BMJ Open

Understanding how and why audits work in improving the quality of hospital care: a realist review protocol

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-015121
Article Type:	Protocol
Date Submitted by the Author:	10-Nov-2016
Complete List of Authors:	Hut-Mossel, Lisanne; University of Groningen, University Medical Centre Groningen, Centre of expertise on Quality and Safety Welker , Gera ; University of Groningen, University Medical Centre Groningen, Centre of expertise on Quality and Safety Ahaus, Kees; University of Groningen, University Medical Centre Groningen, Centre of Expertise on Quality and Safety; University of Groningen, Faculty of Economics and Business, Department Operations Gans, Rob; University of Groningen, University Medical Centre Groningen, Department of Internal Medicine
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Research methods, Health policy
Keywords:	Realist synthesis, Realist review, AUDIT, Quality improvement, Clinical audit < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Hospital Care

Title

Understanding how and why audits work in improving the quality of hospital care: a realist review protocol

Authors

Lisanne Hut-Mossel[1], Gera Welker[1], Kees Ahaus[1, 2] and Rijk Gans[3].

University of Groningen, University Medical Centre Groningen.

Author Note

1 University of Groningen, University Medical Centre Groningen, Centre of Expertise on Quality and Safety, Groningen, the Netherlands.

2 University of Groningen, Faculty of Economics and Business, Department Operations, Groningen, the Netherlands.

3 University of Groningen, University Medical Centre Groningen, Department of Internal Medicine, Groningen, the Netherlands.

Corresponding author

Lisanne Hut-Mossel

Huispostcode LA10

PO Box 30 001

9700 RB Groningen

The Netherlands.

Tel: +31623473381

Correspondence to p.a.mossel@umcg.nl

Word Count (excluding title page, abstract, references, figures and tables): 3413

ABSTRACT

Introduction Many types of audits are commonly used in healthcare to promote quality improvements. However, the evidence on the effectiveness of audits is mixed. The objectives of this proposed realist review are (i) to understand how and why audits might, or might not, work in terms of delivering the intended outcome of improved quality of healthcare and (ii) to examine under what circumstances audits could potentially be effective by formulating and refining underlying programme theories. This protocol will provide the rationale for using a realist review approach and outline the method.

Methods and Analysis This review will be conducted using an iterative four-stage approach. The first and second step have already been executed. The first step was to develop an initial programme theory based on the literature that explains how audits are supposed to work. Second, a systematic literature search was conducted using relevant databases. Third, data will be extracted and coded for concepts relating to context, outcomes and their interrelatedness. Finally, the data will be synthesised in a five-step process: (1) organising the extracted data into evidence tables, (2) theming, (3) formulating chains of inference from the identified themes, (4) linking the chains of inference and formulating CMO configurations and (5) refining the initial programme theory. The reporting of the review will follow the RAMESES publication standards.

Ethics and Dissemination This review does not require formal ethical approval. A better understanding of how and why these audits work, and how context impacts their effectiveness, will inform stakeholders in deciding how to tailor and implement audits within their local context. We will use a range of dissemination strategies to ensure that findings from this realist review are broadly disseminated to academic and non-academic audiences.

Trial registration number This systematic review protocol is registered on the PROSPERO database (registration number CRD42016039882).

Keywords Realist synthesis, Realist Review, Audit, Quality Improvement, Clinical Audit, Hospital Care

STRENGTHS AND LIMITATIONS OF THIS STUDY

- This review goes beyond considering the effectiveness of audits by building an understanding of how and why audits work within various contexts;
- This review uses a systematic screening protocol;
- The main limitation is that realist reviews are dependent upon the transparency and adequacy of the reporting of individual studies by the original authors, and this may hamper a full understanding of how and why audits are effective.

INTRODUCTION

In recent decades, quality and safety issues have become increasingly important in healthcare because of their direct effect on both clinical outcomes and patient satisfaction. However, healthcare still suffers from a quality gap between the ideal care, in line with the best available medical evidence, and the actual care provided to patients [1]. To close this gap, health authorities and organisations currently prioritise quality improvement (QI) strategies, which are seen as systematic, data-driven monitoring and evaluation activities to improve healthcare quality [2]. A widely used QI strategy is the audit.

In this review protocol, we will focus on audits that address quality. Such audits are commonly used to promote quality improvements by monitoring, controlling and/or changing healthcare processes and healthcare providers' performance [3]. However, it is unlikely that audits work in the same way in every setting. Accordingly, it is important to understand how and why audits might lead to quality improvements. A realist review, as outlined in this protocol, will contribute to this understanding.

The range of possible audits can be roughly divided into (1) external audits, used to gain insight into hospitals' compliance with external criteria (e.g. accreditation, certification, external peer reviews), (2) internal audits, often in preparation for an external audit, and (3) clinical audits, carried out on a voluntary basis by healthcare professionals [3, 4]. Although there are differences, such as in the scope and the approach used, in the various types of audits, they all serve the same objective – which is to improve quality.

External audits are used to assess certain dimensions or characteristics of a healthcare providing organisation against specified standards [5]. As such, the implementation of an external audit requires an external standard and collaboration from beyond the hospital - and this distinguishes them from internal audits.

Internal audits are conducted by internal auditors of the hospital's own organisation, such as quality officers or healthcare professionals from another department than the one being audited to guarantee some level of independent judgement. Internal audits vary in purpose. On the one hand, healthcare organisations use internal audits to continuously improve the quality of healthcare. In this way, one could expect that, compared with external audits, threats to quality can be more quickly revealed, allowing the organisation to regularly adapt its processes to improve quality at the local level. Alternatively, internal audits may be used in the framework of external audits and are conducted to avoid performance standards dropping between two external audits. These audits are designed to evaluate and improve the effectiveness of the organisation's quality management system and focus more on organisational conditions and less on the behaviour of healthcare professionals and patient outcomes [4].

Clinical audits differ from other types of audits in that they are mostly undertaken and initiated by healthcare professionals who evaluate their own practices and work together to bring about improvements in their daily practice through the systematic evaluation of the healthcare provided [6]. As such, clinical audits do not necessarily use external criteria [7].

A considerable amount of literature addresses the effectiveness of audits and reports mixed results [8-10]. A systematic review on audit and feedback [11] demonstrated a positive overall effect of audits on clinical practice. Further, the authors noted differences in the design and the effectiveness of audits. This variety can be attributed to at least two issues. First, audits are used to improve specific aspects of healthcare and can be targeted at different levels. For example, external audits are performed to induce changes at the organisational level (e.g. in organisational policy or procedures) whereas clinical audits are performed to alter local healthcare practices (e.g. clinical day-to-day practices or local guidelines). Second, audits are used in different contexts, and this considerably complicates the evaluation of their effects.

For example, an audit could be effective in one organisation, or department, but not in another because of, for instance, the amount of support offered for quality improvements, as an element of their differing contexts (see supplementary file 1). The literature on QI strategies recognises that the mixed effects are partly due to the differing contexts in which interventions are planned [12, 13].

The variety in the targeted levels of audits, together with the heterogeneity of their contexts, suggests that it is unlikely that audits work in the same way in every setting. This creates challenges when attempting to synthesise evidence in a systematic review. Given this situation, more information about why and how audits work is needed [14, 15]. A detailed understanding of the contextual factors and the mechanisms that influence the effectiveness of audits is a prerequisite for understanding the mechanisms through which audits might lead to quality improvements. More importantly, a better understanding of 'how and why audits might work' will inform decision-making on how to tailor quality improvements at the local level.

A useful approach for explaining how and why audits might work, and investigating the interactions between context, mechanism and outcome, is the use of a realist review [16, 17]. The value of a realist review is that it is concerned with *how* an intervention works, rather than *whether* an intervention works, which is the focus of the conventional systematic review approach. Furthermore, the realist review methodology is specifically designed to cope with the intervention heterogeneity (in both the chosen study design and the used outcome measures) present in previous research on audits. Finally, this method is appropriate for the current research because audits are complex context-sensitive interventions [8, 13, 18]. Within the past decade, similar studies in other contexts have used realist reviews to understand how complex interventions work and are put into practice (e.g. [19]).

The objectives of the current review are (i) to understand how and why audits might, or might not, work in producing the intended outcome of improved quality of care and (ii) to examine under what circumstances audits could potentially be effective by formulating and refining underlying programme theories. Consequently, this review focusses on three research questions:

1. What are the mechanisms through which audits deliver their intended outcomes?
2. What contextual factors determine whether the identified mechanisms produce the intended outcomes of audits?
3. In what circumstances (i.e. which combination(s) of mechanisms and context) are audits most likely to be effective?

METHOD

The realist review aims to clarify from observed data the outcomes (O) of particular interventions in relation to context (C) and mechanisms (M). This 'CMO' configuration is based on the assumption that an intervention in a specific context (C) leads to mechanisms (M) that generate an outcome (O). Consequently, the underlying mechanisms can be expected to produce a broad range of different outcomes (O) when performed in different contexts (C) [16, 17, 20]. One of the key outputs of a realist review is the development of programme theories that set out how and why an intervention is thought to 'work' to generate certain outcomes [21].

Study design

This review follows Pawson's steps for conducting realist reviews, namely: (1) clarifying the scope, and programme theory development; (2) searching for evidence; (3) appraising primary studies and extracting data; and (4) synthesising evidence and drawing conclusions [16]. The reporting of the review will follow the 'Realist and Meta-Review Evidence Synthesis: Evolving Standards' (RAMESES) publication standards [21].

In line with these standards, data extraction and synthesis will be an interpretive process, driven by reflection and discussion by the review team [21]. This process requires repeated reading of primary studies because, as the synthesis progresses, new or refined elements of theory are expected to emerge. The protocol outlined below was written after the first steps had already been initiated or completed. Accordingly, both the past tense (steps that have been completed) and the future tense (steps that have yet to be initiated) are used.

1. Scope of the review and programme theory development

The first step of this review process has already been executed with the aim of building a programme theory that would explain how and why audits might work. The unit of analysis in a realist review is not the intervention itself, but the contexts, mechanisms and outcomes that underpin the intervention. Given this situation, the initial step in formulating a programme theory draws on the literature on the effectiveness of QI strategies. As audits are QI strategies, we would assume that the contexts, mechanisms and outcomes uncovered in the QI literature might also play a role in the effectiveness of audits. The initial programme theory explains how audits are supposed to work by framing the interrelationships between context, mechanism and outcome (see Figure 1) [16]. An exploration of programme theories was initiated through on-going conversations within the review team and by a preliminary search of the literature. In addition, key terms were defined to guide the review and to ensure a common understanding (see Supplementary File 1).

After a number of iterations and discussions, we developed an initial programme theory regarding how and why audits might work. This suggests that having an organisational culture that is supportive of quality improvement, a leadership committed to quality and previous audit experiences are important contextual factors in the success of an audit [22].

These contextual factors trigger mechanisms, including a focus on continuous improvement rather than auditing for compliance [23, 24], bottom-up initiatives as a prerequisite for ownership [9, 25] and the active involvement of healthcare professionals in audit processes [11, 23, 26], that in turn lead to improvements in the quality of healthcare.

This initial programme theory provides only a provisional structure for the review, and additional contextual factors, mechanisms and outcomes will be identified as the review progresses. The initial programme theory will be expanded, tested and refined using data from studies included in the review.

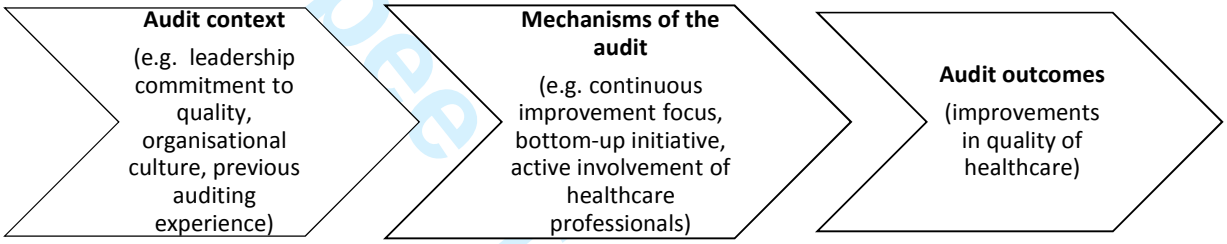

Figure 1. Initial programme theory for the effectiveness of an audit

2. Search for evidence

As a second step, a search strategy was developed and performed in collaboration with an experienced university librarian. To ensure that all relevant articles were identified, a systematic literature search was conducted in MEDLINE, Embase, PsycINFO, Academic Search Premier, Business Source Premier, EmeraldInsight, Cochrane Library and Web of Science for the period 2005 - 2015. These databases were selected because they contain the core of quality and patient safety studies in the field of healthcare management as well as the biomedical view on quality of healthcare. The search strategy was developed first for MEDLINE and later adapted for searching the other databases (see Supplementary File 2).

The search included appropriate indexing terms (i.e. MeSH terms and keywords) on descriptors of audits (e.g. clinical audit, accreditation, certification, peer review, quality improvement, quality assurance), outcomes (e.g. efficiency, effectiveness, improvement), and hospital care (e.g. academic medical centres, health organisations). The reference lists of the uncovered review articles were studied to identify additional primary studies.

A realist synthesis approach to searching for evidence is iterative and evolves as the understanding of the subject matter deepens. Consequently, as the review progresses further, we will also search for unpublished and grey literature (e.g. websites, national guidelines, policy documents and information reported in specialist conferences) on the assumption that the literature on this topic may be diverse and dispersed. In addition, our expectation is that not all the included publications will adequately report on all aspects of an audit. We will therefore identify papers, and other research outputs, that relate to the same study by using 'cluster' searching [27]. For example, a search can be based on the members of a research team of an included article to identify all other refereed journal articles and related documents. Further, an additional iterative search may be necessary if it is determined that more data are required to refine a specific part of the programme theory, or if new prospective theories are identified during data extraction or synthesis.

3. Appraise primary studies and extract data

The selection of appropriate primary studies has already been executed. First, one reviewer (LH) identified and removed duplicates. Next, two reviewers (LH and GW) independently screened all titles and abstracts for suitability for inclusion. The focus was on empirical studies that evaluated the effects of audits in hospital settings within high-income countries, without restrictions on the type of study design. Only studies published in English were included to avoid misinterpretation of the content of an article due to language barriers (see Table 1).

Second, the full texts of a random ten percent of these articles were independently reviewed by LH and GW and retained if they were deemed relevant (i.e. that the article could provide data on the context and mechanisms of an audit). One reviewer (LH) completed the remaining 90%.

