International Journal for Quality in Health Care Volume 32 Supplement 1 January 2020 pp. 1–103

A WORLD OF RESOURCES FOR QUALITY IN HEALTH CARE

Deepening our Understanding of Quality in Australia (DUQuA)

International Journal for Quality in Health Care

Volume 32 Supplement 1 January 2020

ISSN 1353-4505 (PRINT)
ISSN 1464-3677 (ONLINE)

academic.oup.com/intqhc

ISQua International Society for Quality in Health Care

OXFORD UNIVERSITY PRESS

OXFORD

International Journal for Quality in Health Care

The Official Journal of the International Society for Quality in Health Care (ISQua)

Funding
1. School of Nursing, Midwifery and Health Sciences, University College Dublin, Belfield, Dublin 4, Ireland.
2. Mater Lean Academy, Transformation Office, Mater Misericordiae University Hospital, Eccles Street, Dublin 7, Ireland.

International Journal for Quality in Health Care

Aims and scope

The purpose of this journal is to make available to a worldwide readership activities and research related to quality and safety in health care. The Journal publishes papers in all disciplines related to the quality of health care, including health services research, health care evaluation, health economics, utilization review, cost containment and nursing care research, as well as aspects of clinical research related to quality of care.

This peer-reviewed journal is truly interdisciplinary and includes contributions from representatives of all health professions, such as doctors, nurses, quality assurance professionals, managers, politicians, social workers and therapists, as well as researchers from health-related backgrounds.

The Journal accepts original articles describing empirical studies and their results. Descriptions of the organization and process of health care quality evaluation and improvement are also published, as are editorials and letters dealing with the principles, ethics and philosophy of health care evaluation.

The Journal contains news of the International Society for Quality in Health Care (ISQua), including announcements of forthcoming courses, symposia and congresses. As the official organ of the Society, it serves as a forum for communication on Society activities.

SUBSCRIPTIONS

A subscription to the *International Journal for Quality in Health Care* comprises ten issues. Our prices include Standard Air as postage outside of the UK. Airmail rates are available on request. *International Journal for Quality in Health Care* Advance Access contains papers that have been finalised, but have not yet been included within the issue. Advance Access is updated daily.

Annual Subscription Rate (Volume 32, ten issues, 2020)

Institutional: Print edition and site-wide online access: £911.00/$1736.00/€1368.00, Print edition only: £840.00/$1594.00/€1259.00
Site-wide online access only: £662.00/$1293.00/€993.00
Personal: Print edition and individual online access: £143.00/$277.00/€214.00
Please note that commencing September 2007, all orders delivered outside Europe and North America (USA and Canada only) will be renewed and invoiced in GBP sterling. There may by other subscription rates available for a complete listing, please visit http://intqhc.oxfordjournals.org/subscriptions/.
The current year and two previous years' issues are available from Oxford University Press. Previous volumes can be obtained from the Periodicals Service Company, 11 Main Street, Germantown, NY 12526 USA. E-mail: psc@periodicals.com. Tel: +1 (518) 537 4700. Fax: +1 (518) 537 5899.
For further information, please contact: Journals Customer Service Department, Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, UK. E-mail: jnls.cust.serv@oup.com. Tel (and answerphone outside normal working hours): +44 (0)1865 353907. Fax: +44 (0)1865 353485. **In the US, please contact:** Journals Customer Service Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513, USA. E-mail: jnlorders@oup.com. Tel (and answerphone outside normal working hours): 800 852 7323 (toll-free in USA/Canada). Fax: 919 677 1714. **In Japan, please contact:** Journals Customer Service, Oxford University Press, 4-5-10-8F Shiba, Minato-ku, Tokyo 108-8386, Japan. E-mail: custserv.jp@oup.com. Tel: (03) 5444 5858. Fax: (03) 3454 2929. **Methods of payment:** Full pre-payment in the correct currency is required for all orders. Payment should be in US dollars for orders being delivered to the USA or Canada; Euros for orders being delivered within Europe (excluding the UK); GBP sterling for orders being delivered elsewhere (i.e. not being delivered to USA, Canada, or Europe). All orders should be accompanied by full payment and sent to your nearest Oxford Journals office. Subscriptions are accepted for complete volumes only. Orders are regarded as firm, and payments are not refundable. Our prices include Standard Air as postage outside of the UK. Claims must be notified within four months of despatch/order date (whichever is later). Subscriptions in the EEC may be subject to European VAT. If registered, please supply details to avoid unnecessary charges. For subscriptions that include online versions, a proportion of the subscription price may be subject to UK VAT. Subscribers in Canada, please add GST to the prices quoted. Personal rate subscriptions are only available if payment is made by personal cheque or credit card, delivery is to a private address, and is for personal use only. The following methods of payment are accepted: (i) Cheque (payable to Oxford University Press, to Oxford University Press, Cashiers Office, Great Clarendon Street, Oxford OX2 6DP, UK) in GB£ Sterling (drawn on a UK bank), US$ Dollars (drawn on a US bank), or EU€ Euros. (ii) Bank transfer to Barclays Bank Plc, Oxford Group Office, Oxford (bank sort code 20-6518) (UK), overseas only Swift code BARC GB 22 (GB£ Sterling to account no. 70299332, IBAN GB89BARC2065187$ 0299332; US$ Dollars to account no. 66014600, IBAN GB27BARC20651866014600; EU€ Euros to account no. 78923655, IBAN GB16BARC20651878923655). (iii) Credit card (Mastercard, Visa, Switch or American Express). Due to excessive charges, subscription agents paying by credit card will forfeit their agency discount.
The *International Journal for Quality in Health Care* (ISSN 1353 4505) is published bimonthly in February, April, June, August, October and December by Oxford University Press, Oxford, UK. The *International Journal for Quality in Health Care* is distributed by Mercury Media Processing, LLC, 1850 Elizabeth Ave., Suite #C, Rahway, NJ 07065, USA. Periodicals postage paid at Rahway, NJ and at additional entry points.
US POSTMASTER: send address changes to the *International Journal for Quality in Health Care*, c/o Mercury Media Processing, 1634 E. Elizabeth Ave., Linden, NJ 07036, USA.

Environmental and ethical policies

Oxford Journals, a division of Oxford University Press, is committed to working with the global community to bring the highest quality research to the widest possible audience. Oxford Journals will protect the environment by implementing environmentally friendly policies and practices wherever possible. Please see http://www.oxfordjournals.org/ethicalpolicies.html for further information on environmental and ethical policies.

Supplements, reprints and corporate sales

For requests from industry and companies regarding supplements, bulk article reprints, sponsored subscriptions, translation opportunities for previously published material, and corporate online opportunities, please email special.sales@oup.com, fax +44 (0) 1865 353774, or visit www.oxfordjournals.org/jnls/sales.

DOIs (Digital Object Identifiers)

For information about DOIs and to resolve them, please visit http://www.doi.org/.

Permissions

For information on how to request permissions to reproduce articles/information from this journal, please visit www.oxfordjournals.org/jnls/permissions/.

Advertising

Advertising, inserts and artwork enquires should be addressed to Advertising and Special Sales, Oxford Journals, Oxford University Press, Great Clarendon Street, Oxford, OX2 6DP, UK. Tel: +44 (0) 1865 354767; Fax: +44 (0) 1865 353774; E-mail: jnlsadvertising@oup.com.

Disclaimer

Whilst every effort is made by the publishers and editorial board to see that no inaccurate or misleading data, opinion or statement appears in this journal, they wish to make it clear that the data and opinions appearing in the articles and advertisements herein are the sole responsibility of the contributor or advertiser concerned. Accordingly, the publishers, the editorial board and editors and their respective employees, officers and agents accept no responsibility or liability whatsoever for the consequences of any such inaccurate or misleading data opinion or statement.

Drug and dosage selection: the authors have made every effort to ensure the accuracy of the information herein, in particular with regard to drug selection and dose. However, appropriate information sources should be consulted, especially for new or unfamiliar drugs or procedures. It is the responsibility of every practitioner to evaluate the appropriateness of a particular opinion in the context of actual clinical situations and with due consideration to new developments.

Join ISQua!

Go to www.isqua.org and receive the Journal free of charge as an ISQua member! Members of the International Society for Quality in Health Care (ISQua) receive the Journal—including online access to all archived and current issues—as a membership benefit.

Individual membership

Individual membership is available to any person who is interested in or working in the field of quality improvement in health care. Annual individual membership rates for 2019: 150 for upper income countries, €112.50 for middle income countries, €75.00 for lower income countries to include online-only access to the journal (an extra €50 to be paid if you wish for hard copies). Student and Retired €75.00.

Institutional membership

Institutional membership is offered by the Society to any institution or society with similar aims and objectives to ISQua. Annual institutional membership rates for 2019: €1200 for upper income countries, €900 for middle income countries and €700 for lower income countries.

Membership enquiries

Full membership details and an online membership application form for credit card payment can be found at ISQua's website: www.isqua.org
Membership enquiries should be sent to: **Chief Executive Officer,** International Society for Quality in Health Care (ISQua), Huguenot House, 35 - 38 St Stephens Green, Dublin, D02 NY63, Ireland. Tel: +353 1 670 6750. E-mail: info@isqua.org

Typeset by SPi Global Manila Philippines. Printed by Bell & Bain, Glasgow, UK.

International Journal for Quality in Health Care

Volume 32 Supplement 1 January 2020

Deepening our Understanding of Quality in Australia (DUQuA)

Scan to view this journal on your mobile device

OXFORD UNIVERSITY PRESS

ISQua International Society for Quality in Health Care

International Journal for Quality in Health Care, 2020, 32(S1), 1–7
doi: 10.1093/intqhc/mzz102
Perspectives on Quality

Perspectives on Quality

Bending the quality curve

JEFFREY BRAITHWAITE[1], ROBYN CLAY-WILLIAMS[1], NATALIE TAYLOR[2,3], HSUEN P. TING[1], TERESA WINATA[1], GASTON ARNOLDA[1], ROSA SUNOL[4,5], OLIVER GRÖNE[6,7], CORDULA WAGNER[8,9], NIEK S. KLAZINGA[10], LIAM DONALDSON[7], and S. BRUCE DOWTON[11]

[1]Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Road, Macquarie University, NSW, Sydney, 2109, Australia, [2]Cancer Research Division, Cancer Council NSW, 153 Dowling St, Woolloomooloo, NSW 2011, Australia, [3]Faculty of Health Sciences, University of Sydney, Camperdown, Sydney, NSW, 2006, Australia, [4]Avedis Donabedian Research Institute (FAD), Universitat Autonoma de Barcelona, Bellaterra, Barcelona, 08193, Spain, [5]Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Barcelona, Spain, [6]OptiMedis AG, Burchardstraße 17, Hamburg, 20095, Germany, [7]London School of Hygiene and Tropical Medicine, University of London, Keppel St, Bloomsbury, London, WC1E 7HT, United Kingdom, [8]Netherlands Institute of Health Services Research (NIVEL), Otterstraat 118, CR Utrecht, 3513, The Netherlands, [9]Amsterdam UMC, VU University Medical Centers, De Boelelaan, 1117, Amsterdam, [10]Amsterdam UMC, Academic Medical Centre, University of Amsterdam, Meibergdreef 9, 1105 AZ, The Netherlands, and [11]Office of the Vice Chancellor, Macquarie University, NSW, Sydney, 2109, Australia

Address reprint requests to: Jeffrey Braithwaite, Centre for Healthcare Resilience and Implementation Science (CHRIS), Australian Institute of Health Innovation, Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia. Tel: +61 2 9850 2401; Fax: +61 2 9850 2499; E-mail: jeffrey.braithwaite@mq.edu.au

Editorial Decision 7 September 2019; Accepted 12 September 2019

Abstract

With this paper, we initiate the Supplement on Deepening our Understanding of Quality in Australia (DUQuA). DUQuA is an at-scale, cross-sectional research programme examining the quality activities in 32 large hospitals across Australia. It is based on, with suitable modifications and extensions, the Deepening our Understanding of Quality improvement in Europe (DUQuE) research programme, also published as a Supplement in this Journal, in 2014. First, we briefly discuss key data about Australia, the health of its population and its health system. Then, to provide context for the work, we discuss previous activities on the quality of care and improvement leading up to the DUQuA studies. Next, we present a selection of key interventional studies and policy and institutional initiatives to date. Finally, we conclude by outlining, in brief, the aims and scope of the articles that follow in the Supplement. This first article acts as a framing vehicle for the DUQuA studies as a whole. Aggregated, the series of papers collectively attempts an answer to the questions: what is the relationship between quality strategies, both hospital-wide and at department level? and what are the relationships between the way care is organised, and the actual quality of care as delivered? Papers in the Supplement deal with a multiplicity of issues including: how the DUQuA investigators made progress over time, what the results mean in context, the scales designed or modified along the way for measuring the quality of care, methodological considerations and provision of lessons learnt for the benefit of future researchers.

Key words: quality management, national standards, accreditation, quality improvement

Table 1 Australia's health and health system: profile and characteristics

Australia	
Population[a]	25 203 198
Indigenous population[b]	3.3%
Land mass[a]	7 692 024 km^2
GDP (per capita)[a]	US$ 62 765
Urbanisation[b]	71% live in cities
Australia's health	
Average life expectancy[b]	80.4 (males); 84.6 (females) born in 2016
Number of births[b]	309 000 per year (2015 data)
Leading cause of death[b]	Coronary heart disease (males); dementia and Alzheimer's disease (females)
Chronic conditions[b]	50% have at least 1/8 chronic conditions
Disability[b]	18% of the population
Overweight[b]	63% of adults are overweight or obese
Australia's health system	
Hospitals[b]	701 public hospitals, 630 private hospitals (2015–2016)
Average Length of Stay (admitted patients)[b]	5.7 days (public); 5.2 days (private)
Health expenditure, proportion of GDP[c]	9.25%
Admitted patients[b]	6.3 million
Emergency Department presentations (2016–2017)[b]	7.8 million

[a]Based on 2019 data. World Population Review (2019) Australia Population, 2019. [Available at: http://worldpopulationreview.com/countries/australia-population/].
[b]Australian Institute of Health and Welfare (2018) Australia's health 2018: in brief. AIHW, Canberra, Australia. ISBN: 978-1-76054-377-8.
[c]Based on 2016 data. The World Bank (2019) Current health expenditure (% of GDP) [Available at: https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS].

Introduction

This article frames the studies that follow, which collectively form a report on the Deepening our Understanding of Quality in Australia (DUQuA) programme. The Supplement is an in-depth examination, across 12 articles, of a 5-year effort to examine the quality of care in 32 large hospitals, geographically spread over the Australian States and Territories. It follows the landmark Deepening our Understanding of Quality improvement in Europe (DUQuE) seven-country study. We begin with a summary profile of the Australian health system (Table 1) and then contextualise the 11 subsequent articles in the Supplement with a brief outline of our view of where the research on the quality of care sits at the moment.

Australia is a relatively wealthy country with a well-funded and organised health system. About two-thirds of the care provided is publicly funded. Some 9.25% of gross domestic product (GDP) is spent on healthcare, and while the population is healthy with internationally benchmarked above life expectancy, it is ageing and there are patients with chronic conditions, disabilities and obesity. The leading causes of death are coronary heart disease in males and dementia and Alzheimer's disease in females.

The arc of improvement

Internationally as well as in Australia, clinicians, their patients and those accountable for healthcare delivery want to see the curve of care arc toward improvement. A healthy healthcare system contributes better outcomes for the population it serves and gives patients enhanced experiences. Architects of well-designed health systems will also preside over the well-being of staff who provide care and seek to reduce costs wherever possible. This ideal has been termed the quadruple aim [1]. The premise is that all four goals can be achieved simultaneously. The first three take care on an improvement gradient, while the cost curve bends downwards (Figure 1) [2].

What needs to happen for this overarching goal of systems improvement along multiple dimensions to be realised? There have been many responses. Donaldson [3] evocatively contended that we need for healthcare to become 'an organisation with a memory': to have a better handle on learning from past failures and adverse events. Berwick [4] argued that we should forge a new, more moral epoch which he labelled Era 3: one predicated on better cultures, interprofessional working and commitment to openness and improvement, transforming from Era 1 (the age of professional dominance) and Era 2 (the age of accountability, scrutiny, bureaucracy and measurement). Other thought leaders are building a multi-faceted evidentiary model for getting what we know into routine practice, which has come to be called implementation science [5–7]. In implementation science, emphasis is shifting from generalisable interventions towards the importance of context [8]. Hollnagel, Braithwaite and colleagues, taking a complex systems approach [9], see that much effort to date has centred on trying to stamp out harm after the event, which has been summarised in the motif 'Safety-I'. They have promoted a more proactive approach to quality and safety, bringing into sharper focus how everyday performance succeeds more often than it fails (labelled 'Safety-II'), and recognising that human variability can be an important contributor to success in complex systems. For them, the key task then shifts to encouraging health systems to be more resilient, by learning from the variability of daily common practice—often described simply as how things go right—and figuring out ways to support, augment or encourage how people succeed.

Progress to date

Despite these attractive ideas and frameworks for change, progress in the quality of care has not followed the curves in Figure 1 for most system-level improvement initiatives. Furthermore, the shape of the healthcare cost curve has been pointing upwards, not downwards, for decades. In particular, large-scale change has proven infuriatingly difficult to orchestrate [10]. In one notable example of the lack of improvement despite multi-pronged initiatives, Landrigan and colleagues [11] reported a study of the rates of harm in 10 hospitals

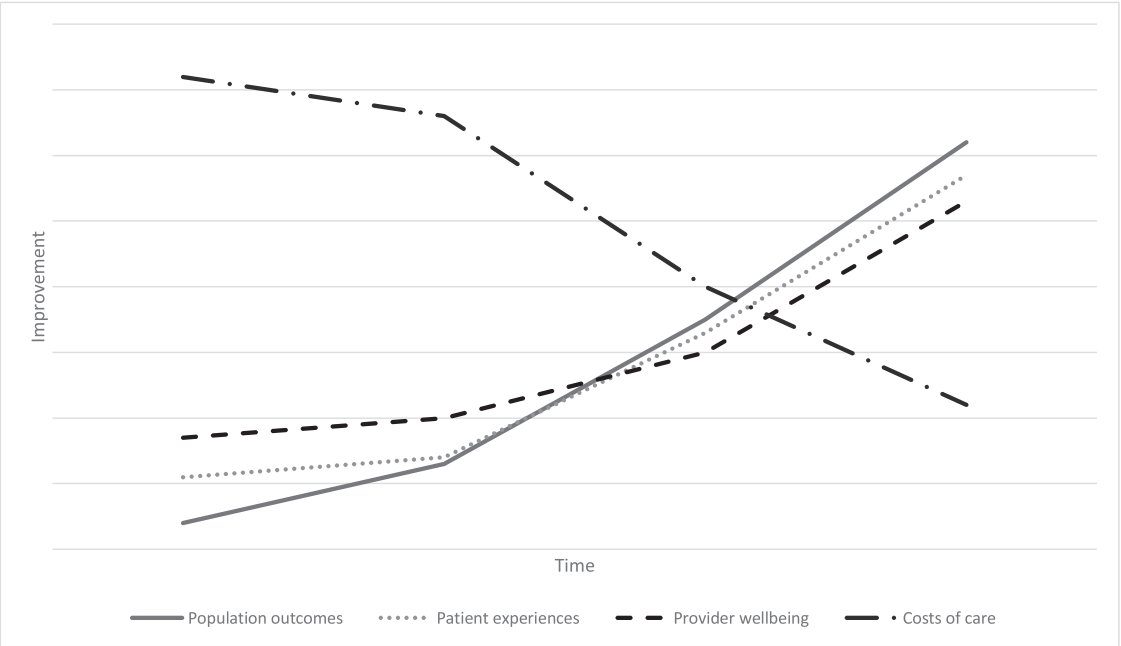

Figure 1 Bending quality and cost along an improvement gradient. Source: Authors' conceptualisation, hypothetical data.

in North Carolina, USA, between 2002 and 2007, which had participated extensively in local and national improvement campaigns (Figure 2). Harm remained static over the 6-year period despite these endeavours.

In another system-wide initiative, this time in the UK, Benning and Lilford led an evaluation of a concerted effort to improve care for patients (The Health Foundation Safer Patients Initiative) in a group of intervention and control hospitals [12]. The results of the improvement efforts were not significantly different between the two groups (Figures 3 and 4). The figures show change in *Clostridioides difficile* and Methicillin-resistant *Staphylococcus aureus* (MRSA) in the control and intervention hospitals; both groups fluctuated over the 8-year period, but at the same rate. Therefore, the net effect of the intervention group was no additional gains in performance.

Whether or not the advocates of safer, higher quality care focus in the future on reducing harm or doing more things right, or both, they will need to base their endeavours on at least two strategies. One is designing and implementing interventions to improve the quality of care, which have greater success than the Landrigan *et al*. [11] and Benning *et al*. [12] studies. The other is making meaningful and accurate measurement of that quality so that progress can be tracked over time. On interventions, there has been much resourcefulness exercised to design theoretical and practical projects by which care quality can be enhanced. Some initiatives have made gains, but these have often been context-dependent; from a litany of examples, prominent under the Safety-I banner is incident reporting systems, bundles to reduce catheter-related bloodstream infections, hand hygiene campaigns, root cause analyses, checklists, and computerised alerts and reminders. In one celebrated example in The Netherlands, adverse events were reduced by 30% across the country [13]. The Quality and Safety in Europe by Research study, another pan-European program of work, conducted in a similar time-frame to DUQuE, looked at healthcare quality practices and policies in five European countries [14, 15]. The study focused on macro-, meso- and micro-level relationships of care, including conducting longitudinal case studies in hospitals via 389 interviews with healthcare practitioners and 803 hours of observations. Safety-II initiatives include tools for understanding how care frequently goes right, looking to support the ability of systems to perform well under varying conditions. The Resilience Assessment Grid is one of these tools, identifying which efforts to improve in terms of four potentials: to respond, to monitor, to learn and to anticipate [16]. The Functional Resonance Analysis Method is another, which enables performance variability—a naturally recurring property of activities in complex systems—to be modelled and understood [17]. This systems view will become important as we traverse the articles that follow in this Supplement.

Policy and institutional responses

Outside the research domain, at the macro-level of systems, policymakers have tried to regulate or guide clinicians' behaviours on the front lines of care such that practices are safe, and care is of high quality. But it is true to say that top-down policy-designed mandates, prescriptions or encouragement have also not created the progress we would like them to have made [18–20]. Meso-level management has also struggled at the district, hospital and community levels to embrace, adopt or take up—or otherwise reinforce—policymakers' prescriptions or researchers' findings. Although many clinicians on the front lines have become accepting in principle over the last two decades of the quality improvement agenda, they have often exercised autonomy over their activities rather than simply be compliant with the multiplicity of rules, regulations, policies and procedures that have been enacted [21]. Many have also been too busy providing direct care to embrace improvement. Some clinicians have argued that quality improvement has become more bureaucratic, burdensome and time-consuming over time.

Among varied responses, many health systems have encouraged clinicians to take up managerial roles [22]. And consumers have entered the field of quality improvement more recently, becoming increasingly vocal and activated, and have claimed the right to be

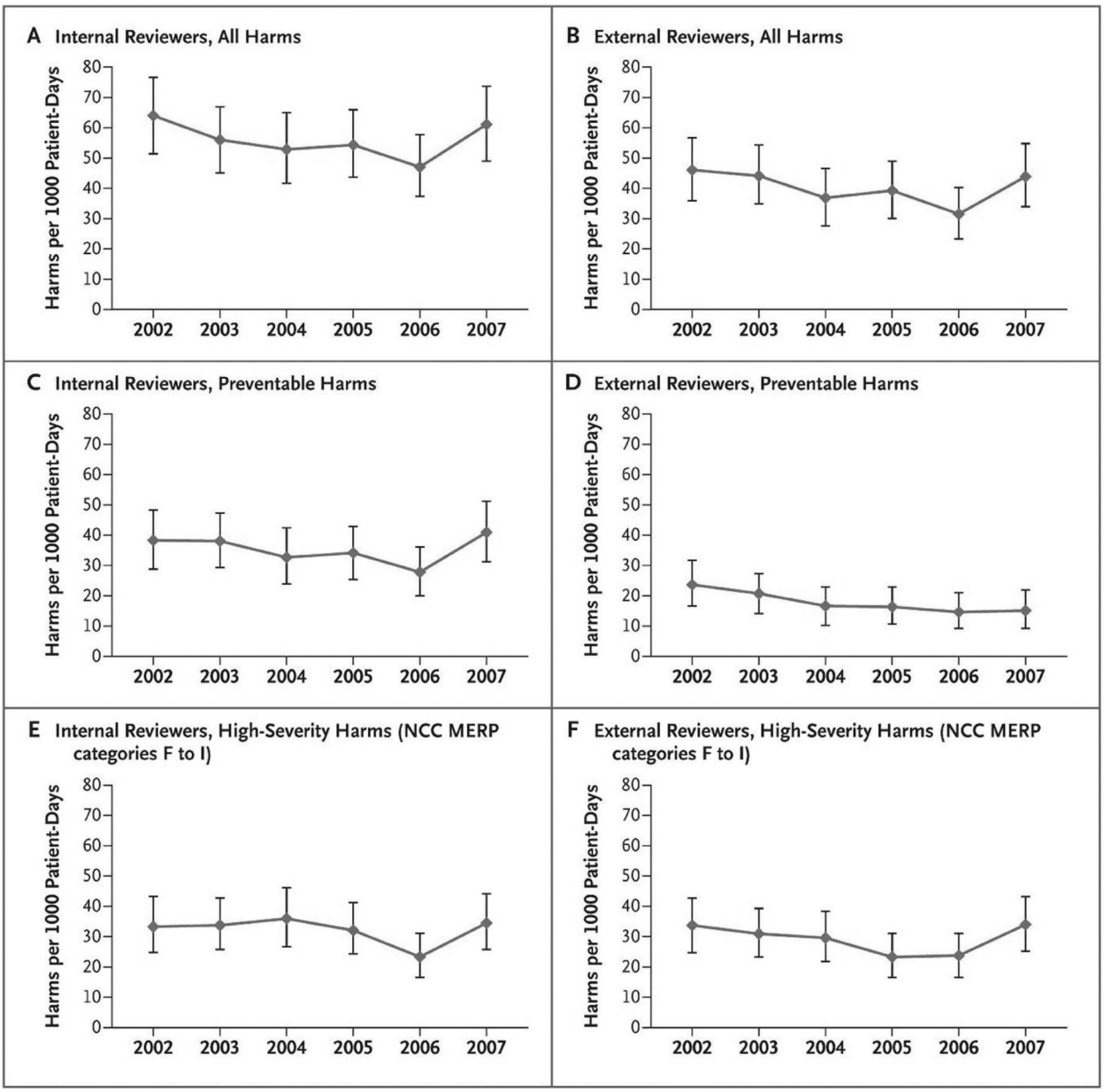

Figure 2 Rates of all harms, preventable harms and high-severity harms per 1000 patient-days, identified by internal and external reviewers in 10 North Carolina hospitals, 2002–2007. Source: Landrigan *et al*. [11].

more than just an input to care decisions, but an active participant in the quality of care that is delivered to them—the 'co-creators' rather than 'passive recipients' of care [23].

All-in-all, despite increasing interest from all these stakeholders over two and a half decades, progress has been painfully slow. One missing piece of the jigsaw is to understand how quality is enacted on the ground in acute settings across-the-board, and for this, we need to dig deeper.

The present research agenda

To this end, articles reported in this Supplement have been configured as a series of studies under the DUQuA programme of research [24] funded by the Australia's National Health and Medical Research Council. These observational studies measure the quality of care in a sample of large Australian hospitals, cross-sectionally. That said that the program of work included an action-research strategy to provide benchmarked feedback reports to each of the 32 study hospitals (article 10) [25], designed to stimulate targeted internal discussions based on hospital-specific findings and provide a platform for improvement.

In terms of its pedigree, DUQuA follows, with appropriate modifications to the research design, earlier results reported in this Journal in 2014 of the DUQuE programme of research funded by the European Union's 7th Research Framework Programme [26, 27]. Methodologically, while the DUQuA studies are based on the original DUQuE design, they depart from that template in a number of important ways. There are multiple reasons for these departures,

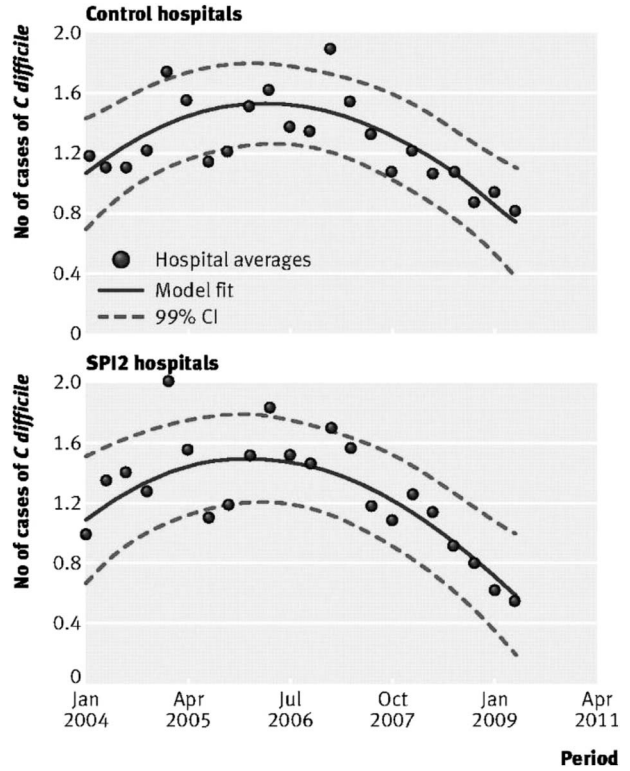

Figure 3 Rate of *C. difficile*, control and intervention hospitals, UK Safer Patients Initiative, 2004–2011.
Source: Benning *et al*. [12].

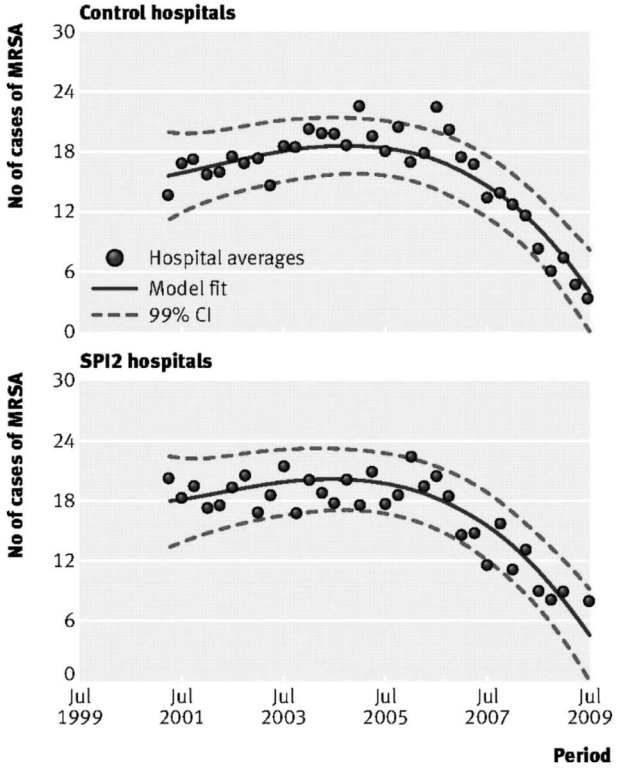

Figure 4 Rate of MRSA, control and intervention hospitals, UK Safer Patients Initiative, 2004–2011.
Source: Benning *et al.* [12].

including that methods have advanced, localised modifications were needed and the DUQuA research team added questions along the way. Additionally, DUQuA took into account aspects of the Australian health system and developed and validated scales for those who might need rigorous measurement tools in the future. DUQuE studies were primarily quantitative. DUQuA, too, is quantitative but also includes qualitative studies—for example, on benchmarking and surveying of hospitals. We acknowledge that alternative methodological approaches may offer other advantages and be helpful in disentangling the complex relationships between organisations, quality and patient outcomes.

The DUQuA studies in outline

Across the pages of this Supplement, we will see how the investigators examined two broad questions that the DUQuA team have been grappling with for the last 5 years. The first is: what is the relationship between strategies to manage the quality of care at the organisational and departmental levels in hospitals? The second is: what are the relationships between the way care is organised, and the actual quality of care delivered to patients?

In the article immediately following this one (article 2) [28], the DUQuA investigators take an overarching look at the answers to these questions. The next four contributions [29–32] trace how the project made progress in understanding organisational- and departmental-level quality factors, outcomes and cultures of care. article 3 [29] presents the scales for studying organisational- and department-level quality; article 4 [30] presents results on the extent to which organisation-level quality management systems influence department-level quality; article 5 [31] examines the relationships between the key variables: quality management systems, safety culture and leadership, and patient outcomes in Emergency Departments; and article 6 [32] looks at how we refined and validated a questionnaire scale to examine clinician safety culture and leadership. article 7 [33] changes the focus, examining clinician factors, and article 8 [34], patient factors. The next three articles [25, 35–36] look at methodological issues (article 9) [35], benchmarking of results for the benefits of participating hospitals (article 10) [25] and the use of external surveyors to assess care quality (article 11) [36]. Finally, article 12 [37] concludes the Supplement's work, reflects on the lessons learnt and considers what should happen next as a result of this work.

Conclusion

By way of summarising, and setting up the rest of the Supplement: progress in shifting the quality curves along multiple dimensions in the right direction has been slower than the community wants and patients deserve. Large-scale studies with some celebrated examples [13] have not made the much-anticipated gains. Stakeholders embedded in healthcare (policymakers, managers, clinicians, associated staff and patients) have been active in contributing to the quality enterprise but they, too, desire more improvement—and at a faster pace. Among multiple changes taking place, Safety-II has offered an alternative perspective, clinicians are not always enjoined in the quality enterprise, patients are more involved than in past eras and care is becoming more complex.

Against this backdrop of challenge and change, it is timely to peer inside the quality black box and to look at one country's attempts

to bend the curve in two of four of the quadruple aims: population outcomes for groups of acute patients, and patient experiences, in this case by making a cross-sectional assessment of the care delivered by 32 of Australia's largest hospitals. It is to this task that we now turn.

Contributors

The DUQuA research team consists of experienced researchers, clinicians, biostatisticians, and project managers with expertise in health services research, survey design, and validation, large-scale research and project management, sophisticated statistical analysis, quality improvement and assessment, accreditation, clinical indicators, policy and patient experience. JB conceived the idea to embrace DUQuE for Australia, led the research grant to fund the project, chairs the steering committee, and led the development of the manuscript. NT and RCW co-led the detailed study design, managed the project across time, and contributed to the development of the manuscript. HPT and GA provided statistical expertise for the study design and developed the analysis plan for the manuscript. TW contributed to the logistics of project management, the refinement of measures, and the development of the manuscript. Over the years, CW, NSK, LD and SBD provided advice, expertise and encouragement.

Supplementary material

Supplementary material is available at *INTQHC Journal* online

Acknowledgements

Members of the original DUQuE team (Oliver Gröne, Rosa Suñol) provided extensive input and advice in the initial stages of the project, supporting design modifications from DUQuE and the planned DUQuA study approach. Dr Annette Pantle provided expert advice on revision and development of DUQuA measures, and Professor Sandy Middleton, Associate Professor Dominique Cadillac, Kelvin Hill, Dr Carmel Crock, and Professor Jacqueline Close provided input into the development of our AMI, hip fracture and stroke clinical process indicator lists. Nicole Mealing and Victoria Pye provided statistical advice in the initial phases of the project. We greatly appreciate their efforts.

Funding

National Health and Medical Research Council (NHMRC) Program Grant APP1054146 (CI Braithwaite).

Data sharing statement

Data will be made publicly available to the extent that individual participants or participating hospitals cannot be identified, in accordance with requirements of the approving Human Research Ethics Committees.

References

1. Sikka R, Morath JM, Leape L. The quadruple aim: care, health, cost and meaning in work. *BMJ Qual Saf* 2015;**24**:608–10.
2. Groene O, Pfaff H, Hildebrandt H. Scaling up a population-based integrated health care system: the case of ´Healthy Kinzigtal´ in Germany. In: Braithwaite J, Matsuyama Y, Mannion R, Shekelle P (eds.). *Health Systems Improvement Across the Globe: Success Stories from 60 Countries*. Boca Raton, FL, CRC Press, 2018.
3. Donaldson L. *An Organisation with a Memory: Report of an Expert Group on Learning from Adverse Events in the NHS*. London, UK, Department of Health, 2000.
4. Berwick DM. Era 3 for medicine and health care. *JAMA* 2016;**315**:1329–30.
5. Rapport F, Clay-Williams R, Churruca K *et al.* The struggle of translating science into action: foundational concepts of implementation science. *J Eval Clin Pract* 2018;**24**:117–26.
6. Braithwaite J, Marks D, Taylor N. Harnessing implementation science to improve care quality and patient safety: a systematic review of targeted literature. *Int J Qual Health Care* 2014;**26**:321–9.
7. Taylor N, Clay-Williams R, Hogden E *et al.* High performing hospitals: a qualitative systematic review of associated factors and practical strategies for improvement. *BMC Health Serv Res* 2015;**15**:244.
8. Kringos DS, Sunol R, Wagner C *et al.* The influence of context on the effectiveness of hospital quality improvement strategies: a review of systematic reviews. *BMC Health Serv Res* 2015;**15**:277.
9. Braithwaite J, Churruca K, Ellis LA *et al. Complexity Science in Healthcare—Aspirations, Approaches, Applications and Accomplishments: A White Paper*. Sydney, Australia, Australian Institute of Health Innovation, Macquarie University, 2017.
10. Clay-Williams R, Nosrati H, Cunningham FC *et al.* Do large-scale hospital- and system-wide interventions improve patient outcomes: a systematic review. *BMC Health Serv Res* 2014;**14**:369.
11. Landrigan CP, Parry GJ, Bones CB *et al.* Temporal trends in rates of patient harm resulting from medical care. *N Engl J Med* 2010;**363**:2124–34.
12. Benning A, Dixon-Woods M, Nwulu U *et al.* Multiple component patient safety intervention in English hospitals: controlled evaluation of second phase. *BMJ* 2011;**342**:d199.
13. Baines R, Langelaan M, de M *et al.* How effective are patient safety initiatives? A retrospective patient record review study of changes to patient safety over time. *BMJ Qual Saf* 2015;**24**:561–71.
14. Burnett S, Renez A, Fernandes A *et al.* Prospects for comparing European hospitals in terms of quality and safety: lessons from a comparative study in five countries. *Int J Qual Health Care* 2013;**25**:1–7.
15. Wiig S, Aase K, von Plessen C *et al.* Talking about quality: exploring how 'quality' is conceptualized in European hospitals and healthcare systems. *BMC Health Serv Res* 2014;**14**:478.
16. Hollnagel E. RAG—the resilience analysis grid. In: *Resilience Engineering in Practice: A Guidebook*. Surrey, UK, Ashgate Publishing Ltd, 2011, 275–96.
17. Hollnagel EFRAM. *The Functional Resonance Analysis Method: Modelling Complex Socio-technical Systems*. Boca Raton, FL, CRC Press, 2012.
18. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003;**362**:1225–30.
19. Vincent C. *Patient Safety*. Wiley-Blackwell, West Sussex, UK, 2011.
20. Nolte E. *How Do We Ensure that Innovation in Health Service Delivery and Organization is Implemented, Sustained and Spread?* Copenhagen, DE, World Health Organization, 2018.
21. Lipworth W, Taylor N, Braithwaite J. Can the theoretical domains framework account for the implementation of clinical quality interventions? *BMC Health Serv Res* 2013;**13**:530.
22. Rotar AM, Botje D, Klazinga NS *et al.* The involvement of medical doctors in hospital governance and implications for quality management: a quick scan in 19 and an in depth study in 7 OECD countries. *BMC Health Serv Res* 2016;**16**:160.
23. World Health Organization. *Patient Safety*. Geneva, CH, 2019. https://www.who.int/patientsafety/en/.
24. Taylor N, Clay-Williams R, Hogden E *et al.* Deepening our understanding of quality in Australia (DUQuA): a study protocol for a nationwide, multilevel analysis of relationships between hospital quality management systems and patient factors. *BMJ Open* 2015;**5**:e010349.
25. Hibbert P, Saeed F, Taylor N *et al.* Can benchmarking Australian hospitals for quality identify and improve high and low performers? Disseminating research findings for hospitals. *Int J Qual Health Care* 2020;**32**:84–88.
26. Secanell M, Groene O, Arah OA *et al.* DUQuE project consortium. Deepening our understanding of quality improvement in Europe (DUQuE): overview of a study of hospital quality management in seven countries. *Int J Qual Health Care* 2014;**26**:5–15.

27. Groene O, Sunol R. The investigators reflect: what we have learned from the deepening our understanding of quality improvement in Europe (DUQuE) study. *Int J Qual Health Care* 2014;**26**:2–4.
28. Braithwaite J, Clay-Williams R, Taylor N *et al*. Deepening our Understanding of Quality in Australia (DUQuA): An overview of a nation-wide, multi-level analysis of relationships between quality management systems and patient factors in 32 hospitals. *Int J Qual Health Care* 2020;**32**:8–21.
29. Clay-Williams R, Taylor N, Winata T *et al*. Organization quality systems and department-level strategies: refinement of the Deepening our Understanding in Quality in Australia (DUQuA) organization and department-level scales. *Int J Qual Health Care* 2020;**32**:22–34.
30. Taylor N, Clay-Williams R, Ting HP *et al*. Do organization-level quality management systems influence department-level quality? A cross-sectional study across 32 large hospitals in Australia. *Int J Qual Health Care* 2020;**32**:35–42.
31. Clay-Williams R, Taylor N, Ting HP *et al*. The relationships between quality management systems, safety culture and leadership and patient outcomes in Australian Emergency Departments. *Int J Qual Health Care* 2020;**32**:43–51.
32. Clay-Williams R, Taylor N, Ting HP *et al*. The clinician safety culture and leadership questionnaire: refinement and validation in Australian public hospitals. *Int J Qual Health Care* 2020;**32**:52–59.
33. Clay-Williams R, Taylor N, Ting HP *et al*. Do quality management systems influence clinical safety culture and leadership? A study in 32 Australian hospitals. *Int J Qual Health Care* 2020;**32**:60–66.
34. Taylor N, Clay-Williams R, Ting HP *et al*. Validation of the patient measure of safety (PMOS) questionnaire in Australian public hospitals. *Int J Qual Health Care* 2020;**32**:67–74.
35. Arnolda G, Winata T, Ting HP *et al*. Implementation and data-related challenges in the Deepening our Understanding of Quality in Australia (DUQuA) study: implications for large-scale cross-sectional research. *Int J Qual Health Care* 2020;**32**:75–83.
36. Winata T, Clay-Williams R, Taylor N *et al*. Using accreditation surveyors to conduct health services research: a qualitative, comparative study in Australia. *Int J Qual Health Care* 2020;**32**:89–98.
37. Braithwaite J, Taylor N, Clay-Williams R *et al*. Conclusion: the road ahead: where should we go now to improve healthcare quality in acute settings? *Int J Qual Health Care* 2020;**32**:99–103.

International Journal for Quality in Health Care, 2020, 32(S1), 8–21
doi: 10.1093/intqhc/mzz103
Article

Article

Deepening our Understanding of Quality in Australia (DUQuA): An overview of a nation-wide, multi-level analysis of relationships between quality management systems and patient factors in 32 hospitals

JEFFREY BRAITHWAITE[1], ROBYN CLAY-WILLIAMS[1], NATALIE TAYLOR[2,3], HSUEN P. TING[1], TERESA WINATA[1], EMILY HOGDEN[2], ZHICHENG LI[4], AMANDA SELWOOD[1], MEAGAN WARWICK[1], PETER HIBBERT[1,5], and GASTON ARNOLDA[1]

[1]Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Road, Sydney, NSW 2109, Australia, [2]Cancer Research Division, Cancer Council NSW, 153 Dowling Street, Woolloomooloo, NSW 2011, Australia, [3]Faculty of Health Sciences, University of Sydney, Camperdown, Sydney, NSW 2006, Australia, [4]Faculty of Medicine and Health, University of Sydney, Camperdown, Sydney, NSW 2006, Australia, and [5]Australian Centre for Precision Health, Cancer Research Institute (UniSA CRI), School of Health Sciences, University of South Australia, 101 Currie Street, Adelaide, SA 5000, Australia

Address reprint requests to: Jeffrey Braithwaite, Centre for Healthcare Resilience and Implementation Science (CHRIS), Australian Institute of Health Innovation, Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia. Tel: +61-2-9850-2401; Fax: +61-2-9850-2499; E-mail: jeffrey.braithwaite@mq.edu.au

Editorial Decision 7 September 2019; Accepted 12 September 2019

Abstract

Objective: The Deepening our Understanding of Quality in Australia (DUQuA) project is a multisite, multi-level, cross-sectional study of 32 of the largest hospitals in Australia. This overview examines relationships between (i) organization-level quality management systems and department-level quality management strategies and (ii) patient-level measures (clinical treatment processes, patient-reported perceptions of care and clinical outcomes) within Australian hospitals.

Design: We examined hospital quality improvement structures, processes and outcomes, collecting data at organization, department and patient levels for acute myocardial infarction (AMI), hip fracture and stroke. Data sources included surveys of quality managers, clinicians and patients, hospital visits, medical record reviews and national databases. Outcomes data and patient admissions data were analysed. Relationships between measures were evaluated using multi-level models. We based the methods on the Deepening our Understanding of Quality Improvement in Europe (DUQuE) framework, extending that work in parts and customizing the design to Australian circumstances.

Setting, participants and outcome measures: The 32 hospitals, containing 119 participating departments, provided wide representation across metropolitan, inner and outer regional Australia. We obtained 31 quality management, 1334 clinician and 857 patient questionnaires, and conducted 2401 medical record reviews and 151 external assessments. External data via a secondary source comprised 14 460 index patient admissions across 14 031 individual patients. Associations between hospital, Emergency Department (ED) and department-level systems and strategies and five

patient-level outcomes were assessed: 19 of 165 associations (11.5%) were statistically significant, 12 of 79 positive associations (15.2%) and 7 of 85 negative associations (8.2%).
Results: We did not find clear relationships between hospital-level quality management systems, ED or department quality strategies and patient-level outcomes. ED-level clinical reviews were related to adherence to clinical practice guidelines for AMI, hip fracture and stroke, but in different directions. The results, when considered alongside the DUQuE results, are suggestive that front line interventions may be more influential than department-level interventions when shaping quality of care and that multi-pronged strategies are needed. Benchmark reports were sent to each participating hospital, stimulating targeted quality improvement activities.
Conclusions: We found no compelling relationships between the way care is organized and the quality of care across three targeted patient-level outcome conditions. The study was cross-sectional, and thus we recommend that the relationships studied should be assessed for changes across time. Tracking care longitudinally so that quality improvement activities are monitored and fed back to participants is an important initiative that should be given priority as health systems strive to develop their capacity for quality improvement over time.

Key words: hospital quality management systems, multi-level research, patient level factors, patient safety, hospital performance, quality improvement

Introduction

While the evidence base for quality care has been expanding in recent years [1–4], it remains rare to have an in-depth analysis of the state of healthcare quality across an entire country [5]. Other deficits in quality of care research include that studies are often under-powered, under-resourced or conducted with unvalidated tools [6, 7]. We know less than we would like about the relationships between how hospitals are organized for quality and the resulting patient outcomes, with consequent limitations to our understanding of how we can enact meaningful large-scale change [8, 9]. Preceding the research reported in this Supplement, the Deepening our Understanding of Quality Improvement in Europe (DUQuE) project [10] looked at the relationships between hospital structures, clinical culture, guideline-adherence and patient outcomes in seven countries, focusing on four conditions: acute myocardial infarction (AMI), hip fracture, stroke and deliveries [11–20]. Extending the DUQuE model and refocusing the research design—and building on the need to examine relationships between quality methods and interventions on the one hand and organizational and patient outcomes on the other—we report on the Deepening our Understanding of Quality in Australia (DUQuA) study, seeking to advance international knowledge in this area.

This is a critical area of research endeavour. Millions of Dollars, Yen, RMB, Euros and Pounds are committed to quality improvement structures, initiatives, programmes and projects. But we do not have robust study designs and tools by which to associate these endeavours with the outcomes they are designed to achieve. The DUQuA study aimed to measure and investigate relationships between quality-related measures at organization, department and patient levels within Australian hospitals for three conditions: AMI, hip fracture and stroke. Departing from the earlier European work which did not differentiate between quality strategies used in Emergency Departments (EDs) and inpatient care [21], we incorporated this distinction into our conceptual framework and study design [22, 23]. Three overarching research questions guided our work:

1. Are organization-level systems associated with patient-level outcomes?
2. Are ED strategies associated with patient-level outcomes, after controlling for organizational factors?
3. Are department-level strategies associated with patient-level factors after controlling for organizational and ED factors?

Methods and analysis

Here, we provide a summary of the methods, including changes from the published protocol [24]. These are minor but it is important to articulate them. Figure 1 (based on a simplified directed acyclic graph (DAG) [25]) shows the theorized relationships between the organization, ED, department and patient-level measures in the DUQuA study; the measures at each level are described below. For the outcomes under consideration in this paper, we theorized that the patient-care pathway runs through the ED to the inpatient departments. In a separate analysis of teamwork and leadership [26], however, we theorize a direct relationship between hospital and departmental levels without differentiating between ED and other departments.

Setting and participants

The project targeted all Australian hospitals that met the following criteria: (i) general public hospitals with 200 beds or more, (ii) with an ED and (iii) that regularly admit more than 30 each of AMI, hip fracture and stroke patients over a 3-month period. A group of 78 hospitals meeting these criteria was identified, 70 meeting all criteria [24] with a further eight added as they were close to the specified thresholds with one or more in each of the six Australian states and in the two territories. Within hospitals, participants included healthcare professionals (quality managers and clinicians) and patients.

Recruitment and data collection

A formal invitation, supported by the Royal Australasian College of Medical Administrators, was sent to each hospital's director of medical services or an equivalent senior hospital leader. Of the 78 hospitals approached, 62 initially agreed to participate but 30 were lost due to leadership changes, time-consuming ethics and

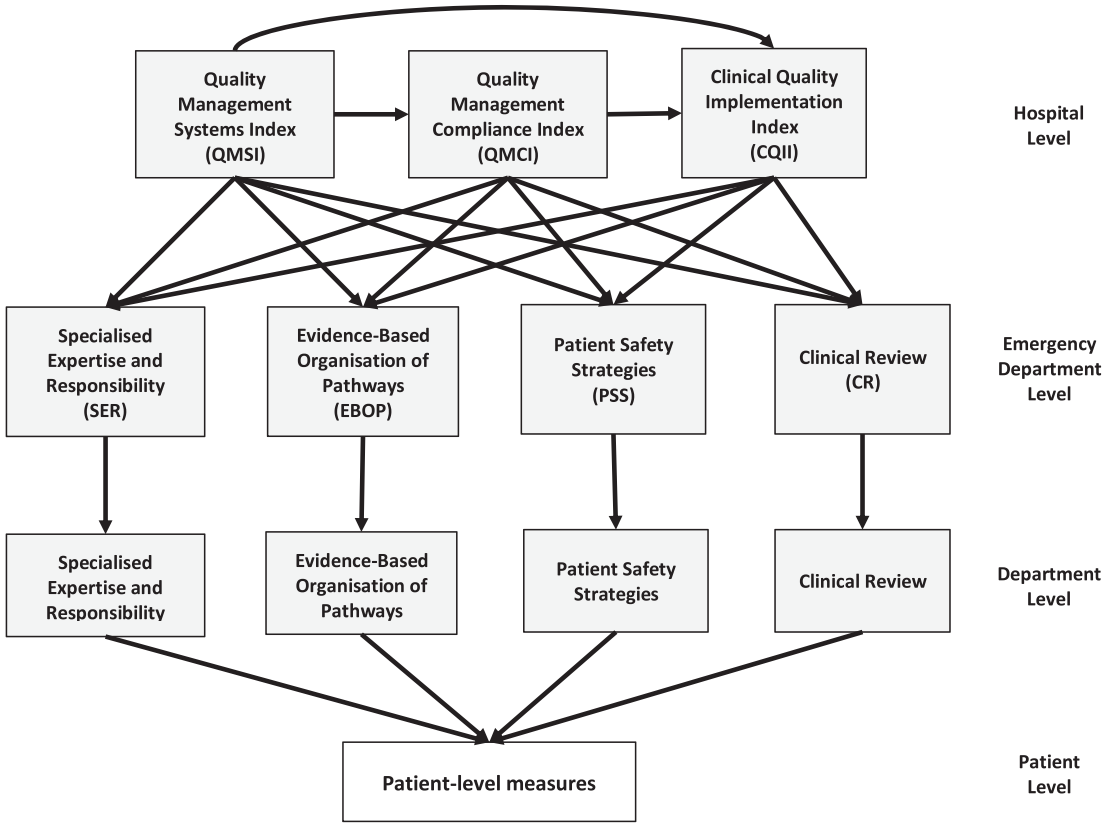

Figure 1 The relationships between multi-level study measures in DUQuA.

governance requirements, concerns about workload or other issues, leaving 32 participating hospitals. We have previously reported on the challenges and expense associated with recruiting and maintaining hospitals in complex, nation-wide studies [27, 28].

At each participating hospital, a senior staff member was designated as the 'local principal investigator' (LPI), acting as a coordinator and point of contact for the study. Through the LPI: (i) the local quality manager was asked to complete an organization-level survey; (ii) site access was provided for accredited hospital surveyors as external assessors to complete two organization-level assessments and four department-level measures (see Supplementary Appendix A for information about training and conducting data collection visits at hospitals) and in 28 hospitals, to undertake clinical audit of medical records; (iii) clinicians were invited to complete a safety culture and leadership survey and (iv) the heads of department responsible for inpatient care of AMI, hip fracture and stroke were asked to arrange for patients to complete a questionnaire about safety and—in four hospitals where this was not undertaken by external assessors (because the hospitals had an audit resource available to perform the task)—assign clinical staff to audit medical records for AMI and hip fracture. For the medical record review, external surveyors and internal auditors were given the same manual, to guide their data abstraction; online training in how to conduct the medical record review was provided to both groups via a webinar. Figure 2 shows key features of the design of the study, including hospital and participant recruitment, ethics and governance processes and measures applied for piloting and data collection. This report uses data from all sources shown in Figure 2, except the clinician and staff questionnaires (shaded); data from these sources were validated [29], and the relationship between organization and clinician-level measures is reported elsewhere in the Supplement [26].

Measures

Measures used in this project are summarized in Table 1, as are data-gathering methods. Where the measure is a scale comprising subscales, the subscales are listed in the table. We briefly summarize the measures below.

Organization-level measures: There were three measures of organization-level quality management systems: the quality management system index (QMSI) [30], the quality management compliance index (QMCI) and the clinical quality implementation index (CQII) [16]. Previously validated by DUQuE [16, 30], these measures were modified for the Australian context [24]; for example: some questions in the QMSI survey are worded differently to align with Australian terminology and processes, and a small number of additional questions were included. The validation of the DUQuA measures is described elsewhere in this Supplement [31].

Department-level measures: 'Department' refers to two sub-levels as depicted in Figure 1: the ED level and the condition-specific department level (AMI/hip fracture/stroke). Four DUQuE measures were again modified to assess department-level quality management strategies in Australian hospitals: Specialised expertise and responsibility (SER), evidence-based organization and pathways (EBOP), patient safety strategies (PSS) and clinical review (CR) [17]. PSS questions were uniform across the three condition-specific departments and

Table 1 DUQuA measures, content and data collection methods

Measures	Content	Data collection methods	Administration system
Organization-level measures			
QMSI	*Eight subscales:* Quality policy; hospital governance board activities; quality resources; quality management; preventive protocols; performance monitoring; internal quality methods for patients	Self-report questionnaire completed by the hospital's quality manager or equivalent	Paper-based questionnaire
QMCI	*Two subscales:* Monitoring patient and professional opinions; quality control and monitoring	External quality assessment by trained healthcare surveyors (site visit)	Paper-based audit forms filled by external assessors
CQII	*Seven subscales:* Preventing and controlling healthcare-associated infections; medication safety; preventing patient falls; preventing pressure injuries; routine assessment and diagnostic testing of patients in elective surgery; safe surgery that includes an approved checklist; recognizing and responding to clinical deterioration in acute health care		
Department-level measures[a]			
SER	Assignment of clinical responsibilities for a condition	External quality assessment by trained healthcare surveyors (site visit)	Paper-based audit forms filled by external assessors
EBOP	Organization of department processes (admission, acute care and discharge to facilitate evidence-based care recommendations)		
PSS	Use of international consensus-based patient safety recommendations		
CR	Integration of audit and systematic monitoring in departmental quality management mechanisms		
Culture and leadership questionnaire[b]	*Three scales:* Teamwork climate; safety climate and leadership	Self-report questionnaire completed by the hospital's doctors, nurses and allied health professionals working in participating departments	Both paper and electronically administered questionnaire
Patient-level measures			
Clinical treatment processes indicators	Nationally recognized process composite indicators based on evidence of impact on patient outcomes	*AMI*: Medical record review by external surveyors (n = 22 hospitals) or local staff (n = 4 hospitals). *Hip fracture*: Medical record review by external surveyors (n = 25 hospitals) or local staff (n = 4). *Stroke*: Patient data retrieved from the NSF registry (n = 29 hospitals) and medical record review by external surveyors (n = 3).	*AMI and hip fracture*: Paper-based medical record review forms filled by hospital staff or external surveyor. *Stroke*: NSF database extract (n = 29 hospitals), and paper-based medical record review forms filled by external surveyor (n = 3).

Continued.

Table 1 Continued

Measures	Content	Data collection methods	Administration system
Nationally collected audit data from AIHW	Readmission to the same hospital within 28 days; Death as the mode of hospital discharge from the index admission; and Length of stay of the index admission	National database.	Database extract provided by AIHW
PMOS	*Nine domains:* Communication and teamwork; organization and care; ward type and layout; access to resources; equipment (design and function); staff training; delays; staff roles and responsibilities; information flow.	Self-report questionnaire completed by eligible patients (sometimes assisted by staff)	Paper-based questionnaire

[a]Department refers to two sub-levels as depicted in Figure 1: the ED level; and the condition-specific department level (AMI/Hip Fracture/Stroke). PSS questions were uniform across the three condition-specific departments and ED. For SER, EBOP and CR, each non-ED department was only asked condition-specific questions, while the ED form had separate questions for each of the three conditions.
[b]Included in this table for completeness. These measures are not further discussed in this report but are discussed elsewhere in this Supplement [26, 29].

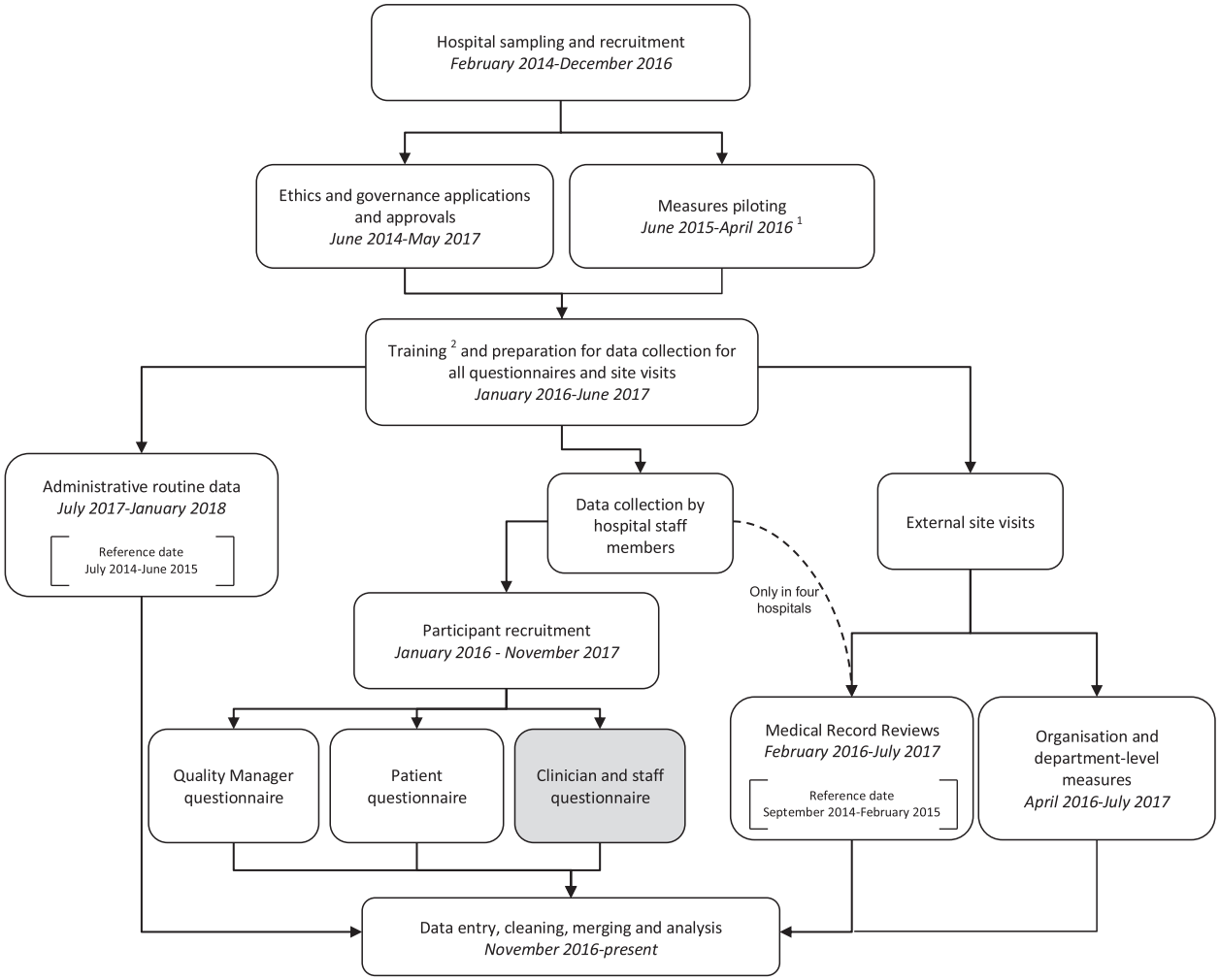

Figure 2 Stages of DUQuA data collection. Legend: Shaded box represents data collected but not used for this report. [1]Measures were piloted at different times: Patient questionnaire (May–July 2015); Medical Records Review forms (November 2015–February 2016); External site visits (Organization and department-level measures, April 2016); Quality manager questionnaire was restricted to critique experts and Clinician and staff questionnaire was previously validated. [2]Training was specifically held for 'External site visits' for healthcare surveyors who were contracted to collect data for 'Organization and department-level measures' at DUQuA hospitals.

