

Quality improvement science

Evaluating improvement

A Niroshan Siriwardena MMedSci PhD FRCGP
Professor of Primary and Prehospital Health Care

Community and Health Research Unit (CaHRU), Faculty of Health and Social Sciences, University of Lincoln, UK

Steve Gillam MD FFPH FRCP FRCGP

Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, UK

ABSTRACT

Evaluating quality improvement interventions requires a variety of methods. These range from quantitative methods, such as randomised controlled trials, to quasi-experimental (controlled before-and-after and interrupted time series) and uncontrolled before-and-after studies, including clinical audits, to determine whether improvement interventions have had an effect. Qualitative methods are often also used to understand how or why an intervention was successful and which components

of a complex or multifaceted intervention were most effective. Finally, mixed methods designs such as action research or case study methods are widely used to design and evaluate improvement interventions.

Keywords: case study methods; evaluation; experimental methods; general practice; primary care; qualitative studies; quality improvement

Introduction

A range of methods has been used to evaluate quality improvement interventions. These can vary in terms of the rigour of the methods used and their ability to attribute improvement to the intervention being proposed. Studies can range in design from randomised controlled trials where attribution is clearer, to other types of experimental methods including quasi-experimental designs, such as non-randomised control group (sometimes call controlled before-and-after) or interrupted time series methods, to uncontrolled before-and-after studies (including clinical audits) where attribution is less certain (Figure 1).¹

Improvement interventions are often complex (that is, multiple rather than single) and pragmatic so that 'real-world' designs are called for, involving evaluation of complex interventions. Improvement often involves a series of interventions including education (of professionals and/or patients), reminders (to professionals and/or patients), audit and feedback or other measures which vary in content, intensity or timing between different intervention sites so that it is not always clear which components in the so-called 'black box' of the intervention are effective.²

In order to understand how or why an intervention works, it is often necessary to use methods such as surveys or qualitative interviews, focus groups, documentary (textual) analysis, observational or ethnographic methods. It may also be necessary to combine quantitative and qualitative methods, for example, with case study methods, or to work with participants to design the evaluation, for example, using action research methods. Quality improvement methods themselves can also be used to evaluate improvement, which adds to the complexities of improvement evaluations.³

Designing evaluations

A starting point for designing an evaluation is the logic model. Logic models can also be used to design improvement interventions by defining the population and problem that the intervention is aimed at, specifying inputs (in terms of resources provided for

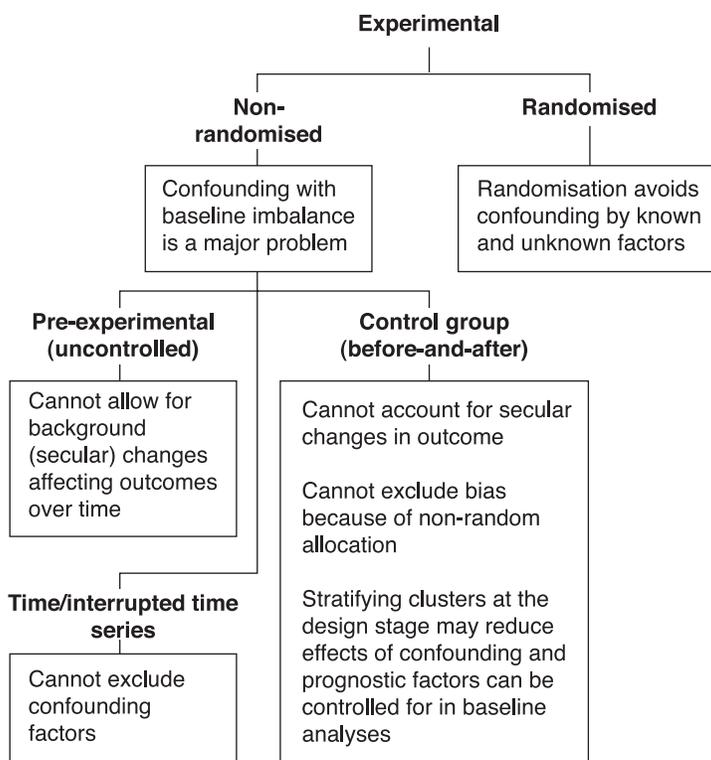


Figure 1 Experimental studies used to evaluate improvement interventions (adapted from Ukuomunne *et al*⁴)

planning, implementation and evaluation), outputs (in terms of healthcare processes implemented and the population that is actually reached) and longer term outcomes measured in terms of health and wider benefits or harms, whether intended or incidental and in the short, medium or long term.³