However, a number of these articles required discussion or joint reading by two reviewers as it was sometimes difficult to decide between inclusion and rejection. Disagreements were recorded and discussed to ensure that decisions made were consistent. When disagreements remained, the matter was resolved by discussion involving the entire review team.

Inclusion criteria
Research on accreditation, certification, peer review/Dutch visitatie model[a] or local clinical audit
Hospital setting
High-income country
Published in English
English abstract available
Description of the medical or technical content
Description of the process of how the audit was conducted
Description of the impact of audit on medical and process outcomes

Table 1. Inclusion criteria. [a] This is a doctors-led and – owned system of peer review designed to assess the quality of care provided by groups of hospital-based medical specialists. Practices are surveyed every 3-5 years by a group of peers [28].

The next section describes activities that have yet to be started. Realist reviews amount to mixed-method reviews in that they incorporate both quantitative and qualitative studies, as well as grey literature. Consequently, different approaches are required to assess the quality of the included studies. Two reviewers (LH and either GW, KA or RG) will assess the quality of each included study. Any disagreement will be resolved through consensus-based group discussions within the review team. First, following realist synthesis principles, the evidence will be appraised using the concept of rigour [16, 17].

Rigour will be assessed by describing fidelity and nuggets (i.e. the potential match with the initial programme theory and valuable observations presented in primary studies), and trustworthiness (i.e. whether the methods used to generate the data are credible and trustworthy) [29, 30]. Second, to make the concept of trustworthiness more concrete, and to ensure transparency in decision-making, the quality of the evidence of each individual study will be presented in the form of an evidence-level table based on criteria established by the Cochrane Effective Practice and Organisation of Care (EPOC) review group (see Table 2) [31, 32]. These criteria range from systematic reviews (A1) to descriptive, non-analytical studies (D).

Level	Description
A1	Systematic review
	Review of data from multiple RCT studies
A2	Randomised trial
	Comparative study with (random) intervention and control group design
B	Controlled trial
	Trial with intervention and control group and comparisons of outcomes
	B1 multiple measurement points
	B2 single measurement point
C	Non-controlled study
	C1 multiple case, multiple measurements points
	C2 multiple case, single measurement point
	C3 single case, multiple measurements point
	C4 single case, single measurement point
D	Descriptive, non-analytical
	D1 multiple projects
	D2 single project
	D3 literature review

Table 2. Levels of evidence quality

In addition, the Quality Improvement Minimum Quality Criteria Set (QI-MQCS) will be used to assess the completeness of the reporting of each study [33]. This tool includes 16 content domains to critically appraise QI intervention publications and determine whether a minimum quality standard has been met (see Table 3).

Domain	Minimum standard (see Supplementary Appendix 2 online in [33])
1. Organisational motivation	Names or describes at least one motivation for the organisation's participation in the intervention
2. Intervention rationale	Names or describes a rationale linking at least one central intervention component to intended effects
3. Intervention description	Describes in detail at least one specific change including the personnel executing the intervention
4. Organisational characteristics	Reports at least two organisational characteristics
5. Implementation	Names at least one approach used to introduce the intervention
6. Study design	Names the study design
7. Comparator	Describes at least one key care process
8. Data source	Describes the data source and defines the outcome of interest
9. Timing	Describes the timing of the intervention and its evaluation to determine the presence of baseline data and the follow-up period after all intervention components have been fully implemented
10. Adherence/fidelity	Reports fidelity information for at least one intervention component, or describes evidence of adherence or of a mechanism ensuring compliance to the intervention
11. Health outcomes	Reports data on at least one health-related outcome
12. Organisational readiness	Reports at least one organisational-level barrier or facilitator
13. Penetration/reach	Describes the proportion of all eligible units that actually participated
14. Sustainability	Describes the sustainability or the potential for sustainability
15. Spread	Describes the potential for spread, existing tools for spread, or spread attempts / largescale rollout
16. Limitations	Reports at least one limitation of the design / evaluation

Table 3. Quality Improvement Minimum Quality Criteria Set (QI-MQCS) domains [33]

Two reviewers (LH and either GW, KA or RG) will independently undertake the data extraction. Following this, the review team will discuss the data extracted so that data are not simply categorised but are used to begin to develop a reasoning that provides input to the final synthesis phase. As the aim of the data extraction process is to evaluate and refine the initial programme theory, the contents of the data extraction form will be developed by the review team based on the content of the initial programme theory. To test the usability of the data extraction form, the tool will be pre-tested on two purposefully selected articles [34]. From each study, general characteristics will be extracted concerning the study setting, the study design (e.g. level of evidence) and the unit of analysis (including type of organisation). Furthermore, relevant sections of the articles, i.e. relating to context, mechanisms and their relationship to the produced outcomes, will be coded. This coding will be both inductive (codes emerge and are created during the data extraction) and deductive (codes created in advance of data extraction and informed by the initial programme theory).

4. Synthesise evidence and draw conclusions

Evidence will be synthesised by examining the relationships between contexts (e.g. organisational culture), mechanisms (e.g. bottom-up initiative) and outcomes (i.e. intended and unintended consequences and the impact of audits) to determine what works, in what circumstances, how and why. Rycroft-Malone and colleagues [34] have developed a five-step approach for a realist synthesis, incorporating the work of Pawson [16], as follows:

1. Organise the extracted data into evidence tables: The data extraction form for each individual study will be summarised and organised into one or more evidence tables. The evidence tables will also include a link back to the source papers.

2. Theme the data: Themes will be developed from the initial codes based on recurring contexts, mechanisms or outcomes. Identified themes will then be discussed among the reviewers, and contrary evidence will be sought.

3. Formulate chains of inference from the identified themes: Through an iterative process, we will search for chains of inference (connections) across extracted data and themes.

For example, the 'leadership and competency' chain of inference might incorporate multiple themes including, for example, active engagement, competencies in quality improvement, strong legitimacy within the organisation, and a sound knowledge of quality issues. In order to support and formulate such chains of inference, patterns of similar mechanisms will be sought across different contexts to see if emerging patterns of outcomes ('demi-regularities') are identified. Cases where the contexts are restrictive, rather than supportive, will be identified and this will help in formulating the chains of inference. Two reviewers (LH and either GW, KA or RG) will jointly formulate the chains of inference, and this information will be shared and discussed in the review team.

4. Link the chains of inference and formulate CMO configurations: The chains of inference will be linked together to develop CMO configurations, which will then be linked back to themes or theories emerging from the literature (e.g. Commitment, Organisational Culture). The CMO configurations will be confirmed by returning to the source evidence. This iterative process will be guided by the research questions and the aims of the review. Following this, the generated CMO configurations will be used to either form new programme theories or to test, refine and supplement the initial programme theory. All these processes will be performed through discussions and agreement within the review team.

5. Refine the initial programme theory: Following the above four steps, a cumulative picture will be developed around the programme theories that summarises the nature of the context, mechanism and outcome, and links to the characteristics of the individual studies included. This cumulative picture will be based on hypotheses. For example, our review may suggest that hospitals that have a supportive culture for quality improvement (context) and that seek the active participation of healthcare professionals in audits (mechanism) generate improved safety as part of the quality of care (outcome). A narrative will be developed around each hypothesis that will describe the characteristics of the supportive evidence.

Pawson et al. (2005) argue that stakeholders should be involved in both the process of confirming the emerging findings and in dissemination activities [16]. To that end, emerging findings, supporting evidence and CMO configurations will be shared and discussed during a focus group session involving researchers, managers, policymakers and clinicians. The focus group will have 10 – 12 participants who will be selected to ensure some degree of homogeneity since this will enable them to share and discuss ideas by having comparable relevant knowledge in the field of audits [35]. This process will help to refine the focus and the presentation of the narrative stemming from the CMO configurations.

ETHICS AND DISSEMINATION

Under the Dutch Law on Medical Research Involving Human Subjects (WMO), this review does not require formal ethical approval. One of the key contributions of this review, compared to the majority of audit evaluations and systematic reviews, is that it focuses on how and why audits might work, rather than just on the impact of audits. To really understand how and why audits might work, or might not, we believe that a clear picture of the underlying processes that lead to the outcomes is essential. By providing this, this review will extend the current literature by providing knowledge on how, and why, audits may lead to sustainable quality improvements.

This review has important practical implications. Along with the increasing emphasis on patient safety and healthcare quality, controlling rising healthcare costs has become a top policy priority in many countries. Research programmes, such as the review proposed here, can provide a basis for identifying appropriate strategies for quality improvements in healthcare. A better understanding of how these audits 'work', and how context might impact on the intended outcome of improved healthcare quality, will inform stakeholders in their decision-making about how to tailor and implement audits within their local context.

We will use various dissemination strategies to ensure that findings from this realist review are broadly disseminated to academic and non-academic audiences. First, we will submit the findings of this realist review to a peer-reviewed journal. In addition, review results will be disseminated through public websites, publications in professional journals and by presenting our work at relevant national and international conferences, and at conferences for practitioners. As part of a more active dissemination strategy, we also intend a follow-up meeting with the focus group participants to discuss the findings and key messages.

Acknowledgement All persons contributing to the manuscript meet the criteria for authorship. There was no external funding for this paper.

Competing interests None.

Funding This research has not received a specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Contributors LH was responsible for project conception, protocol development and the writing and submission of the manuscript. KA, GW and RG were responsible for protocol development and editing the manuscript. All authors have given final approval of the version to be published.

REFERENCES

1 McDonald KM, Chang C, Schultz E. Closing the Quality Gap: Revisiting the State of the Science. 2013;AHRQ Publication No. 12(13)-E017.

2 Shojania KG, McDonald KM, Wachter RM, et al. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol. 1: Series Overview and Methodology). Rockville (MD): UCSF Evidence-based Practice Center 2004.

3 Spencer E, Walshe K. National quality improvement policies and strategies in European healthcare systems, *Qual Saf Health Care* 2009;18 Suppl 1:i22-7.

4 Bohigas L, Heaton C. Methods for external evaluation of health care institutions, *Int J Qual Health Care* 2000;12:231-8.

5 Walshe K, Freeman T, Latham L, et al. Chapter 6. The development of external reviews of clinical governance. In: Anonymous . Clinical governance - from policy to practice. Birmingham, UK: University of Birmingham, Health Services Management Centre 2000.

6 Scrivener R, Morrell C, Baker R, et al. Principles for best practice in clinical audit. Abingdon, UK: Radcliffe Medical Press Ltd 2002.

7 Dixon N. Getting clinical audit right to benefit patients. Romsey, England: Healthcare Quality Quest 2014.

8 Brubakk K, Vist GE, Bukholm G, et al. A systematic review of hospital accreditation: the challenges of measuring complex intervention effects, *BMC Health Serv Res* 2015;15:280,015-0933-x.

9 Flodgren G, Pomey MP, Taber SA, et al. Effectiveness of external inspection of compliance with standards in improving healthcare organisation behaviour, healthcare professional behaviour or patient outcomes, *Cochrane Database Syst Rev* 2011;(11):CD008992. doi:CD008992.

10 Greenfield D, Braithwaite J. Health sector accreditation research: a systematic review, *Int J Qual Health Care* 2008;20:172-83.

11 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.

12 Ovretveit J. Understanding the conditions for improvement: research to discover which context influences affect improvement success, *BMJ Qual Saf* 2011;20 Suppl 1:i18-23.

13 Walshe K, Freeman T. Effectiveness of quality improvement: learning from evaluations, *Qual Saf Health Care* 2002;11:85-7.

14 Shepperd S, Lewin S, Straus S, et al. Can we systematically review studies that evaluate complex interventions? *PLoS Med* 2009;6:e1000086.

15 Walshe K. Understanding what works--and why--in quality improvement: the need for theory-driven evaluation, *Int J Qual Health Care* 2007;19:57-9.

16 Pawson R, Greenhalgh T, Harvey G, et al. Realist review--a new method of systematic review designed for complex policy interventions, *J Health Serv Res Policy* 2005;10 Suppl 1:21-34.

17 Pawson R. Evidence-based policy: a realist perspective. London: Sage 2006.

18 Braithwaite J, Shaw CD, Moldovan M, et al. Comparison of health service accreditation programs in low- and middle-income countries with those in higher income countries: a cross-sectional study, *Int J Qual Health Care* 2012;24:568-77.

19 Mazzocato P, Savage C, Brommels M, et al. Lean thinking in healthcare: a realist review of the literature, *Qual Saf Health Care* 2010;19:376-82.

20 Denyer D, Tranfield D, van Aken J. Developing Design Propositions through Research Synthesis, *Organ Stud* 2008;29:393-413.

21 Wong G, Greenhalgh T, Westhorp G, et al. RAMESES publication standards: realist syntheses, *BMC Med* 2013;11:21,7015-11-21.

22 Kaplan HC, Brady PW, Dritz MC, et al. The influence of context on quality improvement success in health care: a systematic review of the literature, *Milbank Q* 2010;88:500-59.

23 Johnston G, Crombie IK, Davies HT, et al. Reviewing audit: barriers and facilitating factors for effective clinical audit, *Qual Health Care* 2000;9:23-36.

24 Power D, Terziovski M. Quality audit roles and skills: Perceptions of non-financial auditors and their clients, *J Oper Manage* 2007;25:126-147.

25 Greenhalgh T, Robert G, Macfarlane F, et al. Diffusion of innovations in service organizations: systematic review and recommendations, *Milbank Q* 2004;82:581-629.

26 Spurgeon P, Mazelan PM, Barwell F. Medical engagement: a crucial underpinning to organizational performance, *Health Serv Manage Res* 2011;24:114-20.

27 Booth A, Harris J, Croot E, et al. Towards a methodology for cluster searching to provide conceptual and contextual "richness" for systematic reviews of complex interventions: case study (CLUSTER). *BMC medical research methodology* 2013;13:118.

28 Lombarts MJMH, Klazinga NS. Inside self-regulation: peer review (visitatie) by Dutch medical specialists, *Clinical Governance: An International Journal* 2003;8:318-330.

29 Pawson R. Digging for Nuggets: How `Bad' Research Can Yield `Good' Evidence, *International Journal of Social Research Methodology* 2006;9:127-142.

30 Williams L, Rycroft-Malone J, Burton CR, et al. Improving skills and care standards in the support workforce for older people: a realist synthesis of workforce development interventions, *BMJ Open* 2016;6:e011964,2016-011964.