ED. For SER, EBOP and CR, each non-ED department was only asked condition-specific questions, while the ED form had separate questions for each of the three conditions.

Patient-level measures: Clinical process indicators for AMI, hip fracture and stroke were adapted from the DUQuE indicators, in consultation with various expert groups [10, 24]: Stroke indicators were aligned with the National Stroke Foundation (NSF) Clinical Audit and the relevant Australian Commission on Safety and Quality in Health Care (ACSQHC) indicators [32–34]. For AMI, final indicators were chosen to align with the ACSQHC indicator specification standard [35]. The DUQuA final hip fracture indicators were selected after consultation with the Australian and New Zealand Hip Fracture Registry [36]. Supplementary Appendix B presents the final list of DUQuA clinical treatment indicators for the three conditions (Supplementary eTable B1).

Compliance with clinical process indicators for AMI and hip fracture was assessed by medical record audits. For stroke, data were drawn directly from the NSF clinical audit, where available ($n = 29$ hospitals), and otherwise assessed by medical record audit. Records were assessed against the eligibility criteria (Appendix B, Supplementary eTable B2), applied to records of patients admitted to hospital between September 2014 and February 2015; during this time period, 30 eligible records were targeted for extraction, for each condition at each hospital.

To assess patients' perceptions of the factors contributing to patient safety, the patient measure of safety (PMOS) questionnaire, developed and validated in the UK, was adapted [37, 38]. The modified PMOS was piloted in one of the study hospitals, for all three patient study populations and validated [39]. DUQuA again targeted 30 patients per inpatient department, and these were sampled opportunistically amongst eligible patients, depending on the availability of staff to undertake recruitment and to assist with questionnaire completion.

Selected patient-level outcomes were provided from the routinely collected admitted patient data collection maintained by the Australian Institute of Health and Welfare (AIHW). The selected outcomes were: (i) death as the mode of hospital separation in the index admission; (ii) death or readmission within 28 days to the same hospital (a composite measure, as these are competing risks) and (iii) length of stay in the index admission. Index admissions were

Table 2 Comparison of participating and non-participating hospitals

Characteristic		Participating ($N = 32$)		Non-participating ($N = 46$)	
		n	%	*n*	%
State	ACT	2	6.3	0	0.0
	NSW	11	34.4	18	39.1
	NT	1	3.1	0	0.0
	QLD	8	25.0	7	15.2
	SA	1	3.1	3	6.5
	TAS	1	3.1	1	2.2
	VIC	8	25.0	11	23.9
	WA	0	0.0	6	13.0
Hospital peer group	Principal referral	17	53.1	13	28.3
	Public acute group A	14	43.8	28.5[a]	62.0
	Public acute group B	1	3.1	4.5[a]	9.8
Remoteness area	Major cities	26	81.3	33	71.7
	Inner regional	4	12.5	12	26.1
	Outer regional	2	6.3	1	2.2
Average available beds	Less than 200	2	6.3	4	8.7
	200 to less than 500	17	53.1	31	67.4
	500 to less than 1000	11	34.4	11	23.9
	1000 and more	2	6.3	0	0.0

[a]One hospital operated across two campuses, with each campus having a different peer group classification.

sought for the period 1 July 2014 to 30 June 2015, with readmissions tracked to late July 2015. Data on age and sex were provided, to allow analysis controlling for these patient-level factors.

Statistical analysis

The simplified DAG at Figure 1 shows the hypothesized relations amongst variables, identifying the key variables controlled for in the statistical models. For example, to examine whether QMCI has an effect on a patient-level measure, we controlled for QMSI because, following the DUQuE study, it is a predictor of both QMCI and the patient-level measure.

The dataset was processed as detailed in Supplementary Appendix C and analysed using SAS/STAT software version 9.4 (SAS Institute, Cary, NC, USA). General linear mixed models were used to analyse PMOS data. Generalized linear mixed models were applied to length of stay (LOS); with a negative binomial distribution) and to dichotomous outcomes (binomial distribution). The models were adjusted for one random effect (hospital), several fixed effects (hospital peer group categorized as Referral versus Acute [A or B], patient age and sex), and variables designated in Figure 1. A multiple imputation method was used [40] to address missing data at the organization level, described in additional detail in an article discussing methodological issues in the DUQuA study, elsewhere in this Supplement [41]; analysis was repeated for each of the 100 imputations, and the SAS/STAT MIANALYZE procedure was used to obtain the pooled parameter estimates and standard errors. Because of the small number of hospitals involved, and reflecting the study's exploratory purpose, we did not correct for multiple testing, choosing rather to explore patterns of response; this decision is further discussed in a separate paper in this Supplement [41].

Ethical considerations

Ethical approvals were secured from State and Territory human research ethics committees in New South Wales (#14/206), Victoria (#15/36), the Australian Capital Territory (#15/131), South Australia (#15/260), the National Territory (#15-2509), Tasmania (#H0015383) and Queensland (#15/361). Site-specific authorizations, including permission for external assessors (surveyors) to collect data in hospitals, were granted. We complied with confidentiality requirements of national legislation or standards of practice of each jurisdiction. Separate ethics approval was provided by the AIHW (#EO2017/2/315) for patient-level hospital data.

Results

Characteristics of participating and non-participating hospitals are shown in Table 2. Although a wide spread of Australian states and territories are included in the study, the sample over-represents Queensland and under-represents Western Australia, over-represents principal referral hospitals, under-represents inner regional remoteness areas and over-represents hospitals with 500 or more beds. Figure 3 shows the geographical distribution of the participating hospitals.

Table 3 presents the number of measures targeted and achieved for each measure; routinely collected data sourced from AIHW were collected for all eligible index admissions. Completed PMOS surveys were collected for 857 patients (32.5% of target, primarily due to low response rates); medical record reviews were completed for 2401 patients (90.9% of target, primarily due to exclusion of pilot hospital data collected prior to form modification, or to hospitals not contributing to the NSF collection) and the AIHW data (three outcomes; not shown in Table 3) covered 14 460 index admissions across 14 031 patients.

Descriptive statistics for the condition-specific department-level measures (SER, EBOP, PSS and CR) are shown in Appendix D, Supplementary eTable D1. The analysis of patient-level measures adjusts for ED-level measures of SER, EBOP and CR (condition-specific) and for PSS (same for all conditions); the descriptive statistics for ED-level measures are summarized in Supplementary eTable D2.

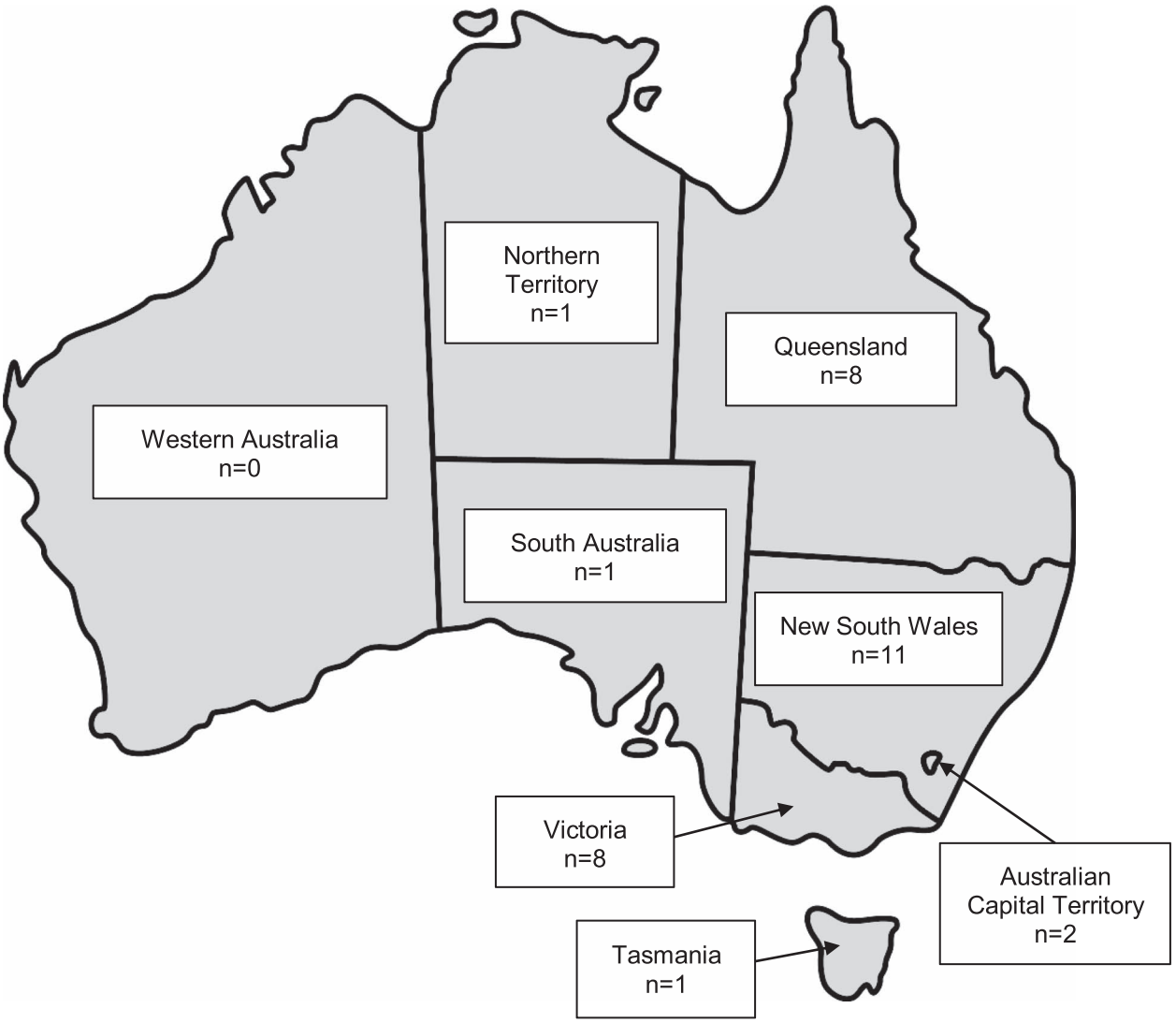

Figure 3 Map of Australia, showing final hospital sample.

Table 3 Number of measures expected and achieved, by data source[a]

Type of measure	Measures	Number expected	Number included in analysis	% of target achieved
Quality manager questionnaire	QMSI	31[b]	31[b]	100
Patient questionnaire	PMOS	2640[c]	857	32.5
Medical record reviews	Clinical treatment processes indicators including National Stroke Foundation registry database	2640[c]	2401	90.9
External visitor assessments	QMCI CQII SER EBOP PSS CR	32 120 	32 119 	100 99.2

[a]Routinely collected administrative data sourced from AIHW is not included in this table as there was no target number. Information was sought for all eligible index admissions.
[b]Two hospitals shared the same quality management system, so there were only 31 responses.
[c]Expected based on 30 records from each of 88 departments for care of patients with AMI ($n = 27$), hip fracture ($n = 29$) and stroke ($n = 32$).

Table 4 Associations between PMOS and hospital/department indices

Index	AMI			Hip fracture			Stroke		
	n	Beta (95% C.I.)	*P*-value	*n*	Beta (95% C.I.)	*P*-value	*n*	Beta (95% C.I.)	*P*-value
QMSI[a]	332	0.016 (−0.013, 0.046)	0.270	253	0.007 (−0.027, 0.041)	0.684	272	−0.005 (−0.039, 0.030)	0.780
QMCI[a,b]	332	−0.050 (−0.233, 0.134)	0.574	253	−0.021 (−0.121, 0.078)	0.660	272	−0.033 (−0.239, 0.172)	0.736
CQII[a,b,c]	332	−0.014 (−0.050, 0.021)	0.423	253	0.016 (−0.023, 0.055)	0.399	272	0.011 (−0.034, 0.056)	0.601
SER ED[a,b,c,d]	332	0.052 (−0.041, 0.145)	0.274	253	−0.039 (−0.125, 0.047)	0.348	272	0.195 (0.089, 0.300)	0.001*
EBOP ED[a,b,c,d]	332	0.051 (−0.194, 0.297)	0.682	253	−0.072 (−0.176, 0.032)	0.158	272	0.065 (−0.141, 0.271)	0.503
PSS ED[a,b,c,d]	332	−0.009 (−0.244, 0.225)	0.937	253	−0.102 (−0.470, 0.267)	0.566	272	−0.242 (−0.636, 0.152)	0.199
CR ED[a,b,c,d]	332	0.044 (−0.011, 0.099)	0.120	253	−0.130 (−0.205, −.054)	0.002*	272	0.052 (−0.023, 0.127)	0.156
SER[a,b,c,d,e]	332	0.005 (−0.107, 0.116)	0.935	253	0.021 (−0.085, 0.127)	0.678	272	−0.002 (−0.152, 0.147)	0.975
EBOP[a,b,c,d,f]	332	0.048 (−0.077, 0.173)	0.455	253	0.051 (−0.070, 0.172)	0.377	272	0.110 (−0.017, 0.236)	0.083
PSS[a,b,c,d,g]	332	0.118 (−0.317, 0.553)	0.595	253	−0.097 (−0.458, 0.265)	0.579	272	−0.195 (−0.591, 0.200)	0.303
CR[a,b,c,d,h]	332	−0.015 (−0.097, 0.067)	0.721	253	0.008 (−0.076, 0.092)	0.848	272	−0.017 (−0.141, 0.107)	0.780

Legend: QMSI (Index 0–24), QMCI (Index 0–8), CQII (Index 0–28), SER (Index 0–4), EBOP (Index 0–4), PSS (Index 0–4), CR (Index 0–4), PMOS (1–5).

[a]Adjusted for random effect (hospital) and fixed effects (hospital peer group, patient sex and age).
[b]Additionally adjusted for fixed effect: QMSI.
[c]Additionally adjusted for fixed effect: QMCI.
[d]Additionally adjusted for fixed effect: CQII.
[e]Additionally adjusted for fixed effect: SER ED.
[f]Additionally adjusted for fixed effect: EBOP ED.
[g]Additionally adjusted for fixed effect: PSS ED.
[h]Additionally adjusted for fixed effect: CR ED.
*Statistically significant at 5%.

Characteristics of patients who completed PMOS, whose medical records were reviewed, or who contributed data to AIHW outcome analyses are summarized in Supplementary eTables D3–D5. Descriptive statistics of patient-level measures (PMOS, record review, AIHW outcomes) can be found in Supplementary eTables D6–D9.

Results of the association for each patient-level measure can be found in Tables 4–8. Each table reports 11 comparisons for each of the three conditions (i.e. 33 in total). Across the five patient-level measures, there were 19 statistically significant associations (19/165 = 11.5%, over twice that expected by chance): two for PMOS, five for guideline adherence; three for mortality; four for mortality or readmission and five for LOS. The statistically significant results do not appear to follow a coherent pattern. For example, higher scores for ED clinical review processes are associated with 'lower' clinical practice guidelines (CPG) adherence for AMI ($P = 0.01$) and 'higher' levels of CPG adherence for hip fracture ($P = 0.02$) and stroke ($P = 0.09$; ns). LOS revealed five statistically significant findings, three of which indicated that higher levels of development of systems or strategies were associated with 'longer' LOS.

Treating shorter LOS as a positive outcome (e.g. due to lower complication rates), we also looked at the pattern of positive and negative outcomes associated with higher index scores: of 164 outcomes, fewer than half (79/164 = 48.2%) were positive. Twelve of the 79 positive associations (15.2%) were statistically significant, and seven of the 85 negative relationships (8.2%) were statistically significant.

Excluding LOS as an outcome, we found that 10 of 67 positive associations (14.9%) and 4 of 65 negative associations (6.2%) were statistically significant. For hospital-level systems, two of 18 positive (11.1%) and one of 18 negative associations (5.6%) were significant; for department-level strategies—ED or condition-specific—after adjusting for hospital-level associations, eight of 49 positive (16.3%) and three of 47 negative associations (6.4%) were statistically significant.

Discussion

In regard to the three overarching research questions provided at the outset, the DUQuA results do not reveal coherent causal relations between hospital-level quality management systems, department or ED quality strategies and the studied patient-level outcomes. This means, on the basis of this overarching study, that we cannot assume there is one best way of organizing for quality care or that one level of managing care (hospital wide, department-level or ED-level) contributes more or less to patient outcomes. This is an important point: often, improvement agents assume that change at policy- [42], meso- [43] or microsystem-level [44] is what is needed to improve the quality of care. We cannot accept this to be the case according to our findings. Consequently, we are not able make recommendations about which way to proceed in organizing or managing care to deliver better outcomes according to these results. It is more persuasive to say that multi-pronged strategies are needed in dealing with the issue of improving care quality in complex settings [45].

Nevertheless, DUQuA has contributed to our knowledge of hospital quality improvement activities in multiple ways. It examined, across 32 of the largest hospitals in one country, structures and overarching quality improvement mechanisms against measures of organizational and patient outcomes. The project adapted and fine-tuned indices for the examination of systems of care, compliance against requirements and improvement interventions at hospital level and applied four indices to explore the effect of quality strategies at department level, separately in EDs and in the inpatient departments in which staff cared for patients with our target conditions. Basing DUQuA on the DUQuE research design but advancing some of its aims has (i) allowed for the modification and testing of the

Table 5 Associations between CPG adherence and hospital/department indices

Index	AMI			Hip fracture			Stroke		
	n	Beta (95% C.I.)	*P*-value	*n*	Beta (95% C.I.)	*P*-value	*n*	Beta (95% C.I.)	*P*-value
QMSI[a]	5552	−0.031 (−0.076, 0.015)	0.186	7549	0.007 (−0.031, 0.045)	0.709	8259	−0.010 (−0.063, 0.042)	0.700
QMCI[a,b]	5552	0.086 (−0.114, 0.287)	0.398	7549	0.092 (−0.004, 0.189)	0.061	8259	−0.005 (−0.131, 0.120)	0.934
CQII[a,b,c]	5552	−0.050 (−0.096, −.003)	0.037*	7549	0.014 (−0.026, 0.054)	0.487	8259	−0.006 (−0.063, 0.051)	0.836
SER ED[a,b,c,d]	5552	−0.110 (−0.250, 0.030)	0.123	7549	0.036 (−0.052, 0.124)	0.419	8259	−0.008 (−0.187, 0.171)	0.931
EBOP ED[a,b,c,d]	5552	−0.306 (−0.616, 0.003)	0.052	7549	0.091 (−0.011, 0.194)	0.080	8259	0.416 (0.164, 0.668)	0.001*
PSS ED[a,b,c,d]	5552	0.082 (−0.295, 0.458)	0.670	7549	0.134 (−0.168, 0.437)	0.384	8259	−0.037 (−0.420, 0.346)	0.851
CR ED[a,b,c,d]	5552	−0.113 (−0.199, −.027)	0.010*	7549	0.105 (0.018, 0.192)	0.018*	8259	0.087 (−0.012, 0.187)	0.086
SER[a,b,c,d,e]	5552	0.076 (−0.079, 0.231)	0.335	7549	0.104 (0.018, 0.191)	0.018*	7961	−0.050 (−0.333, 0.234)	0.731
EBOP[a,b,c,d,f]	5552	−0.115 (−0.275, 0.044)	0.156	7549	−0.021 (−0.151, 0.108)	0.746	7961	0.067 (−0.110, 0.244)	0.461
PSS[a,b,c,d,g]	5552	0.270 (−0.275, 0.816)	0.332	7549	0.270 (−0.051, 0.592)	0.100	7961	−0.363 (−0.821, 0.095)	0.120
CR[a,b,c,d,h]	5552	0.056 (−0.058, 0.170)	0.335	7549	0.059 (−0.028, 0.146)	0.184	7961	−0.019 (−0.176, 0.139)	0.814

Legend: QMSI (Index 0–24), QMCI (Index 0–8), CQII (Index 0–28), SER (Index 0–4), EBOP (Index 0–4), PSS (Index 0–4), CR (Index 0–4), *n* = number of indicators assessed.

[a]Adjusted for random effect (hospital) and fixed effects (hospital peer group, patient sex and age group).
[b]Additionally adjusted for fixed effect: QMSI.
[c]Additionally adjusted for fixed effect: QMCI.
[d]Additionally adjusted for fixed effect: CQII.
[e]Additionally adjusted for fixed effect: SER ED.
[f]Additionally adjusted for fixed effect: EBOP ED.
[g]Additionally adjusted for fixed effect: PSS ED.
[h]Additionally adjusted for fixed effect: CR ED.
*Statistically significant at 5%.

Table 6 Associations between in-hospital mortality and hospital/department indices

Index	AMI			Hip fracture			Stroke		
	n	Beta (95% C.I.)	*P*-value	*n*	Beta (95% C.I.)	*P*-value	*n*	Beta (95% C.I.)	*P*-value
QMSI[a]	8299	−0.012 (−0.059, 0.034)	0.598	2122	0.035 (−0.027, 0.098)	0.264	4039	0.043 (−0.001, 0.088)	0.056
QMCI[a,b]	8299	0.014 (−0.122, 0.150)	0.839	2122	0.066 (−0.119, 0.251)	0.485	4039	−0.038 (−0.170, 0.093)	0.569
CQII[a,b,c]	8299	0.046 (−0.003, 0.095)	0.064	2122	−0.055 (−0.108, −.003)	0.039*	4039	−0.024 (−0.068, 0.019)	0.272
SER ED[a,b,c,d]	8239	0.064 (−0.076, 0.205)	0.369	2119	−0.008 (−0.131, 0.116)	0.903	4039	0.088 (−0.053, 0.228)	0.222
EBOP ED[a,b,c,d]	8239	−0.021 (−0.322, 0.280)	0.891	2119	0.047 (−0.091, 0.185)	0.501	4039	0.029 (−0.209, 0.268)	0.809
PSS ED[a,b,c,d]	8299	0.085 (−0.250, 0.421)	0.619	2122	0.318 (−0.118, 0.754)	0.153	4039	−0.305 (−0.586, −.024)	0.034*
CR ED[a,b,c,d]	8239	0.026 (−0.063, 0.116)	0.562	2119	0.001 (−0.129, 0.131)	0.993	4039	−0.015 (−0.101, 0.072)	0.737
SER[a,b,c,d,e]	7793	0.152 (−0.023, 0.326)	0.088	2108	−0.069 (−0.208, 0.070)	0.332	3968	−0.157 (−0.337, 0.024)	0.089
EBOP[a,b,c,d,f]	7793	0.113 (−0.100, 0.325)	0.300	2108	−0.168 (−0.351, 0.014)	0.071	3968	−0.185 (−0.341, −.029)	0.020*
PSS[a,b,c,d,g]	7793	0.553 (−0.022, 1.129)	0.060	2108	−0.162 (−0.617, 0.292)	0.484	3968	−0.245 (−0.571, 0.081)	0.140
CR[a,b,c,d,h]	7793	0.058 (−0.080, 0.197)	0.406	2108	0.071 (−0.082, 0.223)	0.363	3968	−0.031 (−0.151, 0.088)	0.607

Legend: QMSI (Index 0–24), QMCI (Index 0–8), CQII (Index 0–28), SER (Index 0–4), EBOP (Index 0–4), PSS (Index 0–4), CR (Index 0–4), *n* = number of admissions assessed.

[a]Adjusted for random effect (hospital) and fixed effects (hospital peer group, patient sex and age group).
[b]Additionally adjusted for fixed effect: QMSI.
[c]Additionally adjusted for fixed effect: QMCI.
[d]Additionally adjusted for fixed effect: CQII.
[e]Additionally adjusted for fixed effect: SER ED.
[f]Additionally adjusted for fixed effect: EBOP ED.
[g]Additionally adjusted for fixed effect: PSS ED.
[h]Additionally adjusted for fixed effect: CR ED.
*Statistically significant at 5%.

conceptual model over half a decade later, (ii) facilitated qualitative and quantitative evaluation of the original model and (iii) helped extend its reach into a new design while respecting its overarching strategy and features.

Further, benchmark reports sent to each of the participating hospitals were designed to stimulate quality improvement activity, making this more than simply a cross-sectional research project but a feedback strategy, applying action research techniques. With the

Table 7 Associations between 28-day readmission or in-hospital mortality and hospital/department indices

Index	AMI			Hip fracture			Stroke		
	n	Beta (95% C.I.)	*P*-value	*n*	Beta (95% C.I.)	*P*-value	*n*	Beta (95% C.I.)	*P*-value
QMSI[a]	8089	−0.004 (−0.026, 0.018)	0.736	2103	−0.009 (−0.057, 0.039)	0.718	3924	0.015 (−0.025, 0.056)	0.455
QMCI[a,b]	8089	−0.026 (−0.095, 0.044)	0.465	2103	0.104 (−0.037, 0.244)	0.148	3924	0.005 (−0.116, 0.125)	0.936
CQII[a,b,c]	8089	−0.007 (−0.032, 0.018)	0.567	2103	−0.035 (−0.078, 0.007)	0.100	3924	−0.044 (−0.084, −.004)	0.030*
SER ED[a,b,c,d]	8030	−0.034 (−0.108, 0.041)	0.372	2100	0.062 (−0.032, 0.157)	0.195	3924	0.106 (−0.017, 0.230)	0.092
EBOP ED[a,b,c,d]	8030	0.056 (−0.103, 0.214)	0.490	2100	0.038 (−0.066, 0.141)	0.477	3924	−0.008 (−0.220, 0.205)	0.943
PSS ED[a,b,c,d]	8089	−0.063 (−0.244, 0.118)	0.497	2103	0.029 (−0.303, 0.361)	0.864	3924	−0.284 (−0.542, −.025)	0.031*
CR ED[a,b,c,d]	8030	0.012 (−0.034, 0.058)	0.604	2100	0.004 (−0.095, 0.102)	0.943	3924	−0.004 (−0.082, 0.075)	0.929
SER[a,b,c,d,e]	7586	0.022 (−0.056, 0.100)	0.581	2089	−0.004 (−0.114, 0.106)	0.944	3854	−0.109 (−0.278, 0.060)	0.208
EBOP[a,b,c,d,f]	7586	0.008 (−0.086, 0.102)	0.867	2089	−0.196 (−0.336, −.056)	0.006*	3854	−0.111 (−0.260, 0.038)	0.146
PSS[a,b,c,d,g]	7586	0.331 (0.062, 0.599)	0.016*	2089	0.151 (−0.198, 0.501)	0.396	3854	−0.185 (−0.514, 0.144)	0.270
CR[a,b,c,d,h]	7586	0.032 (−0.023, 0.087)	0.257	2089	0.047 (−0.071, 0.165)	0.436	3854	−0.025 (−0.136, 0.087)	0.664

Legend: QMSI (Index 0–24), QMCI (Index 0–8), CQII (Index 0–28), SER (Index 0–4), EBOP (Index 0–4), PSS (Index 0–4), CR (Index 0–4), *n* = number of admissions assessed.

[a]Adjusted for random effect (hospital) and fixed effects (hospital peer group, patient sex and age group).
[b]Additionally adjusted for fixed effect: QMSI.
[c]Additionally adjusted for fixed effect: QMCI.
[d]Additionally adjusted for fixed effect: CQII.
[e]Additionally adjusted for fixed effect: SER ED.
[f]Additionally adjusted for fixed effect: EBOP ED.
[g]Additionally adjusted for fixed effect: PSS ED.
[h]Additionally adjusted for fixed effect: CR ED.
*Statistically significant at 5%.

Table 8 Associations between LOS and hospital/department indices

Index	AMI			Hip fracture			Stroke		
	n	Beta (95% C.I.)	*P*-value	*n*	Beta (95% C.I.)	*P*-value	*n*	Beta (95% C.I.)	*P*-value
QMSI[a]	8299	−0.012 (−0.030, 0.007)	0.222	2122	−0.026 (−0.051, −.002)	0.035*	4039	0.010 (−0.010, 0.031)	0.321
QMCI[a,b]	8299	−0.026 (−0.079, 0.027)	0.339	2122	−0.013 (−0.081, 0.055)	0.707	4039	−0.030 (−0.088, 0.028)	0.305
CQII[a,b,c]	8299	0.024 (0.003, 0.045)	0.022*	2122	0.016 (−0.012, 0.043)	0.260	4039	0.014 (−0.009, 0.038)	0.221
SER ED[a,b,c,d]	8239	0.068 (0.015, 0.121)	0.012*	2119	−0.057 (−0.110, −.003)	0.038*	4039	−0.027 (−0.096, 0.042)	0.439
EBOP ED[a,b,c,d]	8239	0.172 (0.063, 0.282)	0.002*	2119	−0.015 (−0.080, 0.049)	0.648	4039	−0.039 (−0.149, 0.071)	0.490
PSS ED[a,b,c,d]	8299	0.037 (−0.102, 0.176)	0.602	2122	0.197 (−0.010, 0.404)	0.062	4039	0.083 (−0.069, 0.235)	0.284
CR ED[a,b,c,d]	8239	0.034 (−0.001, 0.069)	0.055	2119	0.004 (−0.059, 0.066)	0.907	4039	0.000 (−0.044, 0.044)	0.998
SER[a,b,c,d,e]	7793	0.039 (−0.029, 0.108)	0.261	2108	0.010 (−0.046, 0.066)	0.726	3968	0.054 (−0.039, 0.146)	0.256
EBOP[a,b,c,d,f]	7793	0.009 (−0.063, 0.081)	0.811	2108	0.067 (−0.011, 0.145)	0.094	3968	0.010 (−0.071, 0.092)	0.804
PSS[a,b,c,d,g]	7793	0.161 (−0.082, 0.403)	0.194	2108	−0.095 (−0.299, 0.109)	0.363	3968	−0.031 (−0.225, 0.163)	0.753
CR[a,b,c,d,h]	7793	0.013 (−0.042, 0.069)	0.636	2108	0.026 (−0.038, 0.090)	0.422	3968	−0.017 (−0.074, 0.041)	0.575

Legend: QMSI (Index 0–24), QMCI (Index 0–8), CQII (Index 0–28), SER (Index 0–4), EBOP (Index 0–4), PSS (Index 0–4), CR (Index 0–4), *n* = number of admissions assessed.

[a]Adjusted for random effect (hospital) and fixed effects (hospital peer group, sex and age group).
[b]Additionally adjusted for fixed effect: QMSI.
[c]Additionally adjusted for fixed effect: QMCI.
[d]Additionally adjusted for fixed effect: CQII.
[e]Additionally adjusted for fixed effect: SER ED.
[f]Additionally adjusted for fixed effect: EBOP ED.
[g]Additionally adjusted for fixed effect: PSS ED.
[h]Additionally adjusted for fixed effect: CR ED.
*Statistically significant at 5%.

completion of the project, the DUQuA team now offers validated tools, new measures, re-confirmation of the overall DUQuE methodology, variations on the original design and data and indices against which future research projects can compare.

Of 165 relationships examined, 19 (11.5%) were statistically significant; as this is over twice the number expected by chance, there is some suggestion that there is a mixture of random and systematic differences. Going further, we therefore examined the data for patterns, without restricting them to statistically significant results; we did not find consistent effects for any of the hospital systems or ED or department-level strategies across the target conditions. *Post hoc* descriptive examination of the constellation of results, however,

suggests a pattern for a higher proportion of statistically significant positive associations to be associated with department-level strategies versus hospital systems, which could be quantitatively explored in future studies with larger numbers of hospitals and qualitatively.

An analysis of a subset of 43 Spanish hospitals that participated in the pan-European Methods of Assessing Response to Quality Improvement Strategies study looked at the relationship between hospitals' quality improvement system maturity, as measured by an index, and a variety of patient outcomes. That study found no statistically significant relationships between the hospital's quality maturity index and risk-adjusted mortality, readmission or LOS [46]. The appropriateness of mortality as an endpoint has been questioned [47], and the difficulty of interpreting LOS has been noted [46]. In the DUQuA study, these issues are compounded by the lack of detailed data to permit patient-level risk adjustment.

The DUQuA study additionally did not find a consistent pattern of association between hospital-level systems and ED and department-level strategies with either compliance with clinical treatment processes indicators or the PMOS. We are unaware of any previous study that examined the impact of these factors on patient measures of safety. European colleagues in DUQuE, however, applying similar methods, found that organization-level quality systems had weak or limited impacts on guideline adherence while department-level strategies had strong effects for several individual indicators in both AMI and stroke (but not hip fracture), concluding that front-line interventions may be more effective [18]. DUQuA looked at a larger number of indicators (11 versus 6 for AMI, 12 versus 4 for hip fracture and 13 versus 5 for stroke), most of which were different to that assessed in DUQuE, and evaluated overall adherence averaged across all indicators for each condition; it is plausible that if examined in much more detail, indicator by indicator, relationships may be revealed.

Of all relationships studied, the most obvious link is between clinical review strategies and the medical record audit for compliance with guidelines (Table 5). No statistically significant results were found for department-level clinical review strategies for each of the three conditions; significant and borderline relationships were found for ED clinical review strategies but these results were in contradictory directions: higher scores for ED clinical review strategies were associated with *lower* CPG adherence for AMI ($P = 0.01$) and *higher* levels of CPG adherence for hip fracture ($P = 0.02$) and stroke ($P = 0.09$; ns). In DUQuE, associations between clinical review strategies and higher adherence were found for two of the AMI indicators, but not for any of the hip fracture or stroke indicators [18].

As to strengths, the DUQuA project enabled large amounts of data on different factors important for quality improvement to be gathered, assessed and mapped against the conceptual framework. It offers tested measures and indices and a platform for future research into hospital quality of care. Limitations to studies such as these generally, and DUQuA specifically, include the challenge of finding sufficient resources to power all aspects of the study. The cross-sectional nature of the design precludes understanding of care processes longitudinally. Sampling is also a perennial difficulty in studies such as these, and the struggle of enrolling hospitals and keeping them in the study for the duration of the ethics and data-gathering phases is particularly challenging. Added to this, there are systematic biases attributable to the quality of data available within the time-frames afforded by available resources, and differences in hospital structures, processes, governance, arrangements and policies, all of which may introduce biases and challenges to study validity and generalizability. For example, the timing of QMSI data collection varied by hospital (from September 2015 to July 2017), depending on the timing of completion of approvals, while the most recent available AIHW patient outcome data were for the period 1 July 2014 to 30 June 2015; while it can be argued that most hospital systems and department-level strategies are generally consistent across these timeframes, it is likely that some measurement error is introduced by these timing differences. Nevertheless, despite multiple limitations, we took steps at every opportunity across the protocol formulation, research development, model refinement, implementation, design and research reporting phases to optimize the robustness and fidelity of our study and ameliorate design, practice and analytic limitations.

Conclusion

In an ideal world data about care processes, pathways and outcomes at hospital, department and patient levels across multiple conditions would be gathered routinely and robustly—as part of formal integrated research projects—and reported on an ongoing basis. This longitudinal dream is currently not a reality, necessitating cross-sectional research such as this, on top of ongoing in-house quality improvement activities. It is our hope that, as research in quality improvement matures and studies such as this sharpen the appetite for ongoing evaluation and assessment of care quality, our longitudinal aspirations will be realized.

Strengths and limitations of this study

- This is an in-depth study examining quality management systems and patient-level measures in a large sample of hospitals, substantially advancing earlier work.
- We consulted widely with Australian and international stakeholders to operationalize definitions and measures in acute myocardial infection (AMI), hip fracture and stroke.
- We found no consistent pattern of results, implying there is no one gateway to improvement efforts.
- The study informs the design of future interventions and quality management systems in Australia and internationally.
- Key challenges for work of this kind include the expensive nature of such studies, sufficient sample sizes, the recruitment of hospitals, complex ethical, governance and site approval processes and attrition of recruited hospitals during the approval process.

Supplementary material

Supplementary material is available at *INTQHC Journal* online.

Acknowledgements

Members of the original DUQuE team (Oliver Gröne, Rosa Suñol) provided extensive input and advice in the initial stages of the project, supporting design modifications from DUQuE and the planned DUQuA study approach. Dr Annette Pantle provided expert advice on revision and development of DUQuA measures, and Professor Sandy Middleton, Associate Professor Dominique Cadillac, Kelvin Hill, Dr Carmel Crock, and Professor Jacqueline Close provided input into the development of our AMI, hip fracture and stroke clinical process indicator lists. Nicole Mealing and Victoria Pye provided statistical advice in the initial phases of the project. We greatly appreciate their efforts.

Funding

National Health and Medical Research Council (NHMRC) Program Grant APP1054146 (CI Braithwaite).

Contributors

The research team consists of experienced researchers, clinicians, biostatisticians and project managers with expertise in health services research, survey design, and validation, large-scale research and project management, sophisticated statistical analysis, quality improvement and assessment, accreditation, clinical indicators, policy and patient experience. JB conceived the idea to do an Australian in-depth study of quality, led the research grant to fund the project, chairs the steering committee and led the development of the manuscript. NT and RCW co-led the detailed study design, managed the project across time and contributed to the development of the manuscript. HPT and GA provided statistical expertise for the study design and developed the analysis plan for the manuscript. TW, EH, ZL, AS, PH and MW contributed to the logistics of project management, the refinement of measures and the development of the manuscript.

Data sharing statement

Data will be made publicly available to the extent that individual participants or participating hospitals cannot be identified, in accordance with requirements of the approving Human Research Ethics Committees.

References

1. Runciman WB, Hunt TD, Hannaford NA *et al*. CareTrack: assessing the appropriateness of healthcare delivery in Australia. *Med J Aust* 2012;**197**:100–5.
2. McGlynn EA, Asch SM, Adams J *et al*. The quality of health care delivered to adults in the United States. *N Engl J Med* 2003;**348**:2635–45.
3. Duckers M, Faber M, Cruijsberg J *et al*. Safety and risk management interventions in hospitals: a systematic review of the literature. *Med Care Res Rev* 2009;**66**:90S–119S.
4. Hollnagel E, Braithwaite J, Wears RL (eds). *Resilient Health Care*. Surrey, UK, Ashgate Publishing Ltd., 2013.
5. Braithwaite J, Matsuyama Y, Mannion R, Johnson J (eds). *Healthcare Reform, Quality and Safety: Perspectives, Participants, Partnerships and Prospects in 30 Countries*. Surrey, UK, Ashgate Publishing Ltd., 2015.
6. Taylor N, Clay-Williams R, Hogden E *et al*. High performing hospitals: a qualitative systematic review of associated factors and practical strategies for improvement. *BMC Health Serv Res* 2015;**15**:244.
7. Clay-Williams R, Nosrati H, Cunningham FC *et al*. Do large-scale hospital-and system-wide interventions improve patient outcomes: a systematic review. *BMC Health Serv Res* 2014;**14**:369.
8. The Research Priority Setting Working Group of the WHO World Alliance for Patient Safety. *Summary of the Evidence on Patient Safety: Implications for Research*. Geneva, CH, World Health Organization, 2008.
9. Groene O, Klazinga N, Wagner C *et al*. Investigating organizational quality improvement systems, patient empowerment, organizational culture, professional involvement and the quality of care in European hospitals: the 'Deepening our understanding of quality improvement in Europe (DUQuE)' project. *BMC Health Serv Res* 2010;**10**:281.
10. Secanell M, Groene O, Arah OA *et al*. Deepening our understanding of quality improvement in Europe (DUQuE): overview of a study of hospital quality management in seven countries. *Int J Qual Health Care* 2014;**26**:5–15.
11. Botje D, Klazinga NS, Suñol R *et al*. Is having quality as an item on the executive board agenda associated with the implementation of quality management systems in European hospitals: a quantitative analysis. *Int J Qual Health Care* 2014;**26**:92–9.
12. Greoene O, Kringos D, Sunol R, On Behalf of the DUQuE Project. *Seven Ways to Improve Quality and Safety in Hospitals. An Evidence-Based Guide*. DUQuE Collaboration, 2014.
13. Groene O, Arah OA, Klazinga NS *et al*. Patient experience shows little relationship with hospital quality management strategies. *PLoS One* 2015;**10**:e0131805.
14. Groene O, Sunol R. Patient involvement in quality management: rationale and current status. *J Health Organ Manag* 2015;**29**:556–69.
15. Wagner C, Groene O, Thompson CA *et al*. DUQuE quality management measures: associations between quality management at hospital and pathway levels. *Int J Qual Health Care* 2014;**26**:66–73.
16. Shaw CD, Groene O, Botje D *et al*. The effect of certification and accreditation on quality management in 4 clinical services in 73 European hospitals. *Int J Qual Health Care* 2014;**26**:100–7.
17. Wagner C, Thompson CA, Arah OA *et al*. A checklist for patient safety rounds at the care pathway level. *Int J Qual Health Care* 2014;**26**:36–46.
18. Sunol R, Wagner C, Arah OA *et al*. Implementation of departmental quality strategies is positively associated with clinical practice: results of a multicenter study in 73 hospitals in 7 european countries. *PLoS One* 2015;**10**:e0141157.
19. Groene O, Sunol R, Klazinga NS *et al*. Involvement of patients or their representatives in quality management functions in EU hospitals: implementation and impact on patient-centred care strategies. *Int J Qual Health Care* 2014;**26**:81–91.
20. Wagner C, Mannion R, Hammer A *et al*. The associations between organizational culture, organizational structure and quality management in European hospitals. *Int J Qual Health Care* 2014;**26**:74–80.
21. Haggerty JL, Reid RJ, Freeman GK *et al*. Continuity of care: a multidisciplinary review. *BMJ* 2003;**327**:1219–21.
22. Kyriacou DN, Ricketts V, Dyne PL *et al*. A 5-year time study analysis of emergency department patient care efficiency. *Ann Emerg Med* 1999;**34**:326–35.
23. Tsai C-L, Magid DJ, Sullivan AF *et al*. Quality of care for acute myocardial infarction in 58 U.S. emergency departments. *Acad Emerg Med* 2010;**17**:940–50.
24. Taylor N, Clay-Williams R, Hogden E *et al*. Deepening our understanding of quality in Australia (DUQuA): a study protocol for a nationwide, multilevel analysis of relationships between hospital quality management systems and patient factors. *BMJ Open* 2015;**5**:e010349.
25. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;**10**:37–48.
26. Clay-Williams R, Taylor N, Ting HP *et al*. Do quality management systems influence clinical safety culture and leadership? A study in 32 Australian hospitals. *Int J Qual Health Care* 2020;**32**:60–66.
27. Braithwaite J, Clay-Williams R, Taylor N. Potential solutions to improve the governance of multicentre health services research. *Med J Aust* 2018;**209**:187–8.
28. Clay-Williams R, Taylor N, Braithwaite J. Potential solutions to improve the governance of multicentre health services research. *Med J Aust* 2018;**208**:152–4.
29. Clay-Williams R, Taylor N, Ting HP *et al*. The relationships between quality management systems, safety culture and leadership and patient outcomes in Australian Emergency Departments. *Int J Qual Health Care* 2020;**32**:43–51.
30. Wagner C, Groene O, Thompson CA *et al*. Development and validation of an index to assess hospital quality management systems. *Int J Qual Health Care* 2014;**26**:16–26.
31. Clay-Williams R, Taylor N, Winata T *et al*. Organization quality systems and department-level strategies: refinement of the Deepening our Understanding in Quality in Australia (DUQuA) organization and department-level scales. *Int J Qual Health Care* 2020;**32**:22–34.
32. Australian Commission on Safety and Quality in Health Care. *Indicator Specification: Acute Stroke Clinical Care Standard*. Sydney, AU, Australian Commission on Safety and Quality in Health Care, 2015.
33. National Stroke Foundation. *Clinical Guidelines for Stroke Management*. Melbourne, AU, 2010.
34. National Stroke Foundation. *National Stroke Audit—Acute Services Organisational Survey Report*. Melbourne, AU, 2013.
35. Australian Commission on Safety and Quality in Health Care. *Indicator Specification: Acute Coronary Syndromes Clinical Care Standard*. Sydney, AU, Australian Commission on Safety and Quality in Health Care, 2014.
36. Australian and New Zealand Hip Fracture Registry (ANZHFR) Steering Group. *Australian and New Zealand Guideline for Hip Fracture Care: Improving Outcomes in Hip Fracture Management of Adults*. Sydney,

Australian and New Zealand Hip Fracture Registry Steering Group, 2014.
37. McEachan RRC, Lawton RJ, O'Hara JK *et al*. Developing a reliable and valid patient measure of safety in hospitals (PMOS): a validation study. *BMJ Qual Saf* 2013;**23**:565–73.
38. Giles SJ, Lawton RJ, Din I, McEachan RRC. Developing a patient measure of safety (PMOS). *BMJ Qual Saf* 2013;**22**:554–62.
39. Taylor N, Hogden E, Clay-Williams R *et al*. Older, vulnerable patient view: a pilot and feasibility study of the patient measure of safety (PMOS) with patients in Australia. *BMJ Open* 2016;**6**:e011069.
40. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY, John Wiley & Sons, 2004.
41. Arnolda G, Winata T, Ting HP *et al*. Implementation and data-related challenges in the Deepening our Understanding of Quality in Australia (DUQuA) study: implications for large-scale cross-sectional research. *Int J Qual Health Care* 2020;**32**:75–83.
42. Bogh SB, Falstie-Jensen AM, Hollnagel E *et al*. Improvement in quality of hospital care during accreditation: a nationwide stepped-wedge study. *Int J Qual Health Care* 2017;**28**:715–20.
43. Jha A, Epstein A. Hospital governance and the quality of care. *Health Aff* 2010;**29**:182–7.
44. Likosky DS. Clinical microsystems: a critical framework for crossing the quality chasm. *J Extra Corpor Technol* 2014;**46**:33–7.
45. Braithwaite J, Churruca K, Ellis LA *et al*. *Complexity Science in Healthcare—Aspirations, Approaches, Applications and Accomplishments. A White Paper*. Sydney, AU, Australian Institute of Health Innovation, Macquarie University, 2017.
46. Groene O, Mora N, Thompson A *et al*. Is the maturity of hospitals' quality improvement systems associated with measures of quality and patient safety? *BMC Health Serv Res* 2011;**11**:344.
47. Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ* 2010;**340**.

International Journal for Quality in Health Care, 2020, 32(S1), 22–34
doi: 10.1093/intqhc/mzz096
Article

Article

Organization quality systems and department-level strategies: refinement of the Deepening our Understanding in Quality in Australia (DUQuA) organization and department-level scales

ROBYN CLAY-WILLIAMS[1], NATALIE TAYLOR[2,3], TERESA WINATA[1], HSUEN P. TING[1], GASTON ARNOLDA[1], and JEFFREY BRAITHWAITE[1,]

[1]Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Road, NSW 2109, Australia, [2]Cancer Research Division, Cancer Council NSW, 153 Dowling St, Woolloomooloo, NSW 2011, Australia, and [3]Faculty of Health Sciences, University of Sydney, Camperdown, Sydney, NSW 2006, Australia

Address reprint requests to: Robyn Clay-Williams, Centre for Healthcare Resilience and Implementation Science (CHRIS), Australian Institute of Health Innovation, Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia. Tel: +61 2 9850 2438; E-mail: robyn.clay-williams@mq.edu.au

Editorial Decision 7 September 2019; Accepted 13 September 2019

Abstract

Objective: The aim of this study was to develop and refine indices to measure organization and care pathway-level quality management systems in Australian hospitals.
Design: A questionnaire survey and audit tools were derived from instruments validated as part of the Deepening Our Understanding of Quality improvement in Europe (DUQuE) study, adapted for Australian hospitals through expert opinion. Statistical processes were used to explore the factor structure, reliability and non-redundancy and descriptive statistics of the scales.
Setting: Thirty-two large Australian public hospitals.
Participants: Audit of quality management processes at organization-level and care pathway processes at department level for three patient conditions (acute myocardial infarction (AMI), hip fracture and stroke) and senior quality manager, at each of the 32 participating hospitals.
Main Outcome Measure(s): The degree of quality management evident at organization and care pathway levels.
Results: Analysis yielded seven quality systems and strategies scales. The three hospital-level measures were: the Quality Management Systems Index (QMSI), the Quality Management Compliance Index (QMCI) and the Clinical Quality Implementation Index (CQII). The four department-level measures were: Specialised Expertise and Responsibility (SER), Evidence-Based Organisation of Pathways (EBOP), Patient Safety Strategies (PSS) and Clinical Review (CR). For QMCI, and for seven out of eight subscales in QMSI, adequate internal consistency (Cronbach's α >0.8) was achieved. For CQII, lack of variation and ceiling effects in the data resulted in very low internal consistency scores, but items were retained for theoretical reasons. Internal consistency was high for CR (Cronbach's α 0.74–0.88 across the three conditions), and this was supported by all item-total correlations exceeding the desired threshold. For EBOP, Cronbach's α was acceptable for hip fracture (0.80) and stroke (0.76), but only moderate for AMI (0.52). PSS and SER scales were retained for theoretical reasons, although internal consistencies were only moderate (SER) to poor (PSS).

Conclusions: The Deepening our Understanding of Quality in Australia (DUQuA) organization and department scales can be used by Australian hospital managers to assess and measure improvement in quality management at organization and department levels within their hospitals and are readily modifiable for other health systems depending on their needs.

Key words: hospital quality management systems, multi-level research, patient level factors, patient safety, hospital performance, quality improvement

Introduction

The 'Deepening our Understanding of Quality in Australia' (DUQuA) study [1] was a 5-year Australia-wide, multi-level, multi-million dollar cross-sectional study aiming to identify how quality management systems (QMS), leadership and culture in Australian hospitals are related to care delivery and patient outcomes for acute myocardial infarction (AMI), stroke and hip fracture. Based on the original, landmark European Union-funded 'Deepening our Understanding of Quality improvement in Europe' (DUQuE) study [2], the rationale for undertaking DUQuA included potential for comparison with the large European sample of 188 hospitals across seven countries, in addition to in-depth understanding of quality management in Australian acute settings. Evidence- or consensus-based measurement tools were designed or modified and then utilized to collect quantitative data on QMS at hospital and care pathway levels, department-level safety culture and leadership amongst clinicians, clinical treatment processes, patient outcomes and patient perceptions of safety. Collection methods included paper-based and electronic surveys, medical record reviews, external audits and accessing national datasets from the National Stroke Foundation registry and the Australian Institute of Health and Welfare. Linear and multilevel modelling was applied to the datasets to identify relationships between quality management, care delivery and patient outcomes. Findings have the potential to influence decision-making and improvements in quality and safety in Australian and international hospitals.

The aim of this study was to develop and validate indices to measure organization and care pathway-level QMS in Australian hospitals. The DUQuA measurement tools were derived from the DUQuE tools and, where possible, the DUQuE terminology retained. For example, the definition of QMS mirrors that of DUQuE: 'a set of interacting activities, methods and procedures used to monitor, control and improve the quality of care' [3]. The organization and department scales described are the same measures used by DUQuE for organization-level QMS, modified for the Australian context as described later in our 'Method' section. The three organization-level measures are the Quality Management Systems Index (QMSI), the Quality Management Compliance Index (QMCI) and the Clinical Quality Implementation Index (CQII). The four measures of quality management activities at department level focus on Specialised Expertise and Responsibility (SER), Evidence-Based Organisation of Pathways (EBOP), Patient Safety Strategies (PSS) and Clinical Review (CR).

This article presents information about the structure, reliability and non-redundancy and descriptive statistics of the QMSI, QMCI and CQII organization scales and the SER, EBOP, PSS and CR department scales. Relations between the organization and department scales and other outcomes are reported elsewhere in this Supplement [4].

Method

Development of the organization scales

The three organization-level measures from DUQuE (QMSI, QMCI and CQII) were modified by the DUQuA team to reflect the evolution of our condition-specific indicators and to ensure relevance to the Australian context, in consultation with national quality assessment experts, such as the Australian Council on Healthcare Standards. QMSI quantifies the managerial aspects of quality management that might influence the implementation of quality systems in hospitals. QMSI consists of 10 subscales: quality policy documents, hospital governance board activities, quality resources, quality management, evidence-based medicine protocols, preventive protocols, internal quality methods for general activities, personnel, clinical practice and patients. Each item is rated on a four-point Likert scale (range 1–4).

QMCI measures the managerial aspects of quality improvement in hospitals on three subscales: quality planning, monitoring patient and professional opinions and quality control and monitoring. CQII quantifies implementation of quality systems at hospital level including whether systems exist, to what extent implementation has been monitored and whether implementation is sustainable. CQII consisted of seven subscales: preventing and controlling healthcare-associated infections, medical safety, preventing patient falls, preventing pressure injuries, routine assessment and diagnostic testing of patients in elective surgery, safe surgery that includes an approved checklist and recognizing and responding to clinical deterioration in acute healthcare. For QMCI and CQII, items are rated on a five-point Likert scale (range 0–4).

The final DUQuA instruments are similar, but not identical, to the tools used in the DUQuE study. For example: some questions in the QMSI survey are worded differently to align with Australian terminology and processes, and a small number of additional questions have been included.

Development of the department scales

Following a similar process, the DUQuA team modified the DUQuE measures [5] (SER, EBOP, PSS and CR) for the Australian context. SER explores how clinical responsibilities were assigned for a particular condition; EBOP measures whether department processes, such as admission, acute care, rehabilitation and discharge, were organized to facilitate evidence-based care recommendations; PSS measures the use of clinical practice guidelines, and CR assesses the integration of audit and systematic monitoring with department quality management mechanisms. SER, CR and PSS used identical measures for the three conditions, while EBOP has the same structure for each condition-specific department (AMI, stroke and hip fracture) but the content varies to reflect the condition-specific evidence recommendation.

There are differences between the DUQuE and DUQuA department scales. As in the QMSI survey, some audit items are worded differently to align with Australian terminology and processes. Unlike DUQuE, the DUQuA project recognizes the Emergency Department (ED) as a common entry point to the hospital for the majority of AMI, hip fracture and stroke patients. ED condition-specific measures were developed for SER, EBOP and CR, to ensure relevance to the ED context; thus, for example, the ED assessment of SER comprised questions for each of AMI, hip fracture and stroke. For ED, the PSS items are all generic (i.e. not condition-specific).

Setting, participant recruitment and data collection

The DUQuA project commenced in February 2014 in 32 large Australian public hospitals, and data collection was completed by November 2017. In each state and territory, general public hospitals that met the following criteria were recruited: (i) with approximately 200 beds or more, (ii) with an ED and (iii) that regularly admit more than 30 each of stroke, AMI and hip fracture patients over a period of 3 months. A group of 78 hospitals were invited to participate, 70 that met all inclusion criteria and eight that were close to the specified thresholds or lacked at least one of the chosen departments; 62 hospitals initially agreed (89%). Due to factors associated with healthcare leadership changes, hospital relocations and obtaining ethics approval, a number of hospitals withdrew; however, 32 hospitals participated until completion of the study. The QMSI questionnaire was assessed by the senior person responsible for the coordination of quality improvement activities in each of the 32 participating hospitals.

Hospitals responding to an initial research invitation were provided with a formal letter detailing the study's background, data collection procedure and timeline. A Local Principal Investigator (LPI) was nominated by each hospital as the study point of contact. The role of the LPI was to maintain collaborative relationships between the hospital staff and the research team, to contact the quality manager to complete the QMSI questionnaire and to coordinate a site visit for accredited surveyors to conduct an external quality assessment (EQA) visit, during which the two organization-level measures (QMCI and CQII) and the four department-level measures (SER, EBOP, PSS and CR) were assessed. Details about surveyors' visits are reported elsewhere in this Supplement [4, Appendix A]. Questionnaires were completed anonymously, and EQA was conducted as a two-day site visit in the hospital, including the three condition-specific wards (AMI, hip fracture and stroke) and the ED. Data collection commenced in February 2016 and concluded in November 2017.

Statistical analysis

DUQuE applied the following rules in assessing reliability: Cronbach's $\alpha > 0.7$ [6] and item-total correlation coefficient > 0.4 [6] were considered acceptable evidence of internal consistency and consistency with the subscale construct, respectively, and Pearson's correlation coefficient < 0.7 between subscale scores was accepted as demonstrating non-redundancy of a subscale [5, 6]. DUQuA adopted the same rules, except that Cronbach's $\alpha > 0.8$ was deemed preferable, based on the work of Nunnally [7]. These thresholds were used to indicate a need for consideration of model re-specification for the three organization-level measures (QMSI, QMCI and CQII). The four department-level measures are single factor scales, so model re-specification was not considered, but we nevertheless report the test-values for internal reliability to inform interpretation of the measures. Further details on statistical analysis methods used for DUQuA have been published elsewhere in this Supplement [8]. Confirmatory Factor Analysis was not feasible because of the small sample size.

Data were analysed in SAS/STAT 9.4 (Cary, North Carolina, USA). Characteristics of the hospitals were summarized. We used a combination of theoretical grounds, item-total correlation and Cronbach's α, to refine the theoretical model.

After reaching the final model structure, each organization-level subscale was calculated as the mean of all non-missing items. If more than 50% of the items within the subscale were missing, the subscale was set to missing. We performed 100 multiple imputations, using Markov Chain Monte Carlo (MCMC) method [9], to impute missing organization subscales. After imputation, we calculated the organization scales by summing all subscales (for QMSI, eight was subtracted from the sum, in line with DUQuE procedures). To assess the degree of redundancy, Pearson's correlation coefficient was calculated between pairs of subscales within each scale; the analysis was repeated 100 times for each imputation and the SAS/STAT MIANALYZE procedure [10] was used to obtain the pooled parameter estimates and standard errors, if any imputed values were involved in the calculations. For department-level and ED-level scales, each department-level scale was the mean of all applicable non-missing items. If more than 50% of the applicable items within the scale were missing, the scale was set to missing.

Results

Characteristics of participants

Although 32 hospitals participated in DUQuA, two hospitals shared a quality manager and QMS, and therefore, 31 QMS questionnaire surveys were included for analysis. All 32 hospital provided audit data. The characteristics of participating hospitals are summarized in Supplementary eTable A1, Appendix A.

Organization scales—structure, reliability and non-redundancy

Missing data. For QMSI, there were no missing data at subscale level. For QMCI, the percentage of missing subscales was 1.6% and ranged from 0 to 3.1% per subscale, affecting one hospital only. For CQII, the percentage of missing subscales was 0.4% and ranged from 0 to 3.1% per subscale; again, only one hospital was affected.

Descriptive statistics for QMSI, QMCI and CQII subscales. The descriptive statistics for each subscale of each of the three organization-level measures are summarized in Table 1. Subscales for all three measures were left-skewed with all but one scale having a mean value above 3.0. For QMSI, the subscale mean ranged from 3.2 to 3.7; for QMCI, both subscales had a mean of 3.5, and for CQII 'Routine assessment and diagnostic testing of patients in elective surgery' had a mean of 2.3 with the remaining subscale means ranging from 3.3 to 3.9. Item-level summary statistics for each measure can be found at Supplementary eTables A2–4, Appendix A.