In an evaluation logic model, we can add to this by specifying the evidence or data to be collected and the method that will be used to analyse the data. For example, the logic model for an evaluation of a national quality improvement collaborative designed to improve care for acute myocardial infarction and stroke in ambulance services is shown in Figure 2.⁵

The figure shows that we collected quantitative data, survey data (pre- and post- intervention), qualitative data from observations and meetings, and analysed these using a mixture of time series, qualitative analysis, pattern matching to link time series and qualitative findings, and comparison of different sites (cross-case synthesis) to develop an explanation of what happened, as well as why and how this came about as a result of the collaborative.

We describe the individual methods used to determine effect sizes of improvement interventions and to understand how or why an intervention was successful or which components of a complex multifaceted intervention were most effective.

Randomised controlled trials

Because improvement interventions usually involve the education of healthcare staff together with other multiple components, the most common type of randomised controlled trial (RCT) used is the cluster randomised controlled trial (CRCT). CRCTs involve the randomisation of practitioners or groups of practitioners (in a practice, organisation or area), rather than individual patients, allocated to an intervention or control group.

CRCTs are used because educational interventions for professionals cannot be switched on and off with different patients, i.e. professionals are not able to implement their learning with one patient randomised to the intervention while forgetting what they have learnt with another patient allocated to a control group.

The unit of analysis in CRCTs can be at the level of the unit of randomisation or at the level of the patient. Although many design flaws of RCTs can also apply to CRCTs (e.g. allocation bias, volunteer bias), there are additional features that should be considered in CRCTs.

These include the potential correlation of outcomes between patients in clusters (termed the intracluster correlation), which occurs because these patients tend to be more similar to each other than to a randomly