31 Cochrane Effective Practice and Organisation of Care (EPOC) Group. Available from: http://epoc.cochrane.org [Accessed 3 October 2016]

32 Everdingen JJEv. Evidence-based richtlijnontwikkeling : een leidraad voor de praktijk. Houten : Bohn Stafleu Van Loghum, 2004:XXXI, 395 p. : ill. ; 24 cm.

33 Hempel S, Shekelle PG, Liu JL, et al. Development of the Quality Improvement Minimum Quality Criteria Set (QI-MQCS): a tool for critical appraisal of quality improvement intervention publications, *BMJ Qual Saf* 2015;24:796-804.

34 Rycroft-Malone J, McCormack B, Hutchinson AM, et al. Realist synthesis: illustrating the method for implementation research, *Implementation Science* 2012;7:33.

35 Fern EF. Advanced focus group research. Thousand Oaks: Canada: Sage publications 2001.

Supplementary File 1: Definition of terms

Context – Context often refers to the 'setting' of programmes and interventions. The literature suggests that differences should be contextualised by considering four levels of the context: (1) the external level (e.g. the wider social, economic or cultural setting); (2) the institutional level (e.g. the organisational culture, local priorities); (3) the interpersonal level (e.g. communication and collaboration); and (4) the individual level (e.g. personal values or knowledge) [1, 2]. Contextual elements can be expected to influence the relationship between audits and their outcomes and, in some cases, the outcomes of audits will influence the context (for example, a culture change may be generated by the outcomes of an audit). Some contextual elements may be essential for the outcome to occur and, because of this, may be confused with mechanisms [3, 4]. To resolve this, this research considers contextual elements as factors that can influence an outcome but are external to the intervention [3].

Mechanism – Mechanisms have been defined as '…underlying entities, processes, or [social] structures which operate in particular contexts to generate outcomes of interest' [5, p.2]. Identifying the mechanisms will advance the synthesis beyond describing 'what happened' to theorizing on 'why' it happened and 'under what circumstances'.

Outcome – Outcomes can be either intended or unintended, can be proximal, intermediate or final, and result from the activation of different mechanisms in different contexts.

Outcome patterns – Also described as 'demi-regularities' in the realist literature [2, 6], these amount to semi-predictable patterns of outcomes. First, 'semi' because variations in patterns of behaviour can only be partly attributed to contextual differences and, second, because individuals will likely, but not always, make similar choices about the resources they will use.

Circumstances – The phrase 'in what circumstances' is interpreted, in realist terms, as meaning 'in what contexts and by what mechanisms'. One has to examine the key contextual conditions that affect the mechanisms, identity in what way those conditions affect the mechanisms, and describe how the interaction between context and mechanisms affects the outcomes.

Context-Mechanism-Outcome (CMO) configurations – The resulting explanations for the observed outcome patterns are formulated as CMO configurations. A sample CMO configuration is as follows: a hospital with a supportive culture for quality improvement implements an audit (context). Subsequently, improvements in care quality are noted (outcome). The reason for this is the active participation of healthcare professionals in the audit process (mechanism).

Programme theory – Programme theory refers to an abstracted description and/or diagram that explains what a programme or intervention comprises of, and how and why it is expected to work. Programme theories are usually described as 'middle-range', meaning that they are 'specific enough to generate propositions that can be tested about aspects of the program but sufficiently abstract to be applicable to other programs' [5].

REFERENCES

1 Macfarlane F, Greenhalgh T, Humphrey C, et al. A new workforce in the making?: A case study of strategic human resource management in a whole-system change effort in healthcare, *Journal of Health Organization and Management* 2011;25:55-72.

2 Pawson R, Greenhalgh T, Harvey G, et al. Realist review--a new method of systematic review designed for complex policy interventions, *J Health Serv Res Policy* 2005;10 Suppl 1:21-34.

3 Marchal B., van Olmen J., Hoeree T., et al. Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research, *Evaluation* 2012;18:192-212.

4 Astbury B, Leeuw F. Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation, *American Journal of Evaluation* 2010;31:363-381.

5 Wong G, Greenhalgh T, Westhorp G, et al. RAMESES publication standards: realist syntheses, *BMC Med* 2013;11:21,7015-11-21.

6 Pawson R. Evidence-based policy: a realist perspective. London: Sage 2006.

Supplementary File 2: Search strategy MEDLINE

MEDLINE (Pubmed)
("Clinical Audit"[Majr:noexp] OR "Medical Audit"[Majr] OR "Nursing Audit"[Majr] OR "Accreditation"[Majr] OR "Certification"[Majr:noexp] OR "Peer Review, Health Care"[Majr] OR ((extern*[tiab] OR internal*[tiab]) AND audit[tiab]) OR medical audit*[tiab] OR clinical audit*[tiab] OR nursing audit*[tiab] OR audit[ti] OR audits[ti] OR accreditat*[ti] OR visitation*[ti]) AND ("Academic Medical Centers"[Mesh:noexp] OR "Hospitals, Teaching"[Mesh] OR "Outpatient Clinics, Hospital"[Mesh:noexp] OR "Hospitals"[Mesh:noexp] OR hospital*[tiab] OR ((health*[tiab] OR clinical[tiab]) AND (organisation*[tiab] OR organization*[tiab] OR center*[tiab] OR centre*[tiab])) OR health sector*[tiab] OR healthcare sector*[tiab] OR health care sector*[tiab]) AND ("Efficiency, Organizational"[Mesh] OR efficien*[tiab] OR effectiveness*[tiab] OR performan*[tiab] OR improvement*[tiab] OR "Quality Improvement"[Mesh:noexp] OR "Quality Assurance, Health Care"[Majr:noexp] OR quality improv*[ti] OR quality assur*[ti]) NOT ("Animals"[Mesh] NOT "Humans"[Mesh]) AND english[la] AND ("last 10 years"[PDat])

RAMESES publication standards: List of items to be included when reporting a realist synthesis

Reporting item	Description of item	Reported on page #
TITLE		
1	In the title, identify the document as a realist synthesis or review	1
ABSTRACT		
2	While acknowledging publication requirements and house style, abstracts should ideally contain brief details of: the study's background, review question or objectives; search strategy; methods of selection, appraisal, analysis and synthesis of sources; main results; and implications for practice.	2
INTRODUCTION		
3 Rationale for review	Explain why the review is needed and what it is likely to contribute to existing understanding of the topic area.	4-6
4 Objectives and focus of review	State the objective(s) of the review and/or the review question(s). Define and provide a rationale for the focus of the review.	6, 7
METHODS		
5 Changes in the review process	Any changes made to the review process that was initially planned should be briefly described and justified.	10
6 Rationale for using realist synthesis	Explain why realist synthesis was considered the most appropriate method to use.	6, 7
7 Scoping the literature	Describe and justify the initial process of exploratory scoping of the literature.	8, 9
8 Searching processes	While considering specific requirements of the journal or other publication outlet, state and provide a rationale for how the iterative searching was done. Provide details on all the sources accessed for information in the review. Where searching in electronic databases has taken place, the details should include, for example, name of database, search terms, dates of coverage and date last searched. If individuals familiar with the relevant literature and/or topic area were contacted, indicate how they were identified and selected.	9, 10, Supp 2

Reporting item	Description of item	Reported on page #
9 Selection and appraisal of documents	Explain how judgements were made about including and excluding data from documents, and justify these.	10-14
10 Data extraction	Describe and explain which data or information were extracted from the included documents and justify this selection.	15
11 Analysis and synthesis processes	Describe the analysis and synthesis processes in detail. This section should include information on the constructs analyzed and describe the analytic process	15-17
RESULTS		
12 Document flow diagram	Provide details on the number of documents assessed for eligibility and included in the review with reasons for exclusion at each stage as well as an indication of their source of origin (for example, from searching databases, reference lists and so on). You may consider using the example templates (which are likely to need modification to suit the data) that are provided.	n/a
13 Document characteristics	Provide information on the characteristics of the documents included in the review.	n/a
14 Main findings	Present the key findings with a specific focus on theory building and testing.	n/a
DISCUSSION		
15 Summary of findings	Summarize the main findings, taking into account the review's objective(s), research question(s), focus and intended audience(s).	n/a
16 Strengths, limitations and future research directions	Discuss both the strengths of the review and its limitations. These should include (but need not be restricted to) (a) consideration of all the steps in the review process and (b) comment on the overall strength of evidence supporting the explanatory insights which emerged. The limitations identified may point to areas where further work is needed.	3, 17
17 Comparison with existing literature	Where applicable, compare and contrast the review's findings with the existing literature (for example, other reviews) on the same topic.	n/a
18 Conclusion and recommendations	List the main implications of the findings and place these in the context of other relevant literature. If appropriate, offer recommendations for policy and practice.	n/a
19 Funding	Provide details of funding source (if any) for the review, the role played by the funder (if any) and any conflicts of interests of the reviewers.	18

From: Wong G, Greenhalgh T, Westhorp G, et al. RAMESES publication standards: realist syntheses, BMC Med 2013;11:21,7015-11-21.

BMJ Open

Understanding how and why audits work: protocol for a realist review of audit programmes to improve hospital care

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-015121.R1
Article Type:	Protocol
Date Submitted by the Author:	18-Jan-2017
Complete List of Authors:	Hut-Mossel, Lisanne; University of Groningen, University Medical Centre Groningen, Centre of expertise on Quality and Safety Welker , Gera ; University of Groningen, University Medical Centre Groningen, Centre of expertise on Quality and Safety Ahaus, Kees; University of Groningen, University Medical Centre Groningen, Centre of Expertise on Quality and Safety; University of Groningen, Faculty of Economics and Business, Department Operations Gans, Rob; University of Groningen, University Medical Centre Groningen, Department of Internal Medicine
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Research methods, Health policy
Keywords:	Realist synthesis, Realist review, AUDIT, Quality improvement, Clinical audit < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Hospital Care

Title

Understanding how and why audits work: protocol for a realist review of audit programmes to improve hospital care

Authors

Lisanne Hut-Mossel[1], Gera Welker[1], Kees Ahaus[1, 2] and Rijk Gans[3].

University of Groningen, University Medical Centre Groningen.

Author Note

1 University of Groningen, University Medical Centre Groningen, Centre of Expertise on Quality and Safety, Groningen, the Netherlands.

2 University of Groningen, Faculty of Economics and Business, Department Operations, Groningen, the Netherlands.

3 University of Groningen, University Medical Centre Groningen, Department of Internal Medicine, Groningen, the Netherlands.

Corresponding author

Lisanne Hut-Mossel

Huispostcode LA10

PO Box 30 001

9700 RB Groningen

The Netherlands.

Tel: +31623473381

Correspondence to p.a.mossel@umcg.nl

Word Count (excluding title page, abstract, references, figures and tables): 4041

ABSTRACT

Introduction Many types of audits are commonly used in hospital care to promote quality improvements. However, the evidence on the effectiveness of audits is mixed. The objectives of this proposed realist review are (i) to understand how and why audits might, or might not, work in terms of delivering the intended outcome of improved quality of hospital care and (ii) to examine under what circumstances audits could potentially be effective. This protocol will provide the rationale for using a realist review approach and outline the method.

Methods and Analysis This review will be conducted using an iterative four-stage approach. The first and second steps have already been executed. The first step was to develop an initial programme theory based on the literature that explains how audits are supposed to work. Second, a systematic literature search was conducted using relevant databases. Third, data will be extracted and coded for concepts relating to context, outcomes and their interrelatedness. Finally, the data will be synthesised in a five-step process: (1) organising the extracted data into evidence tables, (2) theming, (3) formulating chains of inference from the identified themes, (4) linking the chains of inference and formulating CMO configurations and (5) refining the initial programme theory. The reporting of the review will follow the RAMESES publication standards.

Ethics and Dissemination This review does not require formal ethical approval. A better understanding of how and why these audits work, and how context impacts their effectiveness, will inform stakeholders in deciding how to tailor and implement audits within their local context. We will use a range of dissemination strategies to ensure that findings from this realist review are broadly disseminated to academic and non-academic audiences.

Trial registration number This systematic review protocol is registered on the PROSPERO database (registration number CRD42016039882).

Keywords Realist synthesis, Realist Review, Audit, Quality Improvement, Clinical Audit, Hospital Care

STRENGTHS AND LIMITATIONS OF THIS STUDY

- This review goes beyond considering the effectiveness of audits by building an understanding of how and why audits work within various contexts;
- This review uses a systematic screening protocol;
- The main limitation is that realist reviews are dependent upon the transparency and adequacy of the reporting of data on the context, the mechanisms and their relationship to the produced outcomes of individual studies by the original authors. The potential lack of adequate data in this regard might hamper developing a full understanding of how and why audits are effective and might restrict the full development of the programme theory.

INTRODUCTION

In recent decades, quality and safety issues have become increasingly important in hospital care because of their direct effect on both clinical outcomes and patient satisfaction. However, hospital care still suffers from a quality gap between the ideal care, in line with the best available medical evidence, and the actual care provided to patients [1]. To close this gap, health authorities and organisations currently prioritise quality improvement (QI) strategies, which are seen as systematic, data-driven monitoring and evaluation activities to improve the quality of hospital care [2]. A widely used QI strategy within hospitals is the audit.

In this review protocol, we will focus on audits that address quality. Such audits are commonly used within hospital care to promote quality improvements by monitoring, controlling and/or changing healthcare processes and healthcare providers' performance [3]. However, it is unlikely that audits work in the same way in every setting. Accordingly, it is important to understand how and why audits might lead to quality improvements. A realist review, as outlined in this protocol, will contribute to this understanding.

The range of possible audits can be roughly divided into (1) external audits, used to gain insight into hospitals' compliance with external criteria (e.g. accreditation, certification, external peer reviews), (2) internal audits, often in preparation for an external audit, and (3) clinical audits, carried out on a voluntary basis by healthcare professionals [3, 4]. Externally driven audits (i.e. accreditation, certification, external peer reviews and preparatory internal audits) seem to be more strongly anchored in Quality Assurance (QA), referring to initiatives designed to assure compliance with minimum quality standards [5, 6]. Clinical audits, on the other hand, represent a from QA to a Quality Improvement (QI) process, with a focus on seeking to improve care, and prevent poor care. This process takes place continuously as part of everyday routines [6-8]. Although there are

differences, such as in the scope and the approach used, between the various types of audits, they all serve the same objective: to improve the quality of hospital care.

External audits are used to assess certain dimensions or characteristics of a healthcare providing organisation against specified standards [9]. As such, the implementation of an external audit requires an external standard and collaboration from beyond the hospital - and this distinguishes them from internal audits.