QMSI—structure, reliability and non-redundancy. From the original 10-factor theoretical model for QMSI, we merged item Q7.1 to 7.6 from the 'evidence-based medicine protocols' construct with item Q7.7 to 7.13 from the 'preventive protocols' construct to form a 13-item 'preventive protocols' construct. Items Q8.1 to 8.12 originally forming three 'internal quality methods' constructs (general

Table 1 Descriptive statistics of organization subscales

Scale	Subscale	*n*	Mean	SD	Median	Min	Max	IQR[a]
QMSI	Quality policy	31	3.2	0.90	3.3	1.3	4	1
	Hospital governance—board activities	31	3.4	0.53	3.6	2	4	0.8
	Quality resources	31	3.4	0.37	3.3	2.4	4	0.4
	Quality management	31	3.5	0.48	3.7	2	4	0.7
	Preventive protocols	31	3.6	0.40	3.7	2.5	4	0.6
	Internal quality methods—general activities	31	3.4	0.48	3.4	2	4	0.7
	Performance monitoring	31	3.3	0.70	3.3	1.7	4	1.3
	Internal quality methods—patients	31	3.7	0.42	4	3	4	0.5
QMCI	Monitoring patient and professional opinions	31	3.5	0.76	3.8	0.6	4	0.8
	Quality control and monitoring	32	3.5	0.63	3.8	0.8	4	0.7
CQII	Preventing and controlling healthcare-associated infections	32	3.9	0.20	4	3.2	4	0
	Medication safety	32	3.9	0.21	4	3.2	4	0.2
	Preventing patient falls	32	3.9	0.13	4	3.6	4	0.1
	Preventing pressure injuries	32	3.8	0.51	4	1.2	4	0.3
	Routine assessment and diagnostic testing of patients in elective surgery	32	2.3	1.56	2.5	0	4	3.3
	Safe surgery that includes an approved checklist	31	3.3	1.29	4	0	4	1
	Recognizing and responding to clinical deterioration in acute healthcare	32	3.9	0.18	4	3.2	4	0

[a]IQR, Interquartile Range.

activities, personnel and clinical practice) were restructured to 'internal quality methods—general' and 'performance monitoring'. The final eight-factor model is presented in Table 2, along with the Cronbach's α scores and item-total correlations. All subscales except 'internal quality methods—patients' achieved or nearly achieved internal consistency reliability (Cronbach's α >0.8). Twelve out of 51 items did not meet the desired threshold (>0.4) for item-total correlation, to demonstrate internal consistency within the subscale construct. Appendix A, Supplementary eTable A5, shows between subscale correlations for QMSI. Twenty-five pairs of correlations were within (0.16–0.69) the acceptable range of <0.7; the three scales with potential redundancy were 'Quality resources' (0.75), 'Quality management' (0.74) and 'Internal quality methods—general activities' (0.77), each of which correlated with 'Hospital governance—board activities'.

QMCI—structure, reliability and non-redundancy. From the original three-factor theoretical model for QMCI, we moved two items into the 'quality control and monitoring' subscale: item XQ01 from the single item 'quality planning' subscale and XQ08 from 'monitoring patient and professional opinions'. The final two-factor model can be found in Table 3. Both subscales showed adequate internal consistency through Cronbach's α without reaching our preferred threshold of 0.8. One item in each subscale had an item-total correlation <0.4, indicating less than ideal fit with the subscale construct. The between subscale correlation coefficient for QMCI (0.73) indicated non-redundancy.

CQII—structure, reliability and non-redundancy. We retained the original seven-factor theoretical model for CQII. The model can be found in Table 4 along with the Cronbach's α and item-total correlations. Only three factors ('preventing pressure injuries', 'routine assessment and diagnostic testing of patients in elective surgery', and 'Safe surgery that includes an approved checklist') achieved internal consistency through Cronbach's α and item-total correlations. The other four subscales and their constituent items failed to meet preferred thresholds. We explored removing items to improve Cronbach's α, but were unable to significantly improve results. This was not surprising, as the majority of hospitals scored the maximum plausible value of four for these items (the mean of each of these four subscales was 3.9). We decided to retain these items for theoretical reasons (presented in the Discussion); however, the limited range in scale scores meant that numerous item-total correlations were not calculable. Supplementary eTable A6 provides the intra-scale correlation coefficients for CQII. The correlation coefficient between the 'Medication safety' and 'Recognizing and responding to clinical deterioration in acute healthcare' subscales was 0.88, the only subscale pairing indicating potential redundancy.

Department scales—structure, reliability and non-redundancy

Missing data. There were 119 participating departments (27 AMI, 29 hip fracture, 32 stroke and 32 ED). For EBOP and CR, there were no missing data; for SER, missing items were restricted to hip fracture departments, with 1.1% of hip fracture items missing (ranging from 0 to 3.5% per item); and for PSS, the percentage of missing items was 0.4% (ranging from 0 to 3.8%) per item in AMI and 0.4% (ranging from 0 to 3.5%) per item in hip fracture. There were no departments with a scale missing.

Descriptive statistics for SER, EBOP, PSS and CR. Descriptive statistics for SER, EBOP, PSS and CR are summarized in Table 5 and for their items in Supplementary eTable A7. Items were left-skewed with all means calculated at two or above (in a 0–4 plausible range). Consistently across the three departments, PSS had the highest (3.4–3.6) and CR had the lowest (2.2–2.9) mean scores.

Descriptive statistics for ED-level scales. Descriptive statistics for ED-level SER, EBOP, PSS and CR are summarized in Table 6 and for their items in Supplementary eTable A8. Items were left-skewed with ceiling effect. The PSS, which was assessed using generic rather than condition-specific questions, had a mean score of 3.4. Across all three conditions, EBOP consistently had the highest mean scores (2.5–3.6) and CR had the lowest mean score (1.3–2.0); mean scores for hip fracture were lower than for either AMI or stroke.

Table 2 Cronbach's alpha and item-total correlation for QMSI

Subscale and items	Cronbach's α	Item-total correlation
Quality policy	0.805	
Q3.1 A documented quality policy approved by the hospital governance board		0.650
Q3.2 Quality improvement (QI) plan at hospital level (translation of the quality objectives into concrete activities and measures designed to realize the quality policy)		0.614
Q3.3 Report on evaluation of QI activities (focusing on routine quality and safety indicators, e.g. clinical outcomes, finances, human resources and patient satisfaction)		0.700
Hospital governance—board activities	0.779	
Q4.1 The hospital governance board makes it clear what is expected from clinicians in regard to QI		0.529
Q4.2 The hospital governance board has established formal roles for quality leadership (visible in organizational chart)		0.770
Q4.3 The hospital governance board assesses on an annual or bi-annual basis whether clinicians comply with day-to-day patient safety procedures		0.557
Q4.4 The hospital governance board knows and uses performance data for QI		0.516
Q4.5 The hospital governance board monitors the execution of QI plans		0.518
Quality resources	0.755	
Q5.1 Clinicians attend at least one training session a year to further develop their professional expertise		0.057[a]
Q5.2 Clinicians receive information back on the results of their treatment of patients		0.337[a]
Q5.3 Clinicians are encouraged to report incidents and adverse events		0.230[a]
Q5.4 Clinicians' registrations are reviewed by the hospital annually		0.452
Q5.5 The hospital provides training to clinicians		0.485
Q5.6 Clinicians are trained in teamwork		0.507
Q5.7 Middle management (e.g. NUM, Head of Department, etc.) is trained in QI methods		0.676
Q5.8 Clinicians are trained in QI methods		0.763
Q5.9 Clinicians are trained in patient safety procedures		0.535
Quality management	0.910	
Q6.1 Data used from clinical indicators		0.782
Q6.2 Data used from incident reporting system		0.619
Q6.3 Data used from patient interviews or surveys		0.659
Q6.4 Data used from assessment of guideline compliance		0.705
Q6.5 Data used from results of internal audits		0.867
Q6.6 Data used from audits of hand hygiene compliance		0.673
Q6.7 Data used from audits of patient identification		0.858
Preventive protocols	0.797	
Q7.1 An up-to-date hospital protocol for use of prophylactic antibiotics		0.210[a]
Q7.2 An up-to-date hospital protocol for medication reconciliation		0.199[a]
Q7.3 An up-to-date hospital protocol for the handover of patient information to another care unit		0.475
Q7.4 An up-to-date hospital protocol for use of a surgical checklist		0.485
Q7.5 An up-to-date hospital protocol for recognizing and responding to clinical deterioration in acute healthcare		0.137[a]
Q7.6 An up-to-date hospital protocol for routine assessment and diagnostic testing of elective surgery patients		0.597
Q7.7 Prevention of central line infection		0.672
Q7.8 Prevention of surgical site infection		0.642
Q7.9 Prevention of healthcare-associated infections		0.353[a]
Q7.10 Prevention of ventilator-associated pneumonia		0.351[a]
Q7.11 Prevention of medication errors		0.560
Q7.12 Prevention of patient falls		0.696
Q7.13 Prevention of patient pressure injuries		0.611
Internal quality methods—general activities	0.836	
Q8.1 Root-cause analysis of incidents is conducted according to legislation or policy, and within recommended timeframes		0.233[a]
Q8.2 Risk management, consisting of a systematic process of identifying, assessing and taking action to prevent or manage clinical risks in the care process, is undertaken in all units		0.530
Q8.3 Internal audit, consisting of periodical review of all components of the quality system, is undertaken in all units		0.602
Q8.7 Staff workplace satisfaction is measured and monitored at least annually		0.341[a]
Q8.8 Multidisciplinary CR in all units to assess and improve the results of care delivery		0.866

(Continued)

Table 2 Continued

Subscale and items	Cronbach's α	Item-total correlation
Q8.9 Patient record review in all units to determine incidents and priorities for quality improvement		0.653
Q8.10 Development of care pathways or process redesign		0.524
Q8.11 Benchmarking clinical practice against other departments within the hospital		0.655
Q8.12 Benchmarking clinical practice against other hospitals		0.549
Performance monitoring	0.791	
Q8.4 Executive 'walk-arounds' are frequently conducted to identify safety and quality issues		0.669
Q8.5 Performance of individual doctors is monitored at least annually		0.638
Q8.6 Performance of individual nurses is monitored at least annually		0.660
Internal quality methods—patients	0.449	
Q8.13 Patient satisfaction or experience is measured and monitored at least annually		0.303[a]
Q8.14 Periodical evaluation of patient complaints is used to drive improvements		0.303[a]

[a]Item-total correlation coefficient is less than the acceptable value of 0.4.

Table 3 Cronbach's alpha and item-total correlation for QMCI

Scale and items	Cronbach's α	Item-total correlation
Monitoring patient and professional opinions	0.730	
XQ03 The results of patient satisfaction or experience surveys were formally reported to the hospital governance board		0.571
XQ04 The hospital governance board received results of surveys of staff satisfaction		0.509
XQ09 Patient satisfaction or experience are measured and evaluated		0.638
XQ10 Patient complaints and feedback are investigated and acted upon		0.672
XQ11 Staff opinion or perception on organizational quality and safety culture are measured and evaluated		0.239[a]
Quality control and monitoring	0.785	
XQ01 The hospital governance board approved a current program for quality improvement (QI)		0.535
XQ02 The hospital governance board received regular, formal reports on quality and safety (Q&S)		0.334[a]
XQ05 Clinical leaders received regular, formal reports on Q&S		0.715
XQ06 There is an active clinical guideline register		0.706
XQ07 Processes for implementation and evaluation of clinical guidelines against practice		0.532
XQ08 Clinical incidents (adverse events) are analysed and evaluated		0.541

[a]Item-total correlation coefficient is less than the acceptable value of 0.4.

Table 4 Cronbach's alpha and item-total correlation for CQII

Scale and items	Cronbach's α	Item-total correlation
Preventing and controlling healthcare-associated infections (HAI)	0.257	
XC01.1 Existence of a committee responsible for preventing and controlling HAI		NC[a]
XC02.1 Existence of hospital policy or guidelines for preventing and controlling HAI		NC[a]
XC03.1 Monitoring of compliance with policy or guidelines for preventing and controlling HAI		0.192[b]
XC04.1 Sustainability of system for preventing and controlling HAI		0.111[b]
XC05.1 Improvement focus for preventing and controlling HAI		0.315[b]
Medication safety	0.160	
XC01.2 Existence of a committee responsible for medication safety		NC[a]
XC02.2 Existence of hospital policy or guidelines for medication safety		NC[a]
XC03.2 Monitoring of compliance with policy or guidelines for medication safety		0.152[b]
XC04.2 Sustainability of system for medication safety		0.086[b]
XC05.2 Improvement focus for medication safety		0.226[b]
Preventing patient falls	0.177	
XC01.3 Existence of a committee responsible for preventing patient falls		NC[a]
XC02.3 Existence of hospital policy or guidelines for preventing patient falls		0.212[b]
XC03.3 Monitoring of compliance with policy or guidelines for preventing patient falls		NC[a]
XC04.3 Sustainability of system for preventing patient falls		0.220[b]
XC05.3 Improvement focus for preventing patient falls		0.010[b]

(Continued)

Table 4 Continued

Scale and items	Cronbach's α	Item-total correlation
Preventing pressure injuries	0.804	
XC01.4 Existence of a committee responsible for preventing pressure injuries		NC[a]
XC02.4 Existence of hospital policy or guidelines for preventing pressure injuries		0.847
XC03.4 Monitoring of compliance with policy or guidelines for preventing pressure injuries		0.841
XC04.4 Sustainability of system for preventing pressure injuries		0.739
XC05.4 Improvement focus for preventing pressure injuries		0.490
Routine assessment and diagnostic testing of patients in elective surgery (ESP)	0.920	
XC01.5 Existence of a committee responsible for routine assessment and diagnostic testing of ESP		0.678
XC02.5 Existence of hospital policy or guidelines for routine assessment and diagnostic testing of ESP		0.891
XC03.5 Monitoring of compliance with policy or guidelines for routine assessment and diagnostic testing of ESP		0.825
XC04.5 Sustainability of system for routine assessment and diagnostic testing of ESP		0.842
XC05.5 Improvement focus for routine assessment and diagnostic testing of ESP		0.745
Safe surgery that includes an approved checklist	0.956	
XC01.6 Existence of a committee responsible for safe surgery		0.884
XC02.6 Existence of hospital policy or guidelines for safe surgery		0.970
XC03.6 Monitoring of compliance with policy or guidelines for safe surgery		0.961
XC04.6 Sustainability of system for safe surgery		0.872
XC05.6 Improvement focus for safe surgery		0.784
Recognizing and responding to clinical deterioration in acute healthcare (CDAHC)	−0.003	
XC01.7 Existence of a committee responsible for recognizing and responding to CDAHC		NC[a]
XC02.7 Existence of hospital policy or guidelines for recognizing and responding to CDAHC		NC[a]
XC03.7 Monitoring of compliance with policy or guidelines for recognizing and responding to CDAHC		0.129[b]
XC04.7 Sustainability of system for recognizing and responding to CDAHC		−0.104[b]
XC05.7 Improvement focus for recognizing and responding to CDAHC		0.004[b]

[a]NC, not calculated; item-total correlations were not derived for items with zero variance.
[b]Item-total correlation coefficient is less than the acceptable value of 0.4.

Table 5 Descriptive statistics for department-level scales: condition-specific departments

Department	Scale	*N*	Mean	SD	Median	Min	Max
AMI	SER	27	2.9	0.93	2.7	0.7	4
	EBOP	27	3.0	0.77	3.3	1	4
	PSS	27	3.6	0.34	3.6	2.9	4
	CR	27	2.3	1.50	2.7	0	4
Hip fracture	SER	29	2.5	1.16	2.7	0.3	4
	EBOP	29	2.6	1.06	2.6	0.2	4
	PSS	29	3.5	0.37	3.4	2.7	4
	CR	29	2.2	1.60	2.7	0	4
Stroke	SER	31	3.3	0.74	3.7	1.3	4
	EBOP	31	3.2	0.89	3.4	0.8	4
	PSS	31	3.4	0.33	3.3	2.6	4
	CR	31	2.9	1.25	3.3	0	4

Table 6 Descriptive statistics of ED-level scales and items

Condition	Scale	*n*	Mean	SD	Median	Min	Max
Generic	PSS	32	3.4	0.43	3.6	2.3	4
AMI	SER	31	2.5	1.09	2.5	1	4
	EBOP	31	3.6	0.51	3.8	2.3	4
	CR	31	2.0	1.62	2.3	0	4
Hip fracture	SER	31	1.9	1.33	2	0	4
	EBOP	31	2.5	1.27	3	0	4
	CR	31	1.3	1.32	1	0	4
Stroke	SER	32	2.8	1.24	3	0	4
	EBOP	32	3.4	0.72	3.7	1.3	4
	CR	32	2.0	1.64	2	0	4

SER, EBOP, PSS, CR—structure, reliability and non-redundancy. Department-level SER, EBOP, PSS and CR. The models are shown in Table 7 along with the Cronbach's α and item-total correlations. Internal consistency reliability through Cronbach's α was high for CR (0.74–0.88 across the three conditions), and this was supported by all item-total correlations exceeding the desired threshold. For EBOP, Cronbach's α was acceptable for hip fracture (0.80) and stroke (0.76), but only moderate for AMI (0.52); as suggested by these results, item-total correlations were poor for one item in stroke (−0.16) and two items in AMI (−0.06 and 0.19). For PSS, Cronbach's α suggested its constituents were effectively unrelated (−0.12 to 0.17), with virtually all item-total correlations poor. For SER, Cronbach's α showed a moderate association amongst its constituent items (0.32–0.52), again with relatively few item-total correlations above 0.40.

While exploratory factor analysis was used to reduce and determine which items would be aggregated to build a scale for SER and CR, the items comprising EBOP and PSS were determined based on theoretical importance and background knowledge. It was not possible to build one generic scale for the EBOP, because of the different items across pathways. The other scales developed in this analysis used the same items to compute scores for each pathway.

Table 7 Cronbach's alpha and item-total correlation for department-level scales: condition-specific

Department	Scale and items	Cronbach's α	Item-total correlation
AMI	SER	0.524	
	AS01 There is a strategic committee within the hospital responsible for the overall clinical management of AMI		0.458
	AS02 There are clinical leaders with specialist training who are formally recognized as having principal responsibility for overall clinical care of AMI patients		0.287[b]
	AS03 Evidence-based clinical guidelines have been disseminated by the hospital and formally adopted by clinicians for the management of patients with AMI		0.374[b]
	EBOP	0.516	
	AE01 There is a specialist (consultant) doctor available at all times to determine whether fibrinolysis or percutaneous coronary intervention (PCI) is appropriate		−0.063[b]
	AE02 There are written criteria and procedures to ensure appropriate medication is prescribed on discharge		0.188[b]
	AE03 There are written criteria and procedures to ensure arrangements for ongoing care on discharge		0.529
	AE04 There are written criteria and procedures to ensure information on episode of care is provided to usual clinical provider		0.717
	PSS	0.107	
	AP01 Patients are identified by bracelet		−0.038[b]
	AP02 Safety boxes for disposal of injection devices are available		−0.115[b]
	AP03 Promotional hand hygiene reminders are on display in the workplace		0.101[b]
	AP04 Readily accessible hand sanitizer is provided at the point of patient care		−0.088[b]
	AP05 No concentrated potassium chloride (KCl) stored on the ward		−0.088[b]
	AP06 Diagrammatic instructions for resuscitation are displayed in resuscitation areas or attached to crash cart		0.131[b]
	AP07 Each emergency crash cart has a completed checklist of equipment and supplies		0.184[b]
	AP08 There is a system to report clinical incidents (adverse events)		0.212[b]
	AP09 Peer review included analysis of reported clinical incidents (adverse events)		−0.047[b]
	CR	0.883	
	AC01 Peer review within the last 12 months included analysis of clinical indicators for the management of AMI in the ward or department		0.738
	AC02 There was a multidisciplinary review within the last 12 months of practice against the AMI guidelines in the ward or department		0.816
	AC03 Clinicians receive direct feedback on results within the last 12 months, following audit or review of their practice against the AMI guidelines		0.771
Hip Fracture	SER	0.517	
	HS01 There is a strategic committee within the hospital responsible for the overall clinical management of hip fracture		0.266[b]
	HS02 There are clinical leaders with specialist training who are formally recognized as having principal responsibility for overall clinical care of hip fracture patients		0.499
	HS03 Evidence-based clinical guidelines have been disseminated by the hospital and formally adopted by clinicians for the management of patients with hip fracture		0.252[b]
	EBOP	0.803	
	HE01 There is a clear clinical path for management of hip fracture surgery		0.510
	HE02 There are written criteria and procedures to ensure appropriate prophylactic medication is administered		0.645
	HE03 There are written criteria and procedures for patient mobilization following surgery		0.623
	HE04 There are written criteria and procedures to ensure fall prevention assessment is provided post-surgery and prior to discharge		0.634
	HE05 There are written criteria and procedures to ensure treatment for secondary fracture prevention on discharge		0.572
	PSS	0.167	
	HP01 Patients are identified by bracelet		0.129[b]
	HP02 Safety boxes for disposal of injection devices are available		NC[a]
	HP03 Promotional hand hygiene reminders are on display in the workplace		0.099[b]
	HP04 Readily accessible hand sanitizer is provided at the point of patient care		0.163[b]
	HP05 No concentrated potassium chloride (KCl) stored on the ward		0.035[b]

(Continued)

Table 7 Continued

Department	Scale and items	Cronbach's α	Item-total correlation
	HP06 Diagrammatic instructions for resuscitation are displayed in resuscitation areas or attached to crash cart		0.113[b]
	HP07 Each emergency crash cart has a completed checklist of equipment and supplies		0.017[b]
	HP08 There is a system to report clinical incidents (adverse events)		0.028[b]
	HP09 Peer review included analysis of reported clinical incidents (adverse events)		0.012[b]
	CR	0.867	
	HC01 Peer review within the last 12 months included analysis of clinical indicators for the management of hip fracture in the ward or department		0.663
	HC02 There was a multidisciplinary review within the last 12 months of practice against the hip fracture guidelines in the ward or department		0.754
	HC03 Clinicians receive direct feedback on results within the last 12 months, following audit or review of their practice against the hip fracture guidelines		0.832
Stroke	SER	0.318	
	SS01 There is a strategic committee within the hospital responsible for the overall clinical management of stroke		0.187[b]
	SS02 There are clinical leaders with specialist training who are formally recognized as having principal responsibility for overall clinical care of stroke patients		0.041[b]
	SS03 Evidence-based clinical guidelines have been disseminated by the hospital and formally adopted by clinicians for the management of patients with stroke		0.456
	EBOP	0.747	
	SE01 There is a specialist (consultant) doctor available at all times to determine whether intravenous rt-PA is appropriate		−0.160[b]
	SE02 There are written criteria and procedures to ensure that ischaemic stroke patients are screened and assessed for dysphagia		0.657
	SE03 There are written criteria and procedures for allocation of ischaemic stroke patients to the stroke unit		0.614
	SE04 There are written criteria and procedures to ensure appropriate medication is prescribed on discharge		0.589
	SE05 There are written criteria and procedures to ensure arrangements for ongoing care on discharge		0.746
	PSS	−0.120	
	SP01 Patients are identified by bracelet		0.459
	SP02 Safety boxes for disposal of injection devices are available		−0.181[b]
	SP03 Promotional hand hygiene reminders are on display in the workplace		0.129[b]
	SP04 Readily accessible hand sanitizer is provided at the point of patient care		−0.044[b]
	SP05 No concentrated potassium chloride (KCl) stored on the ward		−0.315[b]
	SP06 Diagrammatic instructions for resuscitation are displayed in resuscitation areas or attached to crash cart		−0.112[b]
	SP07 Each emergency crash cart has a completed checklist of equipment and supplies		0.251[b]
	SP08 There is a system to report clinical incidents (adverse events)		0.118[b]
	SP09 Peer review included analysis of reported clinical incidents (adverse events)		0.082[b]
	CR	0.744	
	SC01 Peer review within the last 12 months included analysis of clinical indicators for the management of stroke in the ward or department		0.528
	SC02 There was a multidisciplinary review within the last 12 months of practice against the stroke guidelines in the ward or department		0.685
	SC03 Clinicians receive direct feedback on results within the last 12 months, following audit or review of their practice against the stroke guidelines		0.511

[a]NC, not calculated; item-total correlations were not derived for items with zero variance.
[b]Item-total correlation coefficient is less than the acceptable value of 0.4.

Despite the same items being used across pathways for the quality measure PSS, no generic scale for the four pathways was revealed after factor analysis.

The final models for ED measures are shown in Table 8 along with the Cronbach's α scores and item-total correlations. PSS in the ED, assessed through a single set of generic questions, had only a moderate Cronbach's α (0.44), but was clearly stronger than in the condition-specific departments. For all three conditions, CR in the ED was again the most reliable scale in terms of both Cronbach's α (0.80–0.93) and item-total correlation (all items). EBOP in the ED showed a similar pattern to the condition-specific results, with an acceptable Cronbach's α for hip fracture (0.80) and stroke (0.78), but

Table 8 Cronbach's alpha and item-total correlation for ED-level scales

Condition	Scale and items	Cronbach's α	Item-total correlation
Generic	PSS	0.438	
	EP01 Patients are identified by bracelet		0.087[b]
	EP02 Safety boxes for disposal of injection devices are available		NC[a]
	EP03 Promotional hand hygiene reminders are on display in the workplace		0.183[b]
	EP04 Readily accessible hand sanitizer is provided at the point of patient care		0.454
	EP05 No concentrated potassium chloride (KCl) stored on the ward		−0.118[b]
	EP06 Diagrammatic instructions for resuscitation are displayed in resuscitation areas or attached to crash cart		0.257[b]
	EP07 Each emergency crash cart has a completed checklist of equipment and supplies		0.343[b]
	EP08 There is a system to report clinical incidents (adverse events)		0.484
	EP09 Peer review included analysis of reported clinical incidents (adverse events)		0.405
AMI	SER	0.016	
	ES01 A clinician from the ED is a member of the strategic committee within the hospital responsible for the overall clinical management of AMI		0.009[b]
	ES04 Evidence-based clinical guidelines have been disseminated by the hospital and formally adopted by ED clinicians for the management of patients with AMI		0.009[b]
	EBOP	0.455	
	EE01 There are written criteria and procedures for fast track admission and treatment of patients presenting with acute chest pain		0.399[b]
	EE02 There are written criteria and procedures to ensure that eligible ST-Elevation Myocardial Infarction patients receive fibrinolysis		0.296[b]
	EE03 There is a clear procedure in the ED to enable immediate transport or transfer for Percutaneous Coronary Intervention (PCI) for eligible STEMI patients		0.261[b]
	EE11 There is immediate access in the ED at all times to a specialist (consultant) doctor to determine whether fibrinolysis or PCI is appropriate		0.221[b]
	CR	0.888	
	EC01 Peer review within the last 12 months included analysis of clinical indicators for the management of AMI in the ED		0.848
	EC04 There was a multidisciplinary review within the last 12 months of practice against the AMI guidelines in the ED		0.700
	EC07 ED clinicians receive direct feedback on results within the last 12 months, following audit or review of their practice against the AMI guidelines		0.802
Hip fracture	SER	0.402	
	ES03 A clinician from the ED is a member of the strategic committee within the hospital responsible for the overall clinical management of hip fracture		0.253[b]
	ES06 Evidence-based clinical guidelines have been disseminated by the hospital and formally adopted by ED clinicians for the management of patients with hip fracture		0.253[b]
	EBOP	0.744	
	EE09 There are written criteria and procedures for initial pain score and pain relief for patients with suspected hip fracture		0.595
	EE10 There are written criteria and procedures for fast track admission and treatment of patients presenting with hip fracture		0.595
	CR	0.804	
	EC03 Peer review within the last 12 months included analysis of clinical indicators for the management of hip fracture in the ED		0.596
	EC06 There was a multidisciplinary review within the last 12 months of practice against the hip fracture guidelines in the ED		0.714
	EC09 ED clinicians receive direct feedback on results within the last 12 months, following audit or review of their practice against the hip fracture guidelines		0.675
Stroke	SER	0.407	
	ES02 A clinician from the ED is a member of the strategic committee within the hospital responsible for the overall clinical management of stroke		0.271[b]
	ES05 Evidence-based clinical guidelines have been disseminated by the hospital and formally adopted by ED clinicians for the management of patients with stroke		0.271[b]
	EBOP	0.781	
	EE04 There are written criteria and procedures for ensuring that patients with suspected stroke are screened and assessed for eligibility to receive intravenous rt-PA		0.616

(Continued)

Table 8 Continued

Condition	Scale and items	Cronbach's α	Item-total correlation
	EE05 There are written criteria and procedures to ensure that eligible ischaemic stroke patients receive intravenous rt-PA		0.689
	EE06 There is a clear procedure in the ED to enable immediate transport or transfer for brain imaging (e.g. CT scan, Magnetic Resonance Angiogram)		0.443
	EE07 There are written criteria and procedures to ensure that ischaemic stroke patients are screened and assessed for dysphagia		0.462
	EE08 There are written criteria and procedures for fast track admission and treatment of ischaemic stroke patients		0.774
	EE12 There is immediate access in the ED at all times to a specialist (consultant) doctor to determine whether intravenous rt-PA is appropriate		0.166[b]
	CR	0.927	
	EC02 Peer review within the last 12 months included analysis of clinical indicators for the management of stroke in the ED		0.925
	EC05 There was a multidisciplinary review within the last 12 months of practice against the stroke guidelines in the ED		0.856
	EC08 ED clinicians receive direct feedback on results within the last 12 months, following audit or review of their practice against the stroke guidelines		0.775

[a]NC, not calculated; item-total correlations were not derived for items with zero variance.
[b]Item-total correlation coefficient is less than the acceptable value of 0.4.

only a moderate score for AMI (0.46). The two SER items relating to AMI in the ED were essentially unrelated (Cronbach's α of 0.02) and moderate (0.40–0.41) for the other conditions.

Discussion

Interpretation of results

This study aimed to refine and validate, in the context of the Australian healthcare system, three scales for measuring quality improvement at organization level (QMSI, QMCI and CQII), and four scales for measuring quality improvement at hospital care pathway-level for AMI, hip fracture and stroke conditions (SER, EBOP, PSS and CR).

The final QMSI scale consists of eight subscales, measuring: quality policy, hospital governance board activities, quality resources, quality management, preventive protocols, internal quality methods, general activities and performance monitoring. The final QMCI consists of two subscales, measuring: monitoring patient and professional opinions and quality control and monitoring. The final CQII consists of seven subscales, measuring: preventing and controlling healthcare-associated infections, medical safety, preventing patient falls, preventing pressure injuries, routine assessment and diagnostic testing of patients in elective surgery, safe surgery that includes an approved checklist and recognizing and responding to clinical deterioration in acute healthcare. In addition to the differences with the DUQuE scales noted in the introduction, some items are grouped within different subscales following subscale analysis.

The final CQII scale could not be validated due to the scores clustering at the high end of the scale. This is likely a consequence of mandatory accreditation requirements for Australian hospitals, as there is a strong alignment between the topics addressed by the CQII subscales and accreditation measures. Nevertheless, we believe it is important to retain the CQII, as the performance assessed has been shown via research to be important for patient safety [9, 10], and the scale needs to allow for the prospect that not all future hospitals will be as high performers as those in DUQuA.

Within each scale, we retained some subscales that had Cronbach's α lower than the desired 0.8, where those subscales measured aspects of quality that were considered to be a critical component of hospital quality management, or where the lower correlation could be explained by understanding how hospital care was structured. For example, in the QMSI scale, all subscales but the 'internal quality methods—patients' achieved (or nearly achieved) internal consistency reliability using the cut-off value. This subscale consisted of ratings of whether patient satisfaction or experience is measured and monitored at least annually, and whether periodical evaluation of patient complaints is used to drive improvements. The DUQuE QMSI scale found a similar result for this subscale [11], and was retained for both studies, recognizing the importance of the patient voice in hospital care. There are good reasons to keep items and scales within the measurement instruments because they are theoretically important, even if they did not fulfil the high statistical standards set before the analysis. While this may seem to undermine the importance of these criteria and the methods used, it is of practical relevance for the final scales to be able to measure an adequate range of elements that make up quality management and quality improvement activities in hospitals.

In many scales and subscales we also had some items below our acceptable correlation cut-off of 0.4. Similarly, we retained items that had lower correlation with other items within the same factor where those items were considered to be an important part of hospital activity. For example, within the QMSI subscale of 'preventative protocols', removing the item rating prevention of ventilator-associated pneumonia would have improved Cronbach's α from 0.797 to >0.8, but we resisted this because patients in Australian hospital are normally only ventilated in the Intensive Care Unit, during surgery, and—occasionally—in the ED, potentially explaining why it did not correlate well with other activities that are performed more broadly across the hospital. In another example within the same subscale, we retained the item 'an up-to-date hospital protocol for recognizing and responding to clinical deterioration in acute healthcare' due to its importance in patient care, despite an item-total correlation of only

0.137. This is an element of care that has received strong endorsement from State health departments over the last decade [12, 13], and the data were strongly left skewed due to high performance in this aspect of care in most participating hospitals.

For department-level care pathway scales, the final scales were not modified as a result of our analysis. We developed a generic scale for PSS, but it was not possible to build generic scales for SER, EBOP and CR, as the processes are different depending on the specific condition and the department; for this reason, we also developed condition-specific versions of each of these scales for use in the ED.

Limitations

While most of the scales are based on independently audited data, the QMSI scale is based on perceptions of the quality manager. In mitigation, we worded the questions in the quality manager survey to include only questions on facts, to minimize bias associated with self-report. Additionally, while we have shown how the organization and department scales have been modified from their European antecedents, the DUQuA scales have only been validated for the Australian context, and their generalizability to other settings remains to be demonstrated. Differences between the DUQuA and DUQuE scales must be taken into account when comparing findings from Australian and European hospitals.

Implications for research, policy and practice

It can be difficult for hospitals to access validated measurement tools for assessing quality management that are powerful yet easy to use. The DUQuA scales are now publicly available to fill this niche. The scales could also be used to collect longitudinal data, for example before and after an intervention, used to collect information to assist in designing a quality improvement intervention, or as part of a comparison of hospitals or hospital departments.

At present, the CQII scale is likely to just confirm what is known from accreditation results. More work is required to develop or refine a tool that is more able to discriminate between hospitals with high implementation indices due to participation in external accreditation processes. Collection of data from international hospitals that are not subject to mandatory accreditation may assist this process.

Conclusions

The DUQuA organization and department scales can be used by Australian hospital managers to assess and measure improvement in quality management at organization and department levels within their hospitals and are readily modifiable for other health systems depending on their needs.

Supplementary material

Supplementary material is available at *INTQHC Journal* online.

Acknowledgements

Members of the original DUQuE team (Oliver Gröne and Rosa Suñol) provided extensive input and advice in the initial stages of the project, supporting design modifications from DUQuE and the planned DUQuA study approach. Dr Annette Pantle provided expert advice on revision and development of DUQuA measures, and Professor Sandy Middleton, Associate Professor Dominique Cadillac, Kelvin Hill, Dr Carmel Crock and Professor Jacqueline Close provided input into the development of our stroke, AMI and hip fracture clinical process indicator lists. Nicole Mealing and Victoria Pye provided statistical advice in the initial phases of the project. We greatly appreciate their efforts.

Funding

This research was funded by the National Health and Medical Research Council (NHMRC) Program Grant APP1054146 (CI Braithwaite).

Authors' contributions

The research team consists of experienced researchers, clinicians, biostatisticians and project managers with expertise in health services research, survey design and validation, large-scale research and project management, sophisticated statistical analysis, quality improvement and assessment, accreditation, clinical indicators, policy and patient experience. JB conceived the idea, led the research grant to fund the project and chairs the steering committee. RCW and NT co-led the detailed study design, managed the project across time and contributed to the development of the manuscript. HPT and GA provided statistical expertise for the study design and developed the analysis plan for the manuscript. TW contributed to the logistics of project management, the refinement of measures and the development of the manuscript.

Ethics approval

Ethical approvals were secured from State and Territory human research ethics committees in New South Wales (#14/206), Victoria (#15/36), the Australian Capital Territory (#15/131), South Australia (#15/260), the National Territory (#15-2509), Tasmania (#H0015383) and Queensland (#15/361). Site-specific authorizations, including permission for external researchers to collect data in hospitals, were granted. We complied with confidentiality requirements of national legislation or standards of practice of each jurisdiction.

Data sharing statement

Data will be made publicly available to the extent that individual participants or participating hospitals cannot be identified, in accordance with requirements of the approving Human Research Ethics Committees.

References

1. Taylor N, Clay-Williams R, Hogden E *et al*. Deepening our understanding of quality in Australia (DUQuA): a study protocol for a nationwide, multilevel analysis of relationships between hospital quality management systems and patient factors. *BMJ Open* 2015;**5**:e010349.
2. Secanell M, Groene O, Arah OA *et al*. Deepening our understanding of quality improvement in Europe (DUQuE): overview of a study of hospital quality management in seven countries. *Int J Qual Health Care* 2014;**26**:5–15.
3. Wagner C, Groene O, Thompson C *et al*. Development and validation of an index to assess hospital quality management systems. *Int J Qual Health Care* 2014;**26**:16–26.
4. Braithwaite J, Clay-Williams R, Taylor N *et al*. Deepening our Understanding of Quality in Australia (DUQuA): An overview of a nation-wide, multilevel analysis of relationships between quality management systems and patient factors in 32 hospitals. *Int J Qual Health Care* 2020;**32**:8–21.
5. Wagner C, Thompson CA, Arah OA *et al*. A checklist for patient safety rounds at the care pathway level. *Int J Qual Health Care* 2014;**26**:36–46.
6. Wagner C, Groene O, Dersarkissian M *et al*. The use of on-site visits to assess compliance and implementation of quality management at hospital level. *Int J Qual Health Care* 2014;**26**:27–35.
7. Nunnally JC. An overview of psychological measurement. In: Wolman BB (ed.). *Clinical Diagnosis of Mental Disorders*. Boston, MA, Springer, 1978, 97–146.
8. Arnolda G, Winata T, Ting HP *et al*. Implementation and data-related challenges in the Deepening our Understanding of Quality in Australia

(DUQuA) study: implications for large-scale cross-sectional research. *Int J Qual Health Care* 2020;**32**:75–83.
9. Gilks WR, Richardson S, Spiegelhalter D. *Markov Chain Monte Carlo in Practice*. Boca Raton, FL, Chapman and Hall & CRC Press, 1995.
10. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY, John Wiley & Sons, 2004.
11. Wagner C, Groene O, Thompson CA *et al*. DUQuE quality management measures: Associations between quality management at hospital and pathway levels. *Int J Qual Health Care* 2014;**26**:66–73.
12. Hillman KM, Rubenfeld GD, Braithwaite J. Time to shut down the acute care conveyor belt? *Med J Aust* 2015;**203**:429–30.
13. Hughes C, Pain C, Braithwaite J *et al*. 'Between the flags': implementing a rapid response system at scale. *BMJ Qual Saf* 2014;**23**:714–7.

International Journal for Quality in Health Care, 2020, 32(S1), 35–42
doi: 10.1093/intqhc/mzz104
Article

Article

Do organization-level quality management systems influence department-level quality? A cross-sectional study across 32 large hospitals in Australia

NATALIE TAYLOR[1,2], ROBYN CLAY-WILLIAMS[3], HSUEN P. TING[3], GASTON ARNOLDA[3], TERESA WINATA[3], EMILY HOGDEN[1], and JEFFREY BRAITHWAITE[3]

[1]Cancer Research Division, Cancer Council NSW, 153 Dowling St, Woolloomooloo, NSW 2011, Australia, [2]Faculty of Health Sciences, University of Sydney, Camperdown, Sydney, NSW 2006, Australia, and [3]Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia

Address reprint requests to: Natalie Taylor, Cancer Council NSW, University of Sydney, Cancer Institute NSW Career Development Fellow. Tel: +61(2) 9334 1974, +61 423 388 064; E-mail: natalie.taylor@nswcc.org.au

Editorial Decision 7 September 2019; Accepted 13 September 2019

Abstract

Objective: Little is known about the influence that hospital quality systems have on quality at department level, in Australia and elsewhere. This study assessed the relationships between organizational-level quality management systems, and the extent to which hospital-level quality management systems and department-level quality management strategies are related.
Design: A multi-level, cross-sectional, mixed-method study.
Setting and participants: As part of the Deepening our Understanding of Quality in Australia (DUQuA) project, we invited all large hospitals in Australia (~200 or more beds) which provided acute myocardial infarction (AMI), hip fracture and stroke care. The quality managers of these hospitals were the respondents for one of seven measures of hospital quality management systems and strategies. Data across the six remaining measures were collected through site visits by external surveyors assessing the participating hospitals.
Main outcome measures: Relationships were assessed between three organization-level quality management system measures: a self-report measure assessing organization-level quality activities (quality management systems index, QMSI); externally assessed organization-level compliance to procedures used to plan, monitor and improve quality of care (quality management compliance index, QMCI); and externally assessed implementation of quality systems (clinical quality implementation index, CQII). Associations were also assessed between organization-level quality management systems and department-level quality management strategies: how clinical responsibilities are assigned for a particular condition; whether department organization processes are organized to facilitate evidence-based care recommendations; compliance with selected recommendations of international agencies; and whether clinical reviews are performed systematically.
Results: Of 78 invited hospitals, 32 participated in the study. QMSI was positively associated with QMCI and CQII, but after controlling for QMSI, no relationship was found between QMCI and CQII. There appears to be a cluster of relationships between QMSI and department-level measures, but this was not consistent across all departments.

Conclusion: This is the first national study undertaken in Australia to assess relationships within and between organization-level and department-level quality management systems. These quality management system tools align with many components of accreditation standards and may be useful for hospitals in continuously monitoring and driving improvement.

Key words: organization-level quality management, department-level quality management, national standards, accreditation, quality improvement

Introduction

Requirements to demonstrate hospital service quality have increased nationally and internationally in recent decades due to increased societal attention, a multiplicity of government policies, and the introduction of mandatory hospital accreditation [1–5]. The necessity for quality strategies to be embedded throughout hospitals has led to numerous efforts to develop reliable ways to measure quality management systems at organization and department levels [6–10]. Less is known, however, about the influence that hospital-level quality (e.g. infrastructure, quality improvement processes) has on department-level quality (e.g. organization of department processes, approaches to patient safety). Understanding these relationships is crucial because, for example, it may help us determine whether efforts to improve department-level quality management systems also require investments in improving organization-level quality systems (e.g. governance board quality, quality management, performance monitoring activities) [11, 12]. Similarly, if the emphasis is on ensuring organization-level quality but commitment, responsibility and engagement is not embedded at the department level, we may find lower quality healthcare delivery, patient safety incidents, or worse outcomes for patients [12–14]. If we can pinpoint how organization and department-level performance are related (and beyond this, the links to patient outcomes, as assessed in other reports in this Supplement), this may uncover key areas to focus targeted interventions to raise the standards of quality and safety [15].

Recent work has been undertaken in Europe to explore, using validated measures to gather both internal and external perspectives, relationships among different elements of quality management systems both within and between organization and department levels of hospitals [9, 10, 16]. The Deepening our Understanding of Quality Improvement in Europe (DUQuE) study concluded that, across 77 hospitals in seven countries, it is possible to obtain a comprehensive picture of hospital quality management maturity, incorporating different administrative levels within the hospital, across several hospital departments (acute myocardial infarction [AMI], hip fracture and stroke) [15]. An assessment of the relationships between quality management systems within and between organization and department levels of hospitals has not yet been undertaken in Australia, and rarely elsewhere. Therefore, the aims of this study are to assess: (i) the relationships between organization-level quality systems including the quality management systems index (QMSI), quality compliance and improvement (quality management compliance index [QMCI]) and implementation of clinical quality activities (quality management compliance index [CQII]), and (ii) the extent to which hospital-level quality management systems, and department-level quality management strategies are related across a sample of hospital departments.

Methods

A protocol for the full study has been published [17], and some additional details (e.g. actual recruitment) have already been described in other papers in this Supplement [18–20]; here we provide a summary of the methods relevant to this sub-study, which focuses on three organization-level quality management systems and four department-level quality management strategies, including changes from the published protocol. The seven measures were initially developed for the DUQuE study and adapted and validated by Deepening our Understanding of Quality in Australia (DUQuA) to ensure relevance to the targeted Australian hospitals. Figure 1 shows the relationships between organization and department-level measures in the DUQuA study. These measure levels are described below.

Setting, participants, recruitment and data collection

The study population comprised 78 Australian public hospitals meeting the DUQuA hospital inclusion criteria (~200 beds or more with an ED and 10 or more admissions per month for each of AMI, hip fracture and stroke) [17]. All were invited and 32 received ethics approval and participated.

The senior hospital quality manager at each hospital, or equivalent role-holder, completed an organization-level survey (QMSI) and coordinated with the research team for external surveyors to conduct an on-site assessment for two organization-level quality management systems and four department-level measures of quality management strategies. The hospital quality manager did not receive special training; external surveyors, all experienced hospital evaluators, received research-specific training as detailed in Supplementary Appendix A. All measures are listed in Figure 2 and summarized in Table 1, with an indication of how data was collected. Where the measure is a scale comprising subscales, the subscales are listed in the table.

Organization-level measures

Organization-level quality management systems were assessed using three measures: the QMSI, the QMCI and the CQII. Each of these measures, validated as part of the DUQuE project, was adapted for Australian conditions and validated in its revised form [19]. Full validation properties can be found for these measures in measures in the Supplement [19].

QMSI. This is a measure of the management system aspects of quality that might influence the implementation of quality systems in hospitals. Fifty one questions in the QMSI are used to generate eight quality subscales including quality policy, hospital governance board activities, quality resources, quality management, preventive protocols, performance monitoring and internal quality methods (separately for general activities and patients) [19]. The questionnaire is designed as a four-point Likert scale (range 1–4) for each item, with scores of 4 representing more comprehensive systems. While questionnaire completion represents self-assessment, the classifications required were considered objective, and thus unlikely to be unduly biased by self-assessment. Seven out of eight QMSI subscales achieved adequate internal consistency (Cronbach's $\alpha > 0.8$) during validation.

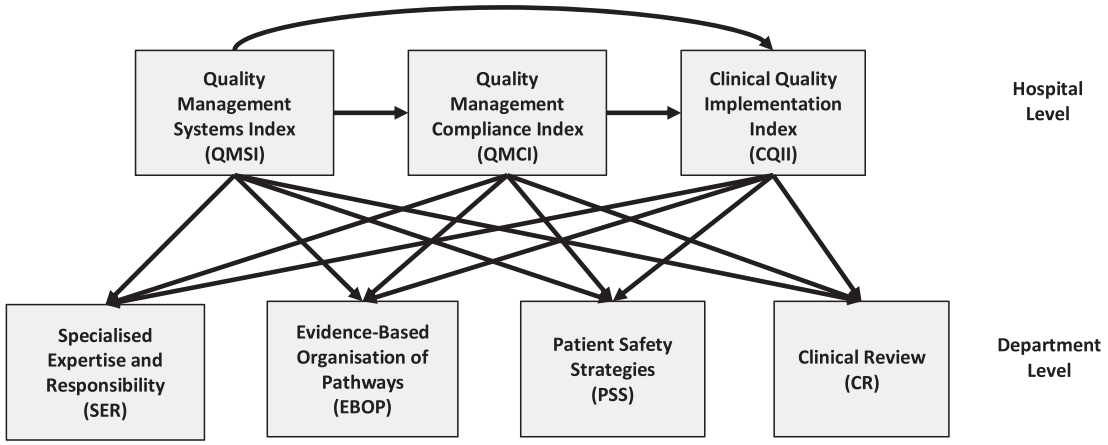

Figure 1 Simplified directed acyclic graph (DAG) showing the relationships between organization-level and department-level measures in DUQuA

Organisation-level measures
a) Quality Management Systems Index (QMSI)
b) Quality Management Compliance Index (QMCI)
c) Clinical Quality Implementation Index (CQII)

Department-level measures
a) Specialised Expertise and Responsibility (SER)
b) Evidence-based Organisation of Pathways (EBOP)
c) Patient Safety Strategies (PSS)
d) Clinical Review (CR)

Figure 2 DUQuA organization and department-level measures list

QMCI. This is a measure of compliance with procedures used to plan, monitor and improve quality of care. The QMCI comprises 11 questions each assessed on a five-point Likert scale (range 0–4), with higher scores reflecting more comprehensive approaches to compliance; scores are used to create two subscales: monitoring of patient and professional opinions; and quality control and monitoring. Assessments are completed during on-site audits completed by trained external surveyors who review and verify relevant evidence of compliance. Adequate internal consistency (Cronbach's $\alpha > 0.8$) was achieved when validating both QMCI subscales.

CQII. This is a measure of evidence of implementation of quality systems at hospital level; including whether systems exist, to what extent implementation has been monitored, and whether implementation is sustainable. The index is based on the results of an on-site audit completed by trained external surveyors, who score each of 35 questions on five-point Likert scales (0–4), with higher scores representing more comprehensive implementation. Seven domains were selected, each producing a separate subscale score: preventing and controlling healthcare associated infections, medication safety, preventing patient falls, preventing pressure injuries, routine assessment and diagnostic testing of patients in elective surgery, safe surgery that includes an approved checklist, recognizing and responding to clinical deterioration in acute healthcare. The surveyors assessed the presence of an active committee for the domain, relevant policy or guidelines and monitoring of compliance, sustainability of the system and the adoption of an improvement focus. Lack of variation and ceiling effects in the data resulted in very low internal consistency scores for CQII, but items were retained for theoretical reasons.

Department-level measures

Four DUQuE measures were refined, and structure and validity tested, to assess department-level quality activities in Australian hospitals: (i) Specialized Expertise and Responsibility (SER) assessed how clinical responsibilities were assigned for a particular condition; (ii) Evidence-Based Organization and Pathways (EBOP) measured if department organization processes, such as admission, acute care, rehabilitation and discharge, were organized to facilitate evidence-based care recommendations; (iii) Patient Safety Strategies (PSS) measured compliance with selected general clinical practice guidelines and (iv) Clinical Review (CR) measured if CRs were performed systematically.

SER, CR and PSS measures were identical for the three conditions, and the EBOP has the same structure for each condition-specific department (AMI, hip fracture and stroke), however the content followed the evidence recommendations specific to each condition. Each of the four indices was scored on a five-point Likert scale (0–4), with higher ratings indicating greater compliance. In line with the experience and advice of the DUQuE research consortium, the four assessments were expected to be completed by surveyors during one- to two-day external on-site assessments.

Statistical analysis

A directed acyclic graph (DAG), a simplified version of which is shown in Figure 1, presents assumed causal relations among variables and thereby determines which confounding variables are included in the statistical models. For example, to examine whether QMCI has an effect on SER, we controlled for QMSI because it is a predictor for both QMCI and SER.

Data were analysed in SAS/STAT software version 9.4 (SAS Institute, Cary, North Carolina, USA). Each organization scale was calculated as the sum of all subscales (for QMSI, 8 was subtracted from the sum, in line with DUQuE procedures) [9]; each subscale was the mean of all non-missing items; subscales with more than 50%

Table 1 DUQuA organizational and department-level measures, data collection methods and its format

Measures	Content	Data collection methods	Format
Organization-level measures			
QMSI	*Eight subscales:* Quality policy; Hospital governance board activities; Quality resources; Quality management; Preventive protocols; Internal quality methods for general activities; Performance monitoring; Internal quality methods for patients	Self-report questionnaire completed by the hospital's Quality Manager or equivalent	Paper-based questionnaire
QMCI	*Two subscales:* Monitoring patient and professional opinions; Quality control and monitoring	External quality assessment by trained healthcare surveyors (site visit)	Paper-based audit forms filled by surveyors
CQII	*Seven subscales:* Preventing and controlling healthcare associated infections; Medication safety; Preventing patient falls; Preventing pressure injuries; Routine assessment and diagnostic testing of patients in elective surgery; Safe surgery that includes an approved checklist; Recognizing and responding to clinical deterioration in acute health care		
Department-level measures (separately for inpatient departments responsible for AMI, hip fracture and stroke)			
SER	Assignment of clinical responsibilities for a condition	External quality assessment by trained healthcare surveyors (site visit)	Paper-based audit forms filled by surveyors
EBOP	Organization of department processes (admission, acute care and discharge to facilitate evidence-based care recommendations)		
PSS	Use of international consensus-based patient safety recommendations		
CR	Integration of audit and systematic monitoring in departmental quality management mechanisms		

missing items were treated as missing, and a value was subsequently imputed. For department scales, each scale was the mean of all applicable non-missing items; scales with more than 50% applicable items missing were treated as missing.

Visual inspection of missing data patterns in the organization-level suggested that data was missing at random. Multiple imputations were, therefore, performed (MI procedure) for the missing organization subscales, repeated 100 times, and scale scores were calculated after imputation. Each imputed dataset was merged with department-level data and general linear models were used to examine the associations between measures as specified in the simplified DAG model in Figure 1, with hospital peer group (referral/acute) also controlled for. Analysis was performed on each imputed dataset and pooled parameter estimates and standard errors were derived, if any imputed values were involved in the calculations.

Results

Participating hospitals over-represent principal referral hospitals and hospitals with over 500 beds and under-represent inner regional remoteness areas and the state of Western Australia (see Appendix B, Supplementary eTable B1). Hospital-level measures were collected from all 32 hospitals; two hospitals shared a single quality management system and governance, so only 31 questionnaires were collected for the QMSI. Missing data for the organization scales ranged from 0% to 3.1% at subscale level. There were 87 department-level measures expected (by condition: 27 AMI, 31 hip fractures and 29 strokes).

Descriptive statistics of the three hospital-level measures, after imputation, are detailed in Table 2, for the overall scale and each component subscale. The mean QMSI score was 19.4 (standard deviation [SD]: 3.1) out of maximum possible of 24; the mean QMCI was 7.1 (SD: 1.3) out of 8; and the mean CQII was 25.1 (SD: 2.9) out of 28. All subscales were rated at three or above with one exception: the CQII subscale for 'Routine assessment and diagnostic testing of patients in elective surgery' had a mean score of 2.3. The median hospital score was the maximum (i.e. 4.0) for one of eight QMSI subscales and for six of seven CQII subscales. All subscales had standard deviations less than one with the exception of two CQII subscales one of which had an out-of-range value, above four, imputed.

Results of the associations between organization-level scales are summarized in Table 3. The value of the betas is an artefact of the scoring ranges of the scales used and is not therefore of importance, excepting that the sign signifies a positive or negative association. QMSI was statistically associated with QMCI ($P = 0.002$) and with CQII ($P = 0.003$) after adjustment for peer group; more specifically, 1-point increases in QMSI is associated with a 0.20 point increase in QMCI (95% CI: 0.07–0.33), and a 0.45 point increase in CQII (95% CI: 0.16–0.74). There was no association between QMCI and CQII ($P = 0.98$) after adjustment for QMSI.

Descriptive statistics of the four department-level measures are summarized in Table 4, separately for AMI, hip fracture and stroke. For each of the three condition-specific departments, the mean scale scores ranged from 2.3 to 3.6 with the highest scores were consistently found for the PSS measure (3.4–3.6), and the lowest scores were for the CR measure (2.2–2.9).

Results of the associations are summarized in Table 5. With 36 associations assessed, two associations would be expected to be statistically significant by chance, one each positive and negative. We found five statistically significant *positive* associations between QMSI and department-level measures: one for SER (stroke); one for EBOP (hip fracture); one for PSS (hip fracture) and two for CR (AMI and stroke). There were no statistically significant associations between QMCI and the four department-level measures, after adjustment for QMSI. There were two statistically significant associations between CQII and the four department-level measures: a negative association with PSS for AMI; and a positive association with CR for hip fracture.

Table 2 Descriptive statistics for hospital-level measures[a]

Index	*n*	Mean	SD	Median	Min	Max[b]	IQR
QMSI	31	19.4	3.14	20.3	10.1	23.9	3.9
Quality policy	31	3.2	0.89	3.3	1.3	4.0	1.0
Hospital governance: board activities	31	3.4	0.52	3.6	2.0	4.0	0.8
Quality resources	31	3.4	0.36	3.3	2.4	4.0	0.4
Quality management	31	3.5	0.48	3.7	2.0	4.0	0.7
Preventive protocols	31	3.6	0.40	3.7	2.5	4.0	0.6
Internal quality methods: general activities	31	3.4	0.47	3.4	2.0	4.0	0.7
Performance monitoring	31	3.3	0.68	3.3	1.7	4.0	1.3
Internal quality methods: patients	31	3.7	0.42	4.0	3.0	4.0	0.5
QMCI	32	7.1	1.27	7.4	1.4	8.2	0.8
Monitoring patient and professional opinions	32	3.5	0.74	3.8	0.6	4.8	0.8
Quality control and monitoring	32	3.5	0.62	3.8	0.8	4.0	0.7
CQII	32	25.1	2.85	25.6	18.8	28.0	4.3
Preventing and controlling healthcare associated infections	32	3.9	0.20	4.0	3.2	4.0	0.0
Medication safety	32	3.9	0.21	4.0	3.2	4.0	0.2
Preventing patient falls	32	3.9	0.13	4.0	3.6	4.0	0.1
Preventing pressure injuries	32	3.8	0.50	4.0	1.2	4.0	0.3
Routine assessment and diagnostic testing of patients in elective surgery	32	2.3	1.54	2.5	0.0	4.0	3.3
Safe surgery that includes an approved checklist	32	3.3	1.25	4.0	0.0	5.0	1.0
Recognizing and responding to clinical deterioration in acute healthcare	32	3.9	0.18	4.0	3.2	4.0	0.0

[a]After multiple imputation techniques applied to missing data.
[b]Out-of-range values were retained to ensure accurate calculation of variance when assessing the association between measures [21].
Legend: IQR = inter-quartile range

Table 3 Associations between organization-level measures

	QMCI			CQII		
Index	*n*	Beta (95% CI)	*P*-value	*n*	Beta (95% CI)	*P*-value
QMSI[a]	32	0.199 (0.071, 0.327)	0.002*	32	0.447 (0.155, 0.740)	0.003*
QMCI[a,b]				32	−0.013 (−0.861, 0.834)	0.975

Legend: QMSI (Index 0–24), QMCI (Index 0–8), CQII (Index 0–28).
[a]Adjusted for hospital peer group; [b]Additionally adjusted for QMSI; *Statistically significant at $P = 0.05$.

Table 4 Descriptive statistics for department-level measures

Department	Index	*n*	Mean	SD	Median	Min	Max	IQR
AMI	SER	27	2.9	0.93	2.7	0.7	4.0	1.7
	EBOP	27	3.0	0.77	3.3	1.0	4.0	1.0
	PSS	27	3.6	0.34	3.6	2.9	4.0	0.6
	CR	27	2.3	1.50	2.7	0.0	4.0	3.3
Hip fracture	SER	29	2.5	1.16	2.7	0.3	4.0	2.0
	EBOP	29	2.6	1.06	2.6	0.2	4.0	1.2
	PSS	29	3.5	0.37	3.4	2.7	4.0	0.4
	CR	29	2.2	1.60	2.7	0.0	4.0	3.0
Stroke	SER	31	3.3	0.74	3.7	1.3	4.0	1.0
	EBOP	31	3.2	0.89	3.4	0.8	4.0	1.4
	PSS	31	3.4	0.33	3.3	2.6	4.0	0.5
	CR	31	2.9	1.25	3.3	0.0	4.0	2.0

Legend: IQR = inter-quartile range

Discussion

To our knowledge, this is the first time the relationship between quality management systems within and between organization and department levels of hospitals has been tested in Australia, and one of only a few internationally. Collecting data from 32 hospitals across five states and two territories, key findings indicate that QMSI (a self-report measure assessing organizational-level quality activities, was positively associated with QMCI and CQII (two measures applied by externally trained surveyors for assessing organizational-level compliance to procedures used to

Table 5 Associations of department-level measures with hospital-level measures

		AMI			Hip fracture			Stroke		
Index	Outcome	*n*	Beta (95% CI)	*P*-value	*n*	Beta (95% CI)	*P*-value	*n*	Beta (95% CI)	*P*-value
QMSI[a]	SER	27	−0.023 (−0.137, 0.091)	0.678	29	0.051 (−0.097, 0.200)	0.486	31	0.090 (0.005, 0.174)	0.038*
	EBOP	27	−0.032 (−0.131, 0.066)	0.504	29	0.120 (0.015, 0.225)	0.027*	31	0.074 (−0.028, 0.177)	0.150
	PSS	27	0.033 (−0.009, 0.074)	0.115	29	0.049 (0.005, 0.094)	0.030*	31	0.020 (−0.021, 0.061)	0.322
	CR	27	0.211 (0.038, 0.384)	0.019*	29	0.081 (−0.073, 0.235)	0.289	31	0.186 (0.047, 0.325)	0.010*
QMCI[a,b]	SER	27	0.216 (−0.271, 0.702)	0.385	29	0.279 (−0.097, 0.655)	0.146	31	−0.010 (−0.376, 0.356)	0.958
	EBOP	27	0.269 (−0.145, 0.683)	0.202	29	−0.005 (−0.282, 0.272)	0.972	31	−0.092 (−0.537, 0.353)	0.685
	PSS	27	−0.055 (−0.234, 0.123)	0.544	29	0.056 (−0.059, 0.170)	0.340	31	0.098 (−0.076, 0.272)	0.271
	CR	27	0.332 (−0.409, 1.072)	0.380	29	0.133 (−0.270, 0.536)	0.517	31	0.161 (−0.438, 0.759)	0.599
CQII[a,b,c]	SER	27	0.085 (−0.042, 0.212)	0.189	29	0.039 (−0.130, 0.208)	0.653	31	0.005 (−0.097, 0.106)	0.928
	EBOP	27	0.044 (−0.067, 0.155)	0.434	29	0.042 (−0.082, 0.166)	0.506	31	0.024 (−0.098, 0.147)	0.696
	PSS	27	−0.051 (−0.094, −.007)	0.023*	29	0.026 (−0.025, 0.077)	0.312	31	0.018 (−0.029, 0.065)	0.452
	CR	27	0.171 (−0.015, 0.358)	0.072	29	0.183 (0.017, 0.349)	0.031*	31	0.087 (−0.075, 0.250)	0.293

Legend: QMSI (Index 0–24), QMCI (Index 0–8), CQII (Index 0–28), SER (Index 0–4), EBOP (Index 0–4), PSS (Index 0–4), CR (Index 0–4).
Adjusted for: [a]hospital peer group; [b]also adjusted for QMSI; [c]also adjusted for QMCI.
*Statistically significant at 5%.

plan, monitor and improve quality of care, and implementation of quality systems), but after controlling for QMSI, no relationship was found between QMCI and CQII. There appears to be a cluster of relationships between QMSI and department-level measures, but this was not consistent across all departments. More specifically, QMSI was associated with how clinical responsibilities are assigned for a particular condition (SER) for stroke, whether department organization processes are organized to facilitate evidence-based care recommendations (EBOP) and compliance with selected recommendations of international agencies (PSS) for hip fracture, and systematic implementation of CR for AMI and stroke. There was no clear evidence of an effect of QMCI or CQII on department-level measures, after adjustment for QMSI.