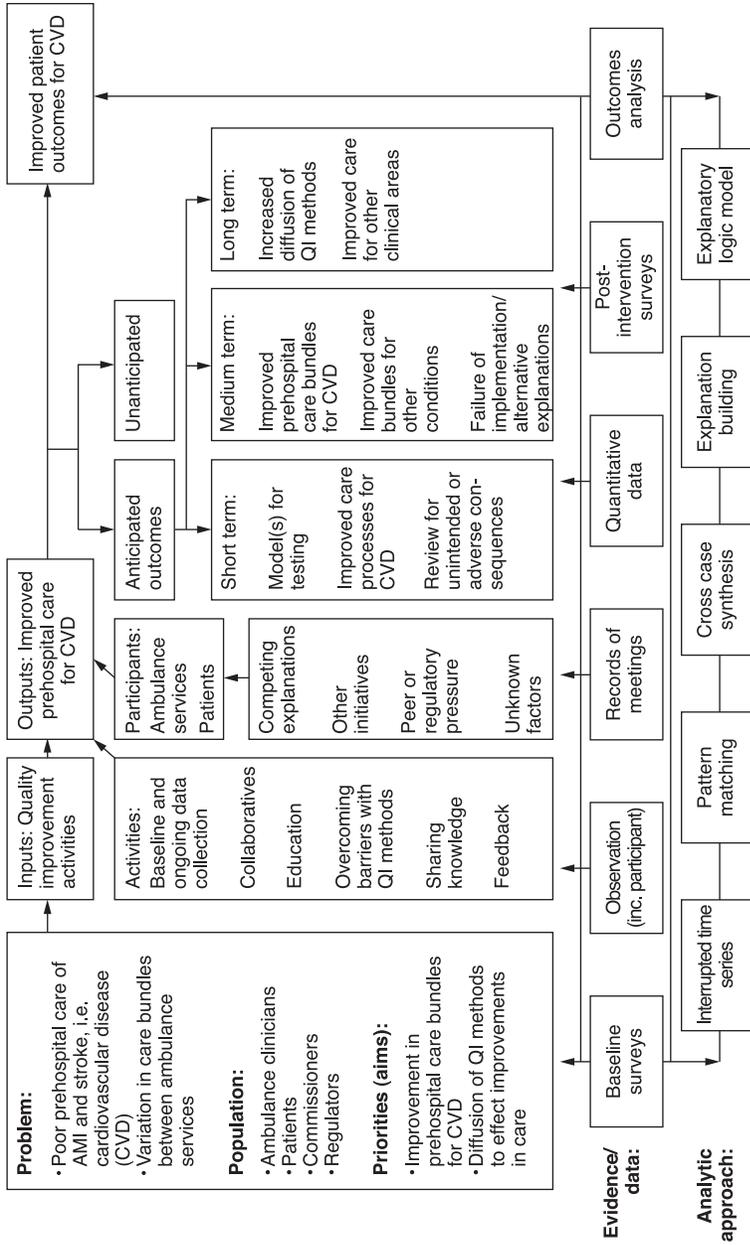


Figure 2 Ambulance Services Cardiovascular Quality Initiative (ASCQI): evaluation logic model⁵

selected patient. There is an additional risk of patients in control clusters receiving the intervention. This can occur because professionals in the intervention arm move to the control cluster (i.e. switch organisations or locations) or because those in the control arm learn about the intervention from colleagues in the intervention arm, an occurrence termed contamination.

An example of a CRCT for an improvement intervention is shown in Box 1. In this example, both the unit of randomisation and analysis was the practice.

Before and after studies

Single group before-and-after (or pre-post intervention) studies without a control group, sometimes termed pre-experimental studies, are often used in improvement studies. An example is shown in Box 2.

Pre-experimental designs suffer from significant and often irremediable flaws. It may be impossible to determine whether an improvement or other change in outcome is due to the intervention itself or to a confounding or alternative explanation, such as an external factor or a natural change over time, referred to as a secular trend. Outcomes may also be altered due to the participants changing their behaviour as a result of being observed (the Hawthorne effect) or due to regression to the mean, where out-

lying variables tend to move towards mean values. However, they may be useful for developing an improvement intervention prior to more rigorous testing.

Quasi-experimental studies

Quasi-experimental trials are more robust than pre-experimental studies, but less so than randomised controlled trials. There are two main types of quasi-experimental study: the non-randomised controlled before-and-after study and the (interrupted) time series study. In the controlled before-and-after design, an intervention is administered to a study group and compared with a control group who continue as usual. An example is shown in Box 3. Confounding may be due to external influences on outcomes occurring between the pre- and post-intervention phases. Potential sources of bias include selection bias from the non-random selection of intervention and control groups or areas leading to baseline imbalance in outcomes of other differences between the two groups. Regression to the mean and differences in secular trends between groups may also occur in such studies.

The interrupted time series design looks at data for the outcome of interest for a period before, during and after the intervention, and therefore takes secular trends into account. However, this design can be affected by

Box 1 Case study: cluster randomised controlled trial of an educational outreach visit to improve influenza and pneumococcal immunisation rates in primary care⁶

Improvement in the delivery of influenza and pneumococcal vaccinations to high-risk groups is an important aspect of preventive care delivered by primary healthcare teams. We aimed to investigate the effect of an educational outreach intervention to primary healthcare teams on influenza and pneumococcal vaccination uptake in high-risk patients.

We used a cluster randomised controlled trial design. The trial involved 30 general practices in the Trent region, UK. Fifteen practices were randomised to the intervention and 15 to the control group after stratifying for baseline vaccination rate. All intervention practices were offered and received an educational outreach visit to primary healthcare teams, in addition to audit and feedback directed at improving influenza and pneumococcal vaccination rates in high-risk groups. Control practices received audit and feedback alone. We measured influenza and pneumococcal vaccination rates in high-risk groups in all practices. Primary outcomes were improvements in vaccination rates in patients aged 65 years and over, and patients with coronary heart disease (CHD), diabetes and a history of splenectomy.

Improvements in pneumococcal vaccination rates in the intervention practices were significantly greater compared with controls in patients with CHD, 14.8% versus 6.5% (risk ratio [RR] = 1.23, 95% confidence interval [CI] = 1.13–1.34) and diabetes, 15.5% versus 6.8% (RR = 1.18, 95% CI = 1.08–1.29), but not splenectomy, 6.5% versus 4.7% (RR = 0.96, 95% CI = 0.6–1.42). Improvements for influenza vaccination were also usually greater in intervention practices, but did not reach statistical significance. The increases for influenza vaccination in intervention versus control practices were: CHD, 18.1% versus 13.1% (RR = 1.06, 95% CI = 0.99–1.12); diabetes, 15.5% versus 12.0% (RR = 1.07, 95% CI = 0.99–1.16), splenectomy 16.1% versus 2.9% (RR = 1.22, 95% CI = 0.78–1.93); and those over 65 years 20.7% versus 25.4% (RR = 0.99, 95% CI = 0.96–1.02).

