNASKED

New arrival in the Department of Health Sciences, Professor Tim Croudace is a psychometric epidemiologist. Connecting the disciplines of psychology, computing, and statistics, his work focuses on reducing the uncertainties associated with the measurement of mental health and well-being. As he explains, "Psychometrics is the statistics of psychological assessment; epidemiology is the understanding of the distribution and determinants of diseases and states of health; psychometric epidemiology is a useful phrase because, by joining these two approaches together, you're developing *new* methods to tackle *new* challenges."

Psychometric assessment has its roots in the field of education and the search for ways of measuring attainment and intelligence in school children. The field initially focused on two main tasks: the construction of measurement instruments, and the development and refinement of theoretical approaches to measurement. More recently, psychometric assessment has been adapted to explore health and concepts such as wellbeing or distress.

As Tim says "if you want to look at something complex like the idea of well-being, you need to approach it by asking sets of questions. In education those questions tend to be in the form of a test, asking about things which are easy or hard. In health, we look at things which are rare or common. We all have common mood experiences: anxiety, tiredness, happiness. But when these become more extreme, for example unusually long periods of misery, it is clearly something we want to be able to distinguish and identify."

Scientists have always categorised things in a bid to reduce uncertainty, and Tim is no exception, but his methods are somewhat novel. In the past, psychologists and epidemiologists have often used paper and pencil questionnaires to collect data on symptoms, feelings or attitudes, and used statistical approaches to see how responses cluster together. These are then – often rather clumsily – grouped into categories with "cut-offs" above which someone is classified as 'depressed', 'anxious', 'confident' or 'happy'. But the methods Tim is developing are more refined. As he says, "The more information we have, the more confident we can be about the classification that we assign to a person. We want to have a technical way of expressing our uncertainty around those classifications

"Psychometrics is the statistics of psychological assessment; epidemiology is the understanding of the distribution and determinants of diseases and states of health; psychometric epidemiology is a useful phrase because... by joining these two approaches together, you're developing *new* methods to tackle *new* challenges."

and, typically, the more questions we ask, the more accurate we can be, the more we can reduce uncertainty. With traditional approaches, we ask everyone the same set of questions but with more sophisticated approaches we can get to the same level of certainty or uncertainty more efficiently, for example by asking questions in particular sequences, and using the answer to one question to change the way we ask the next question."

Tim doesn't just spend his days thinking up better ways of asking questions. He bridges the gap between clinicians with good ideas about the kinds of questions that need to be asked and the statistical programmers who can set up innovative ways to collect data using computer-based questionnaires. Tim and his team analyse the validity and reliability of new instruments – which in turn provides data that can be used to inform new study designs and questions, alternative sequencing of questions, or creative ways in which the technology can be made more efficient.

Tim sees his role as bringing together the clinicians, the software engineers and the statisticians. Working this way he supports new and better methods and collaborations. He tells me "That's what's really exciting in York – I have got everything in place in terms of colleagues and networks, to really advance these methods, and we're constantly looking for case studies and projects, where our work can enhance measurement and provide new insights.

In a recent example, Tim worked with researchers in York and Bradford who were trying to understand whether or not mothers from different ethnic groups responded differently to questions about their mental health. They used a well-established instrument, the General Health Questionnaire-28, which measures overall distress, as well as several sub-scales: somatic symptoms; anxiety and insomnia; social dysfunction; and severe depression. They found that women of different ethnic origin differed in how they responded to questions, and that this made comparisons between them difficult on these different sub-scales. Their work will improve the validity of the measure for future women from different ethnic groups.

As technology develops, the opportunity to ask smarter questions in smarter ways and explore the patterns of peoples' responses is accelerating. "Thoughts, emotions and behaviours are complex, and to understand complex issues you need to be able to measure lots of different dimensions – if you can measure each of them more quickly and more accurately, then you can measure more at the same time."

Tim is building a strong team of researchers, all of whom are excited about the potential of their methods and the value they can bring to health research at York. He's also involved in helping to make York and the other White Rose universities (Leeds and Sheffield) centres of excellence for advanced doctoral training in quantitative research methods, getting the next generation of researchers off to a flying start.

Tim Croudace is Professor of Psychometric Epidemiology at the University of York. He works in both the Hull York Medical School and the Mental Health and Addiction Research Group, Department of Health Sciences. Tim came to York in 2012, prior to that he worked at both the Universities of Nottingham and Cambridge. Tim is also the Post Graduate Research program lead for Health Sciences at York.



Masters in Public Health

Public health concerns continue to include social inequality, economic and environmental changes, political challenges, and issues of human rights. The **Masters in Public Health** course at the University of York offers the chance for students to get a solid grounding in public health through training in public health history and practice, epidemiology and research methods, whilst at the same time ensuring that they will be able to incorporate a wider global vision of public health. The optional modules allow students the choice of focussing on research methodologies, economics and social science in relation to health and global public health and health policy.

The MPH is run in association with the Hull York Medical School (HYMS) and is a member of the Association of Schools of Public Health in the European Region (ASPHER).

Who is the course for?

The public health training offered at York is suitable for students from a wide variety of disciplines who plan to work as public health practitioners, to become researchers, to work in governmental or non-governmental organisations, to go on to study medicine or to pursue PhD studies. It also provides a solid foundation for those planning to take the membership examination of the Faculty of Public Health.

Funding

The NHS, through its regional organisation Health Education Yorkshire and the Humber (HE Y&H), provides a number of fully funded places for part-time masters level programmes.

University of York

www.york.ac.uk/healthsciences/gradschool/masters/public-health/