Internal audits are conducted by internal auditors of the hospital's own organisation, such as quality officers or healthcare professionals from another department than the one being audited to guarantee some level of independent judgement. Internal audits vary in purpose. On the one hand, healthcare organisations use internal audits to continuously improve the quality of healthcare. In this way, one could expect that, compared with external audits, threats to quality can be more quickly revealed, allowing the organisation to regularly adapt its processes to improve quality at the local level. Internal audits are also frequently used in the framework of external audits and are conducted to avoid performance standards dropping between two external audits. These audits are designed to evaluate and improve the effectiveness of the organisation's quality management system and focus more on organisational conditions and less on the behaviour of healthcare professionals and patient outcomes [4].

Clinical audits differ from other types of audits in that they are mostly undertaken and initiated by healthcare professionals. As such, healthcare professionals work together to collect data and evaluate their own practices. Following this, they develop and apply improvements in their daily practices, and then the audit cycle is repeated to demonstrate improved and sustained improvements [7]. As such, clinical audits do not necessarily use external criteria and are not carried out in response to external demands as the initiative comes from the healthcare professionals themselves [10] . A considerable amount of literature addresses the effectiveness of audits and reports mixed results [11-13]. A systematic review on audit and feedback [14] demonstrated a positive overall effect of audits on clinical practice. Further, the authors noted differences in the

design and the effectiveness of audits. This variety can be attributed to at least two issues. First, audits are used to improve specific aspects of healthcare and can be targeted at different levels. For example, external audits are performed to induce changes at the organisational level (e.g. in organisational policy or procedures) whereas clinical audits are performed to alter local healthcare practices (e.g. clinical day-to-day practices or local guidelines). Second, audits are used in different contexts, and this considerably complicates the evaluation of their effects. For example, an audit could be effective in one organisation, or department, but not in another because of, for instance, the amount of support offered for quality improvements, as part of their differing contexts (see supplementary file 1). The literature on QI strategies recognises that the mixed effects are partly due to the differing contexts in which interventions are planned [15, 16].

The variety in the levels of audits, together with the heterogeneity of their contexts, suggests that it is unlikely that audits work in the same way in every setting. This creates challenges when attempting to synthesise evidence in a systematic review. Given this situation, more information about why and how audits work is needed [17, 18]. A detailed understanding of the contextual factors and the mechanisms that influence the effectiveness of audits is a prerequisite for understanding the mechanisms through which audits might lead to quality improvements. More importantly, a better understanding of 'how and why audits might work' will inform decision-making on how to tailor quality improvements at the local level.

A useful approach for explaining how and why audits might work, and investigating the interactions between context, mechanism and outcome, is the use of a realist review [19, 20]. The value of a realist review is that it is concerned with *how* an intervention works, rather than *whether* an intervention works, which is the focus of the conventional systematic review approach. Furthermore, the realist review methodology is specifically designed to cope with the intervention heterogeneity (in both the chosen study design and the used outcome measures) present in previous research on audits. Finally, this method is appropriate for the current research because

audits are complex context-sensitive interventions [11, 16, 21]. Within the past decade, similar studies in other contexts have used realist reviews to understand how complex interventions work and are put into practice (e.g. [22]).

The objectives of the current review are (i) to understand how and why audits might, or might not, work in producing the intended outcome of improved quality of care and (ii) to examine under what circumstances audits could potentially be effective by formulating and refining underlying programme theories. Consequently, this review focusses on three research questions:

1. What are the mechanisms through which audits deliver their intended outcomes?
2. What contextual factors determine whether the identified mechanisms produce the intended outcomes of audits?
3. In what circumstances (i.e. which combination(s) of mechanisms and context) are audits most likely to be effective?

METHOD

A realist review aims to clarify, from observed data, the outcomes (O) of particular interventions in relation to context (C) and mechanisms (M). This 'CMO' configuration is based on the philosophical assumption that an intervention in a specific context (C) evokes mechanisms (M) that generate an outcome (O). Consequently, the underlying mechanisms can be expected to produce a broad range of outcomes (O) when performed in different contexts (C) [19, 20, 23]. The philosophical basis is realism, which is positioned between positivism and constructivism and assumes the existence of an external reality (a 'real world') that is 'filtered' (i.e. perceived, interpreted and responded to) through the inputs of individuals. Consequently, it is not the intervention in and of itself that causes outcomes but the individuals who initiate a process of change and as such have an effect on whether and how the intervention works [20]. One of the key outputs of a realist review is the development

of programme theories that set out how and why an intervention is thought to 'work' to generate certain outcomes [24].

Study design

This review follows Pawson's steps for conducting realist reviews, namely: (1) clarifying the scope, and programme theory development; (2) searching for evidence; (3) appraising primary studies and extracting data; and (4) synthesising evidence and drawing conclusions [19]. The reporting of the review will follow the 'Realist and Meta-Review Evidence Synthesis: Evolving Standards' (RAMESES) publication standards [24]. In line with these standards, data extraction and synthesis will be an interpretive process, driven by reflection and discussion by the review team [24]. This process requires repeated reading of primary studies because, as the synthesis progresses, new or refined elements of theory are expected to emerge. The protocol outlined below was written after the first steps had already been initiated or completed. Accordingly, both the past tense (steps that have been completed) and the future tense (steps that have yet to be initiated) are used.

The review team represents a range of disciplines and professions, which enables us to consider multiple perspectives and insights on the data gathered within this realist review. LH has a nursing background and is a PhD candidate. GW is an implementation fellow and has several years of experience as a quality manager. KA has a background in economics and business, is a professor of healthcare management and has numerous publications related to quality and patient safety. RG is a medical specialist, professor of internal medicine, chair of the Dutch Training Program of Internal Medicine and President of the Dutch Society of Hospital Medicine. He is also involved in the training of hospitalists, who are conducting clinical audits as part of their training. Further, all the members of the review team are experienced in qualitative research.

1. Scope of the review and programme theory development

The first step of this review process has already been executed with the aim of building a programme theory that would explain how and why audits might work. The unit of analysis in a

realist review is not the intervention itself, but the contexts, mechanisms and outcomes that underpin the intervention. Given this situation, the initial step in formulating a programme theory draws on the literature on the effectiveness of QI strategies. As audits are QI strategies, we would assume that the contexts, mechanisms and outcomes uncovered in the QI literature might also play a role in the effectiveness of audits. The initial programme theory explains how audits are supposed to work by framing the interrelationships between context, mechanism and outcome (see Figure 1) [19]. An exploration of programme theories was initiated through on-going conversations within the review team and by a preliminary search of the literature. In addition, key terms were defined to guide the review and to ensure a common understanding (see Supplementary File 1).

After a number of iterations and discussions, we developed an initial programme theory regarding how and why audits might work. This suggests that having an organisational culture that is supportive of quality improvement, a leadership committed to quality and previous audit experiences are important contextual factors in the success of an audit [25]. These contextual factors trigger mechanisms, including a focus on continuous improvement rather than auditing for assurance and compliance [26, 27], bottom-up initiatives as a prerequisite for ownership [12, 28] and the active involvement of healthcare professionals in audit processes [14, 26, 29], that in turn lead to improvements in the quality of healthcare. The World Health Organisation (WHO) describes the quality of healthcare quality as follows: 'quality of care means that a health system should seek to make improvements in six areas or dimensions of quality' [30]. These dimensions are: effectiveness, efficiency, accessibility, acceptability, equity and safety.

This initial programme theory provides only a provisional structure for the review, and additional contextual factors, mechanisms and outcomes will be identified as the review progresses. The initial programme theory will be expanded, tested and refined using data from studies included in the review.

[Please insert Figure 1 here]

2. Search for evidence

As a second step, a search strategy was developed and performed in collaboration with an experienced university librarian. To ensure that all relevant articles were identified, a systematic literature search was conducted in MEDLINE, Embase, PsycINFO, Academic Search Premier, Business Source Premier, EmeraldInsight, Cochrane Library and Web of Science for the period 2005 - 2015. These databases were selected because they contain the core of quality and patient safety studies in the field of healthcare management as well as the biomedical view on quality of healthcare. The search strategy was piloted first in MEDLINE and later adapted for searching the other databases (see Supplementary File 2). The search included appropriate indexing terms (i.e. MeSH terms and keywords) on descriptors of audits (e.g. clinical audit, accreditation, certification, peer review, quality improvement, quality assurance), outcomes (e.g. efficiency, effectiveness, improvement), and hospital care (e.g. academic medical centres, health organisations). The reference lists of the uncovered review articles were studied to identify additional primary studies.

A realist synthesis approach to searching for evidence is iterative and evolves as the understanding of the subject matter deepens. Consequently, as the review progresses further, we will also search for unpublished and grey literature (e.g. websites, national guidelines, policy documents and information reported in specialist conferences) on the assumption that the literature on this topic may be diverse and dispersed. In addition, our expectation is that not all the included publications will adequately report on all aspects of an audit. We will therefore identify papers, and other research outputs, that relate to the same study by using 'cluster' searching [31]. For example, a search can be based on the members of a research team of an included article to identify all other refereed journal articles and related documents. Further, an additional iterative search may be necessary if it is determined that more data are required to refine a specific part of the programme theory, or if new prospective theories are identified during data extraction or synthesis.

3. Appraise primary studies and extract data

The selection of appropriate primary studies has already been executed. First, one reviewer (LH) identified and removed duplicates. Next, two reviewers (LH and GW) independently screened all titles and abstracts for suitability for inclusion. The focus was on empirical studies that evaluated the effects of audits in hospital settings within high-income countries, without restrictions on the type of study design. Only studies published in English were included to avoid misinterpretation of the content of an article due to language barriers (see Table 1).

Second, to ensure consistency of judgement, the full texts of a random ten percent of the articles were independently reviewed by LH and GW and retained if they were deemed relevant (i.e. the article could provide data on the context, mechanisms or outcomes of an audit). One reviewer (LH) reviewed the remaining 90% for their relevancy. In practice, a number of these articles required discussion or joint reading by two reviewers as it was sometimes difficult to decide between inclusion and rejection. Disagreements were recorded and discussed to ensure that decisions were made consistently. When disagreements remained, the matter was resolved through discussion involving the entire review team.

Inclusion criteria
Research on accreditation, certification, peer review/Dutch visitatie model[a] or local clinical audit
Hospital setting
High-income country
Published in English
English abstract available
Description of the medical or technical content
Description of the process of how the audit was conducted
Description of the impact of audit on medical and process outcomes

Table 1. Inclusion criteria.

[a] This is a system of peer review that is led and owned by doctors and designed to assess the quality of care provided by groups of hospital-based medical specialists. Practices are surveyed every 3-5 years by a group of peers [32].

The next section describes activities that have yet to be started. Quality appraisal and data extraction will be undertaken using pre-specified Excel spreadsheets (available on request from the first author). As the aim of the data extraction process is to evaluate and refine the initial programme theory, the contents of the data extraction sheets will be developed by the review team based on the content of the initial programme theory. To test the usability of the data extraction sheets, the file will be pretested on two purposefully selected articles [33]. For each study, the quality will be appraised and general characteristics extracted concerning the study's setting, the unit of analysis (including type of organisation) along with sections of the text that relate to context, mechanisms and their relationship to the produced outcomes.

Realist reviews amount to mixed-method reviews in that they incorporate both quantitative and qualitative studies, as well as grey literature. Consequently, different approaches are required to assess the quality of the included studies. Two reviewers (LH and either GW, KA or RG) will assess the quality of each included study. Any disagreement will be resolved through consensus-based group discussions within the review team. First, following realist synthesis principles, the evidence will be appraised using the concept of rigour [19, 20]. Rigour will be assessed by describing fidelity and nuggets (i.e. the potential match with the initial programme theory and valuable observations presented in primary studies), and trustworthiness (i.e. whether the methods used to generate the data are credible and trustworthy) [34, 35]. Second, to make the concept of trustworthiness more concrete, and to ensure transparency in decision-making, the quality of the evidence of each individual study will be presented in the form of an evidence-level table based on criteria established by the Cochrane Effective Practice and Organisation of Care (EPOC) review group (see Table 2) [36, 37]. These criteria range from systematic reviews (A1) to descriptive, non-analytical studies (D).

Level	Description
A1	Systematic review
	Review of data from multiple RCT studies
A2	Randomised trial
	Comparative study with (random) intervention and control group design
B	Controlled trial
	Trial with intervention and control group and comparisons of outcomes
	B1 multiple measurement points
	B2 single measurement point
C	Non-controlled study
	C1 multiple case, multiple measurements points
	C2 multiple case, single measurement point
	C3 single case, multiple measurements point
	C4 single case, single measurement point
D	Descriptive, non-analytical
	D1 multiple projects
	D2 single project
	D3 literature review

Table 2. Levels of evidence quality

In addition, the Quality Improvement Minimum Quality Criteria Set (QI-MQCS) will be used to assess the completeness of the reporting of each study [38]. This tool includes 16 content domains to critically appraise QI intervention publications and determine whether a minimum quality standard has been met (see Table 3).

Domain	**Minimum standard** (see Supplementary Appendix 2 online in [38])
1. Organisational motivation	Names or describes at least one motivation for the organisation's participation in the intervention
2. Intervention rationale	Names or describes a rationale linking at least one

Domain	Minimum standard (see Supplementary Appendix 2 online in [38])
	central intervention component to intended effects
3. Intervention description	Describes in detail at least one specific change including the personnel executing the intervention
4. Organisational characteristics	Reports at least two organisational characteristics
5. Implementation	Names at least one approach used to introduce the intervention
6. Study design	Names the study design
7. Comparator	Describes at least one key care process
8. Data source	Describes the data source and defines the outcome of interest
9. Timing	Describes the timing of the intervention and its evaluation to determine the presence of baseline data and the follow-up period after all intervention components have been fully implemented
10. Adherence/fidelity	Reports fidelity information for at least one intervention component, or describes evidence of adherence or of a mechanism ensuring compliance to the intervention
11. Health outcomes	Reports data on at least one health-related outcome
12. Organisational readiness	Reports at least one organisational-level barrier or facilitator
13. Penetration/reach	Describes the proportion of all eligible units that actually participated
14. Sustainability	Describes the sustainability or the potential for sustainability
15. Spread	Describes the potential for spread, existing tools for spread, or spread attempts / largescale rollout
16. Limitations	Reports at least one limitation of the design / evaluation

Table 3. Quality Improvement Minimum Quality Criteria Set (QI-MQCS) domains [38]

Two reviewers (LH and either GW, KA or RG) will independently undertake the data extraction and, in this way, data from all the included articles will be extracted by two reviewers. Following this, the review team will discuss the data extracted so that data are not simply categorised but are used to begin to develop a reasoning that provides input to the final synthesis phase. Furthermore, relevant sections of the articles, i.e. relating to context, mechanisms and their relationship to the produced outcomes, will be coded. This coding will be both inductive (codes emerge and are created during the data extraction) and deductive (codes created in advance of data extraction and informed by the initial programme theory).