At the organizational level, in contrast to our findings, DUQuE found positive associations between all three measures [15]. The DUQuE study presented 18 out of 36 significant positive associations between hospital and department-level quality management systems across AMI, hip fracture and stroke: five with QMSI; 11 with QMCI after adjustment for QMSI, and two with CQII after adjusting for both other measures. DUQuA found six positive associations: five with QMSI; none with QMCI (adjusted for QMSI) and one with CQII (adjusted with QMSI and QMCI). The starkest difference between the two studies is in the effect of QMCI (adjusted for QMSI) on the four department-level measures: in DUQuE this is consistent across all but one of 12 department-level measures, while in DUQuA no relationships were found.

There are a number of possible explanations for these differences. First, to ensure the measures were relevant and appropriate for the Australian system, a number of modifications were made for the DUQuA study, including removal of some items pertaining to the assessment of individual staff licences, education, certification, experience and performance reviews, and changes to ensure scale descriptions were explicit. Nonetheless, we have found the measurement properties for all revised measures to be sound [19]. Second, the timeframes for data collection for DUQuE and DUQuA were starkly different. For example, data collection for the entire DUQuE study lasted eight months, whereas—despite the cross-sectional study design intentions—DUQuA data collection took over two years. In DUQuA, the timing on the internally assessed measure (QMSI) and the six externally assessed measures (QMCI, CQII, SER, EBOP, PSS and CR) were not identical; the median timing difference was zero (i.e. same month) but it ranged from QMSI completion 11 months before to eight months after (IQR: 2.5 months before to one month after). As such, changes in personnel, processes and/or resources [22, 23] during this time may have introduced error into the assessment of associations. Third, the DUQuA results are left-skewed in comparison to DUQuE: for QMSI, the mean DUQuA score was 19.4 out of 24 (80.8% of maximum) versus 19.4 out of 27 (71.9%) for DUQuE; for QMCI (DUQuA = 7.1/8 [88.8%] versus DUQuE = 10.4/16 [65.0%]); and CQII (DUQuA = 25.1/28 [89.6%] versus DUQuE = 8.4/14 [60.0%]), the differences are more marked, with the Australian hospitals clustered at higher scores. In Australia, accreditation [24] has been mandatory for all public hospitals since 2013 [25]; importantly, during DUQuE data collection, this was not the case for participating hospitals (of the participating DUQuE hospitals, only 34% were locally accredited) [16]. The extensive preparation that Australian hospitals undergo to ensure systems, processes, and records will meet the accreditation standards prior to assessment (every four years with a biannual shortened assessment) [25] likely reduced the range of index scores for DUQuA, especially for CQII (which contains many accreditation-related features). Finally, hospital sample sizes were smaller in DUQuA ($n = 32$) than in DUQuE ($n = 72$).

Assessing our findings in relation to the wider literature, the stronger and more frequent associations between QMSI and other organization and department-level measures align with previously reported quantitative and qualitative findings. For example, the quality of the hospital board and its associated quality activities—a key component of the QMSI measure—has been positively associated with high performing hospitals [26], the likelihood that hospitals have quality improvement programs (a key component of our CQII measure) [12], and better performance on process measures at care pathway level [27, 28]. Furthermore, the provision of adequate resources to departments for quality and use of quality data to drive improvement by senior management and governing boards (two additional components measured by QMSI) have been reported as key attributes of hospital departments with concrete protocols and practices undertaken to reduce mortality among patients [14].

Limitations of this study include non-concurrent internal and external data collection, and a restricted response range to the

organization-level indices. Non-concurrent collection derives from problems related to addressing site-specific governance processes across our participating hospitals [29], and difficulties arranging access to hospitals for external surveyors as research data collectors (despite these individuals being trained to collect accreditation data in the same hospitals) [30]. The lack of a broad response range for organization-level indices, especially QMSI and CQII, makes it difficult to differentiate between hospitals and thus explore associations; for countries with universal accreditation, more sensitive indices may need to be developed if we are seeking to examine associations between organization and department-level measures of system maturity. Both of these issues were compounded by the low number of hospitals retained in the study—we targeted 78, 62 initially agreed, and 32 completed the study; the design and test combination was therefore likely underpowered for detecting the hypothetical effect sizes of interest [17].

Despite these limitations, unlike unvalidated accreditation assessment tools, the previously validated quality management system assessment tools refined and revalidated for the Australian healthcare system, allow hospitals to reliably and continuously measure, diagnose, benchmark and drive-targeted improvements for quality at the organization and department levels. The alignment of these tools with the assessment criteria used by accreditation agencies in Australia may help with interim monitoring of performance in preparation for accreditation.

Conclusion

This national study of 32 Australian hospitals has demonstrated relationships within and between organization and department-level quality management systems, the strength of which is most visible for relationships involving organizational-level quality activities (QMSI). Some similarities in relationships were found with the earlier DUQuE study, despite differences in hospital accreditation policy between Australia and Europe. These quality management system tools align with many components of accreditation standards and may be useful for hospitals in continuously monitoring and stimulating improvement.

Supplementary material

Supplementary material is available at *INTQHC Journal* online.

Acknowledgements

Members of the original DUQuE team (Oliver Gröne, Rosa Suñol) provided extensive input and advice in the initial stages of the project, supporting design modifications from DUQuE and the planned DUQuA study approach. Dr Annette Pantle provided expert advice on revision and development of DUQuA measures, and Professor Sandy Middleton, Associate Professor Dominique Cadillac, Kelvin Hill, Dr Carmel Crock and Professor Jacqueline Close provided input into the development of our AMI, hip fracture and stroke clinical process indicator lists. Nicole Mealing and Victoria Pye provided statistical advice in the initial phases of the project. We greatly appreciate their efforts. Zhicheng Li and Chrissy Clay also organized and coordinated the external site visits at all participating hospitals with all external researchers/assessors (surveyors).

Funding

This research was funded by National Health and Medical Research Council (NHMRC) Program Grant APP1054146 (CI Braithwaite).

Contributions

The research team consists of experienced researchers, clinicians and biostatisticians with expertise in health services research, survey design and validation, large-scale research and project management, sophisticated statistical analysis, quality improvement and assessment, accreditation, clinical indicators, policy and patient experience. JB led the research grant to fund the project. NT led the study design and development of the manuscript. RCW, HPT and GA co-led the study design and contributed to the development of the manuscript. HPT and GA also provided statistical expertise for the study design and developed the analysis plan for the manuscript. TW and EH contributed to the logistics of project management, the refinement of measures and the development of the manuscript.

Data sharing statement

Data will be made publicly available to the extent that individual participants or participating hospitals cannot be identified, in accordance with requirements of the approving Human Research Ethics Committees.

Ethical consideration

Ethical approvals were gained by the DUQuA team at Human Research Ethics Committees (HRECs) in all regions covering participating hospitals [18]. Site-specific authorisations, including the permission of using external researchers to collect data in hospitals, have also been granted before commencement of research. We complied with confidentiality requirements of national legislation or standards of practice of each jurisdiction.

References

1. Klazinga N. Concerted action programme on quality assurance in hospitals 1990–1993 (COMAC/HSR/QA) global results of the evaluation. *Int J Qual Health Care* 1994;**6**:219–30.
2. Duckett SJ. Assuring hospital standards: The introduction of hospital accreditation in Australia. *Aust J Public Administr* 1983;**42**:385–402.
3. Asgar AH, Delgoshaei B, Kringos DS *et al.* Implementing hospital quality assurance policies in Iran. *Int J Health Care Qual Assur* 2015;**28**.
4. Greenfield D, Braithwaite J. Health sector accreditation research: a systematic review. *Int J Qual Health Care* 2008;**20**:172–83.
5. Legido-Quigley H, McKee M, Walshe K *et al.* How can quality of health care be safeguarded across the European Union? *BMJ* 2008;**336**:920–3.
6. Shaw C. External assessment of health care. *BMJ* 2001;**322**:851–4.
7. Øvretveit J, Gustafson D. Using research to inform quality programmes. *BMJ* 2003;**326**:759–61.
8. Wagner C, Thompson CA, Arah OA *et al.* A checklist for patient safety rounds at the care pathway level. *Int J Qual Health Care* 2014;**26**:36–46.
9. Wagner C, Groene O, Thompson C *et al.* Development and validation of an index to assess hospital quality management systems. *Int J Qual Health Care* 2014;**26**:16–26.
10. Wagner C, Groene O, DerSarkissian M *et al.* The use of on-site visits to assess compliance and implementation of quality management at hospital level. *Int J Qual Health Care* 2014;**26**:27–35.
11. Bradley EH, Curry LA, Webster TR *et al.* Achieving rapid door-to-balloon times. *Circulation* 2006;**113**:1079–85.
12. Weiner BJ, Shortell SM, Alexander J. Promoting clinical involvement in hospital quality improvement efforts: the effects of top management, board, and physician leadership. *Health Serv Res* 1997;**32**:491.
13. Taylor N, Clay-Williams R, Hogden E *et al.* High performing hospitals: a qualitative systematic review of associated factors and practical strategies for improvement. *BMC Health Serv Res* 2015;**15**:244.
14. Curry LA, Spatz E, Cherlin E *et al.* What distinguishes top-performing hospitals in acute myocardial infarction mortality rates? A qualitative study. *Ann Intern Med* 2011;**154**:384–90.

15. Wagner C, Groene O, Thompson CA *et al*. DUQuE quality management measures: associations between quality management at hospital and pathway levels. *Int J Qual Health Care* 2014;**26**:66–73.
16. Shaw CD, Groene O, Botje D *et al*. The effect of certification and accreditation on quality management in 4 clinical services in 73 European hospitals. *Int J Qual Health Care* 2014;**26**:100–7.
17. Taylor N, Clay-Williams R, Hogden E *et al*. Deepening our understanding of quality in Australia (DUQuA): a study protocol for a nationwide, multilevel analysis of relationships between hospital quality management systems and patient factors. *BMJ Open* 2015;**5**:e010349.
18. Braithwaite J, Clay-Williams R, Taylor N *et al*. Deepening our Understanding of Quality in Australia (DUQuA): An overview of a nation-wide, multi-level analysis of relationships between quality management systems and patient factors in 32 hospitals. *Int J Qual Health Care* 2020;**32**:8–21.
19. Clay-Williams R, Taylor N, Winata T *et al*. Organization quality systems and department-level strategies: refinement of the Deepening our Understanding in Quality in Australia (DUQuA) organization and department-level scales. *Int J Qual Health Care* 2020;**32**:22–34.
20. Arnolda G, Winata T, Ting HP *et al*. Implementation and data-related challenges in the Deepening our Understanding of Quality in Australia (DUQuA) study: implications for large-scale cross-sectional research. *Int J Qual Health Care* 2020;**32**:75–83.
21. Rodwell L, Lee KJ, Romaniuk H, Carlin JB. Comparison of methods for imputing limited-range variables: a simulation study. *BMC Med Res Methodol* 2014;**14**:57.
22. Harris C, Green S, Ramsey W *et al*. Sustainability in health care by allocating resources effectively (SHARE) 1: introducing a series of papers reporting an investigation of disinvestment in a local healthcare setting. *BMC Health Serv Res* 2017;**17**:323.
23. Kurnat-Thoma E, Ganger M, Peterson K *et al*. Reducing annual hospital and registered nurse staff turnover—a 10-element onboarding program intervention. *SAGE Open Nursing* 2017;**3**.
24. Australian Commission on Safety and Quality in Health Care. *National safety and quality health service standards*, 2nd edn. Sydney, NSW, 2017.
25. Australian Commission on Safety and Quality in Health Care. NSQHS standards in. *transforming the safety and quality of health care*. Sydney, AU, Australian Commission on Safety and Quality in Health Care, 2013, 2014.
26. Jha A, Epstein A. Hospital governance and the quality of care. *Health Affairs* 2010;**29**:182–7.
27. Jiang HJ, Lockee C, Bass K. Board engagement in quality: findings of a survey of hospital and system leaders. *J Healthcare Manag* 2008;**53**.
28. Jiang H, Lockee C, Bass K *et al*. Board oversight of quality: any differences in process of care. *J Healthcare Manage* 2009;**54**:15–29.
29. Clay-Williams R, Taylor N, Braithwaite J. Potential solutions to improve the governance of multicentre health services research. *Med J Aust* 2018;**208**:152–4.
30. Winata T, Clay-Williams R, Taylor N *et al*. Using accreditation surveyors to conduct health services research: a qualitative, comparative study in Australia. *Int J Qual Health Care* 2020;**32**:89–98.

International Journal for Quality in Health Care, 2020, 32(S1), 43–51
doi: 10.1093/intqhc/mzz105
Article

Article

The relationships between quality management systems, safety culture and leadership and patient outcomes in Australian Emergency Departments

ROBYN CLAY-WILLIAMS[1], NATALIE TAYLOR[2,3], HSUEN P. TING[1], TERESA WINATA[1], GASTON ARNOLDA[1], ELIZABETH AUSTIN[1], and JEFFREY BRAITHWAITE[1]

[1]Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Road Macquarie University, NSW 2109, Australia, [2]Cancer Research Division, Cancer Council NSW, 153 Dowling St, Woolloomooloo, NSW 2011, Australia, and [3]Faculty of Health Sciences, University of Sydney, Camperdown, Sydney, NSW 2006, Australia

Address reprint requests to: Jeffrey Braithwaite, Centre for Healthcare Resilience and Implementation Science (CHRIS), Australian Institute of Health Innovation, Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia. Tel: +61 2 9850 2401, +61 414 812 579; Fax: +61 2 9850 2499; E-mail: jeffrey.braithwaite@mq.edu.au

Editorial Decision 7 September 2019; Accepted 13 September 2019

Abstract

Objective: We aimed to examine whether Emergency Department (ED) quality strategies, safety culture and leadership were associated with patient-level outcomes, after controlling for other organization-level factors, in 32 large Australian hospitals.
Design: Quantitative observational study, using linear and multi-level modelling to identify relationships between quality management systems at organization level; quality strategies at ED level for acute myocardial infarction (AMI), hip fracture and stroke; clinician safety culture and leadership and patient-level outcomes of waiting time and length of stay.
Setting: Thirty-two large Australian public hospitals.
Participants: Audit of quality management processes at organization and ED levels, senior quality manager at each of the 32 participating hospitals, 394 ED clinicians (doctors, nurses and allied health professionals).
Main Outcome Measure(s): Within the multi-level model, associations were assessed between organization-level quality measures and ED quality strategies; organization-level quality measures and ED quality strategies and ward-level clinician measures of teamwork climate (TC), safety climate (SC) and leadership for AMI, hip fracture and stroke treatment conditions; and organization-level quality measures and ED quality strategies and ward-level clinician measures of TC, SC and leadership, and ED waiting time and length of stay (performance).
Results: We found seven statistically significant associations between organization-level quality systems and ED-level quality strategies; four statistically significant associations between quality systems and strategies and ED safety culture and leadership; and nine statistically significant associations between quality systems and strategies and ED safety culture and leadership, and ED waiting time and length of stay.
Conclusions: Organization-level quality structures influence ED-level quality strategies, clinician safety culture and leadership and, ultimately, waiting time and length of stay for patients. By focusing only on time-based measures of ED performance we risk punishing EDs that perform well

on patient safety measures. We need to better understand the trade-offs between implementing safety culture and quality strategies and improving patient flow in the ED, and to place more emphasis on other ED performance measures in addition to time.

Key words: hospital quality management systems, multi-level research, Emergency Departments, safety culture, leadership, quality improvement

Introduction

Both in Australia and internationally, Emergency Departments (EDs) typically measure headline performance in terms of patient flow. In 2017–18 there were >8 million presentations to Australian EDs: a 3.4% increase in presentations overall compared with the previous year [1]. Despite the introduction of National Emergency Access Targets (NEAT) across Australia in 2010, 28% of patients were not 'seen on time' in accordance with Australasian Triage Scale standards, and 29% of visits were not concluded within four hours [1]. Increasing numbers of patient presentations, coupled with limited in-patient bed capacity, can result in long waiting times and prolonged lengths of stay (LOS) in the ED. Overcrowding is becoming more prevalent and has previously been associated with an increase in medical errors [2, 3] and poor patient outcomes [4, 5] including death [6].

Australian EDs are required to report their performance on aspects such as waiting time and LOS, and quantifying how fast patients receive care [1]. When patients first arrive at EDs in Australian hospitals, a nurse allocates them to a triage category depending on the urgency of the presenting condition (resuscitation, emergency, urgent, semi-urgent and non-urgent). Waiting time performance targets are based on this triage category (immediate, within 10, 30, 60 and 120 min, respectively) [7]. Quality improvement strategies and models of care (e.g. Fast track, Clinical Initiatives Nurse) have been developed to reduce the time patients spend in ED [8]. ED performance targets such as the Emergency Treatment Performance (ETP; initially implemented as NEAT) expect 81% of patients presenting will physically leave the ED for admission, referral or discharged home within four hours [7]. Across Australia, the introduction of and compliance with NEAT/ETP has improved the timely provision of urgent care in public hospitals and is associated with reduced in-hospital mortality rates for patients admitted through ED [9]. However, EDs across Australia continue to experience access block (where the patient needs to be admitted, but there are no hospital beds available) and ED overcrowding (where there are no beds available in the ED) [1].

Concentrating on improving patient flow, however, has meant that less attention has been paid to other aspects of ED performance, such as clinician safety culture and leadership, patient experiences, and adherence to quality management practices and clinical guidelines. In the ED, safety culture and leadership measures capture the values, behaviours, perceptions and competencies of the ED staff related to health and safety [10, 11]. Patient and professional experiences can provide insight into the culture of the organization and the potential trade-offs that are being made to meet resource-based performance targets. Associations have been found between patient experience, patient safety and clinical effectiveness [12, 13]. Similarly, accreditation processes tell us about ED compliance with quality management practices and clinical guidelines. We suggest that factors such as safety culture and leadership influence ED performance and provide a richer picture of performance than measures of patient flow alone.

Hospitals are complex organizations with multiple competing targets (e.g. budget, quality and service) [11]. Performance culture, effective and supportive leaders across the hospital, as well as the existence, implementation and sustainability of quality systems are characteristics that are associated with high performing hospitals [11]. We know that patients are more satisfied with their care when staff display positive attitudes and waiting times are shorter [14, 15]. To date, however, research has yet to examine the relationship between hospital and department-level quality strategies with ED patient outcomes, so we do not know to what extent hospital quality management systems influence quality and safety at a department level such as the ED and therefore performance. Access blockages for admitted patients to other wards are the primary cause of overcrowding in the ED [16]. Therefore, in order to address ED overcrowding, it is important to consider organization-level factors that influence bed capacity throughout the hospital.

The 'Deepening our Understanding of Quality in Australia' (DUQuA) study [17] was a 5-year Australia-wide, multi-level, multimillion dollar cross-sectional study exploring how quality management systems, leadership and culture in Australian hospitals are related to care delivery and patient outcomes for acute myocardial infarction (AMI), hip fracture and stroke. Evidence- or consensus-based measurement tools were designed or modified and then utilized to collect quantitative data on quality management systems at hospital and care pathway levels, clinician leadership and culture, clinical treatment processes, patient outcomes and patient perceptions of safety. Because the ED is a common entry point for AMI, hip fracture and stroke in Australian hospitals, we aimed to explore whether ED quality strategies were associated with patient-level outcomes after controlling for other organization-level factors.

Methods

A study protocol for the full study has been published [17]. Here, we provide a summary of the methods relevant to this study, focused on the ED measures, including waiting times and LOS; four department-level care pathway measures for AMI, hip fracture and stroke; and three department-level culture and leadership measures. The three department-level and four care pathway measures were initially developed for the Deepening our Understanding of Quality improvement in Europe (DUQuE) study and adapted and validated by DUQuA to ensure relevance to the targeted Australian hospitals [18]. Figure 1 shows the hypothesised relationships between the ED measures in the DUQuA study. These measures are described below.

Measures

Organization-level measures. The three organization-level measures are the Quality Management Systems Index (QMSI), the Quality Management Compliance Index (QMCI), and the Clinical Quality Implementation Index (CQII). QMSI is a measure of the managerial aspects of quality management, such as developing and implementing

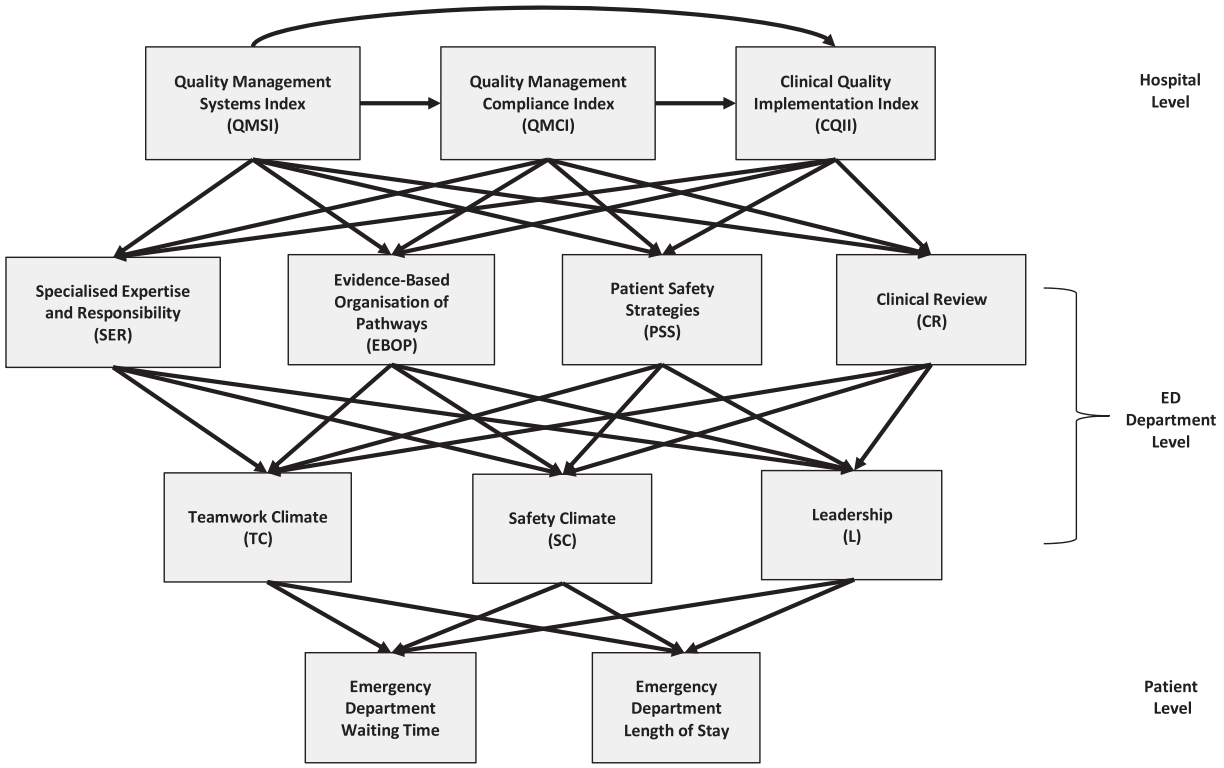

Figure 1 Simplified Directed Acyclic Graph (DAG) showing the relationships between the ED measures in DUQuA.

formally agreed quality policy, and is measured using a questionnaire completed by the quality manager in each hospital. QMSI consists of eight subscales, seven of which achieved adequate internal consistency (Cronbach's $\alpha > 0.8$) during validation [18]. QMCI is a measure of the managerial aspects of quality improvement in hospitals, such as monitoring patient and professional opinions. QMCI consists of two subscales, both of which achieved adequate internal consistency (Cronbach's $\alpha > 0.8$) during validation [18]. CQII is a seven subscale measure of the existence, implementation and sustainability of quality systems, for example preventing and controlling healthcare associated infections. QMCI and CQII are both assessed during an on-site audit by trained external surveyors. When validating CQII, lack of variation and ceiling effects in the data resulted in very low internal consistency scores, but items were retained for theoretical reasons [18]. Detail on refinement and validation of the organization-level measures is presented elsewhere in this Supplement [18].

ED-level care pathway measures. The four measures of quality management activities in the ED are based on the results of an on-site audit completed by trained external surveyors. Specialized Expertise and Responsibility (SER) explores how clinical responsibilities were assigned for a particular condition; Evidence-Based Organization of Pathways (EBOP) measures whether ED processes, such as admission, acute care, rehabilitation and discharge, were organized to facilitate evidence-based care recommendations; Patient Safety Strategies (PSS) measures the use of clinical practice guidelines; and Clinical Review (CR) assesses the extent to which audit and systematic monitoring are integrated in ED quality management mechanisms. Separate condition-specific measures were developed for SER, EBOP and CR, relevant to the ED component of care for each condition; PSS items are all generic, so no condition-specific versions were needed. Refinement and validation methods for the department-level measures are described elsewhere in this Supplement [18]; data for the ED-level care pathway measures is presented at Supplementary eTable 1 in this article's Appendix Material.

ED-level culture and leadership measures. The DUQuA safety culture and leadership scale comprises the following three subscales: teamwork climate (TC), safety climate (SC) and leadership, (L). TC (14 items) is a measure of perceived quality of collaboration between personnel, whereas SC (13 items) measures staff perceptions of the presence of a strong and proactive organizational commitment to patient safety. The leadership scale (eight items) measures clinician perceptions of the vision, inclusiveness and internal and external collaborative behaviours of healthcare leaders in the ED. Refinement and validation of the clinician safety culture and leadership measures is presented elsewhere in this supplement [19]. The three scales are completed by volunteer ED clinicians.

Patient-level measures. Patient-level measures include waiting time to be seen for emergency presentations, and LOS from ED presentation to physical departure. Hospital-level summary data were extracted by the Australian Institute of Health and Welfare (AIHW) from the National Non-admitted Patient Emergency Department Care Database (NNAPEDC) to which all participating hospitals contribute; we used the published results for the participating hospitals for 1 July 2014 to 30 June 2015.

Participants

The study population comprised 78 Australian public hospitals meeting the DUQuA hospital inclusion criteria of approximately 200 beds or more, with an ED, and 10 or more admissions per month for AMI,

hip fracture and stroke [17]. All were invited, and 32 received ethics approval and participated.

Statistical analysis

The simplified Directed Acyclic Graph (DAG) at Figure 1 shows the hypothesized relations among variables, identifying the key variables controlled for in the statistical models. For example, to examine whether QMCI has an effect on ED-level PSS, we controlled for QMSI because the DUQuE study findings indicated that it is a predictor of both QMCI and the ED-level PSS.

The dataset was processed as detailed elsewhere in this Supplement [18] and analysed using SAS/STAT software version 9.4 (SAS Institute, Cary, NC, USA). Where there were no repeated measures (i.e. care pathway quality and clinician measures, and ED performance measures) general linear models were used to analyse the relationships. Models were adjusted for hospital peer group categorized as Referral vs Acute (A or B), in addition to variables in the DAG (Figure 1). For the ED performance measures, we included measures for the five triage categories (resuscitation, emergency, urgent, semi-urgent and non-urgent) and also adjusted for triage category in the models. When examining whether the ED-level clinician measures had an effect on the ED performance measures, we used the median department score for each of the three clinician measures. General linear mixed models were used to analyse repeated measures (in each ED multiple clinicians assessed TC, SC and leadership). The mixed models were adjusted for one random effect (hospital), several fixed effects (hospital peer group, clinician age, gender and profession and self-identified leadership role) and higher-level measures designated in the DAG. A large number of associations were assessed, but no adjustment was made for multiple comparisons; given the small number of hospitals and the resulting lack of statistical power, we preferred to examine the pattern of results, including the number of statistically significant results compared to that expected by chance (rationale for this approach is further discussed in another article in this Supplement) [20].

Multiple imputations [21] were used to address missing data at the organization level, as described in this Supplement [18]. Analysis was repeated for each of the 100 imputation datasets and the SAS/STAT MIANALYZE procedure was applied to obtain pooled parameter estimates and standard errors, if any imputed values were used in the calculations.

Results

Characteristics of participants

Hospitals and EDs. ED measures were collected from all 32 hospitals in the DUQuA study, and from 394 ED clinicians (doctors, nurses and allied health professionals). Characteristics of the participating hospitals can be found in Table 1 (compared to non-participating hospitals in Supplementary eTable 2). A timetable showing when data were collected for each measure can be found in Table 2.

ED clinicians. Characteristics of the responding clinicians are summarized in Table 3. Respondents were more likely to be female and most commonly aged 25–34 years. The proportion of self-identified leaders was 61%, and more than half (52%) of the respondents were doctors. EDs are staffed primarily by doctors and nurses, so there were few allied health professionals.

Table 1 Characteristics of participating hospitals

	Participating ($N = 32$)	
Characteristic	***N***	**%**
State		
ACT	2	6.3
NSW	11	34.4
NT	1	3.1
QLD	8	25.0
SA	1	3.1
TAS	1	3.1
VIC	8	25.0
WA	0	0.0
Hospital peer group		
Principal referral	17	53.1
Public acute group A	14	43.8
Public acute group B	1	3.1
Remoteness area		
Major cities	26	81.3
Inner regional	4	12.5
Outer regional	2	6.3
Average funded beds		
<200	2	6.3
200–<500	17	53.1
500–<1000	11	34.4
1000 and more	2	6.3

Table 2 Timeline for data collection by measures for the DUQuA study

Measures	**Data collection timeline**
Organization level	
1. QMSI	September 2015–August 2017
2. External Quality Assessment (EQA)	April 2016–July 2017
(a) QMCI	
(b) CQII	
Department level	
1. EQA	April 2016–July 2017
(a) SER	
(b) EBOP	
(c) PSS	
(d) CR	
2. Safety culture and leadership measures	June 2016–November 2017
Patient level	
1. Clinical treatment indicators	September 2014–February 2015[a]
2. AIHW inpatient statistic	July 2014–June 2015[a]
3. Patient Measure of Safety (PMOS)	January 2016–November 2017

[a]Reference date of collected data.

Descriptive statistics

ED performance measures. Waiting time and LOS for the ED presentations are summarized in Table 4. Across all hospitals, the median waiting time for resuscitation was zero minutes for all hospitals. The median waiting time for emergency triage was 5.5 min on average

Table 3 Characteristics of ED clinicians

Characteristic	ED ($N = 394$)	
	N	%
Gender		
Male	167	42.4
Female	227	57.6
Age group (years)		
18–24	21	5.3
25–34	160	40.6
35–44	111	28.2
45–54	73	18.5
55–64	25	6.3
65–74	4	1.0
Leadership		
Yes	242	61.4
No	152	38.6
Profession		
Physician	205	52.0
Nurse	176	44.7
Allied health	13	3.3

across 32 hospitals, and the interquartile range was 4–6 min. For the rest of the presentations, the average median waiting time was over 19–27 min: the interquartile range was 12.5–24.0 min for urgent presentations, 19.0–38.0 min for semi-urgent presentations and 18.0–31.5 min for non-urgent presentations. The average median LOS in ED was similar for resuscitation, emergency and urgent triage categories (222–231 min), all longer than semi-urgent (155 min) and non-urgent (101 min) cases. Across all triage categories, the interquartile range was 169–207 min.

ED care pathway measures. As shown in Table 5, PSS and EBOP had mean scores above 3.0 (3.4 and 3.3 respectively). The mean score for SER was 2.4 and for CR was 1.8.

ED safety culture and leadership measures. Descriptive statistics of the clinician-level measures are summarized in Table 6. Mean score was highest for TC (78.5) and around 70 for both SC and leadership (70.6 and 70.4 respectively).

Associations

Effects of organization measures on ED-level measures. In Table 7, we report the effects of the three organization measures on the four ED-level measures. Out of the 12 associations examined, seven were statistically significant. QMSI had positive and statistically significant effect on SER, EBOP and PSS. For every unit increase in QMSI score (range 0–24), SER increased by 0.12; EBOP increased by 0.08 and PSS increased by 0.05 (range 0–4 for department-level scores). QMCI had positive and statistically significant effect on EBOP after adjusting for QMSI. For every unit increase in QMCI score (range 0–8), EBOP increased by 0.20 (range 0–4). CQII had positive and statistically significant effect on SER, EBOP and CR after adjusting for QMSI and QMCI. For every unit increase in CQII score (range 0–28), SER increased by 0.17; EBOP increased by 0.09 and CR increased by 0.24 (range 0–4 for all).

Effects of organization and pathway measures on culture and leadership measures. In Table 8, we report the effects of the three organization measures and the four pathway measures on the safety culture and leadership measures in the ED. Out of the 21 associations examined, four were statistically significant, as compared to one expected by chance. QMSI had positive and statistically significant effect on SC and leadership. For every unit increase in QMSI score, SC and leadership scores each increased by around one unit, on a 0–100 possible score range. EBOP had a negative and statistically significant effect on SC after adjusting for the three organization measures. For every unit increase in EBOP score, SC decreased by around six units (0–100 range). CR had a negative and statistically significant effect on leadership, after adjusting for the three organization measures. For every unit increase in CR score, leadership decreased by almost three units (0–100 range).

Effects of organization, ED level, and ED safety culture and leadership measures on ED performance measures. Table 9 reports the effects of the three organization measures, the four pathway measures and the three culture and leadership measures on the two ED performance measures. Of the 20 associations examined, nine were statistically significant. In accordance with national ED performance targets, we treat shorter median waiting time and median LOS as positive outcomes. QMSI had a positive and statistically significant effect on both the ED performance measures. For every unit increase in QMSI score, which has a possible range of 0–24, the median waiting time decreased by 1.4 min and the median LOS decreased by 2.3 min. QMCI had a negative and statistically significant effect on median LOS after adjusting for QMSI; for every unit increase in QMCI score, which has a possible range of 0–8, the median LOS increased by 4.7 min. PSS, with a possible range of 0–4, had a negative and statistically significant effect on both the ED performance measures after adjusting for the three organization measures. For every unit increase in PSS score, the median waiting time increased by 5 min and the median LOS increased by 18 min. TC had statistically significant effect on both the performance measures after adjusting for the three organization and the four pathway measures. For every unit increase in TC score (0–100 range), the median waiting time decreased by 0.3 min, and the median LOS increased by 0.9 min; returned to its original scale of 1–5 (so that a one-unit change is equivalent to that of the PSS measure), a one-unit change in TC was associated with an 8-min *decrease* in waiting time and a 23-min *increase* in length of stay. SC had a positive and statistically significant effect on both performance measures, after adjusting for the three organization and four pathway measures. For every unit increase in SC score (0–100 range), the median waiting time decreased by 0.3 min; and the median LOS decreased by 1.1 min; returned to its original five-point scale, a one-unit change in SC was equivalent to a 7-min *decrease* in waiting time and a 28-min *decrease* in LOS.

Discussion

Across all hospitals, both the median waiting time for triage category, and LOS, met performance targets. We found seven statistically significant associations between organization-level quality systems and ED-level quality strategies; four statistically significant associations between quality systems and strategies and ED clinician safety culture and leadership; and nine statistically significant associations between quality systems and strategies, ED safety culture and leadership and ED performance. While statistically significant, however, in some cases the size of the effect was small. A greater number of significant associations were found between organization-level measures and ED-level measures than were found for AMI, hip fracture and stroke

Table 4 ED presentations waiting time and length of stay

Triage category (performance target)	*N*	Mean	Median	Min	Max	Q1	Q3
Median waiting time (min)							
Resuscitation (immediate)	32	0.0	0.0	0	0	0.0	0.0
Emergency (<10 min)	32	5.5	6.0	1	10	4.0	6.0
Urgent (<30 min)	32	21.0	19.0	8	47	14.5	24.0
Semi-urgent (<60 min)	32	29.6	26.5	10	72	19.0	38.0
Non-urgent (<120 min)	32	27.1	25.5	11	59	18.0	31.5
Total[a]	32	20.0	17.5	8	48	13.0	27.0
Median length of stay (min)							
Resuscitation (<240 min)	32	235.5	226.0	154	397	210.0	259.0
Emergency (<240 min)	32	239.0	231.0	186	344	215.0	262.5
Urgent (<240 min)	32	221.6	222.0	171	290	194.0	237.0
Semi-urgent (<240 min)	32	160.0	154.5	117	210	141.5	175.5
Non-urgent (<240 min)	32	99.7	100.5	55	138	87.5	109.5
Total[a]	32	187.2	188.5	147	232	169.0	206.5

[a]Includes records for which the triage category was missing/unknown.

Table 5 Descriptive statistics for ED-level measures

Scale[a]	*N*	Mean	SD	Median	Min	Max	Q1	Q3
SER	32	2.4	0.94	2.3	0.7	4	1.9	3.3
EBOP	32	3.3	0.61	3.5	1.9	4	3.1	3.8
PSS	32	3.4	0.43	3.6	2.3	4	3.1	3.8
CR	32	1.8	1.33	1.9	0	4	0.4	2.9

[a]Each scale had a possible range from 0 to 4.

Table 6 Descriptive statistics for ED culture and leadership measures

Scale[a]	*N*	Mean	SD	Median	Min	Max	Q1	Q3
TC	372	78.5	12.21	78.6	21.4	100.0	71.4	87.5
SC	373	70.6	14.45	73.1	21.2	100.0	59.6	80.8
Leadership (L)	391	70.4	18.40	71.9	3.1	100.0	59.4	81.3

[a]Each scale had a possible range from 0 to 100.

Table 7 Associations of ED-level measures with hospital-level measures

	SER			EBOP			PSS			CR		
Predictor	*N*	Beta (95% CI)	*P*-value	*N*	Beta (95% CI)	*P*-value	*N*	Beta (95% CI)	*P*-value	*N*	Beta (95% CI)	*P*-value
QMSI[a]	32	0.121 (0.019, 0.224)	0.022*	32	0.081 (0.014, 0.148)	0.019*	32	0.053 (0.005, 0.101)	0.031*	32	0.073 (−0.082, 0.229)	0.343
QMCI[a,b]	32	−0.055 (−0.339, 0.229)	0.706	32	0.195 (0.024, 0.366)	0.025*	32	0.017 (−0.115, 0.150)	0.797	32	0.175 (−0.252, 0.602)	0.421
CQII[a,b,c]	32	0.169 (0.060, 0.279)	0.002*	32	0.093 (0.026, 0.161)	0.007*	32	0.036 (−0.021, 0.094)	0.217	32	0.236 (0.068, 0.404)	0.006*

Legend: QMSI (Index 0–24), QMCI (Index 0–8), CQII (Index 0–28), SER (Index 0–4), EBOP (Index 0–4), PSS (Index 0–4), CR (Index 0–4).
[a]Adjusted for hospital peer group.
[b–c]Additionally adjusted for: [b]QMSI; and [c]QMCI.
*Statistically significant at 5%.

departments [22]. One possible explanation lies in the hospital-wide focus on the movement of patients as hospital demand for beds increases, and the degree to which the ED is integral to patient flow through the hospital.

We found that by increasing hospital-level quality strategies, ED specialized expertise and responsibility, evidence-based organization of pathways and patient safety strategies also increase. This supports our conjecture that, for the ED, the executives' values, perceptions and operationalization of health and safety strategies flow down and influence how safety is practiced in the department. If the hospital executives do not value and put time into safety strategies, then the underlying message to the wards is that it is not valued and, therefore, does not get done. This is an important finding for the ED, as continued focus on attending to time-based measures of performance

Table 8 Associations of ED culture and leadership measures with hospital and ED-level measures

	TC			SC			L		
Predictor	***N***	**Beta (95% CI)**	***P*-value**	***N***	**Beta (95% CI)**	***P*-value**	***N***	**Beta (95% CI)**	***P*-value**
QMSI[a]	372	0.325 (−0.290, 0.939)	0.283	373	0.976 (0.155, 1.797)	0.022*	391	1.089 (0.179, 2.000)	0.021*
QMCI[a,b]	372	0.132 (−1.647, 1.911)	0.878	373	−0.204 (−2.596, 2.189)	0.862	391	−0.077 (−2.731, 2.577)	0.952
CQII[a,b,c]	372	0.420 (−0.354, 1.194)	0.287	373	0.618 (−0.420, 1.655)	0.243	391	0.246 (−0.942, 1.435)	0.685
SER[a,b,c,d]	372	−0.744 (−3.206, 1.719)	0.554	373	−2.599 (−5.685, 0.486)	0.099	391	−2.626 (−6.165, 0.914)	0.146
EBOP[a,b,c,d]	372	−2.196 (−6.315, 1.922)	0.296	373	−6.106 (−11.15, −1.07)	0.018*	391	−4.607 (−10.66, 1.443)	0.136
PSS[a,b,c,d]	372	0.080 (−5.850, 6.011)	0.979	373	0.059 (−7.563, 7.682)	0.988	391	1.359 (−7.396, 10.11)	0.761
CR[a,b,c,d]	372	−0.928 (−2.696, 0.839)	0.303	373	−1.961 (−4.119, 0.197)	0.075	391	−2.765 (−5.204, −.326)	0.026*

Legend: QMSI (Index 0–24), QMCI (Index 0–8), CQII (Index 0–28), SER (Index 0–4), EBOP (Index 0–4), PSS (Index 0–4), CR (Index 0–4), TC (Index 0–100), SC (Index 0–100), L (Index 0–100).
[a]Adjusted for random effect (hospital), and fixed effects (hospital peer group, clinician gender, age group, profession and leadership).
[b–d]Additionally adjusted for fixed effects: [b]QMSI; [c]QMCI; and [d]CQII.
*Statistically significant at 5%.

Table 9 Associations of ED performance measures with hospital, ED and clinician level measures

	Median waiting time			Median length of stay		
Predictor	***N***	**Beta (95% CI)**	***P*-value**	***N***	**Beta (95% CI)**	***P*-value**
QMSI[a]	160	−1.373 (−1.815, −.931)	0.000*	160	−2.326 (−3.936, −.716)	0.005*
QMCI[a,b]	160	0.218 (−1.055, 1.492)	0.737	160	4.655 (0.103, 9.207)	0.045*
CQII[a,b,c]	160	−0.287 (−0.835, 0.260)	0.304	160	0.743 (−1.235, 2.721)	0.461
SER[a,b,c,d]	160	−0.440 (−2.335, 1.456)	0.649	160	−5.658 (−12.43, 1.112)	0.101
EBOP[a,b,c,d]	160	−1.017 (−4.090, 2.055)	0.516	160	−7.941 (−18.92, 3.036)	0.156
PSS[a,b,c,d]	160	4.935 (1.405, 8.465)	0.006*	160	17.620 (4.881, 30.36)	0.007*
CR[a,b,c,d]	160	0.019 (−1.216, 1.255)	0.975	160	−1.471 (−5.915, 2.973)	0.517
TC[a,b,c,d,e,f,g,h]	145	−0.329 (−0.540, −.117)	0.002*	145	0.913 (0.118, 1.708)	0.024*
SC[a,b,c,d,e,f,g,h]	145	−0.288 (−0.510, −.066)	0.011*	145	−1.134 (−1.953, −.314)	0.007*
L[a,b,c,d,e,f,g,h]	145	−0.211 (−0.428, 0.006)	0.056	145	0.091 (−0.724, 0.906)	0.827

Legend: QMSI (Index 0–24), QMCI (Index 0–8), CQII (Index 0–28), SER (Index 0–4), EBOP (Index 0–4), PSS (Index 0–4), CR (Index 0–4), TC (Index 0–100), SC (Index 0–100), Leadership (Index 0–100).
[a]Adjusted for hospital peer group and triage category.
[b–d]Additionally adjusted for: [b]QMSI; [c]QMCI; [d]CQII; [e]SER; [f]EBOP; [g]PSS; and [h]CR.
*Statistically significant at 5%.

tells the system that time is the most important measure, so tasks are at risk of being performed with time in mind rather than quality of care.

One of the most interesting negative associations, and one which reinforces this conjecture, was that the median waiting time increased by 5 min and the median LOS increased by 18 min for every unit increase in PSS score (range 0–4). Examples of patient safety strategies that comprise the PSS include identifying patients by bracelet, providing readily accessible hand sanitizer and sharps disposal boxes at the point of patient care, ensuring each emergency crash cart has a completed checklist of equipment and supplies, and providing a system to report clinical incidents (adverse events). These specific activities are a general measure of an underlying construct relating to commitment to patient safety, which is associated with greater LOS. While we know that implementation strategies such as these take time, we now have data that shows attention to patient safety items can be translated to longer waits for treatment and longer stays in the ED. This finding highlights the need for measures that place emphasis on or prioritize quality of care rather than being solely time based—in order to provide a richer understanding of how an ED is performing.

There were also significant associations between TC, SC and ED waiting time and LOS performance. A one-unit change in TC (range 0–5) was associated with a 9-min *decrease* in waiting time and a 23-min *increase* in length of stay. This result implies that, with better communication and teamwork, patients are seen quicker but spend longer in the department. There is potential for the increased LOS to be explained by greater attention to care: we know that communication and therefore teamwork takes time—especially if the team member is a phone call away, as is often the case for treatment in an ED. Teamwork and patient safety are related [23, 24], hence longer stays for improved teamwork also aligns with our finding that patient safety strategies are associated with longer stays. Further research is required to better understand this finding.

A one-unit change in SC (range 0–5) was equivalent to a 7-min *decrease* in waiting time and a 28-min *decrease* in LOS. These results are not surprising (except perhaps in their magnitude) as SC covers aspects, such as sufficient staffing to handle the number of patients, which are vital for EDs to function effectively.

Of note was the number of self-identified leaders that participated in our study (61% in the ED). Perhaps it is not surprising that self-identified leaders are more likely to be interested in participating

in research about the performance of the ED and, therefore, self-select into our study. Within an ED, formally appointed leaders are also more likely to be able to find time to participate. While participation of a large proportion of self-identified leaders had potential to introduce selection bias [25], we asked the same questions of leaders and non-leaders in regard to safety culture and leadership. Additionally, we did not find a statistically significant difference in responses between clinicians and clinician-leaders.

Limitations

The DUQuA QMSI and clinician safety culture and leadership scales are based on self-reported data. In mitigation, both scales have been validated [18, 19], and the QMCI and CQII scales are based on external audit of hospital quality processes. There was limited variation in the hospital-level characteristics, and we had unequal numbers of clinicians participating. In addition, we had substantially fewer participating hospitals than the DUQuE study and this limits our statistical power to detect smaller associations, especially when controlling for the multiple factors specified in the DAG. While we assumed (based on other large-scale studies) [26] that characteristics associated with our measures would be relatively stable over time, it should be noted that some measures were collected over a 2-year period. In addition, patient flow measures were collected before safety culture and leadership measures, so it is possible that a change in leadership after patient outcomes were collected may have influenced our results.

The low reliability of the SER and PSS scales, discussed in another paper in this Supplement [18], has potential implications for the current findings, as the low reliability can be interpreted as random measurement error. Where SER and PSS were the exposure variables of interest, the statistical significance of associations (i.e. with clinician scales and ED waiting time and length of stay) are likely to be systematically underestimated by an unknown amount. Random measurement error in SER and PSS considered as covariates, however, has unpredictable impacts on the assessment of other associations (i.e. may either increase or decrease the statistical significance); the statistical significance of associations between clinician scales and ED waiting time and length of stay, which adjust for SER and PSS, should therefore be interpreted with caution.

Implications for research, policy and practice

After nearly 10 years of NEAT, ETP and the 4-hour rule and multiple models of care to address patient flow to reduce overcrowding, it is not enough that EDs do not operate in isolation. This study shows that the hospital-level quality structures affect the delivery of urgent care in the ED. Therefore, future strategies and policies to address patient flow in EDs need to be implemented at the organization level as ED performance is a hospital-wide issue. The study also has implications for the measurement and interpretation of ED performance. Current time-based measures are influenced by patient safety strategies. As such, prioritization of other measures of ED performance are needed to give a complete picture of performance, otherwise we put patient safety at risk by biasing our measures to time.

The findings also have practical significance for training. Implementation of Crew Resource Management-based teamwork training in EDs, for example, has been shown to increase SAQ TC scores by more than one point (on a scale of 1–5), when scores were measured prior to and 8 months following completion of training [27]. Our findings suggest that time spent in training is time saved in practice.

Conclusions

Hospital quality structures influence ED quality strategies and, ultimately, waiting time and length of stay for patients. By focusing on time-based measures, we risk punishing EDs that perform well on patient safety measures. There is a need for organizations to recognize the trade-offs required to meet patient safety standards and that this should be considered when judging performance. Further research is needed to better understand the relationships between quality strategies and patient flow in the ED, and to develop more suitable ED performance measures. In the interim, there is a need for EDs to adopt creative approaches to ensuring waiting times meet standards while adequate safety and quality activities are maintained.

Supplementary material

Supplementary material is available at *INTQHC Journal* online.

Acknowledgements

Members of the original DUQuE team (Oliver Gröne, Rosa Suñol) provided extensive input and advice in the initial stages of the project, supporting design modifications from DUQuE and the planned DUQuA study approach. Dr Annette Pantle provided expert advice on revision and development of DUQuA measures, and Professor Sandy Middleton, Associate Professor Dominique Cadillac, Kelvin Hill, Dr Carmel Crock, and Professor Jacqueline Close provided input into the development of our stroke, AMI, and hip fracture clinical process indicator lists. Nicole Mealing and Victoria Pye provided statistical advice in the initial phases of the project. We greatly appreciate their efforts.

Funding

This research was funded by National Health and Medical Research Council (NHMRC) Program Grant APP1054146 (CI Braithwaite).

Authors' contributions

The research team consists of experienced researchers, clinicians, biostatisticians and project managers with expertise in health services research, survey design, and validation, large-scale research and project management, sophisticated statistical analysis, quality improvement and assessment, accreditation, clinical indicators, policy and patient experience. JB conceived the idea, led the research grant to fund the project and chairs the steering committee. RCW and NT co-led the detailed study design, managed the project across time and contributed to the development of the manuscript. EA assisted in data interpretation, and contributed to development of the manuscript. HPT and GA provided statistical expertise for the study design and developed the analysis plan for the manuscript. TW contributed to the logistics of project management, the refinement of measures and the development of the manuscript.

Ethics approval

Ethical approvals were secured from State and Territory human research ethics committees in New South Wales (#14/206), Victoria (#15/36), the Australian Capital Territory (#15/131), South Australia (#15/260), the National Territory (#15–2509), Tasmania (#H0015383) and Queensland (#15/361). Site-specific authorizations, including permission for external researchers to collect data in hospitals, were granted. We complied with confidentiality requirements of national legislation or standards of practice of each jurisdiction. Separate ethics approval was provided by the AIHW (#EO2017/2/315), for patient-level hospital data.

Data sharing statement

Data will be made publicly available to the extent that individual participants or participating hospitals cannot be identified, in accordance with requirements of the approving Human Research Ethics Committees.

References

1. Australian Institute of Health and Welfare. Emergency department care: Australian hospital statistics 2017-18. Health services series no 89. Cat. No. HSE 216. Canberra: AIHW. https://www.aihw.gov.au/reports/hospitals/emergency-department-care-2017-18/contents/table-of-contents Accessed 28 June 2019.
2. Epstein SK, Huckins DS, Liu SW *et al.* Emergency department crowding and risk of preventable medical errors. *Intern Emerg Med* 2012;**7**:173–80.
3. Trzeciak S, Rivers E. Emergency department overcrowding in the United States: an emerging threat to patient safety and public health. *Emerg Med J* 2003;**20**:402–5.
4. Bernstein SL, Aronsky D, Duseja R *et al.* The effect of emergency department crowding on clinically oriented outcomes. *Acad Emerg Med* 2009;**16**:1–10.
5. Carter EJ, Pouch SM, Larson EL. The relationship between emergency department crowding and patient outcomes: a systematic review. *J Nurs Scholarsh* 2014;**46**:106–15.
6. Richardson DB. Increase in patient mortality at 10 days associated with emergency department overcrowding. *Med J Aust* 2006;**184**:213.
7. Australian Institute of Health and Welfare. Australians health 2018: In brief. Cat. No. AUS 222. Canberra: AIHW. https://www.aihw.gov.au/reports/australias-health/australias-health-2018-in-brief/contents/about (28 June 2019, date last accessed).
8. De Freitas L, Goodacre S, O'Hara R *et al.* Interventions to improve patient flow in emergency departments: an umbrella review. *Emerg Med J* 2018;**35**:626–37.
9. Clinical Access and Redesign Unit. CARU. 2014 [cited 2014 16th January 2015]; Available from: http://www.health.qld.gov.au/caru/improving-services/service-models.asp. (27 June 2019, date last accessed).
10. Sexton JB, Helmreich RL, Neilands TB *et al.* The safety attitudes questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Serv Res* 2006;**6**:44.
11. Taylor N, Clay-Williams R, Hogden E *et al.* High performing hospitals: a qualitative systematic review of associated factors and practical strategies for improvement. *BMC Health Serv Res* 2015;**15**:244.
12. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open* 2013;**3**:e001570.
13. Male L, Noble A, Atkinson J, Marson T. Measuring patient experience: a systematic review to evaluate psychometric properties of patient reported experience measures (PREMs) for emergency care service provision. *Int J Qual Health Care* 2017;**29**:314–26.
14. Sonis JD, Aaronson EL, Lee RY *et al.* Emergency department patient experience: a systematic review of the literature. *Patient Exp J* 2018;**5**:101–6.
15. Taylor C, Benger J. Patient satisfaction in emergency medicine. *Emerg Med J* 2004;**21**:528–32.
16. Sullivan C, Staib A, Griffin B *et al.* The four hour rule: the National Emergency Access Target in Australia, time to review systematic literature review appendix. *Aust Health Rev* 2016;**40**:319.
17. Taylor N, Clay-Williams R, Hogden E *et al.* Deepening our Understanding of Quality in Australia (DUQuA): a study protocol for a nationwide, multilevel analysis of relationships between hospital quality management systems and patient factors. *BMJ Open* 2015;**5**:e010349.
18. Clay-Williams R, Taylor N, Winata T *et al.* Organization quality systems and department-level strategies: refinement of the Deepening our Understanding in Quality in Australia (DUQuA) organization and department-level scales. *Int J Qual Health Care* 2020;**32**:22–34.
19. Clay-Williams R, Taylor N, Ting HP *et al.* The clinician safety culture and leadership questionnaire: refinement and validation in Australian public hospitals. *Int J Qual Health Care* 2020;**32**:52–59.
20. Arnolda G, Winata T, Ting HP *et al.* Implementation and data-related challenges in the Deepening our Understanding of Quality in Australia (DUQuA) study: implications for large-scale cross-sectional research. *Int J Qual Health Care* 2020;**32**:75–83.
21. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ, John Wiley & Sons, 2004.
22. Taylor N, Clay-Williams R, Ting HP *et al.* Do organization-level quality management systems influence department-level quality? A cross-sectional study across 32 large hospitals in Australia. *Int J Qual Health Care* 2020;**32**:35–42.
23. Buljac-Samardzic M, Dekker-van Doorn CM, van Wijngaarden JD *et al.* Interventions to improve team effectiveness: a systematic review. *Health policy* 2010;**94**:183–95.
24. Weaver SJ, Lubomksi LH, Wilson RF *et al.* Promoting a culture of safety as a patient safety strategy: a systematic review. *Ann Intern Med* 2013;**158**:369–74.
25. Valente TW, Pumpuang P. Identifying opinion leaders to promote behavior change. *Health Educ Behav* 2007;**34**:881–96.
26. Clay-Williams R, Nosrati H, Cunningham FC *et al.* Do large-scale hospital-and system-wide interventions improve patient outcomes: a systematic review. *BMC Health Serv Res* 2014;**14**:369.
27. Morey JC, Simon R, Jay GD *et al.* Error reduction and performance improvement in the emergency department through formal teamwork training: evaluation results of the MedTeams project. *BMC Health Serv Res* 2002;**37**:1553–81.

International Journal for Quality in Health Care, 2020, 32(S1), 52–59
doi: 10.1093/intqhc/mzz106
Article

Article

The clinician safety culture and leadership questionnaire: refinement and validation in Australian public hospitals

ROBYN CLAY-WILLIAMS[1], NATALIE TAYLOR[2,3], HSUEN P TING[1], TERESA WINATA[1], GASTON ARNOLDA[1], and JEFFREY BRAITHWAITE[1]

[1]Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia, [2]Cancer Research Division, Cancer Council NSW, 153 Dowling St, Woolloomooloo, NSW 2011, Australia, and [3]Faculty of Health Sciences, University of Sydney, Camperdown, Sydney, NSW 2006, Australia

Address reprint requests to: Jeffrey Braithwaite, Centre for Healthcare Resilience and Implementation Science (CHRIS), Australian Institute of Health Innovation, Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia. Tel: +61 2 9850 2401, +61 414 812 579; Fax: +61 2 9850 2499; E-mail: jeffrey.braithwaite@mq.edu.au

Editorial Decision 7 September 2019; Accepted 13 September 2019

Abstract

Objective: This study aimed to refine and validate a tool to measure safety culture and leadership in Australian hospitals.
Design: The clinician safety culture and leadership questionnaire was constructed by combining and refining the following two previously validated scales: Safety Attitudes Questionnaire and the Leadership Effectiveness Survey. Statistical processes were used to explore the factor structure, reliability, validity and descriptive statistics of the new instrument.
Setting: Thirty-two large Australian public hospitals.
Participants: 1382 clinicians (doctors, nurses and allied health professionals).
Main Outcome Measure(s): Descriptive statistics, structure and validity of clinician safety culture and leadership scale.
Results: We received 1334 valid responses from participants. The distribution of ratings was left-skewed, with a small ceiling effect, meaning that scores were clustered toward the high end of the scale. Using confirmatory factor analysis, we confirmed the structure of the three scales as a combined measure of safety culture and leadership. The data were divided into equal calibration and validation datasets. For the calibration dataset, the Chi-square: df ratio was 4.4, the root mean square error of approximation RMSEA (a measure of spread of the data) was 0.071, the standardized root mean square residual SRMR (an absolute measure of the fit of the data) was 0.058 and the Confirmatory Fit Index (CFI) (another test confirming the fit of the data) was 0.82; while none of the indices suggested good fit, all but CFI fell within acceptable thresholds. All factors demonstrated adequate internal consistency and construct reliability, as desired. All three domains achieved discriminant validity through cross-loadings, meaning that the three domains were determined to be independent constructs. Results for the validation dataset were effectively identical to those found in the calibration dataset.
Conclusions: While the model may benefit from additional refinement, we have validated the tool for measuring clinician safety culture and leadership in our Australian sample. The DUQuA safety

culture and leadership scale can be used by Australian hospitals to assess clinician safety culture and leadership, and is readily modifiable for other health systems depending on their needs.

Key words: hospital quality management systems, Multi-level research, teamwork, safety culture, leadership, quality improvement

Introduction

There is a growing body of evidence linking hospital safety culture with improved patient outcomes [1, 2] and hospital performance [3]. The terms 'safety culture' and 'safety climate' are often used interchangeably in healthcare [4]; however, if we think of 'safety climate' as the perceived value placed on safety within an organization [5], then there are additional dimensions comprising safety culture that include leadership and teamwork [4]. In this framing, a safety culture that contributes to high hospital performance can depend on the combination of a positive safety climate [6], the level of teamwork demonstrated by clinicians [7–9] and effective leadership within the organization [10].