We found that practices where primary care teams received an educational outreach visit demonstrated a significantly greater improvement in uptake in high-risk groups for pneumococcal but not influenza vaccine.

Box 2 Case study: before-and-after study evaluating improvement in influenza and pneumococcal vaccination uptake in high-risk groups in Lincolnshire⁷

The delivery of influenza and pneumococcal vaccine to high-risk groups is an important preventive care responsibility for primary care.

We used a two-stage multi-practice audit of influenza and pneumococcal vaccination rates in high-risk groups before and after graphical anonymised feedback and written advice on improving vaccination rates.

Twenty-two of 105 Lincolnshire practices volunteered to participate. The study period for the baseline data collection was September to December 1998 and re-evaluation took place in January to February 2000 after the next annual influenza vaccination programme. Key measures for improvement were influenza and pneumococcal vaccination rates in high-risk groups, specifically in patients with coronary heart disease, diabetes and post-splenectomy.

A combination of strategies for change was used, including: dissemination of guidelines; advice on setting up disease and vaccine registers; organisational strategies for improving vaccination rates, including call and recall systems; and benchmarking of performance.

For practices participating in both phases of the audit cycle, mean annual influenza vaccination uptake increased by 10.8% (95% CI = 5.3–16.1%, $P = 0.001$) to 74.4% in coronary heart disease patients, by 8.6% (95% CI = 1.5–15.7%, $P = 0.02$) to 70.6% in patients with diabetes, and by 17.3% (95% CI = 4.8–29.8%, $P = 0.01$) in patients post splenectomy. Mean pneumococcal vaccination rates improved by 27.5% (95% CI = 12.6–42.3%, $P = 0.002$) to 58.6% in coronary heart disease patients; by 28.8% (95% CI = 17.2–40.3%, $P < 0.001$) to 64.0% in patients with diabetes, and by 15.9% (95% CI = 1.8–30.1%, $P = 0.03$) in post-splenectomy patients. These improvements occurred prior to the current national programme for influenza vaccination of patients over 65 years old.

Improvements in influenza and pneumococcal vaccination uptake occurred in patients with coronary heart disease, diabetes and post-splenectomy at re-evaluation. Practices were able to achieve and exceed national targets for influenza immunisation of high-risk groups. Quality of care improved through organisational change, audit and feedback with benchmarking of performance.

Box 3 Case study: an evaluation of an educational intervention to reduce inappropriate cannulation and improve cannulation technique by paramedics⁸

Intravenous cannulation enables the administration of fluids or drugs by paramedics in the prehospital setting. Inappropriate use and poor technique carry risks for patients, including pain and infection. We aimed to investigate the effect of an educational intervention designed to reduce the rate of inappropriate cannulation and to improve cannulation technique.

We used a non-randomised control group design, comparing two counties in the East Midlands (UK) as intervention and control areas. The educational intervention was based on Joint Royal Colleges Ambulance Liaison Committee guidance and delivered to paramedic team leaders who cascaded it to their teams. We analysed rates of inappropriate cannulation before and after the intervention using routine clinical data. We also assessed overall cannulation rates before and after the intervention. A sample of paramedics was assessed post intervention on cannulation technique with a 'model' arm using a predesigned checklist.

There was a non-significant reduction in inappropriate (no intravenous fluids or drugs given) cannulation rates in the intervention area (1.0% to 0%) compared with the control area (2.5% to 2.6%). There was a significant ($P < 0.001$) reduction in cannulation rates in the intervention area (9.1% to 6.5%; OR = 0.7, 95% CI = 0.48–1.03) compared with an increase in the control area (13.8% to 19.1%; OR = 1.47, 95% CI = 1.15–1.90), a significant difference ($P < 0.001$). Paramedics in the intervention area were significantly more likely to use correct hand-washing techniques post-intervention (74.5% vs. 14.9%; $P < 0.001$).

loss to follow-up (or attrition), Hawthorne effects, or contamination. An example is shown in Box 4.