4. Synthesise evidence and draw conclusions

Evidence will be synthesised by examining the relationships between contexts (e.g. organisational culture), mechanisms (e.g. bottom-up initiative) and outcomes (i.e. intended and unintended consequences and the impact of audits) to determine what works, in what circumstances, how and why. Rycroft-Malone and colleagues [33] have developed a five-step approach for a realist synthesis, incorporating the work of Pawson [19], as follows:

1. Organise the extracted data into evidence tables: The data extraction sheets from each individual study will be summarised and organised into one or more evidence tables. The evidence tables will also include a link back to the source papers.

2. Theme the data: Themes will be developed from the initial codes based on recurring contexts, mechanisms or outcomes. Identified themes will then be discussed among the reviewers, and contrary evidence will be sought.

3. Formulate chains of inference from the identified themes: Through an iterative process, we will search for chains of inference (connections) across extracted data and themes. For example, the 'leadership and competency' chain of inference might incorporate multiple themes including, for example, active engagement, competencies in quality improvement, strong legitimacy within the organisation, and a sound knowledge of quality issues. First, in order to support and formulate such chains of inference, patterns of similar mechanisms will be sought across different contexts to see if

emerging patterns of outcomes ('demi-regularities') are identified. Second, since we expect context-outcome regularities to be easier to identify than mechanisms, because mechanisms are underlying and hence often unobservable or 'hidden', context-outcome regularities will be used as a basis for uncovering mechanisms [20, 39]. Cases in which the contexts are restrictive or supportive will be identified and this will help in formulating the chains of inference and in recognising and explaining interactions between context, mechanisms and outcomes. Third, we will not overlook the possibility that there may be more than one mechanism in play at the same time. The chains of inference so formulated will function as a basis for the CMO configurations to be developed. Two reviewers (LH and either GW, KA or RG) will jointly formulate the chains of inference, and this information will be shared and discussed in the review team.

4. Link the chains of inference and formulate CMO configurations: The chains of inference will be linked together to develop CMO configurations, which will then be linked back to themes or theories emerging from the literature (e.g. commitment, organisational culture). The CMO configurations will be confirmed by returning to the source evidence. This iterative process will be guided by the research questions and the aims of the review. Following this, the generated CMO configurations will be used to either form new programme theories or to test, refine and supplement the initial programme theory. All these processes will be performed through discussions and agreement within the review team.

5. Refine the initial programme theory: Following the above four steps, a cumulative picture will be developed around the programme theories that summarises the nature of the context, mechanism and outcome, and links to the characteristics of the individual studies included. This cumulative picture will be based on hypotheses. For example, our review may suggest that hospitals that have a supportive culture for quality improvement (context) and that seek the active participation of healthcare professionals in audits (mechanism) generate improved safety as part of the quality of care (outcome). A narrative will be developed around each hypothesis that will describe the characteristics of the supportive evidence.

Pawson et al. (2005) argue that stakeholders should be involved in both the process of confirming the emerging findings and in dissemination activities [19]. To that end, emerging findings, supporting evidence and CMO configurations will be shared and discussed during a focus group session involving researchers, managers, policymakers and clinicians. The focus group will have 10 – 12 participants who will be selected to ensure some degree of homogeneity since this will enable them to share and discuss ideas by having comparable relevant knowledge in the field of audits [40]. This process will help to refine the focus and the presentation of the narrative stemming from the CMO configurations.

ETHICS AND DISSEMINATION

Under the Dutch Law on Medical Research Involving Human Subjects (WMO), this review does not require formal ethical approval. One of the key contributions of this review, compared to the majority of audit evaluations and systematic reviews, is that it focuses on how and why audits might work, rather than just on the impact of audits. To really understand how and why audits might work, or might not, we believe that a clear picture of the underlying processes that lead to the outcomes is essential. By providing this, this review will extend the current literature by providing knowledge on how, and why, audits may lead to sustainable quality improvements.

This review has important practical implications. Along with the increasing emphasis on patient safety and healthcare quality, controlling rising healthcare costs has become a top policy priority in many countries. Research programmes, such as the review proposed here, can provide a basis for identifying appropriate strategies for quality improvements in healthcare. A better understanding of how these audits 'work', and how context might impact on the intended outcome of improved healthcare quality, will inform stakeholders in their decision-making about how to tailor and implement audits within their local context.

It has been argued that the theory and emerging evidence about how best to design audits (and what should be avoided) should be incorporated in the development and reporting of audits [12, 41]. However, such theoretical underpinnings are rarely reported in articles about audits, and this might hamper a full understanding of how and why audits are effective, and further impose restrictions on the ability to fully develop the programme theory and the applicability of the programme theory.

We will use various dissemination strategies to ensure that findings from this realist review are broadly disseminated to academic and non-academic audiences. First, we will submit the findings of this realist review to a peer-reviewed journal. In addition, review results will be disseminated through public websites, publications in professional journals and by presenting our work at relevant national and international conferences, and at conferences for practitioners. The outcomes of this realist review will be disseminated through events organised by The Netherlands Federation of University Medical Centres (Nederlandse Federatie van Universitair Medische Centra) (NFU) and at a national symposium for hospitalists who conduct clinical audits as part of their training. As part of a more active dissemination strategy, we also intend a follow-up meeting with the focus group participants to discuss the findings and key messages.

Acknowledgement All persons contributing to the manuscript meet the criteria for authorship. There was no external funding for this paper.

Competing interests None.

Funding This research has not received a specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Contributors LH was responsible for project conception, protocol development and the writing and submission of the manuscript. KA, GW and RG were responsible for protocol development and editing the manuscript. All authors have given final approval of the version to be published.

REFERENCES

1 McDonald KM, Chang C, Schultz E. Closing the Quality Gap: Revisiting the State of the Science. 2013;AHRQ Publication No. 12(13)-E017.

2 Shojania KG, McDonald KM, Wachter RM, et al. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol. 1: Series Overview and Methodology). Rockville (MD): UCSF Evidence-based Practice Center 2004.

3 Spencer E, Walshe K. National quality improvement policies and strategies in European healthcare systems, *Qual Saf Health Care* 2009;18 Suppl 1:i22-7.

4 Bohigas L, Heaton C. Methods for external evaluation of health care institutions, *Int J Qual Health Care* 2000;12:231-8.

5 Klazinga N. Re-engineering trust: the adoption and adaption of four models for external quality assurance of health care services in western European health care systems, *Int J Qual Health Care* 2000;12:183-9.

6 ISO. ISO 9000:2015. Quality management systems - Fundamentals and vocabulary. Geneva: International Organization for Standardization 2015.

7 Dixon N. Getting clinical audit right to benefit patients. Romsey, England: Healthcare Quality Quest 2014.

8 Balding C. From quality assurance to clinical governance, *Aust Health Rev* 2008;32:383-91.

9 Walshe K, Freeman T, Latham L, et al. Chapter 6. The development of external reviews of clinical governance. In: Anonymous . Clinical governance - from policy to practice. Birmingham, UK: University of Birmingham, Health Services Management Centre 2000.

10 Scrivener R, Morrell C, Baker R, et al. Principles for best practice in clinical audit. Abingdon, UK: Radcliffe Medical Press Ltd 2002.

11 Brubakk K, Vist GE, Bukholm G, et al. A systematic review of hospital accreditation: the challenges of measuring complex intervention effects, *BMC Health Serv Res* 2015;15:280,015-0933-x.

12 Flodgren G, Pomey MP, Taber SA, et al. Effectiveness of external inspection of compliance with standards in improving healthcare organisation behaviour, healthcare professional behaviour or patient outcomes, *Cochrane Database Syst Rev* 2011;(11):CD008992. doi:CD008992.

13 Greenfield D, Braithwaite J. Health sector accreditation research: a systematic review, *Int J Qual Health Care* 2008;20:172-83.

14 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.

15 Ovretveit J. Understanding the conditions for improvement: research to discover which context influences affect improvement success, *BMJ Qual Saf* 2011;20 Suppl 1:i18-23.

16 Walshe K, Freeman T. Effectiveness of quality improvement: learning from evaluations, *Qual Saf Health Care* 2002;11:85-7.

17 Shepperd S, Lewin S, Straus S, et al. Can we systematically review studies that evaluate complex interventions? *PLoS Med* 2009;6:e1000086.

18 Walshe K. Understanding what works--and why--in quality improvement: the need for theory-driven evaluation, *Int J Qual Health Care* 2007;19:57-9.

19 Pawson R, Greenhalgh T, Harvey G, et al. Realist review--a new method of systematic review designed for complex policy interventions, *J Health Serv Res Policy* 2005;10 Suppl 1:21-34.

20 Pawson R. Evidence-based policy: a realist perspective. London: Sage 2006.

21 Braithwaite J, Shaw CD, Moldovan M, et al. Comparison of health service accreditation programs in low- and middle-income countries with those in higher income countries: a cross-sectional study, *Int J Qual Health Care* 2012;24:568-77.

22 Mazzocato P, Savage C, Brommels M, et al. Lean thinking in healthcare: a realist review of the literature, *Qual Saf Health Care* 2010;19:376-82.

23 Denyer D, Tranfield D, van Aken J. Developing Design Propositions through Research Synthesis, *Organ Stud* 2008;29:393-413.

24 Wong G, Greenhalgh T, Westhorp G, et al. RAMESES publication standards: realist syntheses, *BMC Med* 2013;11:21,7015-11-21.

25 Kaplan HC, Brady PW, Dritz MC, et al. The influence of context on quality improvement success in health care: a systematic review of the literature, *Milbank Q* 2010;88:500-59.

26 Johnston G, Crombie IK, Davies HT, et al. Reviewing audit: barriers and facilitating factors for effective clinical audit, *Qual Health Care* 2000;9:23-36.

27 Power D, Terziovski M. Quality audit roles and skills: Perceptions of non-financial auditors and their clients, *J Oper Manage* 2007;25:126-147.

28 Greenhalgh T, Robert G, Macfarlane F, et al. Diffusion of innovations in service organizations: systematic review and recommendations, *Milbank Q* 2004;82:581-629.

29 Spurgeon P, Mazelan PM, Barwell F. Medical engagement: a crucial underpinning to organizational performance, *Health Serv Manage Res* 2011;24:114-20.

30 World Health Organization. Quality of care - A process for making strategic choices in health systems. 2006.

31 Booth A, Harris J, Croot E, et al. Towards a methodology for cluster searching to provide conceptual and contextual "richness" for systematic reviews of complex interventions: case study (CLUSTER). *BMC medical research methodology* 2013;13:118.

32 Lombarts MJMH, Klazinga NS. Inside self-regulation: peer review (visitatie) by Dutch medical specialists, *Clinical Governance: An International Journal* 2003;8:318-330.

33 Rycroft-Malone J, McCormack B, Hutchinson AM, et al. Realist synthesis: illustrating the method for implementation research, *Implementation Science* 2012;7:33.

34 Pawson R. Digging for Nuggets: How `Bad' Research Can Yield `Good' Evidence, *International Journal of Social Research Methodology* 2006;9:127-142.

35 Williams L, Rycroft-Malone J, Burton CR, et al. Improving skills and care standards in the support workforce for older people: a realist synthesis of workforce development interventions, *BMJ Open* 2016;6:e011964,2016-011964.

36 Cochrane Effective Practice and Organisation of Care (EPOC) Group 2013;2016.

37 Everdingen JJEv. Evidence-based richtlijnontwikkeling : een leidraad voor de praktijk. Houten : Bohn Stafleu Van Loghum, 2004:XXXI, 395 p. : ill. ; 24 cm.

38 Hempel S, Shekelle PG, Liu JL, et al. Development of the Quality Improvement Minimum Quality Criteria Set (QI-MQCS): a tool for critical appraisal of quality improvement intervention publications, *BMJ Qual Saf* 2015;24:796-804.

39 Astbury B, Leeuw F. Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation, *American Journal of Evaluation* 2010;31:363-381.

40 Fern EF. Advanced focus group research. Thousand Oaks: Canada: Sage publications 2001.

41 Ivers NM, Sales A, Colquhoun H, et al. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention, *Implement Sci* 2014;9:14,5908-9-14.

Figure legends

Figure 1. Initial programme theory for the effectiveness of an audit

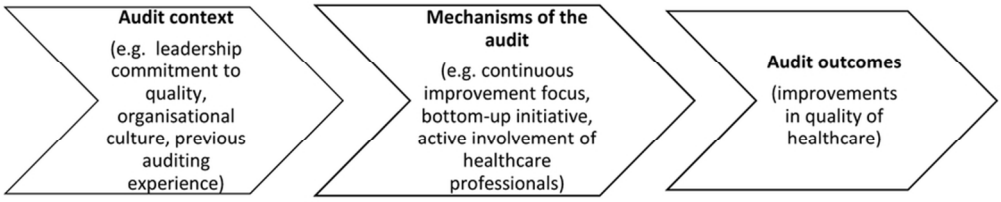

Figure 1. Initial programme theory for the effectiveness of an audit
[Please insert Figure 1 here]
40x9mm (600 x 600 DPI)

Supplementary File 1: Definition of terms

Context – Context often refers to the 'setting' of programmes and interventions. The literature suggests that differences should be contextualised by considering four levels of the context: (1) the external level (e.g. the wider social, economic or cultural setting); (2) the institutional level (e.g. the organisational culture, local priorities); (3) the interpersonal level (e.g. communication and collaboration); and (4) the individual level (e.g. personal values or knowledge) [1, 2]. Contextual elements can be expected to influence the relationship between audits and their outcomes and, in some cases, the outcomes of audits will influence the context (for example, a culture change may be generated by the outcomes of an audit). Some contextual elements may be essential for the outcome to occur and, because of this, may be confused with mechanisms [3, 4]. To resolve this, this research considers contextual elements as factors that can influence an outcome but are external to the intervention [3].

Mechanism – Mechanisms have been defined as '…underlying entities, processes, or [social] structures which operate in particular contexts to generate outcomes of interest' [5, p.2]. Identifying the mechanisms will advance the synthesis beyond describing 'what happened' to theorizing on 'why' it happened and 'under what circumstances'.

Outcome – Outcomes can be either intended or unintended, can be proximal, intermediate or final, and result from the activation of different mechanisms in different contexts.