Over the last decade, a number of systematic reviews have reported the development and psychometric properties of a variety of instruments available to assess different aspects of patient safety climate [4, 11–13]. Of the reviewed instruments, the Safety Attitudes Questionnaire (SAQ) [14] is the most commonly used survey tool to assess healthcare workers' perceptions of patient safety related attitudes in various clinical areas and healthcare settings.

There are many versions of the SAQ, including adaptations of the survey for the Intensive Care Unit [15], surgery [16], Emergency Department (ED) [17], pharmacies [18], ambulatory care [19] and so on. Although varieties of the SAQ can include up to 60 items, a minimum of 30 items is required to assess all six subdomains as follows: teamwork climate (six items), safety climate (seven items), job satisfaction (five items), stress recognition (four items), perceptions of management (four items) and working conditions (four items) [14]. A common version, recommended by the University of Texas where the tool originated, is the Safety Climate questionnaire [20], which consists of 27 items, but measures only two domains—14 items under teamwork climate and 13 items under safety climate. Despite the name, the Safety Climate version includes two of the four questions that comprise the perceptions of management domain from the full 60-item SAQ: the quality of collaboration with other clinical professionals; and the importance of briefing and other communication during handover and shift change, which we believe from the literature to be important for safe care [21, 22].

To measure safety culture in participating hospitals for the Deepening our Understanding of Quality in Australia (DUQuA) study [23], we sought a valid and reliable tool that measured safety culture across the components of teamwork, safety climate and leadership in healthcare. In particular, we needed a measurement instrument that was easy to administer and broadly applicable across a variety of hospital settings. Clinicians in Australian hospitals are already heavily burdened with requests to complete surveys (both mandatory and voluntary). Furthermore, hospital work can be piecemeal, with frequent interruptions [24, 25]. In EDs, for example, clinicians are interrupted on average 6.6 times/h, and when interrupted failed to return to 18.5% of interrupted tasks [25]. Other studies have shown interruptions and multitasking to be prevalent throughout the hospital [24]. Therefore, we needed a questionnaire that could be completed quickly in a single sitting. Although we found an abundance of tools to measure safety culture, teamwork and leadership, we did not find any instrument where all three aspects were measured using a single survey. This paper reports on the development and validation of a tool to measure safety culture and leadership in Australian hospitals.

Method

Participants

DUQuA used purposive sampling. Clinicians were recruited from participating DUQuA hospitals in New South Wales, Victoria, Northern Territory, Queensland, South Australia, Tasmania and the Australian Capital Territory. Local Principal Investigators in each hospital informed the department head and/or clinical leaders and clinical staff on participating wards about the study and invited them to participate in the research through email, workshops and/or meetings and verbal conversations. Participating departments included the ED and departments treating acute myocardial infarction (AMI), stroke and hip fracture. Doctors, nurses or allied health professionals practising in participating departments at least 50% of their work time were eligible to participate. Participants could complete the survey electronically or on paper.

Measures

As with the development of the other scales used in DUQuA, we consulted with lead researchers on the equivalent scale used in the Deepening our Understanding of Quality improvement in Europe (DUQuE) study. In DUQuE [26, 27], teamwork and safety climate were measured using the Teamwork (six items) and Safety Climate (seven items) domains from the SAQ [14]. The composite scale reliability for the SAQ has been reported as 0.90 (Raykov's ρ coefficient), indicating strong reliability [14]. For the DUQuA safety culture measure, we adopted the Australian version of the SAQ [28], but rather than just the 13 questions comprising the safety climate and teamwork climate domains of the SAQ, we used the full 27-item Safety Climate survey. Within this survey, teamwork climate measures interdisciplinary support and collaboration. Safety climate measures clinician and institutional response to error. As this version includes two of the four items from the perceptions of management domain, we added the other two to provide a partial assessment of leadership. To provide a more comprehensive assessment of leadership, we also added Shipton *et al.*'s [29] six-item Leadership Effectiveness Scale (LES). The LES was validated in a survey of ~18,000 employees of the UK National Health Service (Cronbach's $\alpha = 0.92$) [29], which measures staff perceptions of the vision, inclusiveness and internal and external collaborative behaviours of healthcare leaders. Higher scores on the LES have been associated with higher hospital performance and fewer patient complaints [29]. All items were combined into one 35-item DUQuA Safety Culture and Leadership questionnaire, consisting of three factors: teamwork climate ($n = 14$ items), safety climate ($n = 13$) and leadership ($n = 8$).

The 35-item Safety Culture and Leadership Questionnaire (Figure 1) was used to measure the following three domains:

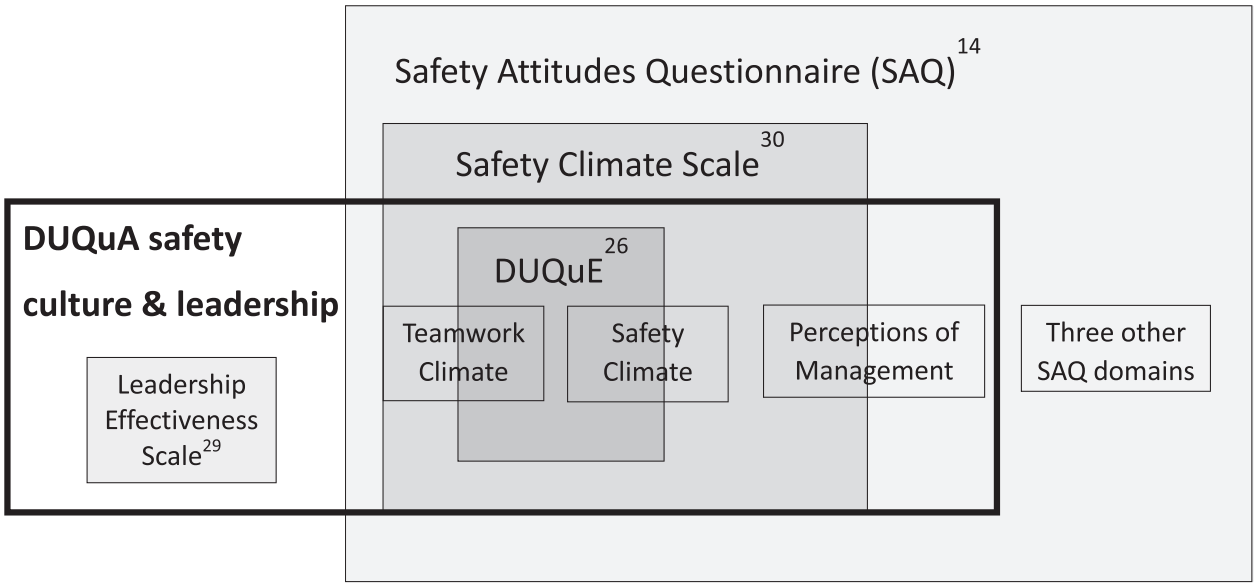

Figure 1 Composition of the DUQuA safety culture and leadership questionnaire.

perceived quality of collaborating between clinicians; organizational commitment to patient safety and management style and the effectiveness of healthcare leaders in their workplace. Each item was scored by clinicians on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Data analysis

Data were analysed in IBM SPSS Amos 25, IBM SPSS Statistics 25 (both Armonk, New York, USA) and SAS/STAT 9.4 (Cary, North Carolina, USA). Demographic characteristics of the clinicians were analysed and the scores of the items and domains were calculated. Scores on negatively-worded items were first reversed and item scores were rescaled from a one to five-point scale to a zero to 100-point scale, as per published procedures for analysing the SAQ [30]. Observations with any factor with >50% missing items were deleted. Missing items were examined using Little's test [31], i.e. a non-significant test result indicates data are missing completely at random (MCAR); a significant test result indicates that data are not MCAR and leads to visual examination of the data to determine if the missingness pattern indicates that data are missing at random (MAR) or missing not at random (MNAR). Imputation using the Expectation-Maximization (EM) method assumes that the missing data pattern is at least MAR [32]. Domain scores were calculated as the mean score of the items that made up that domain.

The sample was randomly split into halves. Confirmatory factor analysis (CFA) was used to test the three-domain theoretical framework with the calibration sample and reassessed in the validation sample. CFA determines how well the items represent the domains in our safety culture and leadership questionnaire. Maximum likelihood estimation was used to perform CFA [33]. This method assumes multivariate normality, which was assessed using Mardia's coefficients of multivariate skewness and kurtosis, with statistically significant results for these tests indicating non-normality. In the event of multivariate non-normality, appropriate transformations were attempted.

To assess convergent validity (whether the items in the scale load onto the intended factor), we followed guidance from a review by Hooper *et al*. [34]. We used three recommended absolute fit indices (Chi-square score, root mean square error of approximation [RMSEA] and standardized root mean squared residual [SRMR]) and one incremental fit index (comparative fit index [CFI]) to assess model fit [35]. Absolute fit indices compare a pre-specified model to no model, while incremental fit indices compare the pre-specified model to a model where all items are uncorrelated (a worst case). Each of the indices has strengths and weaknesses.

First, we consider the absolute fit indices, all of which assess different measures of the differences between the sample data and the hypothesized model. A statistically significant Chi-square statistic (P <0.05) indicates poor model fit. As the chi-square statistic is routinely statistically significant with large sample sizes, we added the Chi-square:degrees of freedom (df) ratio, which is insensitive to sample size, with a desirable value <2 [36], but with values of <5 considered acceptable [37]. Chi-square is also sensitive to non-normality, leading to higher scores even when the model is appropriately specified [38]. The calculation of RMSEA includes the chi-square score, the df and the sample size; while RMSEA calculation controls for large sample size, inflation of the chi-square statistic due to non-normality can lead to higher values. RMSEA <~0.06 [39] has been suggested as indicating a good fit, but lower thresholds of <0.08 [40] have also been proposed. SRMR compares the observed and hypothesized models and is sensitive to small sample size; good fitting models of SRMR have values of <0.05 [41, 42], with values <0.08 deemed acceptable [39].

The sole incremental fit index, CFI assesses the Chi-square (adjusted for df) as a ratio of the target to the null model [43]. In a simulation study, CFI appears to be most resilient to deviations from normality and it is known to be robust to small sample size [44]. A CFI of 0.95 or above is considered a good fit, but earlier recommendations suggested >0.90 is acceptable [39].

The thresholds used to assist with assessing model fit for convergent validity, as a signal for a possible need for model review were: Chi-square:df >2 [37], CFI <0.95, RMSEA >0.06

and SRMR >0.05. Where model review was indicated, modification was guided by modification indices, factor loadings and multiple R-squared. Items with factor loading of <0.32 [36] and multiple R-square <0.2 [34] were considered for deletion or model respecification, with the decision to modify principally guided by theoretical considerations.

To assess internal consistency reliability (whether individual subscale items demonstrate good relationships with one another), Cronbach's α and construct reliability were used. Cronbach's α of >0.8 [45] was used to indicate adequate internal consistency, and construct reliability of >0.6 [46, 47] was considered adequate. To assess discriminant validity (whether theoretically different subscales are distinguishable from one another), we used cross-loadings; each item's factor loadings on the assigned construct should be higher than loading on any other constructs [48].

Results

Questionnaire forms were received from 1382 clinicians. A total of 48 responses (3.5%) were excluded due to insufficient data. Each participating department (n = 120, comprising 32 ED, 27 AMI, 29 hip fracture and 32 stroke) was expected to undertake 30 clinician assessments (i.e. 3600 in total).

Characteristics of the participants

Characteristics of the participants are summarized in Table 1. There were 1334 valid responses (37.1% of target) from 31 hospitals. Responses were received from 25 departments that treated AMI patients (n = 309 participants); 26 departments treating hip fracture (n = 307); 25 departments treating stroke (n = 324) and 29 EDs (n = 394). Almost half of the participants were under 35 years old and 61% were female. The majority of the participants were either a physician (38%) or nurse (52%), and almost half identified themselves as leaders.

Table 1 Summary of clinician characteristics

Characteristic	*n*	Percent (%)
Gender		
Male	517	38.8
Female	817	61.2
Age group (years)		
18–24	105	7.9
25–34	524	39.3
35–44	323	24.2
45–54	246	18.4
55–64	118	8.8
65–74	18	1.3
Leader		
Yes	631	47.3
No	703	52.7
Profession		
Physician	511	38.3
Nurse	695	52.1
Allied Health	128	9.6
Department		
AMI	309	23.2
Hip Fracture	307	23.0
Stroke	324	24.3
ED	394	29.5

Missing data

Out of the 1334 valid responses, the percentage of missing data was 0.5% (range = 0–1.4% per item). Little's MCAR test [31] was significant (P-value <0.001), indicating data were potentially not MCAR; however, visual inspection of missing data patterns suggested the data were MAR, permitting imputation.

Descriptive statistics

The descriptive statistics of the factor and item scores are summarized in Table 2. The mean for the 35 items ranged from 57.4 to 87.3. The mean domain scores were 78.9 for teamwork climate, 71.7 for safety climate and 67.9 for leadership. Item distributions were mostly left-skewed and kurtotic, resulting in significant variation from multivariate normality (both Mardia's coefficients of multivariate skewness and kurtosis had P <0.001). Transformations were attempted but did not achieve multivariate normality.

Safety Culture and Leadership scale: structure and validity

Calibration dataset. We used CFA to confirm the structure for the three scales as a combined measure of safety culture and leadership. Only one item failed to meet our preferred criteria: item Q95.9 'I know the first and last names of all the staff I worked with during my last shift' had factor loading <0.32 and multiple R-squared <0.2. The item was considered for deletion, but retained as deletion of the item did not significantly improve the model fit indices and it did not make theoretical sense to respecify the item in the model, as it has been part of the SAQ, and its antecedent the Flight Management Attitudes Questionnaire (FMAQ), for over two decades (note: the SAQ version was derived from the FMAQ item 'working here is like being part of a large family') [49]. We, therefore, retained the pre-specified theoretical model.

Table 3 summarizes the key model fit statistics in the calibration dataset. The Chi-square P-value was <0.001, the Chi-square/df ratio was 4.4, the CFI was 0.82 and RMSEA was 0.07 and the SRMR was 0.06; except for CFI, all model fit indices were within the acceptable target. Convergent validity measures are shown in Table 4. All factors demonstrated adequate internal consistency and construct reliability. All three pairs of domains achieved discriminant validity through cross-loadings (Supplement eTable A1).

Validation dataset. The final model (eFigure A1) is reported in eTable A1. Fit statistics are shown detailed in Table 3; the results are effectively identical to those found in the calibration dataset, with all model fit indices except for CFI met the acceptable range. The convergent reliability statistics are detailed in Table 4 with, once again, virtually identical results. All domain pairs achieved discriminant validity through cross-loadings (eTable A2).

Discussion

Interpretation of results

We sought to refine and validate a tool to measure teamwork climate, safety climate and leadership in Australian hospitals. Our tool was developed from a strong theoretical base, informed by the evidence underpinning the SAQ and LES surveys from which it was derived.

Table 2 Descriptive statistics of domain and item scores

Factor and items	*n*	Mean	SD	Median	Min	Max	Floor[a] (%)	Ceiling[a] (%)
Teamwork climate	1237	78.9	12.57	80.4	17.9	100		
Q95.1 Nurse input (defined as the views and suggestions about patient care made by nurses) is well received in my clinical area	1322	83.4	19.28	75	0	100	1	48
Q95.2 In my clinical area, it is difficult to speak up if I perceive a problem with patient care[b]	1321	79.1	25.07	75	0	100	2	45
Q95.3 Decision-making in my clinical area utilizes input from relevant staff	1324	83.5	19.61	75	0	100	1	48
Q95.4 The staff in my department work together as a well-coordinated team	1321	84.3	19.75	100	0	100	1	51
Q95.5 Disagreements in my clinical area are resolved appropriately (i.e. not who is right but what is best for the patient)	1325	75.9	22.21	75	0	100	2	32
Q95.6 I am frequently unable to express disagreement with senior staff[b]	1327	69.1	28.77	75	0	100	4	31
Q95.7 It is easy for staff here to ask questions when there is something that they do not understand.	1326	85.8	19.58	100	0	100	1	56
Q95.8 I have the support I need from other staff to care for patients	1327	84.7	19.11	100	0	100	1	51
Q95.9 I know the first and last names of all the staff I worked with during my last shift	1325	58.8	35.44	75	0	100	14	27
Q95.10 Important issues are well communicated at shift changes	1316	74.1	21.31	75	0	100	1	26
Q95.11 Briefing staff before the start of a shift (i.e. to plan for possible contingencies) is important for patient safety	1328	87.3	19.09	100	0	100	1	62
Q95.12 Briefings are common in my clinical area	1329	75.8	24.78	75	0	100	2	37
Q95.13 I am satisfied with the quality of collaboration that I experience with doctors in my clinical area	1332	77.1	22.36	75	0	100	1	34
Q95.14 I am satisfied with the quality of collaboration that I experience with nurses in my clinical area	1331	83.1	17.85	75	0	100	0	44
Safety climate	1251	71.7	13.91	73.1	17.3	100		
Q96.1 The levels of staffing in my clinical area are sufficient to handle the number of patients	1328	57.4	30.98	75	0	100	11	16
Q96.2 I would feel safe being treated here as a patient	1325	76.9	23.36	75	0	100	2	37
Q96.3 I am encouraged by my colleagues to report any patient safety concerns I may have	1325	83.7	19.10	75	0	100	1	49
Q96.4 Staff frequently disregard rules or guidelines (e.g. handwashing, treatment protocols/clinical pathway, sterile field, etc.) that are established for my clinical area[b]	1325	74.3	27.00	75	0	100	3	38
Q96.5 The culture in my clinical area makes it easy to learn from the errors of others	1333	69.7	22.51	75	0	100	3	19
Q96.6 I receive appropriate feedback about my performance	1326	69.7	25.10	75	0	100	3	24
Q96.7 Medical errors are handled appropriately here	1330	77.7	21.01	75	0	100	1	35
Q96.8 I know the proper channels to direct questions regarding patient safety in my clinical area	1330	82.6	18.49	75	0	100	1	43
Q96.9 In my clinical area, it is difficult to discuss errors[b]	1327	72.9	26.45	75	0	100	3	33
Q96.10 Hospital management does not knowingly compromise the safety of patients	1323	67.3	29.36	75	0	100	6	30
Q96.11 The hospital is doing more for patient safety now, than it did one year ago	1317	62.9	23.48	50	0	100	3	15
Q96.12 Leadership is driving us to be a safety-centred hospital	1329	68.5	23.27	75	0	100	2	21
Q96.13 My suggestions about safety would be acted upon if I expressed them to management	1326	66.8	24.56	75	0	100	3	20
Leadership	1313	67.9	19.09	68.8	0	100		
Q97.1 Hospital administration supports my daily efforts	1330	61.0	25.75	75	0	100	5	14
Q97.2 I am provided with adequate, timely information about events in the hospital that might affect my work	1329	65.8	24.47	75	0	100	3	17
Q97.3 The leadership team in my department describes exciting new opportunities for the organization	1331	66.9	24.23	75	0	100	3	19
Q97.4 The leadership team in my department proposes new and creative ideas for improving services or processes	1331	69.4	23.78	75	0	100	3	22
Q97.5 The leadership team in my department is effectively leading the organization to meet patient needs and care for patient safety	1334	74.6	22.05	75	0	100	2	28
Q97.6 The leadership team in my department takes account of both service requirements and staff needs when implementing major changes	1330	69.3	24.41	75	0	100	3	22
Q97.7 The leadership team in my department builds strong and positive relationships with the community	1332	67.2	23.19	75	0	100	2	19
Q97.8 The leadership team in my department builds strong, co-operative links with other organizations	1332	68.2	22.43	75	0	100	2	19

[a]The % floor/ceiling represents the percentage of all records with the lowest/highest score possible.
[b]Negatively-worded item; the score has been reversed.

Table 3 Final model fit statistics on calibration and validation dataset

	Preferred target	Acceptable target	Calibration $n = 667$	Validation $n = 667$
Chi-square	NA	NA	2433.7	2461.7
DF	NA	NA	557	557
P-value	>0.05	NA	<0.001	<0.001
CMIN/DF	<2	<5	4.4	4.4
RMSEA	<~ 0.06	≤0.08	0.071	0.072
SRMR	<0.05	≤0.08	0.058	0.062
CFI	>0.95	>0.9	0.82	0.81

NA, Not applicable.

Table 4 Convergent validity measures on calibration and validation dataset

	Calibration		Validation	
Factor	Construct reliability	Cronbach's α	Construct reliability	Cronbach's α
Teamwork climate	0.86	0.83	0.85	0.82
Safety climate	0.84	0.83	0.84	0.82
Leadership	0.92	0.92	0.92	0.92

Table 5 SAQ component comparison with other studies

SAQ domain	DUQuA ($n = 1334$) mean (SD)	DUQuE [26] ($n = 8525$) mean (SD)	South Australia [28] ($n = 10468$) mean (SD)
Teamwork climate	82.3 (14.2)	71.8 (17.5)	77.8 (14.2)
Safety climate	76.2 (14.7)	68.6 (16.8)	78.8 (13.0)
Perceptions of management	62.9 (19.5)	–	68.8 (16.8)

While, in our sample, we found the distribution of ratings to be asymmetrical, and skewed toward higher scores, it is a common occurrence for SAQ data in healthcare to be distributed this way. Previous studies [44, 50, 51] have shown that multivariate non-normality such as this leads to higher Chi-square and RMSEA scores, even when the models were correctly specified, potentially explaining the failure to meet the desired thresholds for chi-square and RMSEA in our study. The CFI indices (0.82–0.83) were lower than the preferred threshold (>0.95) and acceptable range (>0.90). The other indices were all outside the preferred threshold for good fit, but within the acceptable range: chi-square:df ratio <5; RMSEA <0.08; and SRMR <0.08.

In addition, the domains of the proposed 35-item combined scale showed good convergent validity, as consistently demonstrated by construct reliability and Cronbach's α. Item cross-loadings consistently placed the questions in the domains specified in the pre-specified theoretical model. Although the model may benefit from additional refinement, we found the model to be adequate to assess clinician safety culture and leadership for DUQuA.

We were able to compare our DUQuA findings with those from DUQuE and also from a previous South Australian SAQ study [28], by calculating aggregated scores for the three SAQ domains of teamwork climate (six items), safety climate (seven items) and perceptions of management (four items): see Table 5. While there is some variation, the scores for teamwork climate, safety climate and perceptions of management are broadly comparable across the three studies. This aligns with other research on the SAQ, which has found variation within countries to be higher than variation between countries for the SAQ [14, 52].

Limitations

The DUQuA safety culture and leadership scale is based on self-reported data, with its inherent limitations. While we have shown the genesis of the scale, and its development from the internationally validated SAQ and LESs, the DUQuA version of this scale has only been validated for the Australian context. Differences between the DUQuA and DUQuE scales must be taken into account when comparing findings from Australian and European hospitals. While our final scale consists of 35 items, it assesses both safety culture and leadership in one instrument; additionally, we found it took clinicians only 5–10 min to complete.

Implications for research, policy and practice

It can be difficult for hospitals to access validated measurement tools for assessing healthcare safety culture that are powerful, yet easy to use. This leads to a variety of unvalidated tools being used in healthcare, making results difficult to interpret and to compare. In contrast, and as a consequence of our assessment, we are aware of the statistical limitations of the DUQuA clinician safety culture and leadership tool, and hence we have direction for its targeted refinement. Furthermore, whereas there are a number of tools validated for healthcare that measure safety culture [4, 11, 12, 53] and teamwork [13], ours is the first we are aware of the combines safety culture and leadership into one short survey. The scales could be used to collect longitudinal data to assess change over time, to select the focus for an intervention, to assess change associated with an intervention, or used as part of a comparison of hospitals or hospital departments.

Conclusions

The DUQuA safety culture and leadership scale is available to assist hospital managers to measure and improve safety culture at organizational and departmental levels within their hospitals. The tool is modifiable for other healthcare organizations and systems, depending on their needs.

Supplementary material

Supplementary material is available at *INTQHC Journal* online.

Acknowledgements

Members of the original DUQuE team (Oliver Gröne, Rosa Suñol) provided extensive input and advice in the initial stages of the project, supporting design modifications from DUQuE and the planned DUQuA study approach. Dr Annette Pantle provided expert advice on revision and development of DUQuA measures, and Professor Sandy Middleton, Associate Professor Dominique Cadillac, Kelvin Hill, Dr Carmel Crock and Professor Jacqueline Close provided input into the development of our stroke, AMI and hip fracture clinical process indicator lists. Nicole Mealing and Victoria Pye provided statistical advice in the initial phases of the project. We greatly appreciate their efforts.

Funding

This research was funded by National Health and Medical Research Council (NHMRC) Program Grant APP1054146 (CI Braithwaite).

Contributions

The research team consists of experienced researchers, clinicians, biostatisticians and project managers with expertise in health services research, survey design and validation; large-scale research and project management, sophisticated statistical analysis, quality improvement and assessment, accreditation, clinical indicators, policy and patient experience. JB conceived the idea, led the research grant to fund the project, and chairs the steering committee. RCW and NT co-led the detailed study design, managed the project across time and contributed to the development of the manuscript. HPT and GA provided statistical expertise for the study design and developed the analysis plan for the manuscript. TW contributed to the logistics of project management, the refinement of measures and the development of the manuscript.

Ethics approval

Ethical approvals were secured from State and Territory human research ethics committees in New South Wales (#14/206), Victoria (#15/36), the Australian Capital Territory (#15/131), South Australia (#15/260), the National Territory (#15–2509), Tasmania (#H0015383) and Queensland (#15/361). We complied with confidentiality requirements of national legislation or standards of practice of each jurisdiction.

Data-sharing statement

Data will be made publicly available to the extent that individual participants or participating hospitals cannot be identified, in accordance with requirements of the approving Human Research Ethics Committees.

References

1. Braithwaite J, Herkes J, Ludlow K *et al*. Association between organisational and workplace cultures, and patient outcomes: systematic review. *BMJ Open* 2017;**7**:e017708.
2. Odell DD, Quinn CM, Matulewicz RS *et al*. Association between hospital safety culture and surgical outcomes in a statewide surgical quality improvement collaborative. *J Am Coll Surg* 2019; S1072–7515.
3. Mannion R, Davies HT, Marshall M. Cultural characteristics of "high" and "low" performing hospitals. *J Health Organ Manag* 2005;**19**:431–9.
4. Halligan M, Zecevic A. Safety culture in healthcare: a review of concepts, dimensions, measures and progress. *BMJ Qual Saf* 2011;**20**: 338–43.
5. Neal A, Griffin MA, Hart PM. The impact of organizational climate on safety climate and individual behavior. *Saf Sci* 2000;**34**:99–109.
6. Singer S, Lin S, Falwell A *et al*. Relationship of safety climate and safety performance in hospitals. *Health Serv Res* 2009;**44**:399–421.
7. Buljac-Samardzic M, Dekker-van Doorn CM, van Wijngaarden JD *et al*. Interventions to improve team effectiveness: a systematic review. *Health Policy* 2010;**94**:183–95.
8. Manser T. Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. *Acta Anaesthesiol Scand* 2009;**53**: 143–51.
9. O'Leary KJ, Sehgal NL, Terrell G *et al*. High performance teams and the Hospital of the Future Project Team. Interdisciplinary teamwork in hospitals: a review and practical recommendations for improvement. *J Hosp Med* 2012;**7**:48–54.
10. Taylor N, Clay-Williams R, Hogden E *et al*. High performing hospitals: a qualitative systematic review of associated factors and practical strategies for improvement. *BMC Health Serv Res* 2015;**15**:244.
11. Alsalem G, Bowie P, Morrison J. Assessing safety climate in acute hospital settings: a systematic review of the adequacy of the psychometric properties of survey measurement tools. *BMC Health Serv Res* 2018;**18**: 353.
12. Colla J, Bracken A, Kinney L *et al*. Measuring patient safety climate: a review of surveys. *BMJ Qual Saf* 2005;**14**:364–6.
13. Valentine MA, Nembhard IM, Edmondson AC. Measuring teamwork in health care settings: a review of survey instruments. *Med Care* 2015;**53**:e16–30.
14. Sexton JB, Helmreich RL, Neilands TB *et al*. The safety attitudes questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Serv Res* 2006;**6**:44.
15. Thomas EJ, Sexton JB, Helmreich RL. Discrepant attitudes about teamwork among critical care nurses and physicians. *Crit Care Med* 2003;**31**:956–9.
16. Flin R, Yule S, McKenzie L *et al*. Attitudes to teamwork and safety in the operating theatre. *Surgeon* 2006;**4**:145–51.
17. Kaissi A, Johnson T, Kirschbaum MS. Measuring teamwork and patient safety attitudes of high-risk areas. *Nurs Econ* 2003;**21**:211.
18. Nordén-Hägg A, Sexton JB, Kälvemark-Sporrong S *et al*. Assessing safety culture in pharmacies: the psychometric validation of the safety attitudes questionnaire (SAQ) in a national sample of community pharmacies in Sweden. *BMC Clin Pharmacol* 2010;**10**:8.
19. Modak I, Sexton JB, Lux TR *et al*. Measuring safety culture in the ambulatory setting: the safety attitudes questionnaire—ambulatory version. *J Gen Intern Med* 2007;**22**:1–5.
20. Hutchinson A, Cooper K, Dean J *et al*. Use of a safety climate questionnaire in UK health care: factor structure, reliability and usability. *BMJ Qual Saf* 2006;**15**:347–53.
21. McCulloch P, Rathbone J, Catchpole K. Interventions to improve teamwork and communications among healthcare staff. *Bri J Surg* 2011;**98**:469–79.
22. Sutcliffe KM, Lewton E, Rosenthal MM. Communication failures: an insidious contributor to medical mishaps. *Acad Med* 2004;**79**:186–94.
23. Taylor N, Clay-Williams R, Hogden E *et al*. Deepening our understanding of quality in Australia (DUQuA): a study protocol for a nationwide,

multilevel analysis of relationships between hospital quality management systems and patient factors. *BMJ Open* 2015;**5**:e010349.
24. Kalisch BJ, Aebersold M. Interruptions and multitasking in nursing care. *Jt Comm J Qual Patient Saf* 2010;**36**:126–32.
25. Westbrook JI, Woods A, Rob MI *et al.* Association of interruptions with an increased risk and severity of medication administration errors. *Arch Intern Med* 2010;**170**:683–90.
26. Kristensen S, Hammer A, Bartels P *et al.* Quality management and perceptions of teamwork and safety climate in European hospitals. *International J Qual Health Care* 2015;**27**:499–506.
27. Secanell M, Groene O, Arah OA *et al.* Deepening our understanding of quality improvement in Europe (DUQuE): overview of a study of hospital quality management in seven countries. *International J Qual Health Care* 2014;**26**:5–15.
28. Braithwaite J, Westbrook M, Pirone C *et al. Staff Survey on Patient Safety. South Australian Council for Safety and Quality in Health Care and South Australian Department of Health*. Sydney, AU, Australian Institute of Health Innovation, Sydney, AU, 2009.
29. Shipton H, Armstrong C, West M *et al.* The impact of leadership and quality climate on hospital performance. *Int J Qual Health Care* 2008;**20**:439–45.
30. Sexton JB, Thomas EJ. *The Safety Climate Survey: Psychometric and Benchmarking Properties*. Austin, TX, The University of Texas, 2005.
31. Little RJ. A test of missing completely at random for multivariate data with missing values. *J Am Stat Assoc* 1988;**83**:1198–202.
32. Howell D. The analysis of missing data. In: Outhwaite W, Turner S (eds.). *Handbook of Social Science Methodology*. London, UK, SAGE Publishing, 2008.
33. Mardia KV. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 1970;**57**:519–30.
34. Hooper D, Coughlan J, Mullen M. Structural equation modelling: guidelines for determining model fit. *EJBRM* 2008;**6**:53–60.
35. Kline RB. *Principles and Practice of Structural Equation Modeling*. New York, NY, The Guilford Press, 2015.
36. Tabashnick BG, Fidell LS. *Using Multivariate Statistics*, 5th edn. Boston, MA, Ally & Bocon, 2007.
37. Wheaton B, Muthen B, Alwin DF *et al.* Assessing reliability and stability in panel models. *Sociol Methodol* 1977;**8**:84–136.
38. McIntosh CN. Rethinking fit assessment in structural equation modelling: a commentary and elaboration on Barrett. *Pers Individ Differ* 2007;**42**:859–67.
39. Hu Lt BPM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;**6**:1–55.
40. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychol Methods* 1996;**1**:130.
41. Byrne BM. *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming*. Mahwah, NJ, Lawrence Erlbaum Associates, 1998.
42. Diamantopoulos A, Siguaw J. *Introducing LISREL: a Guide for Structural Equation Modeling*. SAGE Publications, Cornell University, CA, 2000.
43. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull* 1990;**107**:238.
44. Hutchinson SR, Olmos A. Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. *Struct Equ Modeling* 1998;**5**:344–64.
45. Nunnally JC. An overview of psychological measurement. In: Wolman BB (ed.). *Clinical Diagnosis of Mental Disorders*. Boston, MA, Springer, 1978, 97–146.
46. Bagozzi RP, Yi Y. On the evaluation of structural equation models. *J Acad Market Sci* 1988;**16**:74–94.
47. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *J Market Res* 1981;**18**: 39–50.
48. Chin WW. The partial least squares approach to structural equation modeling. In: Marcoulides GA (ed.). *Methodology for Business and Management: Modern Methods for Business Research*. Mahwah, NJ, Lawrence Erlbaum Associates Publishers, 1998, 295–336.
49. Helmreich RL, Merritt AC, Sherman PJ, Gregorich SE, Wiener EL. *The flight management attitudes questionnaire (FMAQ)*. Austin, TX: University of Texas NASA/UT/FAA. 1993; 9394.
50. Muthén B, Kaplan D. A comparison of some methodologies for the factor analysis of non-normal Likert variables. *Br J Math Stat Psychol* 1985;**38**:171–89.
51. Potthast MJ. Confirmatory factor analysis of ordered categorical variables with large models. *Br J Math Stat Psychol* 1993;**46**:273–86.
52. Kristensen S, Christensen KB, Jaquet A *et al.* Strengthening leadership as a catalyst for enhanced patient safety culture: a repeated cross-sectional experimental study. *BMJ Open* 2016;**6**:1–10.
53. Singla AK, Kitch BT, Weissman JS *et al.* Assessing patient safety culture: a review and synthesis of the measurement tools. *J Patient Saf* 2006;**2**:105–15.

International Journal for Quality in Health Care, 2020, 32(S1), 60–66
doi: 10.1093/intqhc/mzz107
Article

Article

Do quality management systems influence clinical safety culture and leadership? A study in 32 Australian hospitals

ROBYN CLAY-WILLIAMS[1], NATALIE TAYLOR[2,3], HSUEN P. TING[1], GASTON ARNOLDA[1], TERESA WINATA[1], and JEFFREY BRAITHWAITE[1]

[1]Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Road, NSW 2109, Australia, [2]Cancer Research Division, Cancer Council NSW, 153 Dowling St, Woolloomooloo, NSW 2011, Australia, and [3]Faculty of Health Sciences, University of Sydney, Camperdown, NSW, 2006, Australia

Address reprint requests to: Jeffrey Braithwaite, Centre for Healthcare Resilience and Implementation Science (CHRIS), Australian Institute of Health Innovation, Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia. Tel: +61-2-9850-2401; Fax: +61-2-9850 2499; E-mail: jeffrey.braithwaite@mq.edu.au

Editorial Decision 7 September 2019; Accepted 12 September 2019

Abstract

Objective: This study aimed to explore the associations between the organization-level quality arrangements, improvement and implementation and department-level safety culture and leadership measures across 32 large Australian hospitals.
Design: Quantitative observational study, using linear and multi-level modelling to identify relationships between quality management systems and clinician safety culture and leadership.
Setting: Thirty-two large Australian public hospitals.
Participants: Quality audit at organization level, senior quality manager at each participating hospital, 1382 clinicians (doctors, nurses and allied health professionals).
Main outcome measures: Associations between organization-level quality measures and department-level clinician measures of teamwork climate, safety climate and leadership for acute myocardial infarction (AMI), hip fracture and stroke treatment conditions.
Results: We received 1332 valid responses from participants. The quality management systems index (QMSI, a questionnaire-based measure of the hospitals' quality management structures) was 'positively' associated with all three department-level scales in the stroke department, with safety culture and leadership in the emergency department, but with none of the three scales in the AMI and hip fracture departments. The quality management compliance index (QMCI, an external audit-based measure of the quality improvement activities) was 'negatively' associated with teamwork climate and safety climate in AMI departments, after controlling for QMSI, but not in other departments. There was no association between QMCI and leadership in any department, after controlling for QMSI, and there was no association between the clinical quality implementation index (CQII, an external audit-based measure of the level of implementation of quality activities) and any of the three department-level scales in any of the four departments, after controlling for both QMSI and QMCI.
Conclusions: The influence of organization-level quality management systems on clinician safety culture and leadership varied depending on the hospital department, suggesting that whilst there

was some consistency on patient safety attitudes and behaviours throughout the organizations, there were also other factors at play.

Key words: hospital quality management systems, multi-level research, teamwork, safety culture, leadership, quality improvement

Introduction

A large body of literature has emerged over the last two decades on the topic of culture in healthcare. Examination of the relationship between culture and safety is ongoing [1–3]. We know that culture is related to hospital performance [4, 5], and there is evidence that culture is related to patient outcomes [6, 7]. Hospitals are under constant pressure from governments, other funders and the public to create a safety culture that reduces medical errors and adverse patient outcomes. One of the common methods for achieving such a safety culture is through implementation of quality management strategies [8, 9]. The Deepening Our Understanding of Quality improvement in Europe (DUQuE) study provided the first evidence that teamwork climate, safety climate and quality management systems are related [10]. What we do not fully understand however, despite the large amount of effort and significant resources routinely devoted to quality improvement in hospitals, is whether the extent to which the way quality is enacted in the hospital is related to the safety culture and leadership of the clinicians (that is, the doctors, nurses and allied health professionals) who work there.

There is a current focus on engaging doctors in leadership activities, including in the safety and quality of care [11]. Yet we do not know to what extent clinician leadership is related to quality management in hospitals. To help answer these questions, our study aimed to explore the relationships between the organization-level quality improvement and department-level safety culture and leadership measures across 32 large Australian hospitals.

Methods

A study protocol for the full study has been published elsewhere [12], and additional details have been described in other papers in this supplement [13]; here we provide a summary of the methods relevant to this sub-study, which focuses on three organization-level quality management systems and three department-level safety culture and leadership measures, including changes from the published protocol. The three organization-level measures were initially developed for the DUQuE study and adapted and validated by the Deepening our Understanding of Quality in Australia (DUQuA) team to ensure relevance to the targeted Australian hospitals [14]. Figure 1 shows the hypothesised relationships between the organization-level and department-level safety culture and leadership measures in the DUQuA study. These measures and levels are described below.

Setting, participants, recruitment and data collection

The study population comprised 78 Australian public hospitals meeting the DUQuA hospital inclusion criteria (200 beds or more with an emergency department (ED) and 10 or more admissions per month for each of acute myocardial infarction (AMI), hip fracture and stroke) [12]. All were invited, and 32 received ethics approval and participated.

The local quality manager at each hospital, or equivalent role-holder, completed an organization-level survey (the quality management systems index, QMSI) and coordinated with the research team for external surveyors to conduct an on-site assessment for two organization-level quality management systems. The hospital quality manager did not receive special training; external surveyors—all experienced hospital evaluators—received project-specific training as detailed in Supplementary Appendix A. The heads of department responsible for inpatient care of AMI, hip fracture and stroke, and the ED, were asked by the quality manager to invite clinicians (doctors, nurses and allied health professionals) in their departments to complete a safety culture and leadership questionnaire producing three measures: teamwork climate (TC), safety climate (SC) and leadership (L). Figure 2 provides an overview of data collection in the DUQuA study, including hospital and participant recruitments, ethics and governance processes and data collection; aspects that are relevant for this study are shaded.

Organization-level measures

Organization-level quality management systems were assessed using three measures: the QMSI, the quality management compliance index (QMCI) and the clinical quality implementation index (CQII). Each of these measures, validated as part of the DUQuE project, was adapted for Australian conditions and validated in its revised form [13, 14].

QMSI is a measure based on a survey completed by the quality manager in each hospital, and provides a proxy measure for the managerial aspects of quality management that might influence the implementation of quality systems in hospitals. The DUQuA QMSI consisted of eight subscales: quality policy, hospital governance board activities, quality resources, quality management, preventive protocols, internal quality methods (general and patient activities) and performance monitoring. Respondents were asked to rate each item on a four-point Likert scale (range 1–4). For QMSI, seven out of eight subscales achieved adequate internal consistency (Cronbach's $\alpha > 0.8$) during validation.

QMCI is a measure based on the results of an on-site audit completed by trained external surveyors [15] and provides a proxy measure for the managerial aspects of quality improvement in hospitals. The DUQuA QMCI consisted of two subscales: monitoring patient and professional opinions and quality control and monitoring. Items were rated on a five-point Likert scale (range 0–4). The choice of the questionnaire items was based on expert opinion from individuals with many years of experience in hospital performance evaluation during accreditation and certification audits. For QMCI, adequate internal consistency (Cronbach's $\alpha > 0.8$) was achieved when validating both subscales.

CQII is also measured using an on-site audit completed by trained external surveyors [15] and provides evidence of implementation of quality systems at hospital level including whether systems exist, to what extent implementation has been monitored and whether implementation is sustainable. The validated DUQuA CQII consists of seven subscales: controlling healthcare-associated infections, medical safety, preventing patient falls, preventing pressure injuries, routine assessment and diagnostic testing of patients in elective surgery, safe surgery that includes an approved checklist and recognizing and

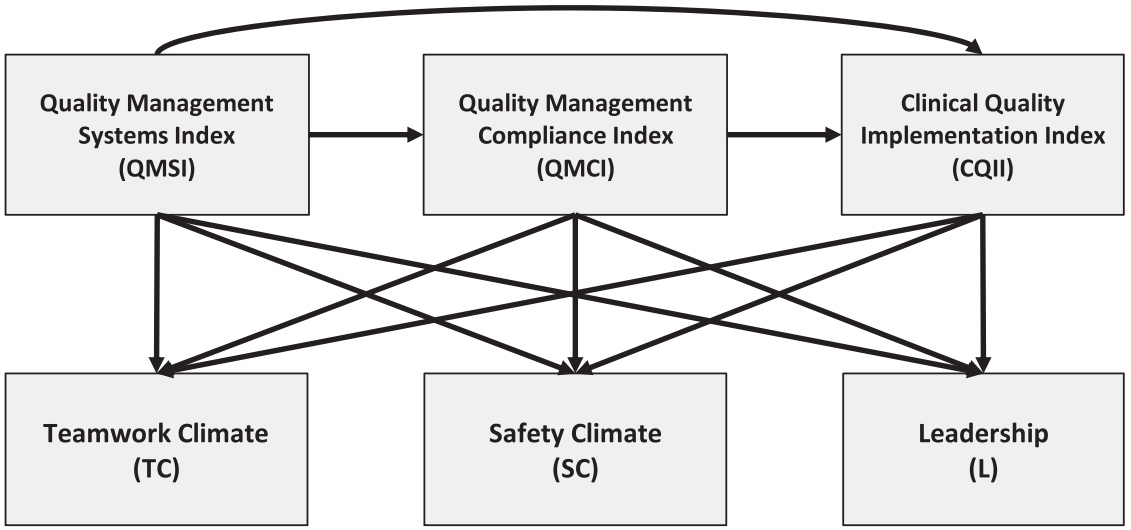

Figure 1 Simplified directed acyclic graph (DAG) showing the relationships between the organization level and safety culture and leadership measures in DUQuA.

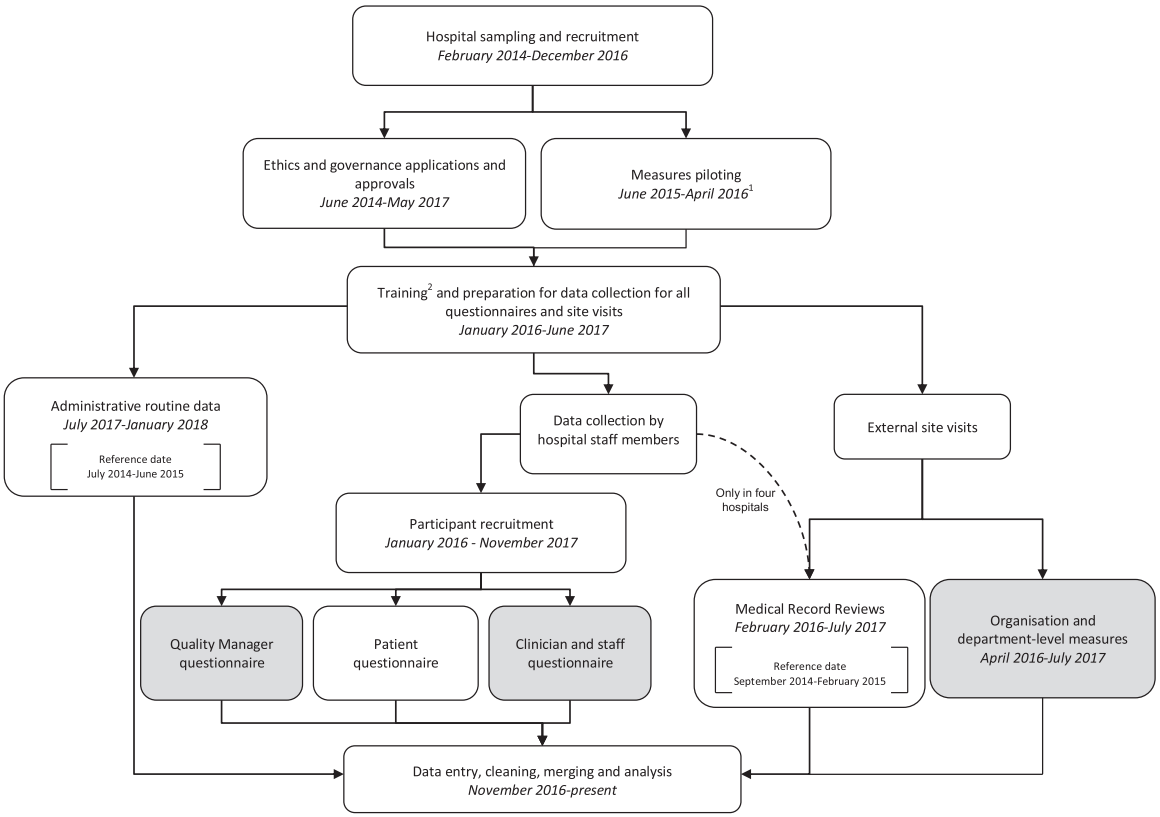

Figure 2 Stages of DUQuA data collection. Legend: shaded box represents relevant data collected for this article.[1]Measures were piloted at different times: patient questionnaire (May–July 2015); medical record review forms (November 2015–February 2016); external site visits (Organization and department-level measures, April 2016); quality manager questionnaire was restricted to critique experts and clinician and staff questionnaire was previously validated.[2]Training was specifically held for 'External site visits' for healthcare surveyors who were contracted to collect data for 'Organisation and department-level measures' at DUQuA hospitals

responding to clinical deterioration in acute healthcare. Items were rated on a five-point Likert scale (range 0–4). When validating CQII, lack of variation and ceiling effects in the data resulted in very low internal consistency scores, but items were retained for theoretical reasons.

Department-level safety culture and leadership measures

The DUQuA-validated safety culture and leadership scale consists of three subscales: teamwork climate, safety climate and leadership. The scale was refined and validated [16] by combining the teamwork

climate, safety climate and perceptions of management items from the safety attitudes questionnaire, developed and validated by Sexton *et al.* [17] with six items from the leadership effectiveness scale, developed and validated by Shipton *et al.* [18]. Teamwork climate (14 items) is a measure of perceived quality of collaboration between personnel, and safety climate (13 items) measures staff perceptions of what a strong and proactive organizational commitment to patient safety entails. The leadership scale (eight items) measures clinician perceptions of the vision, inclusiveness and internal and external collaborative behaviours of healthcare leaders in their department. Items were rated on a five-point Likert scale (range 0–4).

Statistical analysis

A directed acyclic graph (DAG), a simplified version of which is shown in Figure 1, presents assumed causal relations amongst variables and thereby determines which confounding variables are included in the statistical models. For example, to examine whether QMCI has an effect on 'teamwork climate,' we controlled for QMSI because it is assumed to be a predictor for both QMCI and 'teamwork climate.'

Data were analysed in SAS/STAT software version 9.4 (SAS Institute). Each organization scale was calculated as the sum of all subscales (for QMSI, 8 was subtracted from the sum, in line with DUQuE procedures); each subscale was the mean of all items; subscales with more than 50% missing items were treated as missing. Each of the department-level safety culture and leadership scales was calculated as the mean of all items; if the response to any item was missing, the scale was set to missing. In line with published procedures for the Safety Attitudes Questionnaire (SAQ) [10, 19], these scales were rescaled from 1–5 to 0–100.

Visual inspection of missing data patterns in the organization-level items suggested that data were missing at random. Multiple imputation was therefore performed (MI procedure) for the organization subscales, repeated 100 times, and scale scores were calculated after imputation. Each imputed dataset was merged with department-level data and general linear mixed models (MIXED procedure) [20] were used to examine the associations between department-level measures (assessed at clinician level) and hospital-level measures. The models were adjusted for random effect (hospital) and fixed effects (hospital peer group categorized as Referral vs. Acute (A or B), clinician age, gender and self-identified leadership role and profession), in addition to any variables specified in the DAG model (Figure 1). Analysis was performed on each imputation and the MIANALYZE procedure was used to obtain the pooled parameter estimates and standard errors, if any imputed values were involved in the calculations.

Results

Characteristics of participants

Hospitals and departments. Largely due to the inclusion criteria for the DUQuA study, participating hospitals over-represent principal referral hospitals and hospitals with over 500 beds and under-represent inner regional remoteness areas and the state of Western Australia (see Supplementary eTable B1, Appendix B). Hospital-level measures were collected from all 32 hospitals; two hospitals shared a single quality management system, so only 31 questionnaires were collected for the QMSI. Missing data for the organization scales ranged from 0 to 9.4% at item level and from 0 to 3.1% at subscale level. Each participating department ($n = 105$, comprising 29 ED, 25 AMI, 26 hip fracture and 25 stroke) was invited to undertake 30 clinician assessments (i.e. 3150 in total); 1332 clinicians (42.3% of target) provided usable data.

Clinicians. Characteristics of the responding clinicians are summarized in Table 1, separately for each department. Respondents were more likely to be female, especially in the stroke department (70%, compared with 58–59% in other departments), and most commonly aged 25–34 years. The proportion of self-identified leaders varied by department, ranging from 39% in the stroke inpatient department to 61% in the ED. Around half of all respondents were nurses, with allied health staff ranging from 3% of respondents in ED to 19% in stroke. This is broadly representative of the proportions of allied health to other clinical staff working in these departments.

Descriptive statistics

Descriptive statistics of the hospital-level measures have been published elsewhere in this supplement [14] and are included as Supplementary eTable B2, for convenience. The mean QMSI score was 19.4 (standard deviation [SD]: 3.1) out of a maximum possible of 24, the mean QMCI was 7.1 (SD: 1.3) out of a maximum possible of 8 and the mean CQII was 25.1 (SD: 2.9) out of a maximum possible of 28.

Descriptive statistics of the department-level safety culture and leadership scales are summarized in Table 2. Data were left-skewed, with a ceiling effect, meaning that scores were clustered towards the high end of the scale. Mean scores were consistently highest for teamwork climate [M = 77 (SD = 13.8)—M = 81 (SD = 11.4)] and lowest for leadership [(M = 63 (SD = 20.5)—M = 70 (SD = 18.4)].

Associations between organization-level and clinician scales

Results of the associations between organization-level and department-level safety culture and leadership scales are summarized in Table 3. QMSI was 'positively' associated with all three department-level scales in the stroke departments, with safety culture and leadership in the ED, but with none of the three scales in the AMI and hip fracture departments. For every unit increase in QMSI score (range 0–24): stroke teamwork climate, safety climate and leadership increased by 0.75, 1.18 and 1.51, respectively, and ED safety climate and leadership increased by 0.98 and 1.09 (range 0–100 for all safety culture and leadership measures).

QMCI was 'negatively' associated with teamwork climate and safety climate in AMI departments, after controlling for QMSI, but not in other departments. For every unit increase in QMCI score (range 0–8): AMI teamwork climate decreased by 4.10, and AMI safety climate decreased by 4.00 (range 0–100 for both). There was no association between QMCI and leadership in any department, after controlling for QMSI, and there was no association between CQII and any of the three department-level scales in any of the four departments, after controlling for both QMSI and QMCI.

Discussion

Interpretation of results

Our study aimed to explore the relationships between the organization-level quality improvement and department-level safety culture and leadership measures across 32 large Australian hospitals. A positive association between quality improvement and leadership and safety culture measures would suggest that attitudes and behaviours around quality and safety are consistent throughout the organization, from those who provide leadership and direction on quality management in the hospital through to frontline clinicians.

Table 1 Characteristics of clinicians

Characteristic		AMI ($N = 307$)		Hip fracture ($N = 307$)		Stroke ($N = 324$)		ED ($N = 394$)	
		n	%	*n*	%	*n*	%	*n*	%
Gender	Male	128	41.7	125	40.7	98	30.2	167	42.4
	Female	179	58.3	182	59.3	226	69.8	227	57.6
Age group (years)	18–24	22	7.2	34	11.1	28	8.6	21	5.3
	25–34	119	38.8	107	34.9	136	42.0	160	40.6
	35–44	66	21.5	73	23.8	74	22.8	111	28.2
	45–54	63	20.5	57	18.6	52	16.0	73	18.5
	55–64	32	10.4	31	10.1	30	9.3	25	6.3
	65–74	5	1.6	5	1.6	4	1.2	4	1.0
Leadership	Yes	122	39.7	143	46.6	124	38.3	242	61.4
	No	185	60.3	164	53.4	200	61.7	152	38.6
Profession	Physician	119	38.8	111	36.2	78	24.1	205	52.0
	Nurse	173	56.4	160	52.1	183	56.5	176	44.7
	Allied health	15	4.9	36	11.7	63	19.4	13	3.3

Table 2 Descriptive statistics for department-level culture and leadership scales

Department	Index	*n*	Mean	SD	Median	Min	Max	IQR
AMI	TC	290	77.1	13.75	78.6	17.9	100.0	17.9
	SC	294	71.2	14.36	72.1	17.3	98.1	19.2
	L	300	63.4	20.47	65.6	0.0	100.0	25.0
Hip	TC	284	79.2	12.67	80.4	28.6	100.0	17.9
fracture	SC	281	72.5	13.32	73.1	21.2	100.0	17.3
	L	304	68.3	19.07	68.8	0.0	100.0	21.9
Stroke	TC	293	81.1	11.37	82.1	26.8	100.0	14.3
	SC	303	72.7	13.26	71.2	21.2	100.0	19.2
	L	318	68.8	17.90	70.3	3.1	100.0	25.0
Emergency	TC	372	78.5	12.21	78.6	21.4	100.0	16.1
department	SC	373	70.6	14.45	73.1	21.2	100.0	21.2
	L	391	70.4	18.40	71.9	3.1	100.0	21.9

Legend: TC, teamwork climate; SC, safety climate; L, leadership; IQR, inter-quartile range.

The mean for safety climate was fairly consistent across the three condition pathways (range 71.2–72.7), suggesting that the perceptions of staff about what is required in terms of a strong proactive commitment to safety has a degree of homogeneity. However, teamwork climate and leadership both varied, with stronger mean teamwork and leadership scores in stroke departments (81.1 and 68.8, respectively), than in hip fracture (79.2 and 68.3) or AMI (77.1 and 63.4). Leadership scores, whilst lower than teamwork climate and safety climate across all departments, are particularly low in the departments that treat AMI. An explanation may be found in the way AMI is managed in hospitals. Unlike the large, multidisciplinary teams that manage hip surgery and stroke, treatment of AMI is driven primarily by standardized protocols [21, 22]. AMI patients are treated initially in quiet beds with high nurse-to-patient ratios, with staff focussing on the monitor. Junior staff who work in very small teams with individual patients, or one-on-one with consultant supervisors, may not have the same awareness of leadership and teamwork issues as experienced on a busy open stroke or hip fracture ward. The clinician respondents, for example, may not see the work and leadership behind the need for aspirin, beta-blockers and calcium channel antagonists on transfer out of the cardiac care unit. They could simply follow the protocol.

QMSI appears to be important for safety culture and leadership, but not in all departments. The association between QMSI and safety culture and leadership was stronger across the medical departments of stroke (all statistically significant) and AMI (not statistically significant, but suggestive) than in the surgical departments where hip fracture is treated (where no significant association was found). Perhaps this is an indication of differences in the relationships between quality managers and frontline clinicians working in medicine when compared with those working in surgery. A previous study on relationships between safety culture and patient outcomes for surgical teams [23], for example, found that teamwork climate and safety climate did not correlate with risk-adjusted outcomes. This has not been found to be the case for hospitals in general [6]. Another confounder for hip fracture is that care is divided amongst two distinct activities—surgery itself, and post-surgical care in the ward—and this results in more variation in responses for hip fracture when the scores are combined. A recent study on safety culture in surgery that used the SAQ measure, for example, found more positive responses for operating room teams [24].

Unexpectedly, we found a negative relationship between QMCI and teamwork climate and safety climate in AMI. QMCI is an audit measure of the degree to which patient and professional opinions

Table 3 Associations of clinician level measures with hospital level measures

Index	Outcome	AMI			Hip fracture			Stroke			ED		
		n	Beta (95% C.I.)	*P*-value	*n*	Beta (95% C.I.)	*P*-value	*n*	Beta (95% C.I.)	*P*-value	*n*	Beta (95% C.I.)	*P*-value
QMSI[a]	TC	290	0.767 (−0.106, 1.640)	0.082	284	0.276 (−0.365, 0.918)	0.385	293	0.753 (0.186, 1.321)	0.011*	372	0.325 (−0.290, 0.939)	0.283
	SC	294	0.775 (−0.041, 1.592)	0.061	281	0.382 (−0.342, 1.105)	0.291	303	1.179 (0.344, 2.014)	0.008*	373	0.976 (0.155, 1.797)	0.022*
	L	300	0.956 (−0.016, 1.927)	0.054	304	0.686 (−0.481, 1.853)	0.238	318	1.513 (0.384, 2.643)	0.011*	391	1.089 (0.179, 2.000)	0.021*
QMCI[a,b]	TC	290	−4.082 (−7.296, −.868)	0.016*	284	0.110 (−2.139, 2.359)	0.916	293	0.244 (−1.843, 2.330)	0.807	372	0.132 (−1.647, 1.911)	0.878
	SC	294	−3.998 (−6.745, −1.25)	0.008*	281	0.742 (−1.718, 3.202)	0.528	303	0.297 (−2.911, 3.505)	0.848	373	−0.204 (−2.596, 2.189)	0.862
	L	300	−2.645 (−6.456, 1.167)	0.161	304	1.215 (−3.019, 5.449)	0.548	318	0.201 (−4.146, 4.549)	0.923	391	−0.077 (−2.731, 2.577)	0.952
CQII[a,b,c]	TC	290	−0.017 (−0.891, 0.857)	0.970	284	0.338 (−0.226, 0.902)	0.240	293	−0.195 (−0.783, 0.394)	0.517	372	0.420 (−0.354, 1.194)	0.287
	SC	294	−0.385 (−1.121, 0.351)	0.305	281	−0.019 (−0.669, 0.631)	0.955	303	−0.559 (−1.422, 0.304)	0.204	373	0.618 (−0.420, 1.655)	0.243
	L	300	−0.461 (−1.509, 0.587)	0.389	304	0.329 (−0.772, 1.430)	0.559	318	−0.642 (−1.823, 0.538)	0.286	391	0.246 (−0.942, 1.435)	0.685

Legend: QMSI (Index 0–24), QMCI (Index 0–8), CQII (Index 0–28), TC (Index 0–100), SC (Index 0–100), L (Index 0–100).
[a] Adjusted for random effect (hospital) and fixed effects (hospital peer group, clinician gender, age, leadership and profession).
[b] Additionally adjusted for fixed effect (QMSI).
[c] Additionally adjusted for fixed effect (QMCI).
*Statistically significant at 5%.

are monitored and the monitoring of quality control measures such as implementation of guidelines. In Australia, AMI guidelines are applied nationally and supported by external bodies, such as the National Heart Foundation of Australia and Cardiac Society of Australia and New Zealand [21]. Whilst this is more likely to result in highly standardized treatment for patients [22], it is possible that this aspect of QMCI is independent of other department-level measures such as clinician safety culture and leadership. In any event, this finding suggests that further investigation is required. We found no other associations between QMCI and safety culture and leadership measures.