Qualitative methods

Although experimental methods can show the extent of any change resulting from an improvement initiative,

Box 4 Case study: investigation of the effect of a countywide protected learning time scheme on prescribing rates of ramipril using an interrupted time series design⁹

Protected learning time (PLT) schemes were set up in primary care across the UK with little published evidence of their effectiveness.

We wished to investigate the effect of a PLT intervention for general practice in Lincolnshire, UK to increase prescribing of ramipril for prevention of cardiovascular outcomes.

We used a quasi-experimental, interrupted time series design. We analysed prescribing data one year before and after the education investigating change in rate of increase of prescribing of ramipril, whether change in prescribing was related to postulated explanatory variables and to determine intervention costs.

The primary outcome was the rate of change of ramipril (10 mg) prescription items 12 months after compared with before the educational intervention. Secondary outcomes included cost. Following education, ramipril prescribing at the therapeutic dosage increased significantly (OR = 1.50, 95% CI = 1.07–1.93) by 52 345 items (31 132 items at 10 mg) at a cost of £292 000 to £460 000 depending on the drug formulation. This occurred despite a background of secular change.

Most practices were represented by GPs, nurses or both during the education. Single-handed GPs were less likely to attend. Practices showed considerable variation in response to the educational intervention. The only predictor of whether practices increased in prescribing rate after the education was whether a practice nurse had undertaken specific diabetes training. Total list size, dispensing, training or single-handed status and GP attendance did not predict a change in prescribing.

We concluded that PLT schemes can contribute to beneficial changes in prescribing across a large geographical area.

Box 5 Case study: qualitative interview study of practitioners exploring drivers for change in primary care of diabetes following a protected learning time educational event¹⁰

A number of protected learning time schemes have been set up in primary care across the UK but there has been little published evidence of their impact on processes of care. We undertook a qualitative study to investigate the perceptions of practitioners involved in a specific educational intervention in diabetes as part of a protected learning time scheme for primary healthcare teams, relating to changing processes of diabetes care in general practice.

We undertook semi-structured interviews of key informants from a sample of practices stratified according to the extent to which they had changed behaviour in prescribing of ramipril and diabetes care more generally, following a specific educational intervention in Lincolnshire, UK.

Interviews sought information on facilitators and barriers to change in organisational behaviour for the care of diabetes.

An interprofessional protected learning time scheme event was perceived by some, but not all participants as bringing about changes in processes for diabetes care. Participants cited examples of change introduced partly as a result of the educational session. This included using ACE inhibitors as the first line for patients with diabetes who developed hypertension, increased use of aspirin, switching patients to glitazones, and conversion to insulin either directly or by referral to secondary care. Other reported factors for change, unrelated to the educational intervention, included financially driven performance targets, research evidence and national guidance.

Facilitators for change linked to the educational session were peer support and teamworking supported by audit and comparative feedback.

A protected learning time scheme, using interprofessional learning, local opinion leaders and early implementers as change agents may have influenced changes in systems of diabetes care in selected practices but also how other confounding factors played an important part in changes that occurred in practice.

they cannot explain why or how the change occurred without using qualitative methods (Box 5). Qualitative methods can take the form of interviews (of patients or practitioners or both), focus groups and observations

including ethnographic methods and these can provide in-depth information about how and why an improvement intervention might be working.

Box 6 Case study: effect of a national quality improvement collaborative on prehospital care for acute myocardial infarction and stroke in England⁵

Previous studies have shown wide variations in prehospital ambulance care for acute myocardial infarction (AMI) and stroke. We aimed to evaluate the effectiveness of implementing a Quality Improvement Collaborative (QIC) for improving ambulance care for AMI and stroke.

We used an interrupted time series design to investigate the effect of a national QIC on change in delivery of care bundles for AMI (aspirin, glyceryl trinitrate, pain assessment and analgesia) and stroke (face–arm–speech test, blood pressure and blood glucose recording) in all English ambulance services between January 2010 and February 2012. Key strategies for change included local quality improvement (QI) teams in each ambulance service supported by a national co-ordinating expert group that conducted workshops educating staff in QI methods to improve AMI and stroke care. Expertise and ideas were shared between QI teams who met together at three national workshops, between QI leads through monthly teleconferences, and between the expert group and participants. Feedback was provided to services using annotated control charts.