Outcome patterns – Also described as 'demi-regularities' in the realist literature [2, 6], these amount to semi-predictable patterns of outcomes. First, 'semi' because variations in patterns of behaviour can only be partly attributed to contextual differences and, second, because individuals will likely, but not always, make similar choices about the resources they will use.

Circumstances – The phrase 'in what circumstances' is interpreted, in realist terms, as meaning 'in what contexts and by what mechanisms'. One has to examine the key contextual conditions that affect the mechanisms, identity in what way those conditions affect the mechanisms, and describe how the interaction between context and mechanisms affects the outcomes.

Context-Mechanism-Outcome (CMO) configurations – The resulting explanations for the observed outcome patterns are formulated as CMO configurations. A sample CMO configuration is as follows: a hospital with a supportive culture for quality improvement implements an audit (context). Subsequently, improvements in care quality are noted (outcome). The reason for this is the active participation of healthcare professionals in the audit process (mechanism).

Programme theory – Programme theory refers to an abstracted description and/or diagram that explains what a programme or intervention comprises of, and how and why it is expected to work. Programme theories are usually described as 'middle-range', meaning that they are 'specific enough to generate propositions that can be tested about aspects of the program but sufficiently abstract to be applicable to other programs' [5].

REFERENCES

1 Macfarlane F, Greenhalgh T, Humphrey C, et al. A new workforce in the making?: A case study of strategic human resource management in a whole-system change effort in healthcare, *Journal of Health Organization and Management* 2011;25:55-72.

2 Pawson R, Greenhalgh T, Harvey G, et al. Realist review--a new method of systematic review designed for complex policy interventions, *J Health Serv Res Policy* 2005;10 Suppl 1:21-34.

3 Marchal B., van Olmen J., Hoeree T., et al. Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research, *Evaluation* 2012;18:192-212.

4 Astbury B, Leeuw F. Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation, *American Journal of Evaluation* 2010;31:363-381.

5 Wong G, Greenhalgh T, Westhorp G, et al. RAMESES publication standards: realist syntheses, *BMC Med* 2013;11:21,7015-11-21.

6 Pawson R. Evidence-based policy: a realist perspective. London: Sage 2006.

Supplementary File 2: Search strategy MEDLINE

MEDLINE (Pubmed)
("Clinical Audit"[Majr:noexp] OR "Medical Audit"[Majr] OR "Nursing Audit"[Majr] OR "Accreditation"[Majr] OR "Certification"[Majr:noexp] OR "Peer Review, Health Care"[Majr] OR ((extern*[tiab] OR internal*[tiab]) AND audit[tiab]) OR medical audit*[tiab] OR clinical audit*[tiab] OR nursing audit*[tiab] OR audit[ti] OR audits[ti] OR accreditat*[ti] OR visitation*[ti]) AND ("Academic Medical Centers"[Mesh:noexp] OR "Hospitals, Teaching"[Mesh] OR "Outpatient Clinics, Hospital"[Mesh:noexp] OR "Hospitals"[Mesh:noexp] OR hospital*[tiab] OR ((health*[tiab] OR clinical[tiab]) AND (organisation*[tiab] OR organization*[tiab] OR center*[tiab] OR centre*[tiab])) OR health sector*[tiab] OR healthcare sector*[tiab] OR health care sector*[tiab]) AND ("Efficiency, Organizational"[Mesh] OR efficien*[tiab] OR effectiveness*[tiab] OR performan*[tiab] OR improvement*[tiab] OR "Quality Improvement"[Mesh:noexp] OR "Quality Assurance, Health Care"[Majr:noexp] OR quality improv*[ti] OR quality assur*[ti]) NOT ("Animals"[Mesh] NOT "Humans"[Mesh]) AND english[la] AND ("last 10 years"[PDat])

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Checklist item	Page
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	The report is identified as a protocol of a systematic review	1 (identified as a realist review)
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	n/a
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	2
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	1
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	19
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	n/a
Support:			
Sources	5a	Indicate sources of financial or other support for the review	18
Sponsor	5b	Provide name for the review funder and/or sponsor	18
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	18

Section and topic	Item No	Checklist item	Page
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	4-7
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	7 (did not use PICO as using a realist review)
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	11-12
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	10-11
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	10 and Supplementary File 2
Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	10-11
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	11
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	12-15
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	12, 15

Section and topic	Item No	Checklist item	Page
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	15-17
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	12-14
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesise	n/a
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I2, Kendall's τ)	n/a
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	n/a
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	n/a
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	n/a
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	n/a

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.

BMJ Open

Understanding how and why audits work: protocol for a realist review of audit programmes to improve hospital care

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-015121.R2
Article Type:	Protocol
Date Submitted by the Author:	24-Feb-2017
Complete List of Authors:	Hut-Mossel, Lisanne; University of Groningen, University Medical Centre Groningen, Centre of expertise on Quality and Safety Welker , Gera ; University of Groningen, University Medical Centre Groningen, Centre of expertise on Quality and Safety Ahaus, Kees; University of Groningen, University Medical Centre Groningen, Centre of Expertise on Quality and Safety; University of Groningen, Faculty of Economics and Business, Department Operations Gans, Rob; University of Groningen, University Medical Centre Groningen, Department of Internal Medicine
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Research methods, Health policy
Keywords:	Realist synthesis, Realist review, AUDIT, Quality improvement, Clinical audit < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Hospital Care

Title

Understanding how and why audits work: protocol for a realist review of audit programmes to improve hospital care

Authors

Lisanne Hut-Mossel[1], Gera Welker[1], Kees Ahaus[1, 2] and Rijk Gans[3].

University of Groningen, University Medical Centre Groningen.

Author Note

1 University of Groningen, University Medical Centre Groningen, Centre of Expertise on Quality and Safety, Groningen, the Netherlands.

2 University of Groningen, Faculty of Economics and Business, Department Operations, Groningen, the Netherlands.

3 University of Groningen, University Medical Centre Groningen, Department of Internal Medicine, Groningen, the Netherlands.

Corresponding author

Lisanne Hut-Mossel

Huispostcode LA10

PO Box 30 001

9700 RB Groningen

The Netherlands.

Tel: +31623473381

Correspondence to p.a.mossel@umcg.nl

Word Count (excluding title page, abstract, references, figures and tables): 4041

ABSTRACT

Introduction Many types of audits are commonly used in hospital care to promote quality improvements. However, the evidence on the effectiveness of audits is mixed. The objectives of this proposed realist review are (i) to understand how and why audits might, or might not, work in terms of delivering the intended outcome of improved quality of hospital care and (ii) to examine under what circumstances audits could potentially be effective. This protocol will provide the rationale for using a realist review approach and outline the method.

Methods and Analysis This review will be conducted using an iterative four-stage approach. The first and second steps have already been executed. The first step was to develop an initial programme theory based on the literature that explains how audits are supposed to work. Second, a systematic literature search was conducted using relevant databases. Third, data will be extracted and coded for concepts relating to context, outcomes and their interrelatedness. Finally, the data will be synthesised in a five-step process: (1) organising the extracted data into evidence tables, (2) theming, (3) formulating chains of inference from the identified themes, (4) linking the chains of inference and formulating CMO configurations and (5) refining the initial programme theory. The reporting of the review will follow the RAMESES publication standards.

Ethics and Dissemination This review does not require formal ethical approval. A better understanding of how and why these audits work, and how context impacts their effectiveness, will inform stakeholders in deciding how to tailor and implement audits within their local context. We will use a range of dissemination strategies to ensure that findings from this realist review are broadly disseminated to academic and non-academic audiences.

Trial registration number This systematic review protocol is registered on the PROSPERO database (registration number CRD42016039882).

Keywords Realist synthesis, Realist Review, Audit, Quality Improvement, Clinical Audit, Hospital Care

STRENGTHS AND LIMITATIONS OF THIS STUDY

- This review goes beyond considering the effectiveness of audits by building an understanding of how and why audits work within various contexts;
- This review uses a systematic screening protocol;
- The main limitation is that realist reviews are dependent upon the transparency and adequacy of the reporting of data on the context, the mechanisms and their relationship to the produced outcomes of individual studies by the original authors. The potential lack of adequate data in this regard might hamper developing a full understanding of how and why audits are effective and might restrict the full development of the programme theory.

INTRODUCTION

In recent decades, quality and safety issues have become increasingly important in hospital care because of their direct effect on both clinical outcomes and patient satisfaction. However, hospital care still suffers from a quality gap between the ideal care, in line with the best available medical evidence, and the actual care provided to patients [1]. To close this gap, health authorities and organisations currently prioritise quality improvement (QI) strategies, which are seen as systematic, data-driven monitoring and evaluation activities to improve the quality of hospital care [2]. A widely used QI strategy within hospitals is the audit.

In this review protocol, we will focus on audits that address quality. Such audits are commonly used within hospital care aiming to promote quality improvements by evaluating the delivered care against standards, controlling and/or changing healthcare processes and healthcare providers' performance [3]. However, it is unlikely that audits work in the same way in every setting. Accordingly, it is important to understand how and why audits might lead to quality improvements. A realist review, as outlined in this protocol, will contribute to this understanding.

The range of possible audits can be roughly divided into (1) external audits, used to gain insight into hospitals' compliance with external criteria (e.g. accreditation, certification, external peer reviews), (2) internal audits, often in preparation for an external audit, and (3) clinical audits, carried out on a voluntary basis by healthcare professionals [3, 4]. Although there are differences, such as in the scope and the approach used, between the various types of audits, they all serve the same objective: to improve the quality of hospital care.

Externally driven audits (i.e. accreditation, certification, external peer reviews and preparatory internal audits) seem to be more strongly anchored in Quality Assurance (QA), referring to initiatives designed to assure compliance with minimum (national) quality standards [5, 6]. These external audits are used to assess certain dimensions or characteristics of a healthcare providing

organisation against specified standards [7]. As such, the implementation of an external audit requires an external standard and collaboration from beyond the hospital - and this distinguishes them from internal audits.

Internal audits are conducted by internal auditors of the hospital's own organisation, such as quality officers or healthcare professionals from another department than the one being audited to guarantee some level of independent judgement. Internal audits are used to evaluate the delivered care against standards with different purposes. On the one hand, healthcare organisations use internal audits to continuously improve the quality of healthcare. In this way, one could expect that, compared with external audits, threats to quality can be more quickly revealed, allowing the organisation to regularly adapt its processes to improve quality at the local level. Internal audits are also frequently used in the framework of external audits and are conducted to avoid performance standards dropping between two external audits. These audits are designed to evaluate and improve the effectiveness of the organisation's quality management system and focus more on organisational conditions and less on the behaviour of healthcare professionals and patient outcomes [4].

Clinical audits differ from other types of audits in that they are mostly undertaken and initiated by healthcare professionals. Moreover, clinical audits represent a shift from QA to a Quality Improvement (QI) process, with a focus on seeking to improve care, and prevent poor care. This process takes place continuously as part of everyday routines [6-8]. As such, healthcare professionals work together to collect data and evaluate their own practices. Following this, they develop and apply improvements in their daily practices, and then the audit cycle is repeated to demonstrate improved and sustained improvements [8]. As such, clinical audits do not necessarily use external criteria and are not carried out in response to external demands as the initiative comes from the healthcare professionals themselves [9] . A considerable amount of literature addresses the effectiveness of audits and reports mixed results [10-12]. A systematic review on audit and feedback [13] demonstrated a positive overall effect of audits on clinical practice. Further, the authors noted

differences in the design and the effectiveness of audits. This variety can be attributed to at least two issues. First, audits are used to improve specific aspects of healthcare and can be targeted at different levels. For example, external audits are performed to induce changes at the organisational level (e.g. in organisational policy or procedures) whereas clinical audits are performed to alter local healthcare practices (e.g. clinical day-to-day practices or local guidelines). Second, audits are used in different contexts, and this considerably complicates the evaluation of their effects. For example, an audit could be effective in one organisation, or department, but not in another because of, for instance, the amount of support offered for quality improvements, as part of their differing contexts (see supplementary file 1). The literature on QI strategies recognises that the mixed effects are partly due to the differing contexts in which interventions are planned [14, 15].

The variety in the levels of audits, together with the heterogeneity of their contexts, suggests that it is unlikely that audits work in the same way in every setting. This creates challenges when attempting to synthesise evidence in a systematic review. Given this situation, more information about why and how audits work is needed [16, 17]. A detailed understanding of the contextual factors and the mechanisms that influence the effectiveness of audits is a prerequisite for understanding the mechanisms through which audits might lead to quality improvements. More importantly, a better understanding of 'how and why audits might work' will inform decision-making on how to tailor quality improvements at the local level.

A useful approach for explaining how and why audits might work, and investigating the interactions between context, mechanism and outcome, is the use of a realist review [18, 19]. The value of a realist review is that it is concerned with *how* an intervention works, rather than *whether* an intervention works, which is the focus of the conventional systematic review approach. Furthermore, the realist review methodology is specifically designed to cope with the intervention heterogeneity (in both the chosen study design and the used outcome measures) present in previous research on audits. Finally, this method is appropriate for the current research because

audits are complex context-sensitive interventions [10, 15, 20]. Within the past decade, similar studies in other contexts have used realist reviews to understand how complex interventions work and are put into practice (e.g. [21]).

The objectives of the current review are (i) to understand how and why audits might, or might not, work in producing the intended outcome of improved quality of care and (ii) to examine under what circumstances audits could potentially be effective by formulating and refining underlying programme theories. Consequently, this review focusses on three research questions:

1. What are the mechanisms through which audits deliver their intended outcomes?
2. What contextual factors determine whether the identified mechanisms produce the intended outcomes of audits?
3. In what circumstances (i.e. which combination(s) of mechanisms and context) are audits most likely to be effective?

METHOD

A realist review aims to clarify, from observed data, the outcomes (O) of particular interventions in relation to context (C) and mechanisms (M). This 'CMO' configuration is based on the philosophical assumption that an intervention in a specific context (C) evokes mechanisms (M) that generate an outcome (O). Consequently, the underlying mechanisms can be expected to produce a broad range of outcomes (O) when performed in different contexts (C) [18, 19, 22]. The philosophical basis is realism, which is positioned between positivism and constructivism and assumes the existence of an external reality (a 'real world') that is 'filtered' (i.e. perceived, interpreted and responded to) through the inputs of individuals. Consequently, it is not the intervention in and of itself that causes outcomes but the individuals who initiate a process of change and as such have an effect on whether and how the intervention works [19]. One of the key outputs of a realist review is the development

of programme theories that set out how and why an intervention is thought to 'work' to generate certain outcomes [23].