We also found no associations between CQII and teamwork climate, safety climate and leadership. CQII has a restricted range, however, that may have limited our ability to find any relationships. We hypothesize that this is because the seven subscales in CQII measure factors that are specifically assessed as part of hospital accreditation. Achieving the minimum accreditation standards on each of these factors is mandatory for all public hospitals in Australia, and hence the data for this scale were strongly left skewed with an equally strong ceiling effect, meaning that scores were clustered towards the high end of the scale.

DUQuE also found positive associations between QMSI and both teamwork climate and safety climate, with stronger relationships found when the safety culture survey was completed by leaders than when completed by frontline clinicians [10]. The DUQuE researchers did not publish data on associations between safety culture and QMCI or CQII.

Of note was the number of self-identified leaders participating in our study (39–61%, depending on department). Perhaps it is not surprising that self-identified leaders are more likely to be interested in enrolling in research about the performance of their organization and therefore self-selected into our study and, if they are also organization-designated leaders, are more likely to be able to organize their time to allow for participation. A known disadvantage of self-identified leaders is that they have the potential to introduce selection bias [25]. For our study, however, we asked the same questions of leaders and non-leaders in regard to safety culture and leadership. Additionally, an examination of our results did not reveal a statistically significant difference in responses between clinicians and clinician-leaders.

Limitations

The DUQuA QMSI and clinician safety culture and leadership scales are based on self-reported data. In mitigation, both scales have been validated [14, 16] and the QMCI and CQII scales are based on external audit of hospital quality processes. Although the two studies are based on similar data collection instruments and statistical analysis procedures, differences between the DUQuA and DUQuE scales must be taken into account when comparing findings from Australian and European hospitals. The low response rate in some hospitals may have introduced selection bias.

Conclusions

The influence of organization-level quality management systems on clinician safety culture and leadership varied depending on the hospital department, suggesting that whilst there was some consistency in patient safety attitudes and behaviours throughout the organizations, there were likely other factors at play. The negative influence of quality monitoring and collection of patient and professional opinions on clinician teamwork climate and safety climate in AMI departments

suggests a need for further investigation. Finally, the CQII requires further refinement to be a useful measure for discriminating performance across Australian public hospitals.

Supplementary material

Supplementary material is available at *International Journal for Quality in Health Care* online.

Acknowledgements

Members of the original DUQuE team (Oliver Gröne, Rosa Suñol) provided extensive input and advice in the initial stages of the project, supporting design modifications from DUQuE and the planned DUQuA study approach. Dr Annette Pantle provided expert advice on revision and development of DUQuA measures, and Professor Sandy Middleton, Associate Professor Dominique Cadillac, Kelvin Hill, Dr Carmel Crock, and Professor Jacqueline Close provided input into the development of our stroke, AMI, and hip fracture clinical process indicator lists. Nicole Mealing and Victoria Pye provided statistical advice in the initial phases of the project. We greatly appreciate their efforts.

Funding

This research was funded by the National Health and Medical Research Council (NHMRC) Program Grant APP1054146 (CI Braithwaite).

Contributions

The research team consists of experienced researchers, clinicians, biostatisticians and project managers with expertise in health services research, survey design and validation, large-scale research and project management, sophisticated statistical analysis, quality improvement and assessment, accreditation, clinical indicators, policy and patient experience. JB conceived the idea to embrace DUQuE for Australia, led the research grant to fund the project and chairs the steering committee. RCW and NT co-led the detailed study design, managed the project across time and contributed to the development of the manuscript. HPT and GA provided statistical expertise for the study design and developed the analysis plan for the manuscript. TW contributed to the logistics of project management, the refinement of measures and the development of the manuscript.

Ethics approval

Ethical approvals were secured from State and Territory human research ethics committees in New South Wales (#14/206), Victoria (#15/36), the Australian Capital Territory (#15/131), South Australia (#15/260), the National Territory (#15-2509), Tasmania (#H0015383) and Queensland (#15/361). Site-specific authorizations, including permission for external researchers to collect data in hospitals, were granted. We complied with confidentiality requirements of national legislation or standards of practice of each jurisdiction.

Data sharing statement

Data will be made publicly available to the extent that individual participants or participating hospitals cannot be identified, in accordance with requirements of the approving Human Research Ethics Committees.

References

1. Morello RT, Lowthian JA, Barker AL *et al.* Strategies for improving patient safety culture in hospitals: a systematic review. *BMJ Qual Saf* 2013;**22**:11–8.
2. Nieva VF, Sorra J. Safety culture assessment: a tool for improving patient safety in healthcare organizations. *Qual Saf Health Care* 2003;**12**:17ii–23.
3. Weaver SJ, Lubomksi LH, Wilson RF *et al.* Promoting a culture of safety as a patient safety strategy: a systematic review. *Ann Intern Med* 2013;**158**:369–74.
4. Mannion R, Davies HT, Marshall M. Cultural characteristics of "high" and "low" performing hospitals. *J Health Organ Manag* 2005;**19**:431–9.
5. Taylor N, Clay-Williams R, Hogden E *et al.* High performing hospitals: a qualitative systematic review of associated factors and practical strategies for improvement. *BMC Health Serv Res* 2015;**15**:244.
6. Braithwaite J, Herkes J, Ludlow K *et al.* Association between organisational and workplace cultures, and patient outcomes: systematic review. *BMJ Open* 2017;**7**:e017708.
7. Mardon RE, Khanna K, Sorra J *et al.* Exploring relationships between hospital patient safety culture and adverse events. *J Patient Saf* 2010;**6**:226–32.
8. Campione J, Famolaro T. Promising practices for improving hospital patient safety culture. *Jt Comm J Qual Patient Saf* 2018;**44**:23–32.
9. Muething SE, Goudie A, Schoettker PJ *et al.* Quality improvement initiative to reduce serious safety events and improve patient safety culture. *Pediatrics* 2012;**130**:e423.
10. Kristensen S, Hammer A, Bartels P *et al.* Quality management and perceptions of teamwork and safety climate in European hospitals. *Int J Qual Health Care* 2015;**27**:499–506.
11. Dickinson H, Ham C, Snelling I, Spurgeon P. *Are We There Yet? Models of Medical Leadership and their Effectiveness: An Exploratory Study. NIHR Service Delivery and Organisation Programme*, 2013.
12. Taylor N, Clay-Williams R, Hogden E *et al.* Deepening our understanding of quality in Australia (DUQuA): a study protocol for a nationwide, multilevel analysis of relationships between hospital quality management systems and patient factors. *BMJ Open* 2015;**5**:e010349.
13. Braithwaite J, Clay-Williams R, Taylor N *et al.* Deepening our Understanding of Quality in Australia (DUQuA): An overview of a nation-wide, multilevel analysis of relationships between quality management systems and patient factors in 32 hospitals. *Int J Qual Health Care* 2020;**32**:8–21.
14. Clay-Williams R, Taylor N, Winata T *et al.* Organization quality systems and department-level strategies: refinement of the Deepening our Understanding in Quality in Australia (DUQuA) organization and department-level scales. *Int J Qual Health Care* 2020;**32**:22–34.
15. Wagner C, Groene O, Dersarkissian M *et al.* The use of on-site visits to assess compliance and implementation of quality management at hospital level. *Int J Qual Health Care* 2014;**26**:27–35.
16. Clay-Williams R, Taylor N, Ting HP *et al.* The clinician safety culture and leadership questionnaire: refinement and validation in Australian public hospitals. *Int J Qual Health Care* 2020;**32**:52–59.
17. Sexton JB, Helmreich RL, Neilands TB *et al.* The safety attitudes questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Serv Res* 2006;**6**:44.
18. Shipton H, Armstrong C, West M, Dawson J. The impact of leadership and quality climate on hospital performance. *Int J Qual Health Care* 2008;**20**:439–45.
19. Sexton JB, Thomas EJ. *The Safety Climate Survey: psychometric and benchmarking properties. Technical Report 03-03*. The University of Texas Center of Excellence for Patient Safety Research and Practice, 2005.
20. Littell R, Milliken G, Stroup W *et al. SAS for Mixed Models*, 2nd edn. Cary, NC, SAS Institute, 2006.
21. Chew DP, Scott IA, Cullen L *et al.* National Heart Foundation of Australia and Cardiac Society of Australia and new Zealand: Australian clinical guidelines for the management of acute coronary syndromes 2016. *Med J Aust* 2016;**205**:128–33.
22. Runciman WB, Hunt TD, Hannaford NA *et al.* CareTrack: assessing the appropriateness of health care delivery in Australia. *Med J Aust* 2012;**197**:549.
23. Davenport DL, Henderson WG, Mosca CL *et al.* Risk-adjusted morbidity in teaching hospitals correlates with reported levels of communication and collaboration on surgical teams but not with scale measures of teamwork climate, safety climate, or working conditions. *J Am Coll Surg* 2007;**205**:778–84.
24. Odell DD, Quinn CM, Matulewicz RS *et al.* Association between hospital safety culture and surgical outcomes in a Statewide surgical quality improvement collaborative. *J Am Coll Surg* 2019;**229**:175–183.
25. Valente TW, Pumpuang P. Identifying opinion leaders to promote behavior change. *Health Educ Behav* 2007;**34**:881–96.

International Journal for Quality in Health Care, 2020, 32(S1), 67–74
doi: 10.1093/intqhc/mzz097
Article

Article

Validation of the patient measure of safety (PMOS) questionnaire in Australian public hospitals

NATALIE TAYLOR[1,2], ROBYN CLAY-WILLIAMS[3], HSUEN P. TING[3], TERESA WINATA[3], GASTON ARNOLDA[3], EMILY HOGDEN[1], REBECCA LAWTON[4], and JEFFREY BRAITHWAITE[3]

[1]Cancer Research Division, Cancer Council NSW, 153 Dowling St, Woolloomooloo, NSW 2011, Australia, [2]Faculty of Health Sciences, University of Sydney, Camperdown, Sydney, NSW 2006, Australia, [3]Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Road, NSW 2109, Australia, and [4]School of Psychology, University of Leeds, Leeds LS2 9JT, UK

Address reprint requests to: Natalie Taylor, Cancer Research Division, Cancer Council NSW, Dowling St., Woolloomooloo, NSW 2011, Australia. Tel: +61(2) 9334 1974, E-mail: natalie.taylor@nswcc.org.au

Editorial Decision 7 September 2019; Accepted 12 September 2019

Abstract

Objective: Patients can provide a unique perspective on the safety of care in hospitals. Understanding that the extent to which the way hospitals are organized for quality and safety is associated with patient perceptions of care is becoming increasingly valued and necessary for the direction of targeted interventions across healthcare systems. The UK-developed patient measure of safety (PMOS) assesses eight domains of ward safety from the patient point of view and has recently been adapted and piloted in Australia. The aim of this study is to test the psychometric properties of PMOS-Australia (PMOS-A) amongst a large cohort of hospitalized patients.
Design: Cross-sectional questionnaire validation assessment.
Setting and participants: As part of the DUQuA project, the PMOS-A survey was distributed within acute myocardial infarction, hip fracture and stroke departments across 32 large public hospitals in Australia. Patients could complete the PMOS-A independently, or request the assistance of a family member/guardian, or staff on the wards—space was included to record mode of completion.
Main outcome measures: Confirmatory factor analysis (CFA) was undertaken on a calibration sample to generate the model, and a validation sample was used to cross-validate the model. A subset of only those participants who received assistance for PMOS-A completion was also tested using CFA on a calibration and validation sample. Model fit indices (chi-square to degrees of freedom ratio [Chi-square:DF], root mean square error of approximation [RMSEA], comparative fit indices [CFI], standardized root mean squared residual [SRMR]), Cronbach's α, average inter-item correlations, construct reliability and cross-loadings were examined with reference to recommended thresholds to establish the extent of convergent validity and discriminant validity. A marker of criterion validity was assessed through testing associations between the PMOS-A and adherence to clinical guidelines.
Results: Across the calibration and validation samples of the full ($N = 911$) and assisted completers only subset ($N = 490$), three (Chi-square:DF, SRMR, RMSEA) of the four indices consistently or almost always met thresholds for acceptable model fit. CFI indices did not meet the recommended limits (0.72–0.78, against a target > 0.9). Positive relationships were found for all tests between

PMOS-A and adherence to clinical guidelines, and these were significant when assessed in the calibration datasets for the full and assisted completion samples.
Conclusion: A sufficiently reliable and valid measure of patient perceptions of safety has been developed. These findings should provide adequate support to justify the use of this measure to assess patient perceptions of safety in Australian hospitals and can be modified for use elsewhere.

Key words: patient perceptions of safety, vulnerable population, quality improvement, patient safety, patient outcomes

Introduction

The importance of accounting for the patient view when measuring the quality and safety of healthcare has become widely recognized over the past decade [1–4]. Not only can patient perceptions provide unique 'fly on the wall' insights into routine care activities [5], exceptional practice [6] and mistakes ranging in severity of consequences [7], this information—if reliable—can contribute to decisions about key areas of healthcare needing improvement and can be used to corroborate other markers of hospital quality and safety [8]. Understanding that the extent to which the way hospitals are organized for quality and safety (including factors such as provision of resources, processes and practice and organization and ward culture) is associated with patient perceptions of care, in addition to more traditional patient outcomes (e.g. length of stay, readmission, mortality), is becoming increasingly valued and necessary for the direction of targeted interventions across healthcare system levels [4, 9, 10].

Despite this need, progress towards reliably measuring patient perceptions of quality and safety has been slow due to the cross-cutting challenges facing researchers and healthcare practitioners. Broadly, these challenges can be represented by logistics (recruitment, tool administration, patient, carer and staff time) [11, 12], patient background and culture (e.g. demographics, literacy and language) [13–15], patient health status (e.g. too unwell and cognitively impaired) [12, 16] and differences in individual understanding or interpretation (e.g. of technical or interpersonal concepts) [17], values (e.g. expectations) [17, 18] and beliefs and attitudes (e.g. reluctance to provide negative feedback through anticipation of subsequent worse care) [16]. Furthermore, whilst it may be relatively straightforward to measure patients' general feelings of safety, as some suggest we should do [19], this more generalized approach may not help us in knowing what to improve.

Whilst logistics such as time (for healthcare professional staff) or access to healthcare organizations (for researchers) can present obstacles for collecting patients' views [12, 16, 20], the potential value in assisting patients to complete tools to understand their perceptions of quality and safety during their time in hospital may warrant additional effort. Assisting patients may help to ensure some of the literacy, understanding and interpretation issues are avoided or clarified and provides an opportunity to address existing beliefs and attitudes that may be misguidedly influencing a patient's decision not to engage with a survey or their approach to answering particular questions. Providing or offering to provide assistance, however, is only likely to be feasible and worthwhile if a pragmatic and meaningful measure exists which is able to: be integrated into routine patterns of practice, take very little time for patients to complete, and effectively pin-point specific well- or poor-performing areas of quality and/or safety for the purposes of targeted intervention development [11].

The patient measure of safety (PMOS) aims to meet these needs. Originally developed and its psychometric properties tested and validated in the UK [21, 22], the PMOS is a 35-item measure assessing eight domains of ward safety from the patient point of view. Its benefits over other patient-based measures of experience, quality and safety (e.g. Patient Perception of the Quality of Nursing Care [23], UK National Health Service inpatient survey [24] and Nordic Patient Experiences Questionnaire [25]) include that the domains are based on a validated framework of the factors that contribute to patient safety [26], and hospitals and wards can use it as a diagnostic guide for the design of patient safety improvement interventions by front-line staff. To date, it has been administered as both a self- and assisted-completion (by researchers, research nurses and volunteers) tool [27–29]. Recently, the PMOS has demonstrated feasibility and initial validation in Australia with an older, vulnerable population [16]. However, 'PMOS-Australia' (PMOS-A) model specification via confirmatory statistical methods has not yet been assessed, nor have any differences been tested for the reliability of the PMOS-A between self- versus assisted-completers.

The Deepening our Understanding of Quality in Australia (DUQuA) study design required—as part of a suite of outcome measures—a measure of patient perception of safety [30]. These outcomes were then explored in relation to how hospitals are organized for quality and safety. The findings would then provide direction for targeted interventions as part of the feedback promised to participating hospitals. The PMOS-A pilot and feasibility study [16] demonstrated promise for an adapted version of the originally developed PMOS [22] for older, vulnerable patients in Australia and, as such, it was selected to be rolled out to acute myocardial infarction (AMI), hip fracture and stroke patients across the 32 participating hospitals (associations between the PMOS-A and other DUQuA measures are presented elsewhere [31]. The primary aim of this study was to test the PMOS-A for convergent validity (whether the items on the scale load onto the intended appropriate factor), internal consistency reliability (whether individual subscale items demonstrate good relationships with one another), discriminant validity (whether theoretically different subscales are distinguishable from one another) and criterion validity (whether PMOS-A scores were related to other markers of hospital quality). The secondary aim was to test these properties restricted to the sub-sample of assisted PMOS-A completers.

Methods

The UK-validated PMOS questionnaire assesses patient perceptions of the factors contributing to patient safety [21]. Cronbach's α of 0.66–0.89 was found for each of eight subscales, and test-retest reliability ($r = 0.75$) was reported in McEachan et al. [21] to establish the questionnaire's reliability.

The UK-developed and validated PMOS has been previously refined for an older Australian population, using a think aloud and pilot test across AMI, hip fracture and stroke patients in one Australian hospital [16]. This PMOS-A measure contains 43 items to assess local, latent and situational factors across nine key domains: communication and team work, organization and care planning, ward type and layout, access to resources, equipment (design and function), staff training, delays, roles and responsibilities and information flow. Key adaptations from the original UK PMOS include the re-wording of 12 negatively-worded items, such that they were framed positively but retained the original meaning, to assist with ease of understanding. Each item is answered on a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Fourteen items were framed negatively (in comparison to the originally developed UK PMOS measure, in which there were 24 negatively-worded items). Respondents could select 'not applicable (NA)' or 'I prefer not to answer' as an option and record a comment against any item. Below we describe the methods to re-validate PMOS-A within the DUQuA study population.

Setting and participants

As part of the DUQuA project, which employed a cross-sectional research design, the PMOS-A survey was distributed within AMI, hip fracture and stroke departments across 32 large public hospitals in Australia to measure patient perceptions of safety as an outcome. Data were collected between February 2016 and November 2017. In line with recommendations, a minimum sample size of 250 participants was required for planned factor analysis [32]. Inclusion criteria stipulated that patients should: be aged 18 years and over, be admitted to a participating ward, have received notification of discharge or judged by a member of care team to have spent enough time on the ward to answer the PMOS-A, be physically present on the ward (i.e. not already left following discharge) and be willing to give written informed consent (or willing to verbally agree for a carer/relative to provide this on their behalf).

Recruitment and data collection

Between January 2016 and November 2017, the DUQuA Local Principal Investigator in each hospital coordinated with medical staff on participating wards to administer the PMOS-A questionnaire to the patients—a new mode of administering PMOS-A compared to previous approaches [28, 29]. Ward staff used the inclusion criteria to identify eligible patients. In the previously reported pilot study for PMOS-A, the majority of participants requested assistance for PMOS-A completion [16]. As such, this was the originally prescribed approach for administering the PMOS-A as part of the DUQuA study, particularly given the majority of patients in this sample were likely to be > 65; however, during a feasibility assessment with one hospital (for all DUQuA study measures) [30], feedback indicated that achieving assisted completion for all PMOS-A participants via hospital ward staff was impractical. It was therefore decided that patients could complete the PMOS-A independently, or request the assistance of a family member/guardian, or staff on the wards—space was included to record mode of completion. Patients who had not received treatment for AMI, hip fracture or stroke specifically but had been treated on these wards for other conditions could be included in the sample. Patients were not approached if they were assessed by staff to be: too ill, suffering from physical or emotional distress, unable to communicate in English by themselves or transferred from another hospital and did not receive most of their care in the current hospital.

Statistical analysis

Data were analysed in IBM SPSS Amos 25, IBM SPSS Statistics 25 (both Armonk, New York, USA) and SAS/STAT 9.4 (Cary, North Carolina, USA). Given the need to provide hospitals with the option for assisted versus self-completion, we first analysed the full sample and then repeated the analysis for a sub-sample restricted to assisted patients.

'NA' responses were imputed with a 'neither agree or disagree' response. 'Prefer not to answer' and blank responses were treated as missing. Records were deleted from validation if they contained any subscale with more than 50% missing items. Scores for negatively worded items were reversed. Descriptive statistics were summarized for the item and subscale scores, prior to imputation of missing data; subscale scores were calculated as the mean of subscale items. Missing items were examined using Little's test [33]: a non-significant test result indicates data are missing completely at random (MCAR); a significant test result indicates that data are not MCAR and leads to visual examination of the data to determine if the missing pattern indicates that data are missing at random (MAR) or missing not at (MNAR). Imputation using the expectation-maximization (EM) method assumes that the missing data pattern is at least MAR [34]. For validation purposes, missing items were imputed via EM separately for the full sample and the assisted sub-sample [35]; this approach is described in additional detail in articles (Articles 2 and 9) discussing methodological issues in the DUQuA study, elsewhere in this Supplement material [31, 36].

After imputation, records were randomly split into calibration and validation datasets (half and half). Confirmatory factor analysis (CFA) was used to test and refine the PMOS-A nine-domain theoretical framework in the calibration datasets and the refined model subsequently assessed in the validation datasets (convergent validity) [37]. Maximum Likelihood (ML) estimation was used to perform CFA [38]. This method assumes multivariate normality, which was assessed using Mardia's coefficients of multivariate skewness and kurtosis [39], with statistically significant results for these tests indicating non-normality. In the event of multivariate non-normality, appropriate transformations were attempted.

To examine convergent validity we used chi-square statistics, comparative fit index (CFI), root mean square error of approximation (RMSEA) and standardized root mean squared residual (SRMR) to assess model fit [40]. Chi-square statistics assume multivariate normality; deviation leads to inflation of the value even when the model is appropriately specified [41]. Chi-square is also sensitive to sample size [42, 43], which is addressed by using the chi-square:DF ratio. The preferred thresholds adopted were: chi-square P-value > 0.05 [44]; Chi-square:DF < 2 (a conservative signal, but values of Chi-square:DF < 5 have been considered acceptable) [45], CFI > 0.95 (values >0.90 considered acceptable), RMSEA < 0.06 (values <0.08 considered acceptable) and SRMR < 0.05 (values <0.08 considered acceptable) [46]. Model fit improvement in the calibration dataset was guided by a combination of theoretical considerations, modification indices, factor loadings and multiple R-squared. Items with a factor loading of < 0.32 and multiple R-square < 0.2 were considered for deletion or model re-specification [47, 48]; the final decisions were determined by theoretical considerations and internal group agreement.

Internal consistency reliability was measured using Cronbach's α, construct reliability and average inter-item correlations.

Table 1 Summary of patient characteristics

Characteristic	All patients ($N = 911$)[a]		Assisted completion ($N = 490$)		Independent completion ($N = 397$)	
	n	%	*n*	%	*n*	%
Gender						
Male	575	63.1	294	60.0	270	68.0
Female	336	36.9	196	40.0	127	32.0
Age group (years)						
18–24	18	2.0	12	2.4	6	1.5
25–34	36	4.0	14	2.9	19	4.8
35–44	62	6.8	20	4.1	42	10.6
45–54	114	12.5	45	9.2	69	17.4
55–64	200	22.0	93	19.0	102	25.7
65–74	251	27.6	143	29.2	103	25.9
75–84	167	18.3	114	23.3	45	11.3
85+	63	6.9	49	10.0	11	2.8
Department						
AMI	346	38.0	149	30.4	189	47.6
Hip fracture	273	30.0	161	32.9	101	25.4
Stroke	292	32.1	180	36.7	107	27.0

[a]Includes 24 patients that did not designate assisted or self-completion.

Cronbach's α of >0.8 was used to indicate adequate internal consistency [49], construct reliability of >0.6 was considered adequate [50] and the acceptable range of average inter item correlations is 0.15–0.5 [51]. To assess discriminant validity, we used cross loadings, whereby the item factor loadings on the assigned construct should be higher than loading for other constructs [37].

An overall score for the PMOS-A was calculated using the mean of the nine subscales. While we did not have datasets to correlate individual level patient data with individual level medical records or patient outcomes, it was deemed valuable to assess a marker of criterion validity. Therefore, the median PMOS-A score for each ward in a hospital was correlated with percent adherence for clinical process indicators within the same ward. Valid PMOS-A responses for all wards were included unless data were not collected/usable in the criterion validity measures (e.g. incomplete medical record review for assessment of clinical guideline adherence). For the criterion validity assessment, no minimum requirement was set for the number of PMOS-A responses per ward because of the small sample sizes resulting from halving the data into calibration and validation datasets, especially for analysis of the sub-group of assisted completers; any random measurement error that results will bias the correlations towards zero.

Results

PMOS-A forms were received from 965 patients, with 54 (5.6%) excluded due to insufficient data. Whilst we cannot provide a specific response rate, a total of 2640 patients (30 patients per ward across 88 wards) were targeted, providing a proxy response rate of 34.5% across all wards. Of the included 911 responses, 24 forms did not include information on whether PMOS completion was assisted or undertaken independently—these respondents are included in the overall analysis but removed for sub-analyses. Characteristics of the included patients are summarized in Table 1: overall (911), for the 490 who were assisted and for the 397 who self-completed. The 911 patients originated from 21 AMI departments ($n = 346$ patients), 21 hip fracture departments ($n = 273$) and 21 stroke departments ($n = 292$), across 27 hospitals. Of the 911 patients, 575 (63.1%) were male and 481 (52.8%) were aged 65 years and over. In the subset of assisted completers, the 490 patients had an average age of 66.6 years and originated from 21 AMI, 18 hip fracture and 19 stroke departments, across 27 hospitals.

Descriptive statistics and missing data analysis

Descriptive statistics of the subscale and item scores are summarized for the full sample and subset of assisted completers in Supplementary eTables A1 and B1, respectively. For both samples, distributions appeared to be left-skewed (i.e. scores are generally moderately positive) with ceiling effects, and Mardia's coefficient of multivariate kurtosis and skewness were both statistically significant (P-value < 0.001). Transformations were attempted, but did not result in multivariate normality, so untransformed data were used. Data were missing for 1.7% (range = 0.3–5.5% per item) and 0.7% of the items (range = 0.0–2.5% per item) across the full and subset of assisted completers, respectively. For both samples, Little's test was significant (P-value < 0.001), indicating that data were not MCAR. However, given the overall low proportion of missing data (<5%) and visual inspection of missing data patterns across both datasets suggesting the data were MAR, it was deemed unlikely that there were any variables whose pattern of missing values may have influenced the scale variables, therefore justifying EM imputation [52].

For assessment of criterion validity in the validation and calibration datasets, the number of wards included in the analysis were as follows: full sample: calibration = 56 and validation = 53; assisted sample: calibration = 48 and validation = 47. The mean (SD; range) number of PMOS responses per ward were as follows: full sample: calibration = 6.8 (4.05; 1–16) and validation = 7.2 (4.44; 1–20); assisted sample: calibration = 4.5 (3.52; 1–17) and validation = 4.6 (4.03; 1–17).

PMOS scores

Across the full dataset (unimputed scores) on average people answered three or four out of 43 questions negatively (once all

Table 2 Final model fit statistics

Index	Desired value	Acceptable range	Full sample		Subsample: assisted patients	
			Calibration $N = 455$	Validation $N = 456$	Calibration $N = 245$	Validation $N = 245$
Chi-square	na	na	2834.7	2888.8	2085.4	1940.5
DF	na	na	824	824	824	824
P-value	>0.05	Not specified	<0.001	<0.001	<0.001	<0.001
Chi-square/DF	<2	<5	3.44	3.51	2.53	2.36
RMSEA	<~ 0.06	≤0.08	0.073	0.074	0.079	0.075
SRMR	<0.05	≤0.08	0.070	0.074	0.083	0.075
CFI	>0.95	Not specified	0.77	0.78	0.72	0.76

items were reverse scored, a low score, i.e. ≤ 2/5 indicates a negative answer). Individual item analysis revealed scores between mean (SD) = 3.35 (1.28; 'The following aspects of the ward made it uncomfortable for me: Noise levels') and mean (SD) = 4.58 (0.72; 'I was always treated with dignity and respect'). The extent to which items were answered with 'NA' ranged from 0 (2 items) to 168 (1 item), with 'A doctor changed my plan of care and other staff knew about it' receiving 168 NAs. Amongst the nine key domains measured in the PMOS-A, staff training (mean = 4.35, SD = 0.62) scored most favourably, and the ward type and layout (mean = 3.89, SD = 0.68) subscale scored least favourably. The ward type with the lowest average PMOS-A score was hip fracture for both the full sample (mean = 3.96, SD = 0.43) and the assisted sample (mean = 3.92, SD = 0.42); the ward type with the highest average PMOS score was AMI for both the full sample (mean = 4.18, SD = 0.46) and the assisted sample (mean = 4.14, SD = 0.44).

Full sample: model specification and validation

Calibration. Item 15 ('The ward was able to deal with my treatment needs') and item 16 ('Staff were prompt in answering my buzzer'), both from 'Ward type and layout', had a factor loading of < 0.32 and multiple *R*-squared value of < 0.2. Using a combination of theoretical grounds and modification indices, we re-specified item 15 to 'Organisation and care planning', and item 16 to 'Delays'. Negatively worded items 34 and 41 had a factor loading of < 0.32 and a multiple *R*-squared value of < 0.2 but were retained as deletion of the items did not significantly improve the model fit indices. The final model (Supplementary eFigure A1) is reported in Supplementary eTable A1 and the fit statistics in Table 2; CFI was below the threshold of an acceptable model, but chi-square: DF, RMSEA and SRMR all met thresholds designating an acceptable model.

A summary of the statistical reliability and validity threshold assessment for PMOS-A can be found in Table 3. For internal consistency reliability, two factors reached Cronbach's $\alpha > 0.8$, and all nine factors had average item correlation within or close to the acceptable range. Six factors demonstrated adequate construct reliability (Supplementary eTable A2). Of 36 pairs of subscales, 33 pairs achieved discriminant validity through cross loadings (Supplementary eTable A3). The median ward PMOS-A score was positively correlated with percent adherence to clinical guidelines as assessed by medical record review ($r = 0.40$, *P*-value = 0.002).

Validation. The model fit statistics (Table 2) and assessments of reliability and validity were closely comparable to the calibration model. Two factors had Cronbach's $\alpha > 0.8$; all nine factors had average item correlation within the acceptable range, and seven factors had adequate construct reliability (Supplementary eTable A3). Out of the 36 pairs of subscales, 33 pairs achieved discriminant validity through cross loadings (Supplementary eTable A4). The correlation between the PMOS-A and adherence to clinical guidelines was not statistically significant (Table 3).

Subsample of assisted patients: model specification and validation

Calibration. The same items as identified in the full sample (items 15 and 16 from 'Ward type and layout') had factor loadings and multiple *R*-squared values that fell beneath the recommended thresholds, and so using the rationale described earlier, these were re-specified to 'Organisation and care planning', and 'Delays', respectively. Negatively-worded items 30, 34 and 41 had factor loadings and multiple *R*-squared values below the recommended thresholds but were retained as deletion of the items did not improve the model fit indices. The final model, along with summary statistics, is reported in Supplementary eTable B1. The model fit statistics (Table 2) were similar to those for the full sample, demonstrating noticeable improvement on the Chi-square:DF results (though still in the 2–5 range). The internal consistency reliability, construct reliability and discriminant validity statistics are detailed in Supplementary eTables B2 and B3, and alongside patterns demonstrated for the markers of criterion validity (Table 3), these show the same patterns as the full sample.

Validation. The model fit statistics were largely comparable to the calibration sample (Table 2); the internal consistency reliability, construct reliability and discriminant validity statistics (Supplementary eTables B2 and B4) and markers of criterion validity again show the same patterns as those found for the full sample (Table 3).

Discussion

As a component of the overarching DUQuA study [31], we were able to collect PMOS-A questionnaire data from almost 1000 Australian hospital patients, and for the first time, perform CFA to assess and refine the factor structure of the measure across the entire sample, as well as for a sub-sample of assisted completers. Across the calibration and validation samples of the full and assisted completers subset, while the CFI indices (0.72–0.78) were lower than our thresholds indicating an acceptable model (0.90), the remainder were largely within an acceptable range: chi-square:DF ratios (2.5–3.5) were <5

Table 3 Overview of reliability and validity threshold assessments for PMOS

	Full sample		Subsample: assisted patients	
	Calibration	Validation	Calibration	Validation
Construct reliability[a]	6 of 9	7 of 9	6 of 9	6 of 9
Cronbach's α[a]	2 of 9	2 of 9	2 of 9	2 of 9
Average inter-item correlation[a]	8 of 9	8 of 9	8 of 9	9 of 9
Discriminant validity (pairs of factors)[b]	33 of 36	33 of 36	32 of 36	32 of 36
Criterion validity (correlation (r) with clinical guideline adherence)[c]	0.40**	0.14	0.29*	0.28

[a]The number meeting the threshold out of the number of subscales to which they apply.
[b]The number meeting the threshold against the number of item-pairs of possible comparison—see Supplementary eTables Appendix A and B for specific numbers.
[c]Statistical significance: *$P < 0.05$; **$P < 0.01$.

[45], the SRMR indices (0.070–0.083) were within or slightly above the acceptable threshold of < 0.08 and the RMSEA measures (0.073–0.079) were all < 0.08. Previous studies have shown that multivariate non-normality leads to higher RMSEA scores, even when the models were correctly specified [53, 54].

A secondary aim of this study was to assess model fit amongst the subset of assisted completers to explore any differences in the reliability of responses for this group. Results were largely similar across the two sets of analyses. Whilst over half of the full sample were assisted and therefore may contribute to the similarity of results, it is encouraging to see consistency across these groups as this provides some confidence that independent completers are on the whole not interpreting and reporting on items differently to those who were assisted. Furthermore, the average age for assisted and unassisted completers was 66.6 and 59.4 years, respectively. Together, this information may be useful for healthcare professionals and researchers planning to administer the PMOS-A. For example, we know there is a need for reliable reports of perceptions of safety across patients that include older or vulnerable populations [7, 16], but resource and time are scarce. In these instances, distinctions could be made between patient groups based on age and/or extent of vulnerability as to whether assistance for completion is offered at the outset of the interaction.

In comparison to the UK-validated PMOS, which, based on principal components analysis, demonstrated sound psychometric properties, the age of our sample was substantially higher. In the UK study, whereby the PMOS was administered across 11 wards (one of which was paediatric and one did not collect age), the mean age across nine wards was 49.5 years [mean age across wards ranged from 27.3 (maternity) to 61.2 (vascular surgery)]. The mean age in the present study was 59.1—notably older—and, unlike the UK PMOS validation study, all wards consisted of acutely ill patients. These factors may have affected the interpretation of questions or reliability of responses. However, the version (PMOS-A) used in the current study had undergone significant changes to reframe a number of negatively worded items with positive phrasing to simplify questions [55]. Whilst we took these extra steps to overcome some of the challenges reported by patients in our PMOS-A think aloud and pilot study, it is possible that changes to some of the wording affected the meaning or interpretation of items by participants [16].

Our findings illustrate that although the model may benefit from additional refinement, this validated PMOS-A questionnaire is adequate for use as a corroborative measure against another assessment of healthcare quality, including percentage of adherence to clinical guidelines. However, these relationships were based on hospital median scores and should be interpreted with some caution.

As to limitations, the patient population was sampled from only three departments across most of the 32 hospitals; inclusion of additional ward types (e.g. maternity, paediatric and general) could have provided a more representative sample of the patient population as well as a wider age range. Furthermore, the response rate was estimated to be 34.5%, which is low and may have led to selection bias whereby the results are more representative of a subset of the targeted patients (e.g. more motivated, less unwell and more positive) rather than the entire stroke, AMI and hip fracture patient population. The recruitment and administration were largely performed remotely, albeit using detailed instructions from the DUQuA research team. As such, our insights into the rigour with which the protocol was followed are limited. Nonetheless, revealing that it is possible to overcome challenges associated with engaging healthcare professionals from 32 hospitals across five states and two territories to recruit almost 1000 patients to complete the PMOS-A questionnaire is a noteworthy research coordination achievement.

There are a number of encouraging possible implications from this study. First, we have used sophisticated statistical methods to demonstrate the psychometric properties of properties of the PMOS-A, particularly with an older, more vulnerable population—this will be useful for both researchers and healthcare organizations seeking to understand patient perceptions of safety from a multidimensional perspective. Second, validation across independent and assisted completers, with insights into the age-related differences in these samples, provides a justification for considering stratifying approaches to efficiently administering the PMOS-A whilst maintaining reliability of responses. Third, higher PMOS scores were significantly associated with a key quality-related outcome measure—adherence to clinical guidelines; whilst these findings indicate association rather than causation, they are promising given the scepticism around what patients can really tell us about safety [7, 56, 57].

Conclusion

Clearly, more research should be undertaken to fully understand the uses and limitations of the PMOS-A, but after validation across

two countries, including tests to differentiate between older, more vulnerable populations and different modes of administration, a sufficiently reliable and valid measure of patient perceptions of safety has been developed. These findings should provide sufficient support to justify the use of this measure to assess patient perceptions of safety in hospitals. More work is recommended in order to understand the ability of the PMOS-A to diagnose specific areas of safety to drive the development and testing of targeted interventions to enhance the safety of clinical practice.

Supplementary material

Supplementary material is available at *INTQHC Journal* online.

Acknowledgements

Members of the original Deepening our Understanding of Quality Improvement in Europe (DUQuE) team (Oliver Gröne and Rosa Suñol) provided extensive input and advice in the initial stages of the project, supporting design modifications from DUQuE and the planned DUQuA study approach. Dr Annette Pantle provided expert advice on revision and development of DUQuA measures, and Professor Sandy Middleton, Associate Professor Dominique Cadillac, Kelvin Hill, Dr Carmel Crock and Professor Jacqueline Close provided input into the development of our AMI, hip fracture and stroke clinical process indicator lists. Nicole Mealing and Victoria Pye provided statistical advice in the initial phases of the project. We greatly appreciate their efforts. Zhicheng Li and Chrissy Clay also organized and coordinated the external site visits at all participating hospitals with all external researchers/assessors (surveyors).

Funding

This research was funded by the National Health and Medical Research Council (NHMRC) Program Grant APP1054146 (CI Braithwaite).

Authors' contributions

The research team consists of experienced researchers, clinicians and biostatisticians with expertise in health services research, survey design and validation, large-scale research and project management, sophisticated statistical analysis, quality improvement and assessment, accreditation, clinical indicators, policy and patient experience. JB led the research grant to fund the project. NT led the study design and development of the manuscript. RCW, HPT and GA co-led the study design and contributed to the development of the manuscript. HPT, GA and RL also provided statistical expertise for the study design and developed the analysis plan and contributed to the development of the manuscript. TW and EH contributed to the logistics of project management, the refinement of measures and the development of the manuscript.

Data sharing statement

Data will be made publicly available to the extent that individual participants or participating hospitals cannot be identified, in accordance with requirements of the approving Human Research Ethics Committees.

Ethical considerations

Ethical approvals were secured from State and Territory human research ethics committees in New South Wales (#14/206), Victoria (#15/36), the Australian Capital Territory (#15/131), South Australia (#15/260), the National Territory (#15-2509), Tasmania (#H0015383) and Queensland (#15/361). Site-specific authorizations were also granted for all participating hospitals. We complied with confidentiality requirements of national legislation or standards of practice of each jurisdiction. Respondents were anonymous, and participants provided informed consent.

References

1. Darzi A. *High Quality Care for All: NHS Next Stage Review Final Report*, Vol. 2008. London, Department of Health, 2009.
2. Australian Bureau of Statistics. In: Statistics Australian Bureau of Statistics (ed.). *Patient Experiences in Australia: Summary of Findings*. Commonwealth of Australia, Canberra, 2011.
3. Anhang Price R, Elliott MN, Zaslavsky AM *et al.* Examining the role of patient experience surveys in measuring health care quality. *Med Care Res Rev* 2014;**71**:522–54.
4. Secanell M, Groene O, Arah OA *et al.* Deepening our Understanding of Quality Improvement in Europe (DUQuE): overview of a study of hospital quality management in seven countries. *Int J Qual Health Care* 2014;**26**:5–15.
5. Maben J, Peccei R, Adams M *et al. Exploring the Relationship between Patients' Experiences of Care and the Influence of Staff Motivation, Affect and Wellbeing*. Southampton, UK, NIHR Service Delivery and Organization Programme, 2012.
6. Kelly N, Blake S, Plunkett A. Learning from excellence in healthcare: A new approach to incident reporting. *Arch Dis Child* 2016;**101**:788–91.
7. Ward JK, Armitage G. Can patients report patient safety incidents in a hospital setting? A systematic review. *BMJ Qual Saf* 2012;**21**:685–99.
8. Baxter R, Taylor N, Kellar I *et al.* Identifying positively deviant elderly medical wards using routinely collected NHS safety thermometer data: an observational study. *BMJ Open* 2018;**8**:e020219.
9. Groene O, Klazinga N, Wagner C *et al.* Investigating organizational quality improvement systems, patient empowerment, organizational culture, professional involvement and the quality of care in European hospitals: the 'Deepening our Understanding of Quality Improvement in Europe (DUQuE)' project. *BMC Health Serv Res* 2010;**10**:281.
10. Groene O, Arah OA, Klazinga NS *et al.* Patient experience shows little relationship with hospital quality management strategies. *PloS One* 2015;**10**:e0131805.
11. Glasgow RE, Riley WT. Pragmatic measures: what they are and why we need them. *Am J Prev Med* 2013;**45**:237–43.
12. O'Hara JK, Lawton RJ, Armitage G *et al.* The patient reporting and action for a safe environment (PRASE) intervention: a feasibility study. *BMC Health Serv Res* 2016;**16**:676.
13. O'Hara JK, Lawton RJ. At a crossroads? Key challenges and future opportunities for patient involvement in patient safety. *BMJ Qual Saf* 2016;**25**:565–8.
14. Johnstone MJ, Kanitsaki O. Engaging patients as safety partners: some considerations for ensuring a culturally and linguistically appropriate approach. *Health Policy* 2009;**90**:1–7.
15. Coulter A, Cleary PD. Patients' experiences with hospital care in five countries. *Health Aff (Millwood)* 2001;**20**:244–52.
16. Taylor N, Hogden E, Clay-Williams R *et al.* Older, vulnerable patient view: a pilot and feasibility study of the patient measure of safety (PMOS) with patients in Australia. *BMJ Open* 2016;**6**:e011069.
17. Avis M, Bond M, Arthur A. Satisfying solutions? A review of some unresolved issues in the measurement of patient satisfaction. *J Adv Nurs* 1995;**22**:316–22.
18. Wensing M, Elwyn G. Research on patients' views in the evaluation and improvement of quality of care. *Qual Saf Health Care* 2002;**11**:153–7.
19. Barrow E. *Putting the Patient in Patient Safety: Qualitative Exploration and Co-design of a New Paradigm*. London, UK, Imperial College London, 2018.
20. Clay-Williams R, Taylor N, Braithwaite J. Potential solutions to improve the governance of multicentre health services research. *Med J Aust* 2018;**208**:152–4.
21. McEachan RRC, Lawton RJ, O'Hara JK *et al.* Developing a reliable and valid patient measure of safety in hospitals (PMOS): a validation study. *BMJ Qual Saf* 2014;**23**:565–73.

22. Giles SJ, Lawton RJ, Din I, McEachan RRC. Developing a patient measure of safety (PMOS). *BMJ Qual Saf* 2013;**22**:554–62.
23. Dozier AM, Kitzman HJ, Ingersoll GL *et al*. Development of an instrument to measure patient perception of the quality of nursing care. *Res Nurs Health* 2001;**24**:506–17.
24. Care Quality Commission. *National Summary: Inpatient Survey Results 2012*. Newcastle, UK: Picker Institute Europe, 2013.
25. Oltedal S, Garratt A, Bjertnaes O *et al*. The NORPEQ patient experiences questionnaire: data quality, internal consistency and validity following a Norwegian inpatient survey. *Scand J Public Health* 2007;**35**:540–7.
26. Lawton R, McEachan RRC, Giles SJ *et al*. Development of an evidence-based framework of factors contributing to patient safety incidents in hospital settings: a systematic review. *BMJ Qual Saf* 2012;**21**:369–80.
27. Lawton R, O'Hara JK, Sheard L *et al*. Can patient involvement improve patient safety? A cluster randomised control trial of the patient reporting and action for a safe environment (PRASE) intervention. *BMJ Qual Saf* 2017;**26**:622–31.
28. Sheard L, O'Hara J, Armitage G *et al*. Evaluating the PRASE patient safety intervention—a multi-Centre, cluster trial with a qualitative process evaluation: study protocol for a randomised controlled trial. *Trials* 2014;**15**:2282.
29. Louch G, O'hara J, Mohammed MA. A qualitative formative evaluation of a patient-centred patient safety intervention delivered in collaboration with hospital volunteers. *Health Expect* 2017;**20**:1143–53.
30. Taylor N, Clay-Williams R, Hogden E *et al*. Deepening our Understanding of Quality in Australia (DUQuA): a study protocol for a nationwide, multilevel analysis of relationships between hospital quality management systems and patient factors. *BMJ Open* 2015;**5**:e010349.
31. Braithwaite J, Clay-Williams R, Taylor N *et al*. Deepening our Understanding of Quality in Australia (DUQuA): An overview of a nation-wide, multilevel analysis of relationships between quality management systems and patient factors in 32 hospitals. *Int J Qual Health Care* 2020;**32**:8–21.
32. MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. *Psychol Methods* 1999;**4**:84.
33. Little RJ. A test of missing completely at random for multivariate data with missing values. *J Am Stat Assoc* 1988;**83**:1198–202.
34. Howell DC. The analysis of missing data. In: Outhwaite W, Turner S (eds.). *The SAGE Handbook of Social Science Methodology*. London, UK, Sage Publications, 2007.
35. Dong Y, Peng CY. Principled missing data methods for researchers. *Springer Plus* 2013;**2**:222.
36. Arnolda G, Winata T, Ting HP *et al*. Implementation and data-related challenges in the Deepening our Understanding of Quality in Australia (DUQuA) study: implications for large-scale cross-sectional research. *Int J Qual Health Care* 2020;**32**:75–83.
37. Chin WW, Todd PA. On the use, usefulness, and ease of use of structural equation modeling in MIS research: a note of caution. *MIS Q* 1995;**19**:237–46.
38. Hutchinson SR, Olmos A. Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. *Struct Equ Modeling* 1998;**5**:344–64.
39. Mardia KV. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 1970;**57**:519–30.
40. Kline RB. *Principles and Practice of Structural Equation Modeling*, 2nd edn. New York, NY, Guilford Press, 2005.
41. McIntosh CN. Rethinking fit assessment in structural equation modelling: a commentary and elaboration on Barrett (2007). *Pers Individ Dif* 2007;**42**:859–67.
42. Bentler MP, Bonett D. Significance tests and goodness-of-fit in analysis of covariance structures. *Pyschol Bull* 1980;**8**:588–606.
43. Jöreskog KG, Sörbom D. *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language: Scientific Software International Inc*., Chicago, IL: 1993.
44. Barrett P. Structural equation modelling: adjudging model fit. *Pers Individ Dif* 2007;**42**:815–24.
45. Wheaton B, Muthen B, Alwin DF, Summers GF. Assessing reliability and stability in panel models. *Sociol Methodol* 1977;**8**:84–136.
46. Hu L-T, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;**6**:1–55.
47. Tabachnick BG, Fidell LS, Ullman JB. *Using Multivariate Statistics*. Boston, MA, Pearson, 2007.
48. Hooper D, Coughlan J, Mullen M. Structural equation modelling: guidelines for determining model fit. *EJBRM* 2008;**6**: 53–60.
49. Nunnally JC. *Psychometric Theory*, 2nd edn. New York, NY, McGraw-Hill, 1978.
50. Bagozzi RP, Yi Y. On the evaluation of structural equation models. *J Acad Mark Sci* 1988;**16**:74–94.
51. Clark LA, Watson D. Constructing validity: basic issues in objective scale development. *Psychol Assess* 1995;**7**:309.
52. Taylor N, Lawton R, Conner M. Development and initial validation of the determinants of physical activity questionnaire. *Int J Behav Nutr Phys Act* 2013;**10**:74.
53. Muthén B, Kaplan D. A comparison of some methodologies for the factor analysis of non-normal Likert variables. *Br J Math Stat Psychol* 1985;**38**:171–89.
54. Potthast MJ. Confirmatory factor analysis of ordered categorical variables with large models. *Br J Math Stat Psychol* 1993;**46**:273–86.
55. Hancock K, Chenoweth L, Chang E. Challenges in conducting research with acutely ill hospitalized older patients. *Nurs Health Sci* 2003;**5**:253–9.
56. Koutantji M, Davis R, Vincent C, Coulter A. The patient's role in patient safety: engaging patients, their representatives, and health professionals. *Clinical Risk* 2005;**11**:99–104.
57. Entwistle VA, Mello MM, Brennan TA. Advising patients about patient safety: current initiatives risk shifting responsibility. *Jt Comm J Qual Patient Saf* 2005;**31**:483–94.

International Journal for Quality in Health Care, 2020, 32(S1), 75–83
doi: 10.1093/intqhc/mzz108
Perspectives on Quality

Perspectives on Quality

Implementation and data-related challenges in the Deepening our Understanding of Quality in Australia (DUQuA) study: implications for large-scale cross-sectional research

GASTON ARNOLDA[1], TERESA WINATA[1], HSUEN P. TING[1], ROBYN CLAY-WILLIAMS[1], NATALIE TAYLOR[2,3], YVONNE TRAN[1], and JEFFREY BRAITHWAITE[1]

[1]Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Road, NSW 2109, North Ryde, Australia, [2]Cancer Research Division, Cancer Council NSW, 153 Dowling St, Woolloomooloo, NSW 2011, Woolloomooloo, Australia, and [3]Faculty of Health Sciences, University of Sydney, Camperdown, Sydney, NSW 2006, Sydney, Australia

Address reprint requests to: Jeffrey Braithwaite, Centre for Healthcare Resilience and Implementation Science (CHRIS), Australian Institute of Health Innovation, Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia. Tel: +61 2 9850 2401; Fax: +61 2 9850 2499; E-mail: jeffrey.braithwaite@mq.edu.au

Editorial Decision 7 September 2019; Accepted 12 September 2019

Abstract

Healthcare organisations vary in the degree to which they implement quality and safety systems and strategies. Large-scale cross-sectional studies have been implemented to explore whether this variation is associated with outcomes relevant at the patient level. The Deepening our Understanding of Quality in Australia (DUQuA) study draws from earlier research of this type, to examine these issues in 32 Australian hospitals. This paper outlines the key implementation and analysis challenges faced by DUQuA. Many of the logistical difficulties of implementing DUQuA derived from compliance with the administratively complex and time-consuming Australian ethics and governance system designed principally to protect patients involved in clinical trials, rather than for low-risk health services research. The complexity of these processes is compounded by a lack of organizational capacity for multi-site health services research; research is expected to be undertaken in addition to usual work, not as part of it. These issues likely contributed to a relatively low recruitment rate for hospitals (41% of eligible hospitals). Both sets of issues need to be addressed by health services researchers, policymakers and healthcare administrators, if health services research is to flourish. Large-scale research also inevitably involves multiple measurements. The timing for applying these measures needs to be coherent, to maximise the likelihood of finding real relationships between quality and safety systems and strategies, and patient outcomes; this timing was less than ideal in DUQuA, in part due to administrative delays. Other issues that affected our study include low response rates for measures requiring recruitment of clinicians and patients, missing data and a design that necessarily included multiple statistical comparisons. We discuss how these were addressed. Successful completion of these projects relies on mutual and ongoing commitment, and two-way communication between the research team and

hospital staff at all levels. This will help to ensure that enthusiasm and engagement are established and maintained.

Key words: methods, cross-sectional studies, health services research, quality of healthcare, patient safety, hospitals

Introduction

The quality and safety movement began in earnest over two decades ago by estimating the extent of iatrogenic injury [1, 2] and its associated cost, and recommended that healthcare needed to be reconfigured to address the systemic issues that contribute to these problems [3]. Interventions to improve quality and safety have been advocated at system level (e.g. external hospital accreditation) [4, 5], at the level of organizations that manage care (e.g. quality and safety policy and procedures in hospitals) [6, 7] and at the level of care teams (e.g. routine reviews of compliance with clinical guidance in a ward) [8, 9]. Comprehensive implementation and evaluation of these interventions is encouraged.

Compliance with recommended activities comes with an opportunity cost, as hospitals are busy places, often struggling to meet patient needs within limited resources [10]. There is, therefore, a need to understand the effects of these activities, so that we can prioritise those with the greatest potential for improving patient outcomes.

The quality improvement (QI) literature is replete with small or single-ward studies comparing outcomes before and after an intervention. There are potential risks, however, that the results of these studies are not generalisable beyond their immediate environments, and of a publication bias in favour of "good news". Multi-site before-after studies, cluster randomized trials or stepped-wedge studies can address some of these issues, through larger sample sizes and the inclusion of controls. These types of studies tend to focus on a single highly salient subject (e.g. antenatal corticosteroids for preterm birth [11], or attempts to reduce catheter-related infections [12, 13]) to which an enormous amount of energy is dedicated, potentially restricting generalizability beyond that subject.

Few studies have sought to explore the opportunity offered by the observed variation in implementation of quality and safety systems and strategies *in situ*, across organizations. The Methods of Assessing Response to Quality Improvement Strategies (MARQuIS) study, amongst the first large-scale attempts to exploit this opportunity, examined organization quality systems and department quality strategies in 389 hospitals in eight European countries (Spain, France, Poland, Czech Republic, the UK, Ireland, Belgium and the Netherlands) [14, 15]. MARQuIS was followed by the Deepening our Understanding of Quality improvement in Europe (DUQuE) study, which undertook a similar study in 188 hospitals in seven countries (Czech Republic, France, Germany, Poland, Portugal, Spain and Turkey) [16]. These cross-sectional studies at scale measured the degree of implementation of quality and safety policies and strategies at different levels in hospitals, to explore the relations between implementation at different levels, and to identify those measures that are most strongly associated with positive outcomes. The Deepening our Understanding of Quality in Australia (DUQuA) study sought to adapt and extend the DUQuE study for the Australian context, and to modify and extend some of the approaches [17].

The DUQuA study, which ran for five years, was conducted to assess the relationship between quality measures implemented at hospital and department levels, and their relationship with patient outcomes [17, 18]. The study covered three common conditions: acute myocardial infarction (AMI), stroke and hip fracture [19]. In total, we analysed: questionnaire data from 31 quality managers, 857 patients and 1332 clinicians; clinical process audit data for 2401 patients through medical record review and the National Stroke Foundation (NSF) registry; hospital-level and department-level quality assessments data from 151 external assessments; and national audit data for 14 460 index hospital admissions from an Australian Institute of Health and Welfare (AIHW) database [17].

Studies like MARQuIS, DUQuE and DUQuA seek to understand the complex multi-level structure of quality and safety interventions by exploring variation across organisations. Cross-sectional studies encounter substantial problems and constraints. Conducting such studies on a large scale presents logistical and data-related challenges, which add to the difficulty of determining the validity and interpretation of study results.

Here, we examine the challenges experienced in undertaking the DUQuA study, as an end in itself and as a guide to others considering undertaking similar research. We describe relevant experiences to elucidate the challenges for these types of studies in examining complex multi-level relationships. These challenges are separated into two conceptual categories, implementation challenges and analytic issues, for ease of exposition.

Implementation challenges

Numerous implementation headwinds were encountered in undertaking DUQuA, some with implications for interpretation of study results. We discuss a selection of key issues: hospital eligibility; recruitment and retainment of hospitals; ethical and governance approvals; sourcing relevant data routinely collected by external bodies; timing differences in collecting a range of measures across a large number of hospitals; and the response rates achieved. We identify the challenges experienced, our approaches to manage them, and likely implications for future studies.

Defining the eligibility criteria for hospitals. A key challenge was the specification of a study population. Focused on Australia, DUQuA has a smaller population base than MARQuIS or DUQuE. To maximise statistical power, DUQuA needed to enrol as many hospitals as possible, within available resources. The study was restricted to public hospitals because hospitals with Emergency Departments (EDs) providing services for all three target conditions were usually public, and ethics processes were expected to be more standardised. The study was inevitably restricted to larger hospitals for practical reasons: a minimum hospital size was required to ensure that hospitals have the resources to implement a range of quality measures; a minimum number of patients were required for each clinical condition to justify ward-level quality and safety strategies; and a minimum number of staff and patients were required to ensure that desired sample sizes could be achieved.

The specification of these eligibility criteria has implications for the generalizability of study results. DUQuA results cannot necessarily be taken as indicative of performance in smaller hospitals. Importantly, however, whilst the target hospitals make up approximately 20% of all Australian public hospitals providing acute care, they were responsible for over three-quarters of all in-patient admissions [20].

Recruitment of eligible hospitals. Once targeted hospitals were identified, the second challenge was to recruit them. An invitation, supported by the Royal Australasian College of Medical Administrators, was sent to each hospital's director of medical services, or an equivalent senior hospital leader. Seventy-eight hospitals were approached and 62 initially agreed to participate. Ultimately, 30 were lost primarily due to leadership and organisational changes, time-consuming ethics and governance requirements, and concerns about workload, leaving 32 participants.

The 62 hospitals that initially agreed had mixed views of the study; some saw DUQuA as an opportunity to improve their quality management systems and thus considered it a priority, whilst others saw it as a burden. These retrospective reflections were not systematically recorded by hospital, but point to a potential for selection bias. As only 32 of this group proceeded to full participation, it is plausible that these were the hospitals more committed to improving quality and safety, and that their systems and processes differed systematically from the 30 hospitals which ultimately withdrew. Supplementary eTable 1 compares key characteristics of the participating and non-participating hospitals and shows that the sample appears to: over-represent Queensland and under-represent Western Australia; over-represent principal referral hospitals; under-represent inner regional remoteness areas; and over-represent hospitals with 500 or more beds. As a consequence, there are additional limits on generalizability.

Ethical and governance processes. Ethical and governance approval processes represent significant logistical barriers in undertaking multi-site research in Australia. We have previously described these issues in detail [21, 22]. Briefly, research in Australian hospitals must first be reviewed and approved by a properly constituted Human Research Ethics Committee in line with the National Statement on Ethical Conduct in Human Research [23]. At the time of DUQuA implementation, at least one ethics application was required in each of six States and two Territories; whilst there has been some harmonization of ethics application forms across the country in recent decades, a plethora of nuances still remain and each State and Territory had different requirements.

After gaining ethical approvals, projects must then submit a site-specific assessment (SSA) application, which considers matters such as physical resources, staff, insurance and indemnity requirements in addition to ethics affairs. States and Territories have very similar SSA forms, but each hospital has its own variations in documentation, submission processes and panel turnaround time. This is burdensome. On average, eight department head signatures were required per hospital, prior to submission. In addition, most hospitals required research agreements, which had to be prepared by the University's legal Department.

A timeline for ethics and governance approvals for DUQuA is provided in Figure 1. The timeline does not include disruptions that occur after SSAs and research agreements have been completed. For example, a hospital employee must be nominated and assigned as the study site principal investigator (PI) on the ethics and SSA form. High hospital staff turnover meant that some PIs left the study requiring a replacement and amendment to documents.

In short, DUQuA implementation was subjected to an administrative system designed to protect patients participating in interventional clinical trials, creating a large administrative workload. The end result was a diversion of research resources away from study implementation, towards the satisfaction of bureaucratic process, and the demotivation of potential study participants. It is very likely that this contributed to the decision of some hospitals to withdraw, and may have reduced the response rates for data measures from staff within hospitals, and their eagerness to recruit patients. For hospitals that continued with the study, the planned timeline for data collection was markedly delayed.

Sourcing of external data. Only two sets of patent-level data were externally sourced: the NSF routinely collected indicators for stroke in 29 of the participating hospitals; and the AIHW routinely collects information on all patients admitted to Australian hospitals. The process of obtaining data from the AIHW can be time-consuming as AIHW can only provide data with the approval of all States and Territories, and data linkage is additionally restricted to specific authorities [24]. These factors made a request for comprehensive linked patient data infeasible, resulting in uncertainty as to the accuracy of length of stay data; restriction of the mortality measure to death as the mode of hospital separation; and restriction of the hospital readmission measure to 'same-hospital readmission within 28 days'.

Variable timing of data collection. DUQuA sought to measure the effect of organization-level systems, department-level strategies and clinician perceptions of teamwork, safety climate and leadership in their department on patient outcomes. Ideally, data collection would reflect this trajectory, with quality and safety measures collected at one point, and outcomes for a period immediately after.

The timing of individual measures varied widely between hospitals (Figure 2). The peak period for collection of organization, department and clinician-level measures was February to August 2017 but, due to difficulties coordinating hospital-internal activities between different participants and external quality assessors, the quality and safety measures and patient outcomes relate to different time points. For example, hospital 23 collected organizational-level measures in June 2016 and May 2017, department-level measures and clinician surveys in May 2017, and patient surveys in August and September 2017. Other patient-level data, not shown in Figure 2, had pre-specified time periods for all hospitals to align with the NSF data collection period: medical records reviews were based on samples in the time window from September 2014 to February 2015; and AIHW data represented patients discharged between July 2014 and June 2015. These timing differences represent measurement errors with implications for interpretation of study results as discussed below.