We analysed change over time using logistic regression with three predictor variables: time, gender and age. There were statistically significant improvements in care bundles in nine (of 12) participating trusts for AMI (OR = 1.04, 95% CI = 1.04, 1.04), nine for stroke (OR = 1.06, 95% CI = 1.05, 1.07), eleven for either AMI or stroke and seven for both conditions. Overall care bundle performance for AMI increased in England from 43% to 79% and for stroke from 83% to 96%. Successful services all introduced provider prompts and individualised or team feedback. Other determinants of success included engagement with front-line clinicians, feedback using annotated control charts, expert support and shared learning between participants and organisations.

The QIC led to significant improvements in ambulance care for AMI and stroke in England. The use of care bundles as measures, clinical engagement, application of quality improvement methods, provider prompts, individualised feedback and opportunities for learning and interaction within and across organisations helped the collaborative to achieve its aims.

Action research, case study and mixed methods

Evaluations of improvement often involve mixed methods, combining quantitative and qualitative methods to determine both the effect size and determinants of an improvement intervention. Action research studies involve participants to a greater or lesser extent in the conception, design and evaluation of an intervention; they evaluate the effects of an improvement intervention.

Case study methods may be based on a single case or multiple cases.¹¹ They combine methods to develop an explanatory model for why and intervention might work in some cases and not in others. For example, in the Ambulance Services Cardiovascular Quality Initiative (Figure 2 and Box 6), we combined interrupted time series and multiple case study methods, matching the patterns of change in ambulance services with a detailed analysis of changes within each service to develop an explanation of what led to differences in improvement.

REFERENCES

- 1 Siriwardena AN. Experimental methods in health research. In: Saks M, Allsop J (eds). *Researching Health: qualitative, quantitative and mixed methods* (2e). Sage: London, 2013. pp. 263–79.
- 2 Siriwardena AN. The exceptional potential for quality improvement methods in the design and modelling of complex interventions. *Quality in Primary Care* 2008;16:387–9.
- 3 Siriwardena AN. Using quality improvement methods for evaluating healthcare. *Quality in Primary Care* 2009;17:155–9.
- 4 Ukoumunne OC, Gulliford MC, Chinn S, Sterne JAC, Burney PGJ. Methods for evaluating area-wide and organisation-based interventions in health and healthcare: a systematic review. *Health Technology Assessment* 1999;3.
- 5 Siriwardena AN, Shaw D, Essam N *et al*. The effect of a national quality improvement collaborative on prehospital care for acute myocardial infarction and stroke in England. *Implementation Science* 2014;9:17.
- 6 Siriwardena AN, Rashid A, Johnson MR, Dewey ME. Cluster randomised controlled trial of an educational outreach visit to improve influenza and pneumococcal immunisation rates in primary care. *British Journal of General Practice* 2002;52:735–40.
- 7 Siriwardena AN, Hazelwood L, Wilburn T, Johnson MRD, Rashid A. Improving influenza and pneumococcal vaccination uptake in high risk groups in Lincolnshire: a quality improvement report from a large rural county. *Quality in Primary Care* 2003;11:19–28.
- 8 Siriwardena AN, Iqbal M, Banerjee S, Spaight A, Stephenson J. An evaluation of an educational intervention to reduce inappropriate cannulation and im-

prove cannulation technique by paramedics. *Emergency Medicine Journal* 2009;26:831–6.

- 9 Siriwardena AN, Fairchild P, Gibson S, Sach T, Dewey M. Investigation of the effect of a countywide protected learning time scheme on prescribing rates of ramipril: interrupted time series study. *Family Practice* 2007;24:26–33.
- 10 Siriwardena AN, Middlemass JB, Ward K, Wilkinson C. Drivers for change in primary care of diabetes following a protected learning time educational event: interview study of practitioners. *BMC Medical Education* 2008;8:4.
- 11 Yin RK. *Case Study Research: design and methods* (3e). Sage: Thousand Oaks, CA, 2003.

PEER REVIEW

Commissioned; not externally peer reviewed.

CONFLICTS OF INTEREST

None declared.

ADDRESS FOR CORRESPONDENCE

Professor A. Niroshan Siriwardena, Community and Health Research Unit (CaHRU), School of Health and Social Care, College of Social Science, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, UK. Email: nsiriwardena@lincoln.ac.uk

Received 16 February 2014

Accepted 17 February 2014