Study design

This review follows Pawson's steps for conducting realist reviews, namely: (1) clarifying the scope, and programme theory development; (2) searching for evidence; (3) appraising primary studies and extracting data; and (4) synthesising evidence and drawing conclusions [18]. The reporting of the review will follow the 'Realist and Meta-Review Evidence Synthesis: Evolving Standards' (RAMESES) publication standards [23]. In line with these standards, data extraction and synthesis will be an interpretive process, driven by reflection and discussion by the review team [23]. This process requires repeated reading of primary studies because, as the synthesis progresses, new or refined elements of theory are expected to emerge. The protocol outlined below was written after the first steps had already been initiated or completed. Accordingly, both the past tense (steps that have been completed) and the future tense (steps that have yet to be initiated) are used.

The review team represents a range of disciplines and professions, which enables us to consider multiple perspectives and insights on the data gathered within this realist review. LH has a nursing background and is a PhD candidate. GW is an implementation fellow and has several years of experience as a quality manager. KA has a background in economics and business, is a professor of healthcare management and has numerous publications related to quality and patient safety. RG is a medical specialist, professor of internal medicine, chair of the Dutch Training Program of Internal Medicine and President of the Dutch Society of Hospital Medicine. He is also involved in the training of hospitalists, who are conducting clinical audits as part of their training. Further, all the members of the review team are experienced in qualitative research.

1. Scope of the review and programme theory development

The first step of this review process has already been executed with the aim of building a programme theory that would explain how and why audits might work. The unit of analysis in a

realist review is not the intervention itself, but the contexts, mechanisms and outcomes that underpin the intervention. Given this situation, the initial step in formulating a programme theory draws on the literature on the effectiveness of QI strategies. As audits are QI strategies, we would assume that the contexts, mechanisms and outcomes uncovered in the QI literature might also play a role in the effectiveness of audits. The initial programme theory explains how audits are supposed to work by framing the interrelationships between context, mechanism and outcome (see Figure 1) [18]. An exploration of programme theories was initiated through on-going conversations within the review team and by a preliminary search of the literature. In addition, key terms were defined to guide the review and to ensure a common understanding (see Supplementary File 1).

After a number of iterations and discussions, we developed an initial programme theory regarding how and why audits might work. This suggests that having an organisational culture that is supportive of quality improvement, a leadership committed to quality and previous audit experiences are important contextual factors in the success of an audit [24]. These contextual factors trigger mechanisms, including a focus on continuous improvement rather than auditing for assurance and compliance [25, 26], bottom-up initiatives as a prerequisite for ownership [11, 27] and the active involvement of healthcare professionals in audit processes [13, 25, 28], that in turn lead to improvements in the quality of healthcare. The World Health Organisation (WHO) describes the quality of healthcare quality as follows: 'quality of care means that a health system should seek to make improvements in six areas or dimensions of quality' [29]. These dimensions are: effectiveness, efficiency, accessibility, acceptability, equity and safety.

This initial programme theory provides only a provisional structure for the review, and additional contextual factors, mechanisms and outcomes will be identified as the review progresses. The initial programme theory will be expanded, tested and refined using data from studies included in the review.

[Please insert Figure 1 here]

2. Search for evidence

As a second step, a search strategy was developed and performed in collaboration with an experienced university librarian. To ensure that all relevant articles were identified, a systematic literature search was conducted in MEDLINE, Embase, PsycINFO, Academic Search Premier, Business Source Premier, EmeraldInsight, Cochrane Library and Web of Science for the period 2005 - 2015. These databases were selected because they contain the core of quality and patient safety studies in the field of healthcare management as well as the biomedical view on quality of healthcare. The search strategy was piloted first in MEDLINE and later adapted for searching the other databases (see Supplementary File 2). The search included appropriate indexing terms (i.e. MeSH terms and keywords) on descriptors of audits (e.g. clinical audit, accreditation, certification, peer review, quality improvement, quality assurance), outcomes (e.g. efficiency, effectiveness, improvement), and hospital care (e.g. academic medical centres, health organisations). The reference lists of the uncovered review articles were studied to identify additional primary studies.

A realist synthesis approach to searching for evidence is iterative and evolves as the understanding of the subject matter deepens. Consequently, as the review progresses further, we will also search for unpublished and grey literature (e.g. websites, national guidelines, policy documents and information reported in specialist conferences) on the assumption that the literature on this topic may be diverse and dispersed. In addition, our expectation is that not all the included publications will adequately report on all aspects of an audit. We will therefore identify papers, and other research outputs, that relate to the same study by using 'cluster' searching [30]. For example, a search can be based on the members of a research team of an included article to identify all other refereed journal articles and related documents. Further, an additional iterative search may be necessary if it is determined that more data are required to refine a specific part of the programme theory, or if new prospective theories are identified during data extraction or synthesis.

3. Appraise primary studies and extract data

The selection of appropriate primary studies has already been executed. First, one reviewer (LH) identified and removed duplicates. Next, two reviewers (LH and GW) independently screened all titles and abstracts for suitability for inclusion. The focus was on empirical studies that evaluated the effects of audits in hospital settings within high-income countries, without restrictions on the type of study design. Only studies published in English were included to avoid misinterpretation of the content of an article due to language barriers (see Table 1).

Second, to ensure consistency of judgement, the full texts of a random ten percent of the articles were independently reviewed by LH and GW and retained if they were deemed relevant (i.e. the article could provide data on the context, mechanisms or outcomes of an audit). One reviewer (LH) reviewed the remaining 90% for their relevancy. In practice, a number of these articles required discussion or joint reading by two reviewers as it was sometimes difficult to decide between inclusion and rejection. Disagreements were recorded and discussed to ensure that decisions were made consistently. When disagreements remained, the matter was resolved through discussion involving the entire review team.

Inclusion criteria
Research on accreditation, certification, peer review/Dutch visitatie model[a] or local clinical audit
Hospital setting
High-income country
Published in English
English abstract available
Description of the medical or technical content
Description of the process of how the audit was conducted
Description of the impact of audit on medical and process outcomes

Table 1. Inclusion criteria.

[a] This is a system of peer review that is led and owned by doctors and designed to assess the quality of care provided by groups of hospital-based medical specialists. Practices are surveyed every 3-5 years by a group of peers [31].

The next section describes activities that have yet to be started. Quality appraisal and data extraction will be undertaken using pre-specified Excel spreadsheets (available on request from the first author). As the aim of the data extraction process is to evaluate and refine the initial programme theory, the contents of the data extraction sheets will be developed by the review team based on the content of the initial programme theory. To test the usability of the data extraction sheets, the file will be pretested on two purposefully selected articles [32]. For each study, the quality will be appraised and general characteristics extracted concerning the study's setting, the unit of analysis (including type of organisation) along with sections of the text that relate to context, mechanisms and their relationship to the produced outcomes.

Realist reviews amount to mixed-method reviews in that they incorporate both quantitative and qualitative studies, as well as grey literature. Consequently, different approaches are required to assess the quality of the included studies. Two reviewers (LH and either GW, KA or RG) will assess the quality of each included study. Any disagreement will be resolved through consensus-based group discussions within the review team. First, following realist synthesis principles, the evidence will be appraised using the concept of rigour [18, 19]. Rigour will be assessed by describing fidelity and nuggets (i.e. the potential match with the initial programme theory and valuable observations presented in primary studies), and trustworthiness (i.e. whether the methods used to generate the data are credible and trustworthy) [33, 34]. Second, to make the concept of trustworthiness more concrete, and to ensure transparency in decision-making, the quality of the evidence of each individual study will be presented in the form of an evidence-level table based on criteria established by the Cochrane Effective Practice and Organisation of Care (EPOC) review group (see Table 2) [35, 36]. These criteria range from systematic reviews (A1) to descriptive, non-analytical studies (D).

Level	Description
A1	Systematic review
	Review of data from multiple RCT studies
A2	Randomised trial
	Comparative study with (random) intervention and control group design
B	Controlled trial
	Trial with intervention and control group and comparisons of outcomes
	B1 multiple measurement points
	B2 single measurement point
C	Non-controlled study
	C1 multiple case, multiple measurements points
	C2 multiple case, single measurement point
	C3 single case, multiple measurements point
	C4 single case, single measurement point
D	Descriptive, non-analytical
	D1 multiple projects
	D2 single project
	D3 literature review

Table 2. Levels of evidence quality

In addition, the Quality Improvement Minimum Quality Criteria Set (QI-MQCS) will be used to assess the completeness of the reporting of each study [37]. This tool includes 16 content domains to critically appraise QI intervention publications and determine whether a minimum quality standard has been met (see Table 3).

Domain	**Minimum standard** (see Supplementary Appendix 2 online in [37])
1. Organisational motivation	Names or describes at least one motivation for the organisation's participation in the intervention
2. Intervention rationale	Names or describes a rationale linking at least one

Domain	Minimum standard (see Supplementary Appendix 2 online in [37])
	central intervention component to intended effects
3. Intervention description	Describes in detail at least one specific change including the personnel executing the intervention
4. Organisational characteristics	Reports at least two organisational characteristics
5. Implementation	Names at least one approach used to introduce the intervention
6. Study design	Names the study design
7. Comparator	Describes at least one key care process
8. Data source	Describes the data source and defines the outcome of interest
9. Timing	Describes the timing of the intervention and its evaluation to determine the presence of baseline data and the follow-up period after all intervention components have been fully implemented
10. Adherence/fidelity	Reports fidelity information for at least one intervention component, or describes evidence of adherence or of a mechanism ensuring compliance to the intervention
11. Health outcomes	Reports data on at least one health-related outcome
12. Organisational readiness	Reports at least one organisational-level barrier or facilitator
13. Penetration/reach	Describes the proportion of all eligible units that actually participated
14. Sustainability	Describes the sustainability or the potential for sustainability
15. Spread	Describes the potential for spread, existing tools for spread, or spread attempts / largescale rollout
16. Limitations	Reports at least one limitation of the design / evaluation

Table 3. Quality Improvement Minimum Quality Criteria Set (QI-MQCS) domains [37]

Two reviewers (LH and either GW, KA or RG) will independently undertake the data extraction and, in this way, data from all the included articles will be extracted by two reviewers. Following this, the review team will discuss the data extracted so that data are not simply categorised but are used to begin to develop a reasoning that provides input to the final synthesis phase. Furthermore, relevant sections of the articles, i.e. relating to context, mechanisms and their relationship to the produced outcomes, will be coded. This coding will be both inductive (codes emerge and are created during the data extraction) and deductive (codes created in advance of data extraction and informed by the initial programme theory).

4. Synthesise evidence and draw conclusions

Evidence will be synthesised by examining the relationships between contexts (e.g. organisational culture), mechanisms (e.g. bottom-up initiative) and outcomes (i.e. intended and unintended consequences and the impact of audits) to determine what works, in what circumstances, how and why. Rycroft-Malone and colleagues [32] have developed a five-step approach for a realist synthesis, incorporating the work of Pawson [18], as follows:

1. Organise the extracted data into evidence tables: The data extraction sheets from each individual study will be summarised and organised into one or more evidence tables. The evidence tables will also include a link back to the source papers.

2. Theme the data: Themes will be developed from the initial codes based on recurring contexts, mechanisms or outcomes. Identified themes will then be discussed among the reviewers, and contrary evidence will be sought.

3. Formulate chains of inference from the identified themes: Through an iterative process, we will search for chains of inference (connections) across extracted data and themes. For example, the 'leadership and competency' chain of inference might incorporate multiple themes including, for example, active engagement, competencies in quality improvement, strong legitimacy within the organisation, and a sound knowledge of quality issues. First, in order to support and formulate such chains of inference, patterns of similar mechanisms will be sought across different contexts to see if

emerging patterns of outcomes ('demi-regularities') are identified. Second, since we expect context-outcome regularities to be easier to identify than mechanisms, because mechanisms are underlying and hence often unobservable or 'hidden', context-outcome regularities will be used as a basis for uncovering mechanisms [19, 38]. Cases in which the contexts are restrictive or supportive will be identified and this will help in formulating the chains of inference and in recognising and explaining interactions between context, mechanisms and outcomes. Third, we will not overlook the possibility that there may be more than one mechanism in play at the same time. The chains of inference so formulated will function as a basis for the CMO configurations to be developed. Two reviewers (LH and either GW, KA or RG) will jointly formulate the chains of inference, and this information will be shared and discussed in the review team.

4. Link the chains of inference and formulate CMO configurations: The chains of inference will be linked together to develop CMO configurations, which will then be linked back to themes or theories emerging from the literature (e.g. commitment, organisational culture). The CMO configurations will be confirmed by returning to the source evidence. This iterative process will be guided by the research questions and the aims of the review. Following this, the generated CMO configurations will be used to either form new programme theories or to test, refine and supplement the initial programme theory. All these processes will be performed through discussions and agreement within the review team.

5. Refine the initial programme theory: Following the above four steps, a cumulative picture will be developed around the programme theories that summarises the nature of the context, mechanism and outcome, and links to the characteristics of the individual studies included. This cumulative picture will be based on hypotheses. For example, our review may suggest that hospitals that have a supportive culture for quality improvement (context) and that seek the active participation of healthcare professionals in audits (mechanism) generate improved safety as part of the quality of care (outcome). A narrative will be developed around each hypothesis that will describe the characteristics of the supportive evidence.

Pawson et al. (2005) argue that stakeholders should be involved in both the process of confirming the emerging findings and in dissemination activities [18]. To that end, emerging findings, supporting evidence and CMO configurations will be shared and discussed during a focus group session involving researchers, managers, policymakers and clinicians. The focus group will have 10 – 12 participants who will be selected to ensure some degree of homogeneity since this will enable them to share and discuss ideas by having comparable relevant knowledge in the field of audits [39]. This process will help to refine the focus and the presentation of the narrative stemming from the CMO configurations.

ETHICS AND DISSEMINATION

Under the Dutch Law on Medical Research Involving Human Subjects (WMO), this review does not require formal ethical approval. One of the key contributions of this review, compared to the majority of audit evaluations and systematic reviews, is that it focuses on how and why audits might work, rather than just on the impact of audits. To really understand how and why audits might work, or might not, we believe that a clear picture of the underlying processes that lead to the outcomes is essential. By providing this, this review will extend the current literature by providing knowledge on how, and why, audits may lead to sustainable quality improvements.