Response rates. Response rates were different for different measures. As shown in Supplementary eTable 2, response rates (calculated as data available for analysis as a percentage of that targeted) to externally sourced measures (NSF and AIHW data) and single respondent measures (provided by the Quality Manager or the External Quality Assessor) were excellent (99% or above), and the medical record reviews also had high response rates (80% or above). By contrast, completion rates compared to targets were 37% for the clinician survey and 32% for the patient survey; these response rates were non-uniform, with no or few responses from some participating departments. This creates the potential for selection bias, which may in turn undermine the accuracy of estimated relationships between measures.

Analysis-related issues

A study of this scale addresses a multitude of research questions using a variety of analysis techniques. The primary focus of analysis was a series of hierarchical regression analyses, with up to 32 hospital-level measures, 32 ED measures, 87 inpatient department

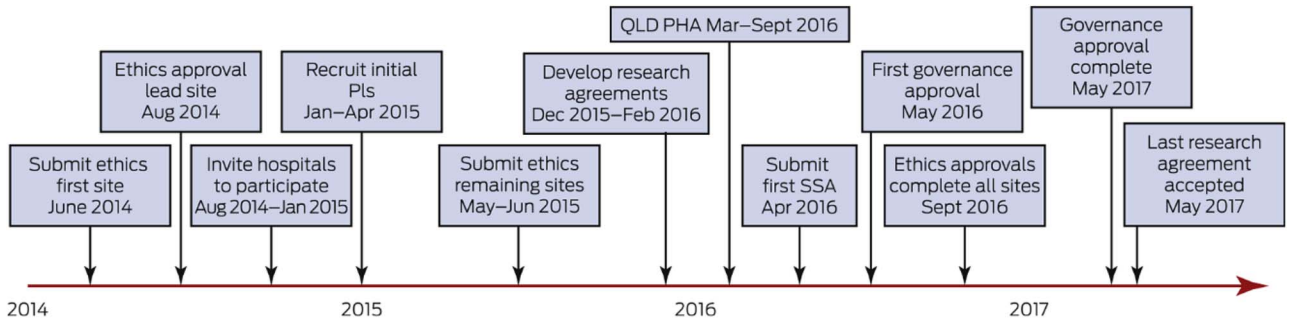

Figure 1 Ethics and governance timeline for the DUQuA study (Source: [21]). Legend: QLD PHA, Queensland Public Health Act 2005.

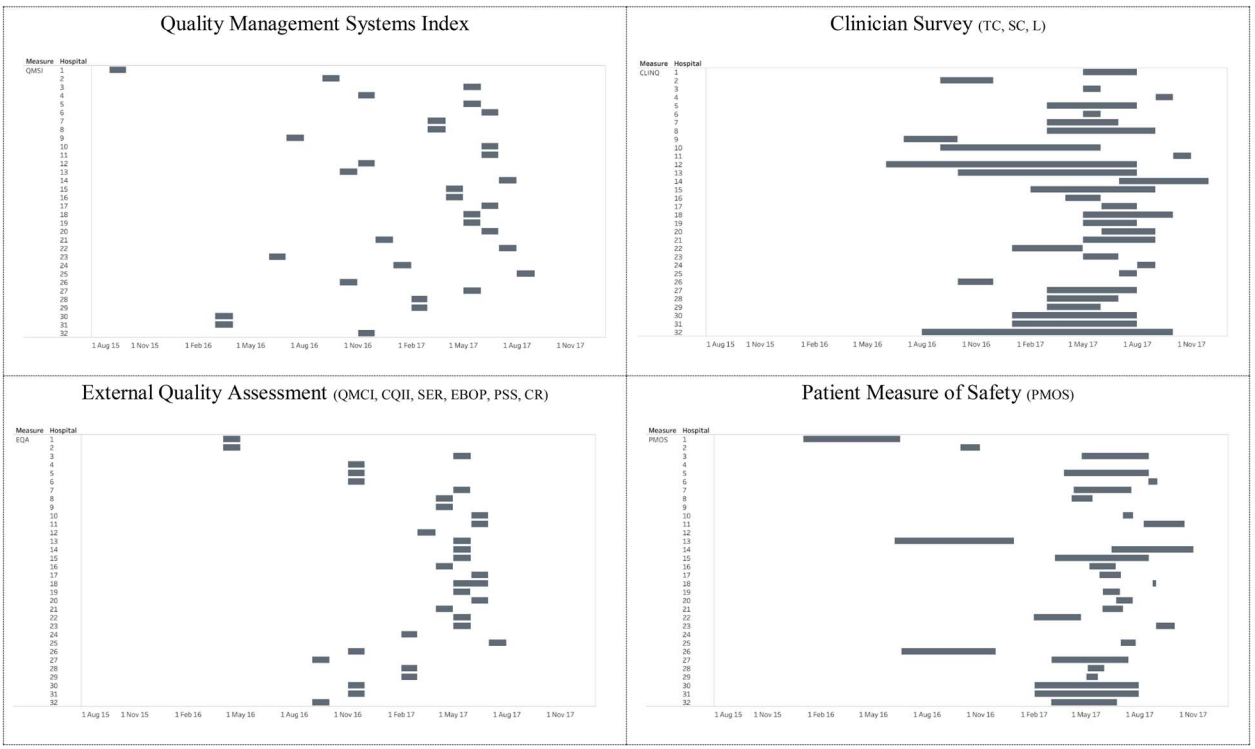

Figure 2 Timeline for data collection by hospital for the DUQuA study. Legend: External Quality Assessment comprised organization-level measures (QMCI=Quality Management Compliance Index, CQII=Clinical Quality Improvement Index), department-level measures (SER=Specialised Expertise and Responsibility, EBOP=Evidence-Based Organisation of Processes, PSS=Patient Safety Strategies, CR=Clinical Review); Clinician survey measures comprised department-level measures (TC=Teamwork Culture, SC=Safety Culture, L=Leadership).

measures (27 AMI, 29 hip fracture and 31 stroke), and multiple patient-level measures within each department [17, 25–27]. *En route* to this goal, we assessed the validity and reliability of a number of scales: three at organisation-level (Quality Management Systems Index [QMSI], Quality Management Compliance Index [QMCI] and Clinical Quality Implementation Index [CQII]) [28]; four assessing department level strategies (Specialised Expertise and Responsibility [SER], Evidence-Based Organisation of Pathways [EBOP], Patient Safety Strategies [PSS] and Clinical Review [CR]) [28]; three assessing department-level clinician safety culture and leadership (Teamwork Climate [TC], Safety Climate [SC], and Leadership [L]) [29]; and one patient-related outcome (Patient Measure of Safety [PMOS]) [30].

We assess challenges encountered during analysis: response rates; missing data; measurement error and multiple comparisons. For each issue, we discuss analysis strategies separately for validation studies and hierarchical models.

Response rates. As discussed above and shown in Supplementary eTable 2, response rates were excellent for single respondent measures but poor where clinicians or patients had to be recruited. For the clinician survey (37% of target), for example, no responses were received from 15 departments (13%) with a median of 13 responses per participating department. For the patient survey, the poor response rate (33%) was compounded by unassisted completion (44%) which was permitted, contrary to previous practise [30], because staff were too busy to assist.

Validation studies: The only validation studies with a low response-rate were the PMOS and clinician surveys [29, 30]. For both, the approximately one-third response rate reduced the sample size to around 1000 records, further reduced to half when split into calibration and validation datasets. In PMOS, to address the issue of self - versus assisted completion, we separately assessed model validity and reliability for all respondents and for the subset of assisted respondents [30].

Hierarchical models: In hierarchical models, a poor response rate at higher levels (e.g. an organization level measure) affects the sample size more severely than at lower levels (e.g. an individual clinician survey response). A single non-responding hospital or department eliminates all lower level measures from the model in a cascade. Poor response to lower level measures only results in elimination of a department or hospital from the model if all measures at that level are missing; for example, a department with no clinician survey measures would result in removal of that department from the model. For PMOS validation, we separately modelled all respondents and compared this to the subset of assisted respondents; in the hierarchical models we used all responses, due to the already low response rate and the similar performance of the two models in validation.

Missing data. Missing data are often an issue even for small studies, so it is not surprising that it was a challenge for a large multi-level research study. Table 1 shows the missing data rates for each measure used in the study, as calculated for the hierarchical models. At item level, the overall average percentage of missing data was consistently <2%. For organization-level measures (QMSI, QMCI and CQII), for example, the subscale was set to missing if more than 50% of the items were missing, and as the scale is the sum of the subscales, all subscales were required for a scale to be calculated. For departmental strategies (SER, EBOP, PSS and CR), more than 50% of the applicable items in a scale had to be missing for the scale to be missing. Missing data for clinician survey data on teamwork and safety culture had multiple respondents per ward, but also used a rule of setting a scale to missing if any item was missing. For PMOS, a subscale was only set to missing if more than 50% of items in the subscale were missing.

Given the sample size of 32 hospitals, preserving every hospital in the analysis was important to maintain statistical power. Even though missing data were minimal for items in organization-level measures, imputation was essential.

Validation studies: For the validation studies, the major challenge with missing data occurred with responses that contained 'Not Applicable' (NA) for some of the scales; this was found in four department strategy measures and PMOS. In validating the departmental strategy measures, NA responses were treated as missing as it indicated that the question could not be answered for that department. In validating PMOS, the NA response was imputed at the midpoint score as this was deemed the implicit value (i.e. neither agree nor disagree). PMOS, additionally, had a unique response category of 'Prefer not to answer' that was treated as missing. When reporting descriptive statistics of item, subscale and scale scores, we reported unimputed data (except for fixed score replacement of NA in PMOS).

When performing Confirmatory Factor Analysis (CFA), to validate the clinician survey measures and PMOS, a single missing item results in list-wise deletion, so an expectation maximisation (EM) method was used to impute all missing items before undertaking the model validation process. EM is a single imputation method, which generates one imputed dataset; for simplicity, we used this method in preparation for CFA in preference to using methods which create multiple imputation datasets (as used in the hierarchical modelling where we were generating pooled estimates). Thus, the fit statistics reflected the EM imputed data.

In validating the three organizational measures and the four department-level strategies, no CFA was undertaken, so internal consistency measures were calculated using unimputed data. In line with DUQuE procedures [31], however, we used multiple imputed data (described below) to calculate the Pearson correlations between subscales of each of the three organizational-level measures.

Hierarchical models: All hierarchical models included the three organizational-level measures. As indicated above, multiple imputation was performed to generate multiple datasets, so that true variance could be estimated. Standard procedure recommends a minimum of 2–10 imputed datasets [32], or a number calculated as 100 times the proportion of missing data [33]. As processing capacity is rapidly ceasing to become a constraint, we conservatively chose to generate 100 imputation datasets despite the small percentage of missing data. When imputing, we followed recommendations to not bound the imputed value to the usual scoring range (e.g. if the usual scoring range is 0–4, the results can be restricted to this range), and not transform non-normal data, as this can bias the variance estimate [34]. These datasets were then each analyzed and the results pooled [32] using the SAS MIANALYZE procedure to ensure variance was reflected, except in situations where the between-imputation variance was zero where only one iteration was required to establish the confidence interval of the estimate.

Measurement error.

Validation studies: The detailed validation results are reported elsewhere [28–30]. In brief, where we assessed model fit statistics only two of the organization-level scales (QMSI, QMCI) were adequate, as CQII results displayed a strong ceiling effect, with four sub-scales all showing means of 3.9/4.0. The four scales assessing departmental strategies (SER, EBOP, PSS and CR) had different department-specific profiles, with CR consistently showing internal reliability across the four departments and EBOP showing internal reliability in most departments, but PSS and SER showing poor or moderate reliability depending on the department. Where CFA was performed (clinician surveys and PMOS) [29, 30], the results did not provide unambiguous support for the pre-specified factor structure, with most fit indices suggesting acceptable fit and none showing good fit.

Because of these limitations, caution is recommended when interpreting the scales. This is also a potential source of random measurement error when the scales are used in the hierarchical models. Random measurement error in exposure or outcome variables always results in a bias towards the finding of statistically non-significant results, increasing the likelihood of false negative findings for outcomes drawn from clinician surveys or PMOS.

Hierarchical models: The potential sources of measurement error varied with the measure. These are discussed individually.

The QMSI survey was undertaken by the Quality Manager at the hospital; in developing this measure, care was therefore taken to ensure that the scoring items were clear and easily delineable, to reduce the potential for biased reporting. For example, in Section 4 'Hospital governance board activities: To what extent do you agree with the statement for your hospital? The hospital governance board has established formal roles for quality leadership (visible in organizational chart)'; this was scored on a four point scale (1 = "Strongly disagree", 2 = "Somewhat disagree", 3 = "Somewhat agree" or 4 = "Strongly agree"). To further minimise the potential for biased reporting, external assessment may have been preferable or, at least, a random sample of QMSI forms could have been separately assessed by an external assessor, to estimate the potential for error. Both options have workload implications.

A number of scales were assessed by external surveyors: the other two organizational-level assessments (QMCI, CQII), and the four department-level strategies. Whilst scoring these scales required the exercise of judgement, they were designed to be objectively scored by staff experienced in external quality assessment. For example, the statement "There is a specialist (consultant) doctor available at all

Table 1 Missing or non-participation data rates in DUQuA measures

Measures		Missing scales/subscales and items[a]			Missing or non-participating hospitals and departments[b] after imputation		
		Items Mean (range)	Rule for setting subscale to missing	Scales or subscales Mean	Hospitals $N = 32$: No. (%)	ED $N = 32$: No. (%)	Clinical wards[c] $N = 88$: No. (%)
Organization-level							
Quality Management Systems Index (QMSI)		0.6% (0.0–6.5)	>50% of items in subscale missing	0.0%	0 (0%)	NA	NA
Quality Management Compliance Index (QMCI)		1.4% (0.0–6.3)	>50% of items in subscale missing	1.6%	0 (0%)	NA	NA
Clinical Quality Implementation Index (CQII)		1.3% (0.0–9.4)	>50% of items in subscale missing	0.4%	0 (0%)	NA	NA
Department level							
Specialized Expertise and Responsibility (SER)	ED	0%	>50% of applicable items in scale missing	0.0%	0 (0%)	0 (0%)	NA
	AMI	0%				NA	0 (0%)
	Hip fracture	1.1% (0.0–3.5)				NA	0 (0%)
	Stroke	0%				NA	1 (3.1%)
Evidence-Based Organisation of Pathways (EBOP)	ED	0%	>50% of applicable items in scale missing	0.0%	0 (0%)	0 (0%)	NA
	AMI	0%				NA	0 (0%)
	Hip fracture	0%				NA	0 (0%)
	Stroke	0%				NA	1 (3.1%)
Patient Safety Strategies (PSS)	ED	0%	>50% of applicable items in scale missing	0.0%	0 (0%)	0 (0%)	NA
	AMI	0.4% (0.0–3.8)				NA	0 (0%)
	Hip fracture	0.4% (0.0–3.5)				NA	0 (0%)
	Stroke	0%				NA	1 (3.1%)
Clinical Review (CR)	ED	0%	>50% of applicable items in scale missing	0.0%	0 (0%)	0 (0%)	NA
	AMI	0%				NA	0 (0%)
	Hip fracture	0%				NA	0 (0%)
	Stroke	0%				NA	1 (3.1%)
Teamwork culture (TC)	ED	0.5% (0.0–1.0)	Any item missing	5.8%	1 (3.1%)	3 (9.4%)	NA
	AMI	0.7% (0.0–1.6)		6.5%		NA	2 (7.4%)
	Hip fracture	0.7% (0.0–1.6)		8.5%		NA	3 (10.3%)
	Stroke	1.4% (0.6–2.8)		10.4%		NA	7 (21.9%)
Safety culture (SC)	ED	0.5% (0.0–1.0)	Any item missing	5.6%	1 (3.1%)	3 (9.4%)	NA
	AMI	0.5% (0.0–1.3)		5.2%		NA	2 (7.4%)
	Hip fracture	0.7% (0.0–2.0)		7.5%		NA	3 (10.3%)
	Stroke	1.3% (0.6–2.5)		7.3%		NA	7 (21.9%)
Leadership (L)	ED	0.4% (0.3–0.8)	Any item missing	1.0%	1 (3.1%)	3 (9.4%)	NA
	AMI	0.7% (0.3–1.6)		3.2%		NA	2 (7.4%)
	Hip fracture	0.1% (0.0–0.3)		1.0%		NA	3 (10.3%)
	Stroke	0.9% (0.6–1.5)		2.8%		NA	7 (21.9%)
Patient level							
Clinical treatment indicators	AMI	NA	NA	NA	0 (0%)	NA	2 (7.4%)
	Hip fracture						1 (3.4%)
	Stroke						3 (9.4%)

Continued.

Table 1 Continued.

Measures		Missing scales/subscales and items[a]			Missing or non-participating hospitals and departments[b] after imputation		
		Items Mean (range)	Rule for setting subscale to missing	Scales or subscales Mean	Hospitals $N = 32$: No. (%)	ED $N = 32$: No. (%)	Clinical wards[c] $N = 88$: No. (%)
AIHW inpatient statistic		NA	NA	NA	0 (0%)	NA	0 (0%)
Patient Measure of Safety	AMI	5.5% (0.6–21.0)	>50% of items in	8.3%	5 (15.6%)	NA	6 (22.2%)
(PMOS)	Hip fracture	5.5% (0.0–21.8)	subscale missing	10.9%		NA	6 (27.6%)
	Stroke	5.8% (0.7–23.3)		9.3%		NA	6 (34.4%)

[a]Rates reflect a denominator calculated after removal of records missing relevant confounders: for PMOS, 1.6% of data were removed because of missing age or sex; for the clinician survey, 3.0% of data were removed due to missing age, sex, profession or leadership role and for clinical treatment indicators, 2.0% of data were removed because of missing age or sex. At organization level (QMSI, QMCI, CQII) and for clinician surveys (TC, SC, L), the percentage missing was limited to data without a recorded response; for SER, EBOP, PSS and CR, the percentage missing uses a numerator of not recorded; for PMOS, the percentage missing uses a numerator of not recorded, responded as 'NA' or 'Prefer not to answer". Note that in validation of PMOS, a lower number was missing as 'NA' was imputed at the midpoint value.

[b]Data are restricted to hospitals and departments removed due to missing data or non-participation in a specific measure (not the response rates reported in Table 1).

[c]Each hospital did not have all three wards. Of 32 hospitals, 27 had AMI wards, 29 had hip fracture wards and 32 had stroke wards. These were the denominators for calculating percentage missing. Some wards elected not to participate in an activity and were treated as missing. For clinical treatment indicators, one pilot AMI ward, one pilot hip fracture ward and three stroke wards that did not have NSF data were treated as missing.

times to determine whether fibrinolysis or PCI is appropriate" was scored of a five point scale ranging from 0 = "No on call specialist doctor(s) specified" to 4 = "Specialist doctor(s) available 24 hours a day seven days a week". Whilst it would be preferable to undertake independent assessments by multiple surveyors, to calculate an agreement coefficient (e.g. kappa), this would have required multiple visits to the same hospital, which was not feasible.

In the clinician survey, an unexpectedly large proportion of respondents with usable data (47%) self-identified as having leadership roles, and this varied enormously by department, from 12 to 100%. No guidance was provided in the data form to define leadership, so the significance of this is unclear, but overall self-identified leaders only had marginally higher scores than non-leaders in their mean scores for TC (80.7 vs 77.3), SC (73.0 vs 70.5) and leadership (69.5 vs 66.5). Confounding by self-identified leadership status was controlled by adjusting for this in hierarchical models.

External assessors also abstracted from the clinical record the relevant clinical audit data in all but four eligible hospital AMI and hip fracture wards, and in three stroke departments that did not contribute data to the NSF registry. Local staff and external surveyors abstracted data from the medical record on a data form, rather than making a specific assessment of each indicator. There was missing data on a number of important variables including, for example, dates and times of admission and separation, and the time at which events occurred and procedures were performed. The assessment of compliance was therefore made using indicator-specific algorithms which restricted data to eligible records, which included rules specifying exclusion of some records with missing data. Separate indicator-specific algorithms then calculated compliance, again with rules to determine the handling of missing data in determining compliance or non-compliance, to permit calculation of rates. Rules for handling missing data may introduce measurement error in the estimated rates of indicator compliance.

The four departments where data were collected by local staff may have differed from external surveyors in the way data were extracted, creating a potential for differential measurement error where this occurred. In hierarchical models assessing audit results for stroke, analysis was restricted to the 29 hospitals with NSF data to prevent this problem. A final potential source of measurement error arises if the indicators used are not representative of important guideline compliance in the department; we undertook extensive consultation [17, 19] to ensure that the assessed guidelines were relevant and important to care of each condition.

The version of the PMOS instrument used by DUQuA comprised 43 questions, to calculate nine sub-scales [30]. The length of the form may discourage attempt of the form or assisted completion and may lead to questions being overlooked or may mitigate against completion. Selection criteria restricted this form to less vulnerable patients, but failure to complete the majority of questions may have led to systematic exclusion of a small group of target patients who do not complete enough of the form.

The data sourced from the AIHW comprised in-hospital mortality, same-hospital readmission within 28 days and length of stay for patients with ICD diagnostic codes identifying the target conditions [17]. Data are organized by patient episode of care, with information about reason for admission and type of separation; multiple episodes of care can be reported for a single patient during a single hospital visit. Analysis was therefore restricted to acute and emergency admissions, and excluded inward/outward acute transfers, and admissions and separations which were administrative in nature. As previously indicated, linking hospital separation with deaths registry

data was not feasible as it would have required a separate ethics application in each participating state and territory. Readmission was restricted to same-hospital readmission as states and territories do not have unique catchment-specific patient identifiers across hospitals, and some form of probabilistic linkage would therefore have been required, delaying provision of the data [24]. As mortality precludes readmission, we assessed readmission or mortality as a composite outcome. Length of stay data was necessarily and additionally restricted to admissions that did not end in death. Hospitals had widely different transfer rates, both inward and outward, so these restrictions were essential.

Disease severity of the selected patients in each participating hospital was quite likely variable, but we lacked data on disease severity at patient level (e.g. an individual risk score for mortality) and were only able to control for this at the gross level of hospital peer group (referral hospital vs public acute hospital). Adjustment at this level does not control for differences in patient disease severity between hospitals within each peer group.

In addition to these measure-specific issues relating to potential measurement error, there was also an important higher-order issue arising from the timing differences. There were broadly five outcome measures evaluated, including clinical treatment indicator compliance by medical record audit; three sets of AIHW recorded clinical outcomes (mortality, mortality or readmission, and length of stay); and PMOS. Medical record audit and AIHW outcomes both, for practical reasons, related to time periods that 'preceded' the date of data collection on measures at organization and department levels by an average of over two years; to the (unknown) extent that these outcome measures change over the 2-year period, this could represent a potentially important source of measurement error reducing the likelihood of finding associations with quality and safety activities. The date of PMOS collection in a hospital approximately coincided with other measures (see Figure 2).

Multiple comparisons versus insufficient sample size. Because of the size of the project, there were statistical analyses performed for a variety of purposes. This immediately raises the problem of multiple comparisons, such that the simple use of $P < 0.05$ is not justified as a signal of a meaningful relationship, as many such findings can arise purely by chance. To address this, we interpreted results within the context of the number of statistically significant findings that would be expected by chance.

Conversely, the sample size at hospital level ($n = 32$) risked the type II error of missing a weaker signal (e.g. significant at $P < 0.1$). To address this, we also looked for patterns of results that were suggestive; for example, were there patterns that were consistently indicative of a relationship in direction and quantum, across all of AMI, hip fracture and stroke or across organization-level versus ED versus ward-level measures?

Implications for future work

We can anticipate over time more large-scale cross-sectional studies exploring the relationship between quality and safety activities and patient-level outcomes. What is needed to improve the conduct of a study of this type?

Working with large numbers of hospitals over a long period of time is logistically challenging. It is clear that systemic changes are required to standardize and simplify both ethics and hospital governance procedures for observational, low risk research. The administrative load of compliance with governance requirements is compounded by frequent senior staff turnover in hospitals. Moreover, there is no clear expectation of hospitals to participate in external multi-hospital projects, nor a clear mechanism to resource this participation. The lack of a structure for health services research stands in contrast to externally funded clinical trials, for example, where hospitals routinely receive some form of compensation for costs associated with data collection for each recruited patient.

Moreover, the drawn-out ethics and governance process led to delays in measurement of quality and safety systems and strategies. This diverted research resources towards the fulfilment of bureaucratic administrative requirements, and increasingly to keeping hospital stakeholders engaged, informed and enthused as the process drew out.

Staff with clinical roles have competing demands for their time, which make it difficult to recruit them as sources of information, and the clinician-patient relationship can create ethical dilemmas for them in recruiting patients. The low response rates that result again raise questions about representativeness. Departments with no responses must be dropped from analysis, and those with low numbers present challenges for hierarchical models. Novel solutions may need to be considered, such as the funding of additional staff on wards during data collection, to recruit and support patients or to allow staff to complete surveys within their work time, confident that their clinical responsibilities are being addressed. In the absence of these, we rely on information sessions to build relationships with clinicians, to enthuse them to add the project to their existing workload.

There are multiple issues relating to measurement. Clear definitions and shorter measurement instruments are required to minimize missing data, to in turn maximise the likelihood of accurate analysis and unbiased results. The development of a suite of valid instruments with strong reliability and validity profiles are essential; DUQuA has contributed to this, but additional work remains. For clinical audit data, we need to compare the results derived from abstraction of data from the medical record, with algorithmic analysis, to those arrived at by an auditor with clinical expertise. For external data with long lead times, we need to bring forward the collection of quality and safety data measures, so that they are contemporaneous.

In the face of these challenges, the key solution to successful programme conclusion remains the quality of relationships with hospital staff, built on a foundation of engagement and nurtured through communication. We achieved much of this in our study and we thank all participating hospitals for their efforts.

Supplementary material

Supplementary material is available at *International Journal for Quality in Health Care* online.

Acknowledgements

Members of the original DUQuE team (Oliver Gröne, Rosa Suñol) provided extensive input and advice in the initial stages of the project, supporting design modifications from DUQuE and the planned DUQuA study approach. Dr Annette Pantle provided expert advice on revision and development of DUQuA measures, and Professor Sandy Middleton, Associate Professor Dominique Cadillac, Kelvin Hill, Dr Carmel Crock, and Professor Jacqueline Close provided input into the development of our stroke, AMI, and hip fracture clinical process indicator lists. Nicole Mealing and Victoria Pye provided statistical advice in the initial phases of the project. We greatly appreciate their efforts.

Funding

This research was funded by National Health and Medical Research Council (NHMRC) Program Grant APP1054146 (CI Braithwaite).

Authors' contributions

The research team consists of experienced researchers, clinicians, biostatisticians, and project managers with expertise in health services research, survey design, and validation, large-scale research and project management, sophisticated statistical analysis, QI and assessment, accreditation, clinical indicators, policy and patient experience. JB conceived the idea, led the research grant to fund the project, and chairs the steering committee. GA, TW and HPT co-led the development of the manuscript. RCW, NT, JB and YT contributed to the development of the manuscript.

Ethics approval

Ethical approvals were secured from State and Territory human research ethics committees in New South Wales (#14/206), Victoria (#15/36), the Australian Capital Territory (#15/131), South Australia (#15/260), the National Territory (#15-2509), Tasmania (#H0015383) and Queensland (#15/361). Site-specific authorisations, including permission for external researchers to collect data in hospitals, were granted. We complied with confidentiality requirements of national legislation or standards of practise of each jurisdiction.

Data sharing statement

Data will be made publicly available to the extent that individual participants or participating hospitals cannot be identified, in accordance with requirements of the approving Human Research Ethics Committees.

References

1. U.S. Institute of Medicine. *To Err is Human: Building a Safer Health System*. Washington, DC, National Academies Press. 1999.
2. Wilson RM, Runciman WB, Gibberd RW *et al.* The quality in Australian health care study. *Med J Aust* 1995;163:458–71.
3. Committee on Quality of Health Care in America and Institute of Medicine. *Crossing the Quality Chasm*. Washington, DC, 2001.
4. Braithwaite J, Greenfield D, Westbrook J *et al.* Health service accreditation as a predictor of clinical and organisational performance: a blinded, random, stratified study. *BMJ Qual Saf* 2010;19:14–21.
5. Braithwaite J, Westbrook J, Pawsey M *et al.* A prospective, multi-method, multi-disciplinary, multi-level, collaborative, social-organisational design for researching health sector accreditation. *BMC Health Serv Res* 2006;6:113.
6. Australian Commission on Safety and Quality in Health Care. *National Safety and Quality Health Service Standards Guide for Hospitals*. Sydney, AU, Australian Commission on Safety and Quality in Health Care. 2017.
7. Vincent C, Amalberti R. *Safer Healthcare: Strategies for the Real World*, 1st edn. Cham, CH, Springer International Publishing, 2016.
8. Braithwaite J, Hibbert P, Jaffe A *et al.* Quality of health care for children in Australia, 2012–2013. *JAMA* 2018;319:1113–24.
9. Runciman WB, Hunt TD, Hannaford NA *et al.* CareTrack: assessing the appropriateness of health care delivery in Australia. *Med J Aust* 2012;197:100–5.
10. Taylor N, Clay-Williams R, Hogden E *et al.* High performing hospitals: a qualitative systematic review of associated factors and practical strategies for improvement. *BMC Health Serv Res* 2015;15:244.
11. Leviton LC, Goldenberg RL, Baker CS *et al.* Methods to encourage the use of antenatal corticosteroid therapy for fetal maturation: a randomized controlled trial. *JAMA* 1999;281:46–52.
12. Pronovost P, Needham D, Berenholtz S *et al.* An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006;355:2725–32.
13. Bion J, Richardson A, Hibbert P *et al.* 'Matching Michigan': a 2-year stepped interventional programme to minimise central venous catheter-blood stream infections in intensive care units in England. *BMJ Qual Saf* 2013;22:110–23.
14. Suñol R, Garel P, Jacquerye A. Cross-border care and healthcare quality improvement in Europe: the MARQuIS research project. *Qual Saf Health Care* 2009;18.
15. Lombarts MJMH, Rupp I, Vallejo P *et al.* Application of quality improvement strategies in 389 European hospitals: results of the MARQuIS project. *Qual Saf Health Care* 2009;18:i28–37.
16. Secanell M, Groene O, Arah OA *et al.* Deepening our understanding of quality improvement in Europe (DUQuE): overview of a study of hospital quality management in seven countries. *Int J Qual Health Care* 2014;26:5–15.
17. Braithwaite J, Clay-Williams R, Taylor N *et al.* Deepening our Understanding of Quality in Australia (DUQuA): An overview of a nation-wide, multilevel analysis of relationships between quality management systems and patient factors in 32 hospitals. *Int J Qual Health Care* 2020;**32**:8–21.
18. Braithwaite J, Clay-Williams R, Taylor N *et al.* Bending the quality curve. *Int J Qual Health Care* 2020;**32**:1–7.
19. Taylor N, Clay-Williams R, Hogden E *et al.* Deepening our understanding of quality in Australia (DUQuA): a study protocol for a nationwide, multilevel analysis of relationships between hospital quality management systems and patient factors. *BMJ Open* 2015;5:e010349.
20. Australian Institute of Health and Welfare. *Australia's Hospitals at a Glance 2016–17*. Canberra, AU, Australian Institute of Health and Welfare, 2018.
21. Clay-Williams R, Taylor N, Braithwaite J. Potential solutions to improve the governance of multicentre health services research. *Med J Aust* 2018;208:152–4.
22. Braithwaite J, Clay-Williams R, Taylor N. Potential solutions to improve the governance of multicentre health services research. *Med J Aust* 2018;209:187–8.
23. National Health and Medical Research Council. *National Statement on Ethical Conduct in Human Research 2007 (Updated 2018)*. Australian Research Council, Universities Australia, Commonwealth of Australia, Canberra, AU, 2007.
24. Mitchell RJ, Cameron CM, McClure RJ, Williamson AM. Data linkage capabilities in Australia: Practical issues identified by a population health research network 'proof of concept project'. *Aust NZ J Public Health* 2015;39:319–25.
25. Clay-Williams R, Taylor N, Ting HP *et al.* Do quality management systems influence clinical safety culture and leadership? A study in 32 Australian hospitals. *Int J Qual Health Care* 2020;**32**:60–66.
26. Clay-Williams R, Taylor N, Ting HP *et al.* The relationships between quality management systems, safety culture and leadership and patient outcomes in Australian Emergency Departments. *Int J Qual Health Care* 2020;**32**:43–51.
27. Taylor N, Clay-Williams R, Ting HP *et al.* Do organization-level quality management systems influence department-level quality? A cross-sectional study across 32 large hospitals in Australia. *Int J Qual Health Care* 2020;**32**:35–42.
28. Clay-Williams R, Taylor N, Winata T *et al.* Organization quality systems and department-level strategies: refinement of the Deepening our Understanding in Quality in Australia (DUQuA) organization and department-level scales. *Int J Qual Health Care* 2020;**32**:22–34.
29. Clay-Williams R, Taylor N, Ting HP *et al.* The clinician safety culture and leadership questionnaire: refinement and validation in Australian public hospitals. *Int J Qual Health Care* 2020;**32**:52–59.
30. Taylor N, Clay-Williams R, Ting HP *et al.* Validation of the patient measure of safety (PMOS) questionnaire in Australian public hospitals. *Int J Qual Health Care* 2020;**32**:67–74.
31. Wagner C, Groene O, Thompson CA *et al.* Development and validation of an index to assess hospital quality management systems. *Int J Qual Health Care* 2014;26:16–26.
32. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY, John Wiley & Sons, 1987.
33. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011;30:377–99.
34. von Hippel PT. Should a normal imputation model be modified to impute skewed variables? *Sociol Methods Res* 2013;42:105–38.

International Journal for Quality in Health Care, 2020, 32(S1), 84–88
doi: 10.1093/intqhc/mzz109
Perspectives on Quality

Perspectives on Quality

Can benchmarking Australian hospitals for quality identify and improve high and low performers? Disseminating research findings for hospitals

PETER HIBBERT[1,2], FAISAL SAEED[3], NATALIE TAYLOR[4,5], ROBYN CLAY-WILLIAMS[1], TERESA WINATA[1], CHRISSY CLAY[1], WADAHA HUSSEIN[6], and JEFFREY BRAITHWAITE[1]

[1]Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Road, North Ryde, NSW 2109, Australia, [2]Australian Centre for Precision Health, Cancer Research Institute (UniSA CRI), School of Health Sciences, University of South Australia, Adelaide, SA 5000, Australia, [3]Safety and Quality Unit, Women's and Children's Hospital, North Adelaide, SA 5006, Australia, [4]Cancer Research Division, Cancer Council NSW, 153 Dowling St, Woolloomooloo, NSW 2011, Australia, [5]Faculty of Health Sciences, University of Sydney, Camperdown, Sydney, NSW 2006, Australia, and [6]Child, Youth and Family Services, Riverwood Community Centre, 151 Belmore Road North, Riverwood, NSW 2210, Australia

Address reprint requests to: Jeffrey Braithwaite, Centre for Healthcare Resilience and Implementation Science (CHRIS), Australian Institute of Health Innovation, Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia. Tel: +61 2 9850 2401; Fax: +61 2 9850 2499; E-mail: jeffrey.braithwaite@mq.edu.au

Editorial Decision 7 September 2019; Accepted 13 September 2019

Abstract

This paper examines the principles of benchmarking in healthcare and how benchmarking can contribute to practice improvement and improved health outcomes for patients. It uses the Deepening our Understanding of Quality in Australia (DUQuA) study published in this Supplement and DUQuA's predecessor in Europe, the Deepening our Understanding of Quality improvement in Europe (DUQuE) study, as models. Benchmarking is where the performances of institutions or individuals are compared using agreed indicators or standards. The rationale for benchmarking is that institutions will respond positively to being identified as a low outlier or desire to be or stay as a high performer, or both, and patients will be empowered to make choices to seek care at institutions that are high performers. Benchmarking often begins with a conceptual framework that is based on a logic model. Such a framework can drive the selection of indicators to measure performance, rather than their selection being based on what is easy to measure. A Donabedian range of indicators can be chosen, including structure, process and outcomes, created around multiple domains or specialties. Indicators based on continuous variables allow organizations to understand where their performance is within a population, and their interdependencies and associations can be understood. Benchmarking should optimally target providers, in order to drive them towards improvement. The DUQuA and DUQuE studies both incorporated some of these

principles into their design, thereby creating a model of how to incorporate robust benchmarking into large-scale health services research.

Key words: hospital quality management systems, patient-level factors, patient safety, hospital performance, quality improvement, benchmarking

Introduction

How do patients know whether they are likely to receive good care at their local hospital? How do clinicians and hospital managers know whether they are delivering good care to patients? How do funders know whether their money is buying good care for patients?

At the core of the answer to these types of questions are data. Frameworks and methods have been developed to measure performance at all levels of the health system from individual clinicians, departments, institutions and systems or countries [1]. Across entire health systems and between them, these can take the form of reports by government or non-government organizations or international organizations such as the Organisation for Economic Cooperation and Development [2–6]. Health service performance is typically measured to monitor aspects such as effectiveness or appropriateness (is care in line with best practice?), safety (will patients be harmed?), efficiency (will care be affordable?) and patient experience (will care be in line with patients' needs, values and beliefs?) [7].

One way of using information to improve quality of care is benchmarking, where the performances of institutions, services or individuals are compared using agreed indicators or standards [8]. The rationale for benchmarking is that institutions will respond positively to being identified as a low outlier, or desire to be a high performer, or both, and patients will be empowered to make choices and 'vote with their feet' to seek care at institutions that are high performers.

Whilst benchmarking has strong face validity, it has been attempted in many jurisdictions with mixed success. At the level of institutions, the jury is out, with a Cochrane review on the public release of performance data on the behaviour of healthcare consumers and providers being equivocal [9]. At the level of clinicians or departments, audit and feedback can lead to small but potentially important improvements in professional practice [10]. However, which type of feedback best optimizes these improvements remains unclear.

The Deepening our Understanding of Quality in Australia (DUQuA) research program is published in this Supplement [11]. In this perspective article, we review DUQuA and its predecessor in Europe, the Deepening our Understanding of Quality improvement in Europe (DUQuE) study [12], in terms of their contributions to benchmarking quality management systems and healthcare outcomes.

What the DUQuA and DUQuE research did and showed

As discussed elsewhere, the DUQuA program is an Australia-wide, National Health and Medical Research Council (NHMRC)-funded research project to identify how hospital quality management systems, leadership and safety culture in Australian hospitals are related to healthcare delivery quality and patient factors [11]. It extends the work undertaken in the DUQuE study, which examined the relationships between quality management systems, clinical processes, and patient factors in 188 hospitals across seven European countries. DUQuA aimed to answer two primary research questions:

- What department-level factors are associated with processes and outcomes for stroke, acute myocardial infarction (AMI) and hip fracture patients?
- What hospital-level factors (including Emergency Department (ED) factors) are associated with processes and outcomes for stroke, AMI and hip fracture patients? How much does each factor contribute to the total variation in outcomes?

This multi-level study involved data collection at organization, ED and care pathway department for AMI, stroke and hip fracture and patient levels for each hospital. Data collection consisted of quality management systems assessments, clinician and patient questionnaires and clinical audits undertaken by experienced External Quality Assessors during a two-day, on-site visit. In total, 32 large acute care hospitals participated in the DUQuA study. At the conclusion of data collection, DUQuA researchers had recruited 2387 healthcare professional and patient participants within the 32 hospitals, resulting in an accumulation of 4934 completed surveys. Analysis consisted of descriptive statistics and linear mixed models. Key features of quality that affect the delivery of care and patient outcomes were benchmarked for each hospital against the DUQuA national sample.

In regard to organization-level factors, Australia's performance was rated higher than the European hospitals that participated in DUQuE across all measures, but had few very high or very low performers. Associations were found between Quality Management infrastructure and Quality Improvement processes, and between Quality Management infrastructure and implementation of clinical quality, but associations were not as strong as those found in the European study. At care pathway level, DUQuA found a cluster of relationships between Quality Management infrastructure and department-level measures of quality (e.g. how clinical responsibilities are assigned for a particular condition; if department processes are organized to facilitate evidence-based care recommendations; compliance with selected recommendations of international agencies and systematic approaches to clinical reviews). However, these were not consistent across AMI, hip fracture and stroke, and in comparison to the DUQuE study, which presented 18 out of 36 significant positive associations between hospital and department-level quality management systems across the three departments, DUQuA found six positive associations.

DUQuA's benchmarking data

In addition to contributing these findings, DUQuA had an agenda of sending each of the participating hospitals a benchmarking report that allowed each hospital to compare its own results with those of the other hospitals in the sample. Information was presented at organizational level with quality management structures (eight indicators), quality improvement processes (two indicators) and clinical

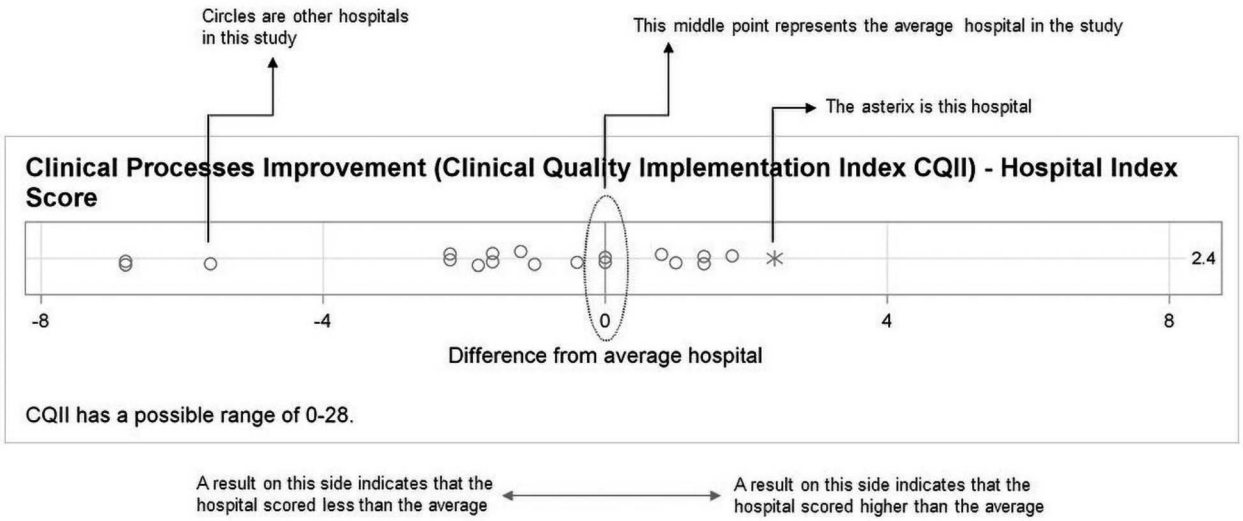

Figure 1 A de-identified example of a hospital-level benchmark used by the DUQuA study.

improvement processes (seven indicators). Care pathway level results were available for AMI, hip fracture and stroke, with each having composite indicators on evidence-based organization of pathways, patient safety strategies, specialized expertise and responsibility and clinical review and then three indicators on culture and leadership, nine on patient measured perceptions of safety and 11–13 on clinical processes. ED results show composite indicators on evidence-based organization of pathways, patient safety strategies, specialized expertise and responsibility and clinical review and three indicators on culture and leadership. Evidence-based recommendations for improvement strategies associated with each data presentation category were also presented, where appropriate.

Figure 1 illustrates the style of presentation of the benchmarking results. Within each graph, the asterisk represented the hospital score and the circles represented the other hospitals in the study. Most graphs compared the difference between the individual hospital and the average hospital in the study: if the difference is zero, then the hospital has the same results as the average study hospital, if the difference is positive, the hospital scored higher than the average hospital and if the difference is negative, the hospital scored lower than the average hospital.

Discussion

Benchmarking is where the performances of institutions or individuals are compared using agreed indicators or standards. It is guided by the principle that performance measurement is centrally concerned with improving practice to achieve better outcomes and not about merely improving numbers [10]. Benchmarking of performance is often based on a logic model, with indicator frameworks that are based on structure, process and outcomes. They are typically created around multiple specialties and patient types, calculated as a continuous variable (as contrasted with a 'yes' or 'no'), and primarily targeted at providers in order to stimulate improvement. Crucially, the purpose of benchmarking should be explicit to the providers of care.

The importance of a logical, acceptable and viable conceptual framework to underpin development of performance indicator sets for performance evaluation is emphasized in the literature [1, 13, 14]. A conceptual framework sets out the rationale and design principles for the indicator set and links it to the wider health system context. This part of the DUQuA program seeks to answer the question 'performance of what—and to what ends?' [15]. Both the DUQuA and DUQuE studies used a validated framework allowing organizations to understand the extent and fidelity of quality improvement activities.

Frameworks for benchmarking should be based on structured models which dictate the data collection. Too often, measurement systems are set up based on what data are collectable or already extant rather than what should be collected. Information that is contained in current information systems such as routinely collected hospital administrative diagnosis-related group-based indicators is prioritized in most health systems, whilst indicators that may be more important but are more difficult to secure are generally not. In the DUQuA study, care pathway data were collected via manual clinical audits. These types of clinical audits are more costly and time-consuming to gather than administrative data but have high utility for clinicians. The introduction of more comprehensive electronic records in the future may make the task of extracting some of the indicator data less problematic and expensive.

A conceptual framework for benchmarking encompassing multiple specialties and diseases is considered to be a key element of health reform over time [16, 17]. Ideally, indicators should be balanced—a combination of structure, process and outcome measures—and developed nationally and locally. They should also be scientifically valid and reliable, consistent over time and geographically be responsive to change over the time period of measurement, be attributable to the organization, or service, and be under an organization's influence to improve [14]. DUQuA and DUQuE provide common frameworks for measuring the quality management systems and are broadly based on a logic model, which helps consistency. For clinical data, DUQuE and DUQuA used a combination of closely linked structure, process and outcome indicators.

DUQuA's and DUQuE's [18] benchmarking data provide each hospital with performance information based on validated scales, with data showing variation. Thus, hospitals are provided with the spectrum of where their performance lies. This is in contrast

to accreditation feedback, which generally places an emphasis on whether an organization has achieved performance on a particular set of indicators or not, rather than on analysing the relationship and association between indicators. Therefore, accreditation is often assessed as either a pass or a fail [19]. Meeting requirements on indicators in isolation can lead to erroneous conclusions and decision-making when the relationships between indicators are not considered, for example, aiming to achieve reduced length of stay without consideration of the impact on an increase in the readmission rate. The DUQuA and DUQuE [18] studies assembled indicator sets that encompass quality systems, process measures, outcomes and patient experience that are comprehensive, results from which can be assessed holistically to better inform decisions and their consequences, particularly where interdependencies manifest or knock-on effects may occur.

The benchmarking data as shown in Figure 1 illustrate how hospitals can compare their performance to the average of their peer group. Alternatively, hospitals can use these data to compare themselves to the best performing hospital in their peer group. Benchmarking provides the best impact when it is designed to facilitate such comparisons. This can then lead to understanding how high performers achieve their results, analysing their processes and structures, and facilitating a learning process, such as by organizing benchmarking workshops or collaborative networks. The Health Roundtable in Australia and New Zealand and the American College of Surgeons National Surgical Quality Improvement Program in the U.S. are two examples of collaborative networks, targeted to hospitals and clinicians, respectively, that use the experiences and practices of high performers as exemplars for innovation and the uptake and spread of successful practices.

Caution is required when benchmarking low or high hospital performers because within-organization variation may be higher than between-organization variation. This is analogous to the finding of wider variations within countries than between them [20]. Hospitals can be highly heterogeneous in their performance which may be dependent on personnel, systems and the culture at ward, specialty and department level [21]. Measuring performance at organization level alone is likely to mask these differences. Benchmarking data is largely used to make management decisions for improving outcomes, yet much improvement happens in the clinical microsystem at the interface between patients, clinicians and processes [22]. Benchmarking data should be specific to the microsystem for it to be effective. Given this and the likelihood of considerable variation within hospitals, a strength of the DUQuA and DUQuE studies was their inclusion of measures at not just hospital level, but also at the level of departments.

Whilst transparency is a desirable outcome of any benchmarking exercise, the question is whether organizational performance data should be made public. Public reporting of hospital performance in terms of patient outcomes has lagged in Australia [23] with available research showing lack of accuracy in data to detect outliers [24]. Public reporting of data can have unintended consequences, such as data 'gaming,' [25, 26] risk aversion behaviour by providers and socioeconomic variations in consumer access to well-performing providers [27]. Damage to an organization can result from public reporting of performance indicators without due attention to the required caveats; organizations can use these same caveats as excuses to avoid internal action (e.g. 'coding errors' were used to vindicate ignoring poor results at Stafford Hospital in England) [28]. As a research study, DUQuA was unlikely to have been able to recruit hospitals without anonymity protections for the benchmarking—both from the individual hospital viewpoint but also ethical considerations, which required anonymity to protect hospital interests.

An interesting question to raise is around purpose: for whose benefit is public reporting? Agencies that publish benchmarking data in Australia are, for example, the Australian Commission on Safety and Quality in Health Care via their Atlas of Variation and state-based systems such as those provided by the Bureau of Health Information in New South Wales. These are designed with health providers and policy makers in mind, rather than consumers, but are public. Similarly, the Australian Health Roundtable benchmark data are designed for providers to discuss results in a closed meeting.

Some systems are directed to patients or consumers where indicators are presented with info-graphics in contrast to the more formal and structured presentation of hard data on provider-focussed websites. Examples include the Australian Institute of Health and Welfare's My Hospital website [29], the Netherlands 'Choose Better' program [30] and in England, NHS Choices [31]. However, research tells us that publishing indicators is more effective in changing behaviour of healthcare workers and management by acting on professional or corporate pride or reputation rather than improving information to healthcare consumers and creating choice [9, 21, 32, 33]. This suggests that providing benchmarking information to healthcare providers using a trusted forum, using an approach such as that taken by DUQuA, is more likely to provide a behavioural response from organizations in contrast to other approaches.

A key theme in the literature is being explicit about the purpose of performance indicators. Are they being used for accountability or for quality improvement? [13, 33, 34]. Examples of public reporting and accountability are the NHS 'star-ratings' system of 2001–2005 and those that use results for non-publicized feedback to organizations to stimulate improvement, such as Germany's voluntary reporting scheme [21, 35]. Developing frameworks whose purpose is to balance accountability ('the dial') with quality improvement ('the tin opener') is difficult, and that clearly articulating the purpose of the framework is necessary. The DUQuA and DUQuE studies were designed as research studies but the benchmarking component of the work is intended to facilitate quality improvement and not to emphasize accountability.

Conclusion

The DUQuA and DUQuE studies provide a timely reminder of the principles of benchmarking and how these can contribute to practice improvement and improved health outcomes. Benchmarking optimally starts with a conceptual framework and is based on a logic model. This framework should drive the selection of indicators to measure performance, rather than their selection being based on what is easy to measure. A Donabedian range of indicators should be chosen, including structure, process and outcomes, created around multiple domains or specialties. Where possible, these indicators should be calculated and displayed as continuous variables to allow organizations to understand where their performance is within the population and their interdependencies and measured associations. Lastly, it should be acknowledged that often, benchmarking is targeted at providers and not consumers in order to stimulate improvement.

Acknowledgements

Members of the original DUQuE team (Oliver Gröne, Rosa Suñol) provided extensive input and advice in the initial stages of the project, supporting design modifications from DUQuE and the planned DUQuA study approach.

Dr Annette Pantle provided expert advice on revision and development of DUQuA measures, and Professor Sandy Middleton, Associate Professor Dominique Cadillac, Kelvin Hill, Dr Carmel Crock, and Professor Jacqueline Close provided input into the development of our stroke, AMI, and hip fracture clinical process indicator lists. Nicole Mealing and Victoria Pye provided statistical advice in the initial phases of the project. We greatly appreciate their efforts.

Funding

This research was funded by NHMRC Program Grant APP1054146 (CI Braithwaite).

Author's contributions

PH wrote the first version of the manuscript. In concert with PH, FS undertook the research on benchmarking and wrote substantial parts of the manuscript. All other authors contributed to writing the manuscript. NT and RCW co-led the detailed study design, managed the project across time and contributed to the development of the manuscript. TW, CC and WH contributed to the logistics of project management, the refinement of measures and the development of the manuscript. JB conceived the idea to embrace DUQuE for Australia, led the research grant to fund the DUQuA project and chairs the steering committee.

References

1. Papanicolas I, Smith PC. Health system performance comparison. An agenda for policy, information and research. In: *European Observatory on Health Systems and Policies Series*, Open University Press, Berkshire, England, 2013.
2. Australian Institute of Health and Welfare. *Australia's Health 2016: In Brief*. Canberra, AU, Australian Institute of Health and Welfare, 2016.
3. Organsation for Economic Co-operation and Development. *Health Expenditure Indicators Paris, FR: Organsation for Economic Co-operation and Development*. 2019. https://www.oecd.org/els/health-systems/health-expenditure.htm.
4. National Health Service England. *NHS Improvement: Performance of the NHS Provider Sector for the Quarter Ended 30 June 2018*. London, UK, NHS Improvement, 2018.
5. Radley DC, McCarthy D, Hayes SL. *Scorecard on State Health System Performance*. Commonwealth Fund, New York, NY, 2018.
6. Steering Committee for the Review of Government Service Provision. *Report on Government Services 2019*. Canberra, AU, Productivity Commission, 2019.
7. Department of Health UK. *High Quality Care for All: NHS Next Stage Review Final Report*. London, UK, Department of Health, 2008.
8. Smith P, Mossialos E, Papanicolas I. *Performance Measurement for Health System Improvement: Experiences, Challenges, Prospects*. Geneva, CH, World Health Organization, 2008.
9. Metcalfe D, Rios Diaz AJ, Olufajo OA *et al.* Impact of public release of performance data on the behaviour of healthcare consumers and providers. *Cochrane Database Syst Rev* 2018;9:CD004538.
10. Ivers N, Jamtvedt G, Flottorp S *et al.* Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;6:CD000259.
11. Braithwaite J, Clay-Williams R, Taylor N *et al.* Deepening our Understanding of Quality in Australia (DUQuA): An overview of a nation-wide, multi-level analysis of relationships between quality management systems and patient factors in 32 hospitals. *Int J Qual Health Care* 2020;**32**:8–21.
12. Secanell M, Groene O, Arah OA *et al.* Deepening our understanding of quality improvement in Europe (DUQuE): overview of a study of hospital quality management in seven countries. *Int J Qual Health Care* 2014;26:5–15.
13. Smith P, Mossialos E, Papanicolas I. *Performance Measurement for Health System Improvement*. Cambridge, UK, 2010.
14. Braithwaite J, Hibbert P, Blakely B *et al.* Health system frameworks and performance indicators in eight countries: a comparative international analysis. *SAGE Open Med* 2017;5:2050312116686516.
15. Arah OA, Westert GP, Hurst J, Klazinga NS. A conceptual framework for the OECD health care quality indicators project. *Int J Qual Health Care* 2006;18:5–13.
16. Adair CE, Simpson E, Casebeer AL *et al.* Performance measurement in healthcare: part II-state of the science findings by stage of the performance measurement process. *Health Policy* 2006;2:56–78.
17. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966;44:166–206.
18. Groene O, Kristensen S, Arah O *et al.* Feasibility of using administrative data to compare hospital performance in the EU. *Int J Qual Health Care* 2014;26:108–115.
19. Winata T, Clay-Williams R, Taylor N *et al.* Using accreditation surveyors to conduct health services research: a qualitative, comparative study in Australia. *Int J Qual Health Care* 2020;**32**:89–98.
20. Groene O, Sunol R, Consortium DUP. The investigators reflect: what we have learned from the deepening our understanding of quality improvement in Europe (DUQuE) study. *Int J Qual Health Care* 2014;26:2–4.
21. Hibbert P, Hannaford N, Long J *et al. Performance indicators used internationally to report publicly on healthcare organisations and local health systems*. Australian Institute of Health Innovation, Sydney, AU, 2013.
22. Nelson EC, Batalden P, Godfrey M. *Quality by design, a clinical microsystems approach*. San Francisco, CA, Jossey-Bass, 2007.
23. Gallagher MP, Krumholz HM. Public reporting of hospital outcomes: A challenging road ahead. *Med J Aust* 2011;194:658–60.
24. Scott IA, Ward M. Public reporting of hospital outcomes based on administrative data: Risks and opportunities. *Med J Aust* 2006;184:571–5.
25. Vallance AE, Fearnhead NS, Kuryba A *et al.* Effect of public reporting of surgeons' outcomes on patient selection, "gaming," and mortality in colorectal cancer surgery in England: Population based cohort study. *BMJ* 2018;k1581:361.
26. Mannion R, Braithwaite J. Unintended consequences of performance measurement in healthcare: 20 salutary lessons from the English National Health Service. *J Intern Med* 2012;42:569–74.
27. Harrion R, Manias E, Mears S *et al. Addressing Unwarranted Variation in Healthcare an Evidence Check Rapid Review*. Sax Institute, Sydney, Australia, 2019.
28. Francis R. *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry*. London, UK, Stationery Office, 2013.
29. Australian Institute of Health and Welfare. *My Hospitals*. Canberra, AU: Australian Institute of Health and Welfare; 2018. https://www.myhospitals.gov.au/.
30. Netherlands Healthcare Institute. *KiesBeter*. Diemen, NL: Netherlands Healthcare Institute; 2018. https://www.kiesbeter.nl/.
31. National Health Service. *NHS Services*. London, UK: National Health Service; 2019. http://www.nhs.uk/Pages/HomePage.aspx.
32. Shekelle P, Lim Y-W, Mattke SCD. *Does Public Release of Performance Results Improve Quality of Care?* London, UK, National Health Foundation, 2008.
33. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. *Med Care* 2003;41:I–30.
34. Raleigh VS, Foot C. *Getting the Measure of Quality: Opportunities and Challenges*. The King's Fund, London, UK, 2010.
35. Arah O, Klazinga N, Delnoij D *et al.* Conceptual frameworks for health systems performance: a quest for effectiveness, quality, and improvement. *Int J Qual Health Care* 2003;15:377–98.

International Journal for Quality in Health Care, 2020, 32(S1), 89–98
doi: 10.1093/intqhc/mzz110
Article

Article

Using accreditation surveyors to conduct health services research: a qualitative, comparative study in Australia

TERESA WINATA[1], ROBYN CLAY-WILLIAMS[1], NATALIE TAYLOR[2,3], EMILY HOGDEN[2], PETER HIBBERT[1], ELIZABETH AUSTIN[1], and JEFFREY BRAITHWAITE[1]

[1]Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Road, NSW 2109, Australia, [2]Cancer Research Division, Cancer Council NSW, 153 Dowling St, Woolloomooloo, NSW 2011, Australia, and [3]Faculty of Health Sciences, University of Sydney, Camperdown, Sydney, NSW, 2006, Australia

Address reprint request to: Teresa Winata, Centre for Healthcare Resilience and Implementation Science (CHRIS), Australian Institute of Health Innovation, Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia. Tel: +61 (2) 9850 2472; E-mail: teresa.winata@mq.edu.au

Editorial Decision 7 September 2019; Accepted 13 September 2019

Abstract

Objective: Healthcare accreditation surveyors are well positioned to gain access to hospitals and apply their existing data collection skills to research. Consequently, we contracted and trained a surveyor cohort to collect research data for the Deepening our Understanding of Quality in Australia (DUQuA) project. The aim of this study is to explore and compare surveyors' perceptions and experiences in collecting quality and safety data for accreditation and for health services research.
Design: A qualitative, comparative study.
Setting and Participants: Ten surveyors participated in semi-structured interviews, which were audio recorded, transcribed and coded using Nvivo11. Interview transcripts of participants were analysed thematically and separately, providing an opportunity for comparison and for identifying common themes and subthemes.
Intervention(s): None.
Main Outcome Measure(s): Topics addressed data collection for healthcare accreditation and research, including preparation and training, structure, organization, attitudes and behaviours of staff and perceptions of their role.
Results: Five themes and ten subthemes emerged from the interviews: (1) overlapping facilitators for accreditation and research data collection, (2) accreditation-specific facilitators, (3) overlapping barriers for accreditation and research data collection, (4) research data collection-specific barriers and (5) needs and recommendations. Subthemes were (1.1) preparation and training availability, (1.2) prior knowledge and experiences; (2.1) ease of access, (2.2) high staff engagement, (3.1) time, (4.1) poor access and structure, (4.2) lack of staff engagement, (4.3) organizational changes; (5.1) short-notice accreditation and (5.2) preparation for future research.
Conclusions: Although hospital accreditation and research activities require different approaches to data collection, we found that suitably trained accreditation surveyors were able to perform both activities effectively. The barriers surveyors encountered when collecting data for research

provide insight into the challenges that may be faced when visiting hospitals for short-notice accreditation.

Key words: accreditation, health services research, qualitative methodology, quality and safety

Introduction

Collecting quality and safety data in hospitals for quality improvement and monitoring, patient safety and for research are essential. However, there are challenges associated with gathering robust and useful supporting data for groups external to the hospital. Challenges include frequent organizational staffing and structural changes, convoluted research ethics and governance processes and institutional barriers to developing and maintaining relationships with healthcare professionals and to timely access to hospital documents [1–3].

Accreditation of healthcare services through external review mechanisms is widely used to assess their compliance with national and international standards and improve the safety and quality of healthcare [4] and is mandatory for public hospitals in Australia [5]. Healthcare accreditation surveyors are central to the process; they are trained to undertake onsite safety and quality assessments in healthcare organizations [6]. The Australian Council on Healthcare Standards' (ACHS') surveyor workforce consists of clinicians, health professionals and consumers with healthcare industry backgrounds. The role of surveyors as both independent assessors of service standards and educators to facilitate continuing quality improvement is becoming increasingly recognized and supported [4, 6].

In contrast, health services researchers collecting safety and quality data across many hospitals face significant difficulties: they must collect data without judgement and need to obtain complex and varied hospital approvals to gain entry to complete the task. In the Deepening our Understanding of Quality in Australia (DUQuA) program, a large, multi-level and cross-sectional study in 32 large Australian public hospitals, ACHS surveyors were engaged to collect research data onsite at study hospitals [7]. Surveyors contracted for DUQuA are uniquely able to compare data collection experiences of research with that of accreditation.

Although several studies have examined the perceptions and experiences of surveyors on the impact of accreditation on quality and patient safety [4, 6, 8–12], no studies have been conducted to qualitatively explore the perspectives of surveyors as data collectors involved in health services research. Qualitative methodology, involving conducting semi-structured interviews and thematic analysis, is utilized in this study as an optimal approach for collecting and analysing data on participants' perspectives, experiences and personal histories. The strength of attaining qualitative data is its ability to provide complex textual descriptions and explanations of how individuals experience and perceive a given research topic; it delivers information about the 'human' side of the issue, such as the often contradictory beliefs, behaviours, emotions, opinions and the relationships of individuals [13]. Exploring and comparing surveyors' perceptions and experiences of accreditation and health services research may enable governments, healthcare providers, clinicians, researchers and healthcare and research managers to understand better how to plan and conduct quality improvement projects and research.

Research aims

Using qualitative methodology, this study aimed to:

1. examine accredited surveyors' experiences and perceptions of data collection processes for both accreditation and health services research and
2. establish key barriers and facilitators through anecdotal experiences associated with engaging surveyors to collect quality and safety data for accreditation and health services research.

Methods

Context

ACHS surveyors contracted for the DUQuA project were trained in research data collection methods prior to hospital data collection visits. Training consisted of a 1-day orientation course and included familiarization with data tools and measures as well as research logistics planning and hospital liaison activities. Details about the training, data collection procedures, timetable and activities and communication tools for the DUQuA hospital visits, are summarized in Supplementary Appendix A. However, delays in research ethics and governance approvals for the project [1, 7] lead to an under-anticipated lag between training and data collection.

During and after data collection, the surveyors provided feedback to the research team though formal feedback forms (see Supplementary Appendix A, eFigure A2) and informal discussions. The reported observations and lessons learned by the surveyors led to the development of this study.

Recruitment and data collection

Sixteen ACHS surveyors contracted for the DUQuA project, who each completed data collection at a minimum of two DUQuA hospitals, were invited to participate in this study. The research investigators (TW, RCW) sent a formal email invitation and a participation information sheet and consent form to those who expressed an interest in participating. A mutually convenient time was arranged with participants, who provided consent to this standalone study, for an individual face-to-face or telephone interview. During the interview, surveyor perceptions were sought on the process of healthcare accreditation and data collection for DUQuA, including preparation, structure, organization, attitudes and behaviours of staff members and perceptions of their responsibilities, as well as their training for both roles. Interview questions, provided in Supplementary Appendix B, were developed from verbal and informal discussion between researchers and surveyors following DUQuA data collection visits. All participants completed a short demographic survey prior to interview sessions.

Data analysis

We conducted semi-structured telephone or face-to-face interviews with participants between 8 March and 30 June 2018. All

interviews were audio recorded, professionally transcribed and coded using Nvivo11. Interview transcripts of participants were analysed thematically and separately, providing an opportunity for comparison, and for identifying common themes and subthemes of individual experiences and perceptions of accreditation and research data collection in Australian health services [14]. Thematic analysis was undertaken to develop key themes relating to their experiences and perceptions; this was done by coding interview data. Coding allowed data to be organized and used to explore connections between data elements and to develop sets of concepts. Once coded, segments of data were then linked in a formal fashion to allow themes to emerge and to determine relationships between different data sets. This is a way of studying real world complex systems such as healthcare [15]. To improve safety provision and quality processes, barriers and enablers associated with conducting accreditation and health services research (such as DUQuA) were also identified. Analyses were completed by two members of the DUQuA researcher team and one external reviewer. Demographic characteristics of the surveyors were assembled from survey data.

This project was approved by the Faculty of Medicine and Health Sciences Low-risk Human Research Ethics Committee (HREC) at Macquarie University on 15 February 2018 (project reference no. 5201701129).

Results

Ten surveyors consented and participated in this study. They were experienced surveyors, having collectively assessed 886 accreditation programs across hospitals in all Australian states and territories. In conducting accreditation visits, five of the ten participants had between 5–10 years' experience, while the remaining five had more than 10 years' experience. Four participants conducted between two and four visits for DUQuA, and the remaining six undertook between five and ten visits (Table 1).

Interview sessions lasted between 25 and 69 min. Transcribed interviews were coded by one researcher (TW). Two interviews were randomly selected and coded by a second researcher (RCW) and themes compared. As the two primary researchers were also members of the DUQuA team, three transcripts were also reviewed and coded by an external reviewer (EA). Minor disagreements were resolved through multiple discussions and consensus reached on themes. Five major themes emerged from the analysis: (1) overlapping facilitators for accreditation and research data collection, (2) accreditation-specific facilitators, (3) overlapping barriers for accreditation and research data collection, (4) research data collection-specific barriers and (5) needs and recommendations. These themes were supported by the following subthemes: (1.1) preparation and training availability, (1.2) prior knowledge and experiences; (2.1) ease of access, (2.2) high staff engagement, (3.1) time, (4.1) poor access and structure, (4.2) lack of staff engagement, (4.3) organizational changes, (5.1) short-notice accreditation and (5.2) preparation for future research. The fifth theme ('needs and recommendations') and its subthemes ('short-notice accreditation' and 'preparation for future research') emerged from the interviews as additions. It was notable that a theme for research-specific facilitators did not emerge. Themes and subthemes are described and reported in greater detail below. Table 2 shows additional, supporting quotes from participants for each theme and subtheme. Figure 1 illustrates the facilitators and barriers for conducting hospital accreditation and health services research.

Table 1 Characteristics of participating surveyors

Characteristic	Number
Age	
35–64 years old	4
>65 years old	6
Gender	
Male	1
Female	9
Highest degree qualification	
Certificate IV	1
Bachelor	1
Postgraduate	8
Experience (time) as surveyors	
5–10 years	5
>10 years	5
Experience (number of accreditation visits) as surveyors	
0–50 visits	3
51–100 visits	2
>100 visits	5
Experience (number of data collection visits) as data collectors	
2–4 visits	4
5–10 visits	6

Theme 1: Overlapping facilitators for accreditation and research data collection

Preparation and training availability

All participants affirmed that being trained as a surveyor provided effective preparations and skills that allowed them to conduct their accreditation survey efficiently. Surveyor training also ensured that surveyors were well-versed in any major changes to hospital protocols or guidelines, which allowed them to apply this knowledge for assessments and reviews of a wide variety of organizational types (refer to Table 2, Code 1.1.A for additional quotes).

> 'You undertake an induction program and then you go on a training survey with a mentor. You have an evaluation performed and after that then you do—you continue on surveying and depending on your feedback you're reappointed every two years.' (Participant 4).

Regarding DUQuA, several participants noted that the orientation day assisted them throughout data collection (Table 2, Code 1.1.R). Safety awareness and hospital navigation were critical components addressed by this training.

> '... we had a fairly strong training program that was adjusted along the way. We were as prepared as we could be utilizing the knowledge and skillset we had of accessing hospitals ...' (Participant 1).

Prior knowledge and experiences

All participants were either executives or clinicians in their previous role. They felt that their prior knowledge and experiences gained throughout their career, including their extensive understanding on how hospital systems and healthcare organizations

Table 2 Additional participant quotes and responses to individual themes and subthemes

CODE	THEME subtheme *(description)*	Participant responses (participant identification number)	
		Hospital accreditation	Health services research (DUQuA)
1.0	Overlapping facilitators		
1.1	Preparation and training availability (*Preparation and training equipped to undertake the activity*)	1.1. A: 'You might have an extra training if there's a new standard or if there's a new accreditation system … I think we get plenty of training.' (Participant 3)	1.1.R '… if you did not have some sort of orientation to it, it would have been difficult to go in without any information at all …' (Participant 7)
1.2	Prior knowledge and experiences (*Knowledge and confidence to undertake the role*)	1.2.A: 'I'd been on the other end, preparing the organization for survey, so you knew what the process was. I'd worked in health for 40 years … I guess being on the other side of it wasn't too hard.' (Participant 5)	1.2.R: 'I was keen to do it, and I had … audited and monitored the data from many clinical records for the process of finding out how organization is traveling compliantly … and then you identify gaps and make improvements.' (Participant 7)
2.0	Accreditation-specific facilitators		
2.1	Ease of access (*Access to relevant evidence, data, staff members and systems within the organization or health provider*)	2.1.A: '… accessing things in the hospital is really very easy. There's an expectation the team will be able to go in and out of both clinical areas, medical records, pathology, whatever it is that needs to be seen.' (Participant 1)	
2.2	High staff engagement (*Flexible engagement between hospital staff members and surveyors*)	2.2. A: 'They allocate one person to you. She or he will accompany you and take you back to the room, because they always give us a room like this … Somebody usually takes you and brings you back. If they do not they might give you a map. … Most of the time you are supported well to get to where you need to go.' (Participant 5)	
3.0	Overlapping barrier		
3.1	Time (*Workload pressure in a set amount of time, delays and gap time between training and undertaking the activity*)	3.1.A: 'If you have got 336 of them to do and there's only one or two of you … you do not get always get the time to see to verify it which I think's risky per se.' (Participant 6)	3.1.R: '… the timelines in some of the places that I went to that there was adequate time at some organizations and your time for other organizations, and that was part of the—the difficulty in navigating some of the paper clinical records to collect the data. So we … very, very busy and had to really try and … We said, how are we going to do this? Because we are not going to manage this in the time allocated. We're going to be here until midnight everyday if we do not find a better way of doing it, and it was because you might be looking to see if they had been referred for physiotherapy prior to discharge, because that was one of the questions, in a hip fracture.' (Participant 7)
4.0	Research data collection-specific barriers		
4.1	Poor access and structure (*Poor organization and structure of data, evidence and tools provided to undertake the activity within the organization*)		4.1. R: '… one is about how hard it was to answer the questions at time and even though we worked really hard on having a tool that was useful, in the end the records were not organized in a way that allowed you to go through in a sequential order. You had to go backwards and forwards, backwards and forwards.' (Participant 1)

(Continued)

Table 2 Continued

CODE	THEME subtheme *(description)*	Participant responses (participant identification number)	
		Hospital accreditation	**Health services research (DUQuA)**
4.2	Lack of staff engagement (*Difficulty in engaging and communicating with hospital staff members*)		4.2. R1: '… the very first one was difficult because we had assumed that the organization would know what they were getting actually but they did not.' (Participant 6) 4.2.R2: '… a couple of sites I went to where I did not even meet the PI. They were not there. They may have delegated to … one of the admin staff. The admin staff knew nothing about running a hospital. They know nothing about how to engage the right people. They're just told to set up this timetable and have these people and then you'd turn up and no one was there, no one knew anything about it.' (Participant 1)
4.3	Organizational changes (*Structural, governance and staff workforce changes of the hospitals*)		4.3. R: 'It was simply because of the length of time and they'd had so much staff changeover there and they also were undergoing a major redevelopment by the time we got there.' (Participant 2)
5.0	Needs and recommendations		
5.1	Short-notice accreditation	5.1. A1: 'Again, I think it shows a lot of maturity in an organization to let you do that, but that's what it should be about. They're trying to get organizations to realize that you are not doing things for accreditation. You're doing things for quality, to make sure that you do the right thing for the patient. We just come in and check your processes. You're not doing the processes for ACHS. A short-notice survey should not matter. You should be able to come in at any time and see what's happening, but not all of the organizations have agreed to it.' (Participant 5) 5.1. A2: 'The current feedback from the teams who have already done it is that it has merit and that the facilities that have opted for that model have been very receptive.' (Participant 1)	
5.2	Preparation for future research		5.2. R: '… I think you should have worked out through ethics approval before you started recruiting people to do the experiment [you could] just have a letter that ACHS were happy for their surveyors to do it when the project was approved or something and you could write your ethics approval around that, that you would use qualified surveyors or something.' (Participant 3)

function, made it easier for them to solve problems during both accreditation surveys and DUQuA data collection (Table 2, Code 1.2.A).

All participants reported they were confident and knowledgeable in navigating through the hospital buildings and wards, and were able to identify who to speak with and were able to communicate

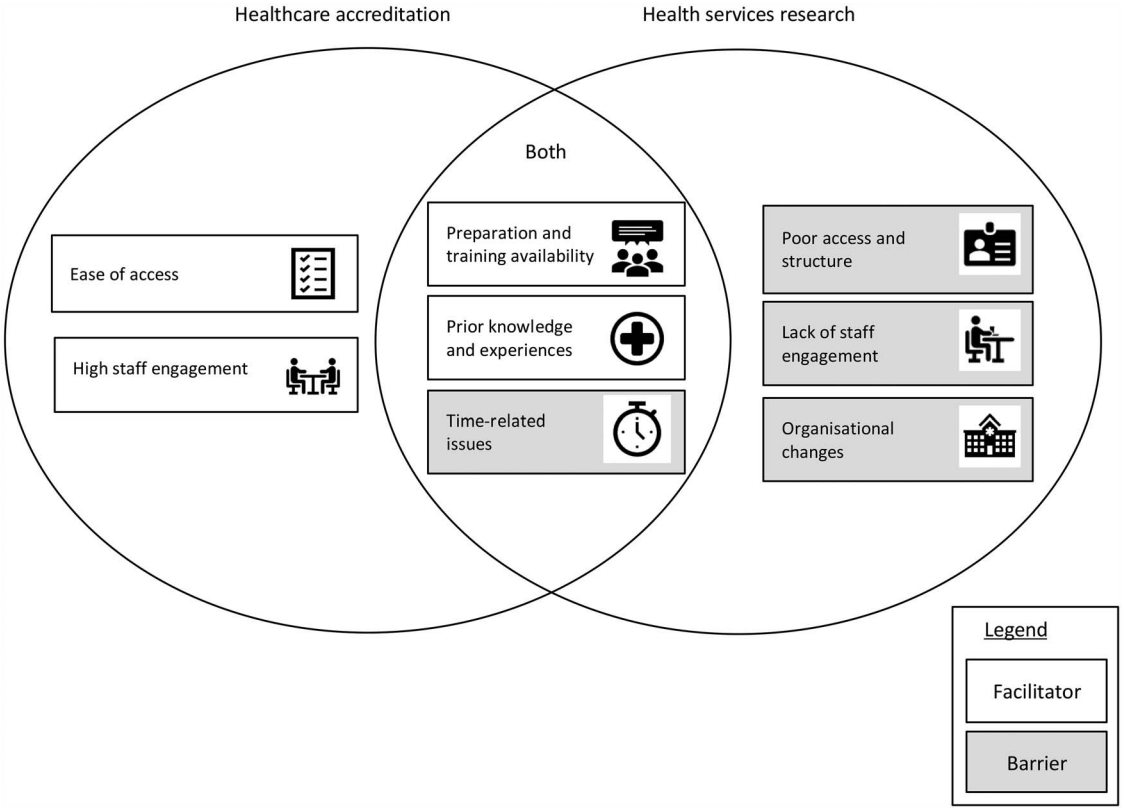

Figure 1 Facilitator and barrier themes for hospital accreditation and health services research.

effectively with staff members, because of their extensive experience working as healthcare surveyors:

> 'Personally, I didn't struggle to solve those problems. I know how to run a hospital. I know who the right people are to connect with … I think that I gained greater insight to the difficulties by being a member of that workforce …' (Participant 1).

> 'I'd worked in health for 40 years … I guess being on the other side of it wasn't too hard.' (Participant 5).

Theme 2: Accreditation-specific facilitators

Ease of access

Participants reported easy access to relevant evidence and data, including patient records, wards and data on safety equipment (Table 2, Code 2.1.A).

Some participants also indicated that access to hospital teams was always straightforward; one interviewee noted that it was common practice to be provided with a hospital coordinator to assist in navigating around the hospital.

> 'They usually allocate a person to each surveyor. It might be their admin assistant or a clinical nurse specialist who hasn't got a clinical role for the day, somebody that they'll allocate to you to take you around.' (Participant 5).

High staff engagement

Staff engagement was considered a major factor for success when undertaking both accreditation and health services research (Table 2, Code 2.2.A). While participants reported a high degree of staff engagement during their accreditation appointments; low staff engagement was reported as barrier to research-related data collection (refer to Lack of staff engagement subtheme below).

> 'The National Standards have engaged clinicians in establishing systems that facilitate accreditation against set standards quite well. Much better than the previous accreditation standards did, and so we do find ourselves meeting with clinicians, particularly in the larger hospitals...' (Participant 2).

Theme 3: Overlapping barriers for accreditation and research data collection

Time

Participants confirmed that the need for effective time management was a major barrier in both accreditation and data collection (Table 2, Code 3.1.A). They mentioned how critical time is for the hospital visits, especially during accreditation where numerous tasks must be coordinated and completed between only two surveyors.

> 'The greatest challenges on a survey are around timetabling and availability of people and/or the timing itself. So, again - an example for me was in a very large team, in a large district, I and another member of the team travelled almost a thousand kilometres, round trip, to see the site. At the last minute on the last day and then when writing up, it was discovered that I had—I didn't know, I had been allocated a particular set of—a particular section of the Standards. No meetings had been scheduled with those people. As

a result, I ran a risk of not having the right information to be able to say whether they met it or they didn't.' (Participant 1).

Similarly, time was repeatedly mentioned by participants as being the biggest obstacle to attaining satisfying and accurate data for DUQuA. Allowance of only 2–3 days to complete data collection was considered insufficient. See Table 2, Code 3.1.R for an additional quote.

'It was too much to do in the time allotted.' (Participant 4).

Additionally, for DUQuA, there was an interval of up to 14 months between the orientation session and data collection due to significant ethics and governance delays and service agreements. This resulted in the loss of prior knowledge and skills that were needed for the research tasks, and in reduced availability of data collectors.

'... we in fact lost surveyors because it had been too long between their training and actually doing it.' (Participant 1).

Theme 4: Research data collection-specific barriers

Poor access and structure

Hospitals plan for accreditation many months in advance. Therefore, having poor access to data, evidence and tools within the organization was rarely experienced in the context of accreditation. In contrast, despite ongoing communication between the research team, surveyors and hospitals for up to 14 months prior to visits, poor organizational planning and access to facilities and resources at hospitals was frequently seen during data collection for DUQuA. Organization tools used to facilitate coordination of hospital visits are shown in Supplementary Appendix A.

'Nine times out of 10 you didn't have a password to access the electronic medical records when you got there.' (Participant 10).

'Sometimes they'd sit you in a very noisy area. That was difficult because you really had to concentrate on the questions.' (Participant 5).

'... the difficulty in navigating some of the paper clinical records to collect the data. So we [were] really time poor, very, very busy and had to really try ...' (Participant 7).

Lack of staff engagement

In DUQuA, the majority of participants reported negative perceptions and experiences with hospital staff members, from lack of support in finding a parking spot, to navigating and searching for the relevant documents. This lack of engagement was interpreted by one surveyor as being due to staff perceptions of this being yet another monitoring exercise:

'... you're coming in, you're [asking] the opinion of people coming in saying, here they come, here they come, checking up on us again.' (Participant 8).

Participants expressed surprise at the difference experienced in hospital staff engagement between the two roles. They mentioned that staff within the organizations to be accredited are notified a couple of months or years before undergoing accreditation and therefore understood the significance and implications of accreditation for patient safety and positive outcomes. In DUQuA, surveyors felt they were mostly viewed as unimportant and inferior to the people within the participating organization. Additionally, participants indicated that, in some cases, key hospital staff were unaware of, or unmotivated to participate in, this research.

'In the vast majority of cases, the organization for the DUQuA study was done by an administrative assistant. Whereas in accreditation it's done by a quality manager. The differences were acute, because usually an administrative manager in the hospital had absolutely no idea of what we were doing. Didn't understand the importance of being able to get access to the medical records, didn't understand how long it would take to interview some of these doctors. They didn't get it. Whereas the quality manager is usually clinical and they understood the requirements, the organization up front and the communication is usually much better between the organization than it was for the DUQuA study. It was really pushing up hill in some organizations.' (Participant 2).

Organizational changes

Although few participants expressed concerns when changes within organizations occurred while surveying (due to the expectations of the surveyors and their willingness to adapt, accreditation normally ran smoothly even when there were changes), organizational changes were problematic when it came to research data collection. Where the research was not deemed important to the participating organization, the data collection task was especially challenging for participants. Issues involved include staffing changes (such as multiple changes in site principal investigators in hospitals) and redevelopment of hospital buildings and management structures (Table 2, Code 4.3.R).

'For various reasons, sometimes it was quite a long time and there'd been quite a lot of staff turnover. Where the organizations were, the same people who were - particularly the clinical lead. If the clinical lead was still there, then it wasn't so bad. But in most cases the clinical lead had moved on to something else. That was the most challenging part.' (Participant 2).

Theme 5: Needs and recommendations from participants

Short-notice accreditation

Participants reported observing a higher level of preparedness for accreditation than for DUQuA. Hospitals consciously plan for accreditation months prior to the visit and this has been identified as an impediment to accurate assessment of more normal day-to-day hospital activity. Australia is moving toward formal, short-notice ('ad hoc') assessments, where hospitals are required to fully comply with the requirements of the National Safety and Quality Health Service (NSQHS) Standards and have in place processes to demonstrate

compliance at any time. Most participants suggested the value of introducing and implementing short-notice accreditation programs as a way of improvement (Table 2, Codes 5.1.A and 5.2.A). They expressed how important it is that every hospital or healthcare organization has ongoing safety and quality governance systems in order to provide positive organizational and patient outcomes.

Some surveyors recommended short-notice programs as a way to encourage organizations to keep updated about any changes that may have happened in their health environment.

> 'I think they would be quite good. I'll tell you why, it's because if the health service is already functioning and prepared for accreditation, they should be able to keep those processes up to date and so they should then be able to set the processes for continually updating their information.' (Participant 9).

A few participants indicated that hospitals should always provide quality and safety services to their patients, so that they are ready and capable to resolve any issues that they may encounter.

Preparation for future research

Conducting nationwide, multi-level health services research, such as DUQuA, is complex. There are numerous factors that must be considered when contracting surveyors as data collectors in research, including the relationships between research team, data collectors/-surveyors and hospital staff team, the clinical and patient level involvement, the organizational governance systems, ethics processes and so forth. Many of the participants offered recommendations for consideration when planning a similar study in the future to involve a surveyor lead in the research-hospital relationship building process, to generate a project issue log and distribute it to the entire team and to finalize all research ethics approvals before bringing surveyors on board. See Table 2, Code 5.2.R.

> '... for this sort of research that involves clinical disciplines and a clinical setting, that you actually have a clinician such as myself on the team, who actually does that negotiation.' (Participant 1).

> '... because there was such a big gap between when we had that initial exposure to what we were required to do and when we went on board, and we all weren't on the same survey, it would have been good if we'd had an issues log along the way. So, just a basic Excel spreadsheet so that you could, if you had an issue in one area where - so, for example, we would probably record antibiotic administration sometimes in theatre, the anaesthetist will write it in the anaesthetic sheet as well as in the medication chart, but just to have some ongoing dialogue between the surveyors of any issue would have been good to have just read so that you had - you could keep track of what was going on.' (Participant 9).

Discussion

This is the first study in Australia to examine and compare qualitatively surveyors' experiences in undertaking accreditation and data collection for a health services research project. We found that, while accreditation and research data collection have facilitators and barriers in common, there were significantly greater barriers to data collection for those in a health services research role.

Participating surveyors affirmed that the facilitating factors, such as ease of access to departments and resources, positive engagement with staff members, training and orientation and prior knowledge and confidence in navigating the healthcare systems, occurred more frequently during the accreditation process than in research activity, and in some research activity cases acted as obstacles. Generally, hospitals can make access to relevant data and evidence available for accreditation surveyors; teams would normally be able to enter clinical areas and access medical records without having to ask hospital staff for specific permissions [16, 17]. Although accreditation surveys collect data using measures that are typically not validated against statistical indices, when cross-referenced against or coupled with results from clinical process indicators, compliance assessment data, and patient-reported outcomes collected by other parties using validated research tools (i.e. DUQuA research) [18–20], there is the potential to obtain a powerful quality and safety snapshot within and between hospitals. The need, therefore, for easier avenues—like those that exist for accreditation surveyors—for multidisciplinary research teams to access, collect and analyse such data is crucial for accurately diagnosing and guiding recommendations for improvement.

Although surveyors presented some positive experiences of data collection for research, these were outweighed by the research-specific challenges they faced. Lack of communication and engagement seem to be major drawbacks. Communication is closely linked with hierarchy, with research showing that the authority gradient (i.e. the psychological distance between individuals and professional groups) can lead to withheld information or information being adapted to suit the recipient [21, 22]. One of the research data collection-specific barriers is limited engagement with hospital staff. Surveyors indicated the key to accessing and collecting accurate, robust hospital data is the ability to meet and speak with key clinical and administrative managers. Not being able to do so caused frustration to surveyors. If, in their capacity as a member of a research team, surveyors can engage in valuable conversation with the organization, with staff at multiple levels, from research governance to care delivery, they can share their knowledge effectively and offer valid feedback to the research team as to what they have witnessed [10, 23].

Implications for practice

For future healthcare research studies involving external data collectors in collaboration with research institutes or universities, explicit, standardized training about what the research entails should be provided. This may include a pilot study whereby data collectors are exposed to the field to perform data collection activities and evaluate these activities [24–26].

Hospital staff are often motivated and well-prepared for accreditation, but participants in this study have indicated that in some cases hospital staff were not motivated for research, and the difference in motivation affected surveyors' research experience, as seen through this project. This concurs with previous research findings that health professionals' attitudes vary between project and activity purposes and measurements [4, 27]. Hospital staff are familiar with and paid to do their role, whereas research is usually a voluntary activity in addition to normal responsibilities. For health services research projects to succeed, it is imperative for any research involvement to be given priority. The organization should be attuned to the research

goals and milestones, and research teams should be explicit about the importance of undertaking this research for the benefit of individual hospitals and beyond.

Time to obtain research ethics and governance approvals were the main barriers in undertaking quality and safety data collection for DUQuA, which is highlighted in the theme 'Time' where delays were encountered by surveyors between research training and data collection period. The recommendation of our cohort is to finalize all ethics and governance applications before involving external researchers. However, the scale, complexity and site-specific variation in approval processes made it difficult for the research team to predict when data collection would start, and this meant that surveyor recruitment and training was conducted much earlier than data collection. Within the current research funding, ethics and governance climate, these circumstances are unlikely to change. Issues around ethics, governance and logistical processes of DUQuA and another research study are reported elsewhere [1, 2].

This unique experience has highlighted that engaging surveyors as data collectors for health services research may be likened to conducting short-notice surveys in hospitals. It mirrors the external evaluation mechanism of ad-hoc accreditation due to its efficiency and capacity to assess facilities in their natural course of operations. Although short-notice accreditation may be more effective and efficient than advance-notification surveys in detecting deficiencies regarding clinical care standards and criteria, the challenges of implementing this evaluation method in hospitals may serve as barriers to survey effectiveness, or impediments to survey accuracy. Our study has presented some of these key barriers through the experiences of our participants who disclosed challenges in conducting data collection for research onsite at hospitals.

Limitations

Limitations of this study include our reliance on surveyors' self-reported experiences. The study was conducted by the same research team as the primary DUQuA study, which may have introduced biases to the data analysis. We attempted to control this by asking a qualitative researcher from a different team to review and analyse a proportion of transcripts for analysis comparison. Although the number of participants was modest, the responses were generally consistent across interviewed surveyors, and data saturation was reached.

Conclusion

This is the first study to investigate differences between experiences of surveyors collecting hospital data for accreditation and research data collection purposes. We found that surveyors encounter more barriers when collecting data for research than for accreditation programs. While participants experienced time pressures associated with both processes, they felt that training and prior knowledge and experience in the healthcare sector eased these pressures and helped them to prepare and conduct data collection. Undertaking data collection for research led surveyors to realize that short-notice accreditation may be a more appropriate way of assessing everyday care in the organization. The use of surveyors in health services research, therefore, has the potential to bring data collection approaches and methodologies into alignment with national quality and safety standards, and facilitate a more robust and holistic approach to healthcare quality improvement.

Supplementary material

Supplementary material is available at *INTQHC Journal* online.

Contributors

The research team consists of experienced researchers with expertise in health services research, qualitative methodology, large-scale research and project management, quality improvement and assessment, accreditation, clinical indicators, policy and patient experience. JB initiated DUQuA and led the research grant to fund the project. TW and RCW co-led the detailed study design, managed the project across time and contributed to the development of the manuscript. NT, EH, PH and EA provided qualitative and critical expertise for the development and revisions of the manuscript.

Acknowledgements

We acknowledge and thank the Australian Council on Healthcare Standards (ACHS) for supporting the Deepening our Understanding of Quality in Australia (DUQuA) study as a whole and for supporting this research. We thank those ACHS surveyors who participated in the interviews and for providing extensive feedback. Our appreciation goes to Ms Zhicheng Li for coordinating and organizing EQA training and data collection visits at hospitals. We also specifically thank Ms Chrissy Clay for coordinating data collection visits.

Funding

This work was funded by National Health and Medical Research Council (NHMRC) Program Grant APP1054146 (CI Braithwaite).

Data sharing statement

Data will be made publicly available to the extent that individual participants or participating hospitals cannot be identified, in accordance with requirements of the approving Human Research Ethics Committees.

References

1. Clay-Williams R, Taylor N, Braithwaite J. Potential solutions to improve the governance of multicentre health services research. *Med J Aust* 2018;**208**:152–4.
2. Foot H, Scott IA, Russell GM *et al.* Ethics and site-specific governance approvals for multi-centre, inter-sector health care research. *Med J Aust* 2018;**209**:175–6.
3. Holden RJ, McDougald Scott AM, Hoonakker PLT *et al.* Data collection challenges in community settings: insights from two field studies of patients with chronic disease. *Qual Life Res* 2015;**24**:1043–55.
4. Greenfield D, Braithwaite J. Health sector accreditation research: a systematic review. *Int J Qual Health Care* 2008;**20**:172–83.
5. Australian Commission on Safety and Quality in Health Care. *National Safety and Quality Health Service Standards*. Sydney, AU, Australian Commission on Safety and Quality in Health Care, 2012.
6. Plebani M. Role of inspectors in external review mechanisms: criteria for selection, training and appraisal. *Clin Chim Acta* 2001;**309**:147–54.
7. Taylor N, Clay-Williams R, Hogden E *et al.* Deepening our understanding of quality in Australia (DUQuA): a study protocol for a nationwide, multilevel analysis of relationships between hospital quality management systems and patient factors. *BMJ Open* 2015;**5**:e010349.
8. Bohigas L, Brooks T, Donahue T *et al.* A comparative analysis of surveyors from six hospital accreditation programmes and a consideration of the related management issues. *Int J Qual Health Care* 1998;**10**:7–13.
9. dos Santos RA, Snell L. Tenorio Nunes MdP. The link between quality and accreditation of residency programs: the surveyors' perceptions. *Medical Education Online* 2017;**1270093**:22.

10. Elnour AA, Hernan AL, Ford D *et al.* Surveyors' perceptions of the impact of accreditation on patient safety in general practice. *Med J Aust* 2014;**201**:s56–9.
11. Ho M-J, Chang H-H, Chiu Y-T, Norris JL. Effects of hospital accreditation on medical students: a national qualitative study in Taiwan. *Acad Med* 2014;**89**:1533–9.
12. Pongpirul K, Sriratanaban J, Asavaroengchai S *et al.* Comparison of health care professionals' and surveyors' opinions on problems and obstacles in implementing quality management system in Thailand: a national survey. *Int J Qual Health Care* 2006;**18**:346–51.
13. Mack N, Woodsong C, MacQueen KM *et al. Qualitative Research Methods: A Data Collector's Field Guide*. Family Health International, North Carolina, USA, 2005.
14. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006;**3**:77–101.
15. Denzin NK, Lincoln YS. *Strategies of Qualitative Inquiry*, 4th edn. USA, SAGE, 2013.
16. Greenfield D, Braithwaite J, Pawsey M. Health care accreditation surveyor styles typology. *Int J Health Care Qual Assur* 2008;**21**:435–43.
17. Greenfield D, Moldovan M, Westbrook M *et al.* An empirical test of short notice surveys in two accreditation programmes. *Int J Qual Health Care* 2011;**24**:65–71.
18. Clay-Williams R, Taylor N, Winata T *et al.* Organization quality systems and department-level strategies: refinement of the Deepening our Understanding in Quality in Australia (DUQuA) organization and department-level scales. *Int J Qual Health Care* 2020;**32**:22–34.
19. Clay-Williams R, Taylor N, Ting HP *et al.* The clinician safety culture and leadership questionnaire: refinement and validation in Australian public hospitals. *Int J Qual Health Care* 2020;**32**:52–59.
20. Arnolda G, Winata T, Ting HP *et al.* Implementation and data-related challenges in the Deepening our Understanding of Quality in Australia (DUQuA) study: implications for large-scale cross-sectional research. *Int J Qual Health Care* 2020;**32**:75–83.
21. Hofstede G. *Culture's Consequences: International Differences in Work-Related Values*, 2nd edn. Beverly Hills, CA, SAGE Publications, 1984.
22. Hofstede G. Dimensionalizing cultures: the Hofstede model in context. Online Readings in Psychology and Culture. 2011;**2**.
23. Nicklin W, Fortune T, van Ostenberg P *et al.* Leveraging the full value and impact of accreditation. *Int J Qual Health Care* 2017;**29**:310–2.
24. Stover B, Lubega F, Namubiru A *et al.* Conducting a large public health data collection project in Uganda: methods, tools, and lessons learned. *J Res Pract* 2018;**14**:M1.
25. Taylor N, Hogden E, Clay-Williams R *et al.* Older, vulnerable patient view: a pilot and feasibility study of the patient measure of safety (PMOS) with patients in Australia. *BMJ Open* 2016;**6**:e011069.
26. Hassan ZA, Schattner P, Mazza D. Doing a pilot study: why is it essential? *Malays Fam Physician* 2006;**1**:70–3.
27. Greenfield D, Pawsey M, Hinchcliff R *et al.* The standard of healthcare accreditation standards: a review of empirical research underpinning their development and impact. *BMC Health Serv Res* 2012;**12**:329.

International Journal for Quality in Health Care, 2020, 32(S1), 99–103
doi: 10.1093/intqhc/mzz098
Perspectives on Quality

Perspectives on Quality

Conclusion: the road ahead: where should we go now to improve healthcare quality in acute settings?

JEFFREY BRAITHWAITE[1], NATALIE TAYLOR[2,3], ROBYN CLAY-WILLIAMS[1], HSUEN P. TING[1], and GASTON ARNOLDA[1]

[1]Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Road, NSW 2109, Australia, [2]Cancer Research Division, Cancer Council NSW, 153 Dowling St., Woolloomooloo, NSW 2011, Australia, and [3]Faculty of Health Sciences, University of Sydney, Camperdown, Sydney, NSW 2006, Australia

Address reprint requests to: Jeffrey Braithwaite, Centre for Healthcare Resilience and Implementation Science (CHRIS), Australian Institute of Health Innovation, Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia.
Tel: +61 2 9850 2401; Fax: +61 2 9850 2499; E-mail: jeffrey.braithwaite@mq.edu.au

Editorial Decision 7 September 2019; Accepted 13 September 2019

Abstract

This final article in our 12-part series articulating a suite of quality improvement studies completes our report on the Deepening our Understanding of Quality in Australia (DUQuA) program of work. Here, we bring the Supplement's key findings and contributions together, tying up loose ends. Traversing the DUQuA articles, we first argued the case for the research, conducted so that an in-depth analysis of one country's health system, completed 5 years after the landmark Deepening our Understanding of Quality Improvement in Europe (DUQuE), was available. We now provide a digest of the learning from each article. Essentially, we have contributed an understanding of quality and safety activities in 32 of the largest acute settings in Australia, developed a series of scales and tools for use within Australia, modifiable for other purposes elsewhere, and provided a platform for future studies of this kind. Our main message is, despite the value of publishing an intense study of quality activities in 32 hospitals in one country, there is no gold standard, one-size-fits-all methodology or guarantee of success in quality improvement activities, whether the initiatives are conducted at departmental, organization-wide or whole-of-systems levels. Notwithstanding this, armed with the tools, scales and lessons from DUQuA, we hope we have provided many more options and opportunities for others going about strengthening their quality improvement activities, but we do not claim to have solved all problems or provided a definitive approach. In our view, quality improvement initiatives are perennially challenging, and progress hard-won. Effective measurement, evaluating progress over time, selecting a useful suite of quality methods and having the persistence to climb the improvement gradient over time, using all the expertise and tools available, is at the core of the work of quality improvement and will continue to be so.

Key words: quality management, national standards, accreditation, quality improvement

Introduction

We began this excursion across the contours, highways and byways of the 5-year Deepening our Understanding of Quality in Australia (DUQuA) research program with an understanding that health systems had not made the progress stakeholders wanted to see in a system striving to provide quality care reliably, effectively and with high levels of fidelity to all who seek it. We have now documented the results of our journey in the Supplement: a multiplicity of projects derived from this ambitious research program examining a sample of the largest hospitals across Australia. The projects have realized a range of useful results. Some of the key outputs, outcomes and learnings are summarized in Table 1.

In this concluding article, we should note at the outset, despite the learnings synthesized in Table 1, that we still have not solved the twin problems of definitively showing how to improve care quality across a health system, or comprehensively demonstrated how to measure systems and processes for improving care quality. What the DUQuA team has contributed, however, is a collection of pointers to where quality improvement activities can best be targeted, and a suite of measurement tools to assist in implementing improvements. We aimed to advance the field methodologically, and tested the original Deepening our Understanding of Quality Improvement in Europe (DUQuE) study in one country, half a decade later.

We have grounds for supporting the proposals that—notwithstanding that system-wide quality is infuriatingly hard to engineer—patients should be in a more prominent place in decisions about their care, department level strategies are at least as likely to achieve results as organization-wide initiatives and multifaceted initiatives and quality bundles rather than stand-alone interventions might help change agents make progress. Quality improvement in our view should be seen much more as a long-term endeavour: less about a clear destination and more about an extended journey of trial and error, test and experimentation. Progress in this field, it seems, will always be hard-won.

DUQuA's legacies

As well as these lessons, DUQuA leaves a legacy of a range of validated scales for assessing aspects of quality of care, and helps dispel the hype, put forward by some commercial or overly enthusiastic interests, that quality improvement is unproblematically achieved by implementing a defined, well-known mix of strategies—just try harder and add root cause analysis, incident reporting, a hand hygiene campaign and some targeted projects using PDSA cycles with control charts of data, and you will be on an improvement gradient with an end point of assured quality of care or even the fabled 'zero harm'. Instead, DUQuA research indicates how uncertain and imprecise this kind of work is—that is to say, that in acute settings there is no compelling or coherent relationship between any particular kinds of quality improvement strategies on the one hand and improved outcomes for patients on the other.

We would also point out that, for all the international endeavours in designing and releasing frameworks, indicators, targets and the like, whether benchmarked, controlled, validated, standardized or otherwise, we do not have widely agreed measures of care, or accepted ways to track the quality of care routinely over time. Notwithstanding this, policymakers, researchers and quality improvement specialists have expended a lot of energy on data and their management: datasets, data refinement, data variation, databases, data comparisons and data analysis. This has led Berwick [1] to argue against 'excessive measurement', with the admonition that we should measure only what is important, and mainly for learning rather than for increasing precision. While we don't fully agree with that conclusion (what is excessive? And when and how, as scientists, should we limit measurement?), we do understand the sentiment—that it is better to get on with the job of improving care, than spending disproportionate time and resources on fine-grained measurement, with precise metrics on things that may not matter. We are mindful of the wise words of one of our heroes, the statistician Tukey: 'An approximate answer to the right question is worth a great deal more than a precise answer to the wrong question' [2, 3].

Yet we also do not have the range of interventions that routinely and reliably work to enable us to say with confidence that we can improve care on the ground, regardless of whether we take a Safety-I, stamping-out-harm approach, or a Safety-II, supporting-things-going-right approach [4]. And we must, surely, when we do improvement work, show what gains we have made compared with an earlier time. Thus, whether judged excessive or not, we need some form of measurement of both processes and outcomes. In essence, what DUQuA and its DUQuE program have shown is that trying to intervene or measure in a complex system, even with a robust approach, a set of tools, a theory of change, statistical capacity and a motivated health system, is among the most complex and thorniest of challenges.

The challenge of quality improvement in complex settings

All in all, despite some progress over the last 25 years, it is not clear cut to argue that organizing effectively for quality improvement or encouraging clinicians to closely follow well-documented patient pathways, or both, will lead inevitably to higher quality, safer or better care or guaranteed outcomes. No one level of improvement in our multifaceted study is associated with any specific improvement in outcomes and no identifiable factors act as a consistent trigger or stimulus for change. One likely reason for this is that we too often in healthcare assume linearity. That is to say, in quality improvement, patient safety or implementation science, there is still a tendency of some people to imagine that there are causal relationships in a logical hierarchy that says if 'X' is done, then 'Y' will occur. Instead, healthcare delivery systems are not well described as a chain, or a pipeline, or via linear models which suggest there are inexorable relationships between inputs, processes to outputs. The reason is that health delivery systems are complex adaptive systems [5–8] with many interacting agents, moving parts, variables and differential subsystems. Behaviours are emergent rather than predictable, much less under strict control. Cultures are path dependent, owing much to their historical antecedents [9, 10]. Healthcare is not a machine with simple coordinating mechanisms, whereby we can switch the quality button to 'on'.

Linear thinking is thus insufficient. This is why three highly influential reports in 2018 [11–13] all argued that systems-level thinking was a crucial enabler for understanding healthcare, and making improvements.

Indeed, healthcare delivery systems are challenging even to model, let alone manage, and their interdependencies are always hard to apprehend. There are variables everywhere and many moving parts such that our studies, and the work of the DUQuE investigators, and the multiplicity of interventional studies that have been developed to try to demonstrate healthcare improvement [14, 15], have encountered difficulties in holding extraneous variables constant and

Table 1 Selected outputs, outcomes and learnings from DUQuA results articles

Article 2: DUQuA: An overview of a nation-wide, multi-level analysis of relationships between quality management systems and patient factors in 32 hospitals	• There is no one-size-fits-all approach to quality, or single gateway to quality improvement • Scales and measures for quality of care are now available, designed for Australian hospitals, and modifiable for other jurisdictions
Article 3: Organization quality systems and department-level strategies: refinement of the Deepening our Understanding in Quality in Australia (DUQuA) organization and department-level scales	• Self-reports and audit measures are now combined in a series of scales to measure quality systems in hospitals • Seven quality systems and strategies have been articulated for measuring quality of care at organization and department level
Article 4: Do organization level quality management systems influence department level quality? A cross-sectional study across 32 large hospitals in Australia	• This is the first time organization-level quality and its influence on department-level quality has been assessed in Australia • While there is no clearly detectable pattern of influence between the two levels, the strength of relationships was most visible for tests involving organization-level quality activities
Article 5: The relationships between quality management systems, safety culture and leadership, and patient outcomes in Australian Emergency Departments	• Hospitals, health departments and governments tend to measure ED performance in terms of patient flow • We show there is a significant trade-off between patient flow and patient safety
Article 6: The clinician safety culture and leadership questionnaire: refinement and validation in Australian public hospitals	• A new, validated instrument for measuring clinician safety culture and leadership scale across the hospital is now released • The scale's development is based on robust theory and evidence-based methods
Article 7: Do quality management systems influence clinical safety culture and leadership? A study in 32 Australian hospitals	• Organization-level quality systems positively influence clinicians' safety culture, and leadership • Mandatory accreditation in Australia results in higher performing hospitals but measuring the influence of quality systems on other hospital factors can be difficult when all hospitals perform equally well
Article 8: Validation of the Patient Measure of Safety (PMOS) questionnaire in Australian public hospitals	• A validated scale incorporating patient perspectives on safe care can now be used • Patients have more positive perceptions of care when clinicians adhere to clinical guidelines
Article 9: Implementation and data-related challenges in the DUQuA study: implications for large-scale cross-sectional research	• Hospital ethics and governance processes present major barriers to studies of this kind and scale • Staff turnover among hospital leadership cohorts and poor hospital capacity to host studies such as DUQuA represents substantial risks to future health services research
Article 10: Can benchmarking Australian hospitals for quality identify and improve high and low performers? Disseminating research findings to hospitals	• DUQuA and DUQuE provide us with the opportunity to benchmark data to improve quality of care • Benchmarking provides a way to communicate with hospitals in their own language, thereby facilitating research translation
Article 11: Using accreditation surveyors to conduct health services research: a qualitative, comparative study in Australia	• Accreditation surveyors encounter substantial barriers when conducting health services research compared with when they are doing accreditation surveys • Some of the barriers encountered when collecting data for research provide insight into the challenges that may be faced when visiting hospitals for short-notice accreditation; nonetheless they are a very useful resource for research of this kind

expecting there will be unambiguous outcomes delivered with observable associations explicitly linking the intervention and those outcomes. We cannot rely on the ratiocination implied by an intervention with a control group, where everything except the independent variable is held constant, and under control. This is why DUQuE reported, and we also report that, despite these two expensive, expansive, deep-dive investigations into quality across entire health systems, there are no clear relationships to which we can point. Instead we document in the articles across this Supplement mixed results, and so do many others when they have tried to demonstrate

that quality improvement activities, even if done well, will result inevitably in better care quality. Put simply, there are many factors at work and multiple mediators in the production of care that must be understood and managed for high-quality, safer care to be produced.

For all that, this postscript of DUQuA and its implications for future change raises the obvious questions of where is it that we have actually made gains, what can we learn from this and where are we heading with quality improvement? What seems clear now that we have stacked up the evidence of both DUQuA and DUQuE is that systems-wide change has proven challenging to orchestrate. That is not to say that there have not been celebrated demonstrations of improvement across multiple sites when concerted efforts have been made. Reductions in central line infections [16, 17], mortality due to rapid response systems [18] and adverse events in The Netherlands [19], to cite only three examples, show that clear benefits are achievable in response to well-designed interventions with sound metrics and robust measuring systems. In other words, some specific initiatives have bent the improvement curve in the right direction. But without taking away from these successes, they are confined to a specific area, issue or concern. Across-the-board, organization-wide, measured, systems-wide improvement is not guaranteed. And even when change is enabled and improvement demonstrated, there is always a risk in a complex adaptive system that the ground will shift, relationships and interdependencies will evolve, cultures and systems will alter or reliability will not be achieved, and the gains might thereby evaporate. Even effective teams, services, departments or whole hospitals might only be a restructure, change in chief executive or external crisis away from losing any advances they have secured.

Ending on a positive note

But we do not want to end on a pessimistic note, and DUQuA does not lead us to that state of mind. Researchers in every discipline, from physics, to geology, to artificial intelligence, to psychology, think they are different, but have broadly carved out a path of progress. Perhaps the situation we are describing in quality of care research is not so different to the way these sciences, and healthcare and medicine have proceeded historically—full of haphazard, convoluted trajectories, infused with uncertainty, with ever-present backstreets, side roads and alleyways, but on a journey of long-term improvement. Before the Age of Sail in the eighteenth century, the causes of scurvy were unknown. Then, despite being demonstrated by James Lind in 1753, it was 42 years before the solution—citrus fruits—was routinely given to sailors as part of their rations by the British Royal Navy. Until then, it was assumed that 50% of a ship's complement would die from scurvy on a long-distance voyage [20, 21]. Every surgeon can recite the story of Sir Joseph Lister, who first published about the benefits of antiseptic techniques in 'The Lancet' of 1867 [22], but was mocked and ridiculed; 'The Lancet' of 1873 even warned against his progressive ideas. And mid-last century, a key breakthrough in public health came in the form of research of Richard Doll and Bradford Hill between 1948 and 1950, culminating in a report in the 'British Medical Journal' demonstrating the link between smoking and lung cancer [23]. Today, many fewer patients smoke, and rates have been falling from the decade of the 1960s—at least in the developed world [24]. What these historical examples tell us is that in the science of health and medical behaviour, we are always dealing with multiple variables, conflicting constraints and many factors that influence causes, effects and consequences. And there is almost always a lag, often measured in decades, between discovery, intervention, take-up, adoption, spread and evaluation of any improvement measure, even the most compelling [25].

Maybe that is the best explanation for where quality improvement is right now: we are dealing with complex multifactorial problems and we will not know the results of our activities until several generations of accumulated experience about what works and what doesn't under a variety of conditions has been assembled. Perhaps we have not had until now a sufficient range of interventions, well-designed and executed studies, *in situ* demonstrations and trial-and-error exemplars to appreciate the landscape of healthcare improvement. If this is the case, we will have to continue to support investigations, whether by analysing and then documenting natural experiments or designing specific experimental interventions. We will also have to see health systems be more willing to facilitate ease of access so studies at scale such as DUQuA can be done more easily [26, 27].

We do now have in our possession as a result of the DUQuA efforts, freshly validated tools and scales for use in Australian healthcare, modifiable to other health systems, to measure and manage quality improvement initiatives at organization, department and Emergency Department levels. There are also tools and scales for key constructs of importance to quality improvement such as culture, leadership, safety culture and patient perspectives. All are available publicly, on open websites [28]. And we have documented in the DUQuA articles the valuable role that can be played by providing feedback via benchmarking, and external surveying of care processes. A parting suggestion is that if we could encourage more people to use the same tools and approaches, we can combine data and compare expertise from many small-scale studies and realize synergies from the collective efforts. This might be a goal worth pursuing.

For all those reasons, despite the slow progress of this field of endeavour, we remain optimistic that, as the discipline matures, and what and how we measure is enhanced, and our interventional expertise advances, we will be able over time to attribute results to specific activities, and climb the improvement gradient. We take heart from the international expertise of colleagues in research and improvement, and from the World Health Organisation to the International Society for Quality in Health Care to the World Bank to the Organisation for Economic Co-operation and Development and beyond who are concertedly working on quality improvement and patient safety. We also take heart from the history of medicine, with the lessons of Lind and scurvy, Lister and antiseptics and Doll and Hill and smoking, all providing more grounds for optimism. That optimism, we hope, is DUQuA's final lesson, and the destination to which our 5 years of research efforts point.

Contributors

The research team consists of experienced researchers, clinicians, biostatisticians and project managers with expertise in health services research, survey design, and validation, large-scale research and project management, sophisticated statistical analysis, quality improvement and assessment, accreditation, clinical indicators, policy and patient experience. JB conceived the idea, led the research grant to fund the project, chairs the steering committee and led the development of this manuscript. NT and RCW co-led the detailed study design, managed the project across time, and contributed to the development of the manuscript. HPT and GA provided statistical expertise for the study design and developed the analysis plan for the manuscript.

Acknowledgements

Members of the original DUQuE team (Oliver Gröne, Rosa Suñol) provided extensive input and advice in the initial stages of the project, supporting design modifications from DUQuE and the planned DUQuA study approach. Dr Annette Pantle provided expert advice on revision and development of DUQuA measures, and Professor Sandy Middleton, Associate Professor Dominique Cadillac, Kelvin Hill, Dr Carmel Crock and Professor Jacqueline Close provided input for the development of our AMI, hip fracture and stroke clinical process indicator lists. Nicole Mealing and Victoria Pye provided statistical advice in the initial phases of the project. We greatly appreciate their efforts.

Funding

This work was funded by National Health and Medical Research Council Program Grant APP1054146 (CI Braithwaite).

Data sharing statement

Data will be made publicly available to the extent that individual participants or participating hospitals cannot be identified, in accordance with requirements of the approving Human Research Ethics Committees.

References

1. Berwick DM. Era 3 for medicine and health care. *JAMA* 2016;**315**:1329–30.
2. Tukey J. The future of data analysis. *Ann Math Statist* 1962;**33**:13.
3. Levitt SD, Dubner SJ. *Superfreakonomics: Global Cooling, Patriotic Prostitutes and Why Suicide Bombers Should Buy Life Insurance*. New York, NY, William Morrow, 2009.
4. Braithwaite J, Wears RL, Hollnagel E. Resilient health care: turning patient safety on its head. *Int J Qual Health Care* 2015;**27**:418–20.
5. Braithwaite J. Changing how we think about healthcare improvement. *BMJ* 2018;**361**:k2014.
6. Braithwaite J, Churruca K, Ellis LA *et al. Complexity science in healthcare–aspirations, approaches, applications and accomplishments: a white paper*. Sydney, Australian Institute of Health Innovation, Macquarie University, 2017.
7. Braithwaite J, Churruca K, Long JC *et al.* When complexity science meets implementation science: a theoretical and empirical analysis of systems change. *BMC Med* 2018;**16**:63.
8. Braithwaite J, Churruca K, Ellis LA. Can we fix the uber-complexities of healthcare? *J Royal Soc Med* 2017;**110**:392–4.
9. Braithwaite J, Herkes J, Ludlow K *et al.* Association between organisational and workplace cultures, and patient outcomes: systematic review. *BMJ Open* 2017;**7**:e017708.
10. Braithwaite J, Hyde P, Pope C (eds). *Culture and Climate in Health Care Organizations*. London, UK, Palgrave Macmillan, 2010.
11. World Health Organization, Organisation for Economic Co-Operation and Development, The World Bank. *Delivering Quality Health Services: A Global Imperative for Universal Health Coverage*. Geneva, CH, World Health Organization, Organisation for Economic Co-operation and Development, and The World Bank, 2018. Available from http://apps.who.int/iris/bitstream/handle/10665/272465/9789241513906-eng.pdf?ua=1.
12. Kruk ME, Pate M, Mullan Z *et al.* Introducing the Lancet Global Health Commission on High-Quality Health Systems in the SDG Era. *Lancet Glob Health* 2017;**5**:e480–1.
13. National Academies of Sciences. *Engineering, and Medicine. Crossing the Global Quality Chasm: Improving Health Care Worldwide*. Washington DC, The National Academies Press, 2018.
14. Landrigan CP, Parry GJ, Bones CB *et al.* Temporal trends in rates of patient harm resulting from medical care. *N Engl J Med* 2010;**363**:2124–34.
15. Benning A, Dixon-Woods M, Nwulu U *et al.* Multiple component patient safety intervention in English hospitals: controlled evaluation of second phase. *BMJ* 2011;**342**:d199.
16. Pronovost P, Needham D, Berenholtz S *et al.* An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006;**355**:2725–32.
17. Bion J, Richardson A, Hibbert P *et al.* 'Matching Michigan': a 2-year stepped interventional programme to minimise central venous catheter-blood stream infections in intensive care units in England. *BMJ Qual Saf* 2013;**22**:110–23.
18. Chen J, Ou L, Hillman KM *et al.* Cardiopulmonary arrest and mortality trends, and their association with rapid response system expansion. *Med J Aust* 2014;**201**:167–70.
19. Baines R, Langelaan M, de Bruijne M *et al.* How effective are patient safety initiatives? A retrospective patient record review study of changes to patient safety over time. *BMJ Qual Saf* 2015;**24**:561–71.
20. Price C. The age of scurvy. *Distillations* 2017;**3**:12–23.
21. Hemilä H. A brief history of vitamin C and its deficiency, scurvy. Finland: 2019. Available from http://www.mv.helsinki.fi/home/hemila/history.
22. Lister J. On the antiseptic principle in the practice of surgery. *Lancet* 1867;**90**:353–6.
23. Doll R, Hill AB. Smoking and carcinoma of the lung. *BMJ* 1950;**2**:739–48.
24. World Health Organization . Prevalence of tobacco smoking. 2019. Available from http://gamapserver.who.int/gho/interactive_charts/tobacco/use/atlas.html.
25. Green LW. Making research relevant: if it is an evidence-based practice, where's the practice-based evidence? *Fam Pract* 2008;**25**:i20–4.
26. Clay-Williams R, Taylor N, Braithwaite J. Potential solutions to improve the governance of multicentre health services research. *Med J Aust* 2018;**208**:154–4.
27. Braithwaite J, Clay-Williams R, Taylor N. Potential solutions to improve the governance of multicentre health services research. *Med J Aust* 2018;**209**:187–8.
28. Australian Institute of Health Innovation. Australian Institute of Health Innovation, world-leading healthcare innovators. 2019. Available from: https://www.mq.edu.au/research/research-centres-groups-and-facilities/healthy-people/centres/australian-institute-of-health-innovation.

The International Society for Quality in Health Care

ISQua works to provide services to guide health professionals, health providers, researchers, agencies, policy makers and consumers to achieve excellence in health care delivery to all people and to continuously improve the quality and safety of care.

ISQua offers a unique opportunity for individuals and institutions with a common interest to share exper tise via an international multidisciplinary forum. **ISQua** is a non-profit, independent organisation, which is suppor ted by its members, including leading quality health care providers and agencies in over 70 countries, together with additional funds from the Irish Government.

ISQua's values

- Transparency
- Integrity
- Diversity
- Ethical
- Evidence based
- Excellence

ISQua's Vision

To be the lead enabler of transformation in healthcare quality and safety globally.

Under Improving Health Care Worldwide

ISQua's focus is on the methodologies needed to facilitate the delivery of safe, quality health care. The Society's links throughout the world are maintained through an active Board of recognized leaders, members, regional and other networks, and other interest groups. Membership of steering groups comprises **ISQua** members who are acknowledged leaders in quality and safety of care.

ISQua's international Accreditation Program brings the world of health care accreditation and external evaluation together by providing:

- An international assessment and accreditation process for health care standards
- An international survey and accreditation process for the performance of accreditation and other external evaluation organisations, and
- An international assessment and accreditation program for surveyor training.

ISQua also hosts an Annual International Accreditation Symposium in association with its Annual Conference.

Education and Professional Development

ISQua conducts an online professional development and education programme for those working in or with an interest in quality and safety in healthcare. The programme features content delivered in various forms provided and presented by world leaders in healthcare. **ISQua** partners with eminent organisations across the world in bringing this programme to par ticipants in over 40 countries.

International Conference Programme

With an annual International Conference, **ISQua** also provides a broad based global meeting point which brings together leaders, interested groups and individuals who are active in the field of quality and safety.

A World-Class Journal

The International Journal for Quality in Health Care, a bi-monthly peer-reviewed journal, is published jointly by **ISQua** and Oxford University Press. The Journal publishes papers in all disciplines engaged in health care quality improvement and research; its Editorial Board members represent countries in all regions of the world. Through the pages of the Journal and also via online electronic access, **ISQua**'s worldwide membership and non-member Journal subscribers receive up-to-date information on current quality and safety in health care activities and research and access to the electronic version of the Journal online.

ISQua Board 2017 – 2019

President Wendy Nicklin, Canada
Immediate Past President Cliff Hughes, Australia
President Elect Jeffrey Braithwaite, Australia
Members Peter Carter, Australia; Stephen Clark, Australia; Asian Society for Quality in Health Care (ASQua): Yuichi Imanaka, Japan; Duncan Inverarity, Ireland; HAS; Bruno Lucet, France; Bhupendra Kumar Rana, India; Council for Health Service Accreditation of Southern Africa (COHSASA): Jacqui Stewart, South Africa; Shin Ushiro, Japan; David Vaughan, Ireland; Accreditation Canada; Leslee Thompson, Canada.

Honorary Advisors: Low and Middle Income Countries; Sheila Leatherman, USA; International Journal for Quality in Health Care, Jack Li, Taiwan.

Chief Executive Officer Peter Lachman

Editor of *The International Journal for Quality in Health Care*, Prof Yu-Chan (Jack) Li, Taiwan, and Honorary Advisor to the Board.

For **ISQua** Membership information check the **ISQua** website: www.isqua.org