This review has important practical implications. Along with the increasing emphasis on patient safety and healthcare quality, controlling rising healthcare costs has become a top policy priority in many countries. Research programmes, such as the review proposed here, can provide a basis for identifying appropriate strategies for quality improvements in healthcare. A better understanding of how these audits 'work', and how context might impact on the intended outcome of improved healthcare quality, will inform stakeholders in their decision-making about how to tailor and implement audits within their local context.

It has been argued that the theory and emerging evidence about how best to design audits (and what should be avoided) should be incorporated in the development and reporting of audits [11, 40]. However, such theoretical underpinnings are rarely reported in articles about audits, and this might hamper a full understanding of how and why audits are effective, and further impose restrictions on the ability to fully develop the programme theory and the applicability of the programme theory.

We will use various dissemination strategies to ensure that findings from this realist review are broadly disseminated to academic and non-academic audiences. First, we will submit the findings of this realist review to a peer-reviewed journal. In addition, review results will be disseminated through public websites, publications in professional journals and by presenting our work at relevant national and international conferences, and at conferences for practitioners. The outcomes of this realist review will be disseminated through events organised by The Netherlands Federation of University Medical Centres (Nederlandse Federatie van Universitair Medische Centra) (NFU) and at a national symposium for hospitalists who conduct clinical audits as part of their training. As part of a more active dissemination strategy, we also intend a follow-up meeting with the focus group participants to discuss the findings and key messages.

Acknowledgement All persons contributing to the manuscript meet the criteria for authorship. There was no external funding for this paper.

Competing interests None.

Funding This research has not received a specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Contributors LH was responsible for project conception, protocol development and the writing and submission of the manuscript. KA, GW and RG were responsible for protocol development and editing the manuscript. All authors have given final approval of the version to be published.

REFERENCES

1 McDonald KM, Chang C, Schultz E. Closing the Quality Gap: Revisiting the State of the Science. 2013;AHRQ Publication No. 12(13)-E017.

2 Shojania KG, McDonald KM, Wachter RM, et al. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol. 1: Series Overview and Methodology). Rockville (MD): UCSF Evidence-based Practice Center 2004.

3 Spencer E, Walshe K. National quality improvement policies and strategies in European healthcare systems, *Qual Saf Health Care* 2009;18 Suppl 1:i22-7.

4 Bohigas L, Heaton C. Methods for external evaluation of health care institutions, *Int J Qual Health Care* 2000;12:231-8.

5 Klazinga N. Re-engineering trust: the adoption and adaption of four models for external quality assurance of health care services in western European health care systems, *Int J Qual Health Care* 2000;12:183-9.

6 ISO. ISO 9000:2015. Quality management systems - Fundamentals and vocabulary. Geneva: International Organization for Standardization 2015.

7 Walshe K, Freeman T, Latham L, et al. Chapter 6. The development of external reviews of clinical governance. In: Anonymous . Clinical governance - from policy to practice. Birmingham, UK: University of Birmingham, Health Services Management Centre 2000.

8 Dixon N. Getting clinical audit right to benefit patients. Romsey, England: Healthcare Quality Quest 2014.

9 Scrivener R, Morrell C, Baker R, et al. Principles for best practice in clinical audit. Abingdon, UK: Radcliffe Medical Press Ltd 2002.

10 Brubakk K, Vist GE, Bukholm G, et al. A systematic review of hospital accreditation: the challenges of measuring complex intervention effects, *BMC Health Serv Res* 2015;15:280,015-0933-x.

11 Flodgren G, Pomey MP, Taber SA, et al. Effectiveness of external inspection of compliance with standards in improving healthcare organisation behaviour, healthcare professional behaviour or patient outcomes, *Cochrane Database Syst Rev* 2011;(11):CD008992. doi:CD008992.

12 Greenfield D, Braithwaite J. Health sector accreditation research: a systematic review, *Int J Qual Health Care* 2008;20:172-83.

13 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.

14 Ovretveit J. Understanding the conditions for improvement: research to discover which context influences affect improvement success, *BMJ Qual Saf* 2011;20 Suppl 1:i18-23.

15 Walshe K, Freeman T. Effectiveness of quality improvement: learning from evaluations, *Qual Saf Health Care* 2002;11:85-7.

16 Shepperd S, Lewin S, Straus S, et al. Can we systematically review studies that evaluate complex interventions? *PLoS Med* 2009;6:e1000086.

17 Walshe K. Understanding what works--and why--in quality improvement: the need for theory-driven evaluation, *Int J Qual Health Care* 2007;19:57-9.

18 Pawson R, Greenhalgh T, Harvey G, et al. Realist review--a new method of systematic review designed for complex policy interventions, *J Health Serv Res Policy* 2005;10 Suppl 1:21-34.

19 Pawson R. Evidence-based policy: a realist perspective. London: Sage 2006.

20 Braithwaite J, Shaw CD, Moldovan M, et al. Comparison of health service accreditation programs in low- and middle-income countries with those in higher income countries: a cross-sectional study, *Int J Qual Health Care* 2012;24:568-77.

21 Mazzocato P, Savage C, Brommels M, et al. Lean thinking in healthcare: a realist review of the literature, *Qual Saf Health Care* 2010;19:376-82.

22 Denyer D, Tranfield D, van Aken J. Developing Design Propositions through Research Synthesis, *Organ Stud* 2008;29:393-413.

23 Wong G, Greenhalgh T, Westhorp G, et al. RAMESES publication standards: realist syntheses, *BMC Med* 2013;11:21,7015-11-21.

24 Kaplan HC, Brady PW, Dritz MC, et al. The influence of context on quality improvement success in health care: a systematic review of the literature, *Milbank Q* 2010;88:500-59.

25 Johnston G, Crombie IK, Davies HT, et al. Reviewing audit: barriers and facilitating factors for effective clinical audit, *Qual Health Care* 2000;9:23-36.

26 Power D, Terziovski M. Quality audit roles and skills: Perceptions of non-financial auditors and their clients, *J Oper Manage* 2007;25:126-147.

27 Greenhalgh T, Robert G, Macfarlane F, et al. Diffusion of innovations in service organizations: systematic review and recommendations, *Milbank Q* 2004;82:581-629.

28 Spurgeon P, Mazelan PM, Barwell F. Medical engagement: a crucial underpinning to organizational performance, *Health Serv Manage Res* 2011;24:114-20.

29 World Health Organization. Quality of care - A process for making strategic choices in health systems. 2006.

30 Booth A, Harris J, Croot E, et al. Towards a methodology for cluster searching to provide conceptual and contextual "richness" for systematic reviews of complex interventions: case study (CLUSTER). *BMC medical research methodology* 2013;13:118.

31 Lombarts MJMH, Klazinga NS. Inside self-regulation: peer review (visitatie) by Dutch medical specialists, *Clinical Governance: An International Journal* 2003;8:318-330.

32 Rycroft-Malone J, McCormack B, Hutchinson AM, et al. Realist synthesis: illustrating the method for implementation research, *Implementation Science* 2012;7:33.

33 Pawson R. Digging for Nuggets: How `Bad' Research Can Yield `Good' Evidence, *International Journal of Social Research Methodology* 2006;9:127-142.

34 Williams L, Rycroft-Malone J, Burton CR, et al. Improving skills and care standards in the support workforce for older people: a realist synthesis of workforce development interventions, *BMJ Open* 2016;6:e011964,2016-011964.

35 Cochrane Effective Practice and Organisation of Care (EPOC) Group 2013;2016.

36 Everdingen JJEv. Evidence-based richtlijnontwikkeling : een leidraad voor de praktijk. Houten : Bohn Stafleu Van Loghum, 2004:XXXI, 395 p. : ill. ; 24 cm.

37 Hempel S, Shekelle PG, Liu JL, et al. Development of the Quality Improvement Minimum Quality Criteria Set (QI-MQCS): a tool for critical appraisal of quality improvement intervention publications, *BMJ Qual Saf* 2015;24:796-804.

38 Astbury B, Leeuw F. Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation, *American Journal of Evaluation* 2010;31:363-381.

39 Fern EF. Advanced focus group research. Thousand Oaks: Canada: Sage publications 2001.

40 Ivers NM, Sales A, Colquhoun H, et al. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention, *Implement Sci* 2014;9:14,5908-9-14.

Figure legends

Figure 1. Initial programme theory for the effectiveness of an audit

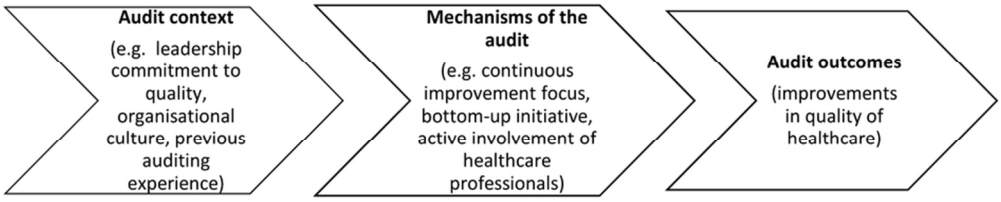

Figure 1. Initial programme theory for the effectiveness of an audit
[Please insert Figure 1 here]
40x9mm (600 x 600 DPI)

Supplementary File 1: Definition of terms

Context – Context often refers to the 'setting' of programmes and interventions. The literature suggests that differences should be contextualised by considering four levels of the context: (1) the external level (e.g. the wider social, economic or cultural setting); (2) the institutional level (e.g. the organisational culture, local priorities); (3) the interpersonal level (e.g. communication and collaboration); and (4) the individual level (e.g. personal values or knowledge) [1, 2]. Contextual elements can be expected to influence the relationship between audits and their outcomes and, in some cases, the outcomes of audits will influence the context (for example, a culture change may be generated by the outcomes of an audit). Some contextual elements may be essential for the outcome to occur and, because of this, may be confused with mechanisms [3, 4]. To resolve this, this research considers contextual elements as factors that can influence an outcome but are external to the intervention [3].

Mechanism – Mechanisms have been defined as '…underlying entities, processes, or [social] structures which operate in particular contexts to generate outcomes of interest' [5, p.2]. Identifying the mechanisms will advance the synthesis beyond describing 'what happened' to theorizing on 'why' it happened and 'under what circumstances'.

Outcome – Outcomes can be either intended or unintended, can be proximal, intermediate or final, and result from the activation of different mechanisms in different contexts.

Outcome patterns – Also described as 'demi-regularities' in the realist literature [2, 6], these amount to semi-predictable patterns of outcomes. First, 'semi' because variations in patterns of behaviour can only be partly attributed to contextual differences and, second, because individuals will likely, but not always, make similar choices about the resources they will use.

Circumstances – The phrase 'in what circumstances' is interpreted, in realist terms, as meaning 'in what contexts and by what mechanisms'. One has to examine the key contextual conditions that affect the mechanisms, identity in what way those conditions affect the mechanisms, and describe how the interaction between context and mechanisms affects the outcomes.

Context-Mechanism-Outcome (CMO) configurations – The resulting explanations for the observed outcome patterns are formulated as CMO configurations. A sample CMO configuration is as follows: a hospital with a supportive culture for quality improvement implements an audit (context). Subsequently, improvements in care quality are noted (outcome). The reason for this is the active participation of healthcare professionals in the audit process (mechanism).

Programme theory – Programme theory refers to an abstracted description and/or diagram that explains what a programme or intervention comprises of, and how and why it is expected to work. Programme theories are usually described as 'middle-range', meaning that they are 'specific enough to generate propositions that can be tested about aspects of the program but sufficiently abstract to be applicable to other programs' [5].

REFERENCES

1 Macfarlane F, Greenhalgh T, Humphrey C, et al. A new workforce in the making?: A case study of strategic human resource management in a whole-system change effort in healthcare, *Journal of Health Organization and Management* 2011;25:55-72.

2 Pawson R, Greenhalgh T, Harvey G, et al. Realist review--a new method of systematic review designed for complex policy interventions, *J Health Serv Res Policy* 2005;10 Suppl 1:21-34.

3 Marchal B., van Olmen J., Hoeree T., et al. Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research, *Evaluation* 2012;18:192-212.

4 Astbury B, Leeuw F. Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation, *American Journal of Evaluation* 2010;31:363-381.

5 Wong G, Greenhalgh T, Westhorp G, et al. RAMESES publication standards: realist syntheses, *BMC Med* 2013;11:21,7015-11-21.

6 Pawson R. Evidence-based policy: a realist perspective. London: Sage 2006.

Supplementary File 2: Search strategy MEDLINE

MEDLINE (Pubmed)
("Clinical Audit"[Majr:noexp] OR "Medical Audit"[Majr] OR "Nursing Audit"[Majr] OR "Accreditation"[Majr] OR "Certification"[Majr:noexp] OR "Peer Review, Health Care"[Majr] OR ((extern*[tiab] OR internal*[tiab]) AND audit[tiab]) OR medical audit*[tiab] OR clinical audit*[tiab] OR nursing audit*[tiab] OR audit[ti] OR audits[ti] OR accreditat*[ti] OR visitation*[ti]) AND ("Academic Medical Centers"[Mesh:noexp] OR "Hospitals, Teaching"[Mesh] OR "Outpatient Clinics, Hospital"[Mesh:noexp] OR "Hospitals"[Mesh:noexp] OR hospital*[tiab] OR ((health*[tiab] OR clinical[tiab]) AND (organisation*[tiab] OR organization*[tiab] OR center*[tiab] OR centre*[tiab])) OR health sector*[tiab] OR healthcare sector*[tiab] OR health care sector*[tiab]) AND ("Efficiency, Organizational"[Mesh] OR efficien*[tiab] OR effectiveness*[tiab] OR performan*[tiab] OR improvement*[tiab] OR "Quality Improvement"[Mesh:noexp] OR "Quality Assurance, Health Care"[Majr:noexp] OR quality improv*[ti] OR quality assur*[ti]) NOT ("Animals"[Mesh] NOT "Humans"[Mesh]) AND english[la] AND ("last 10 years"[PDat])

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Checklist item	Page
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	The report is identified as a protocol of a systematic review	1 (identified as a realist review)
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	n/a
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	2
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	1
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	19
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	n/a
Support:			
Sources	5a	Indicate sources of financial or other support for the review	18
Sponsor	5b	Provide name for the review funder and/or sponsor	18
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	18

Section and topic	Item No	Checklist item	Page
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	4-7
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	7 (did not use PICO as using a realist review)
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	11-12
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	10-11
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	10 and Supplementary File 2
Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	10-11
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	11
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	12-15
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	12, 15

Section and topic	Item No	Checklist item	Page
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	15-17
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	12-14
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesise	n/a
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I2, Kendall's τ)	n/a
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	n/a
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	n/a
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	n/a
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	n/a

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.