



Common Features of Selection Processes of Health System Performance Indicators in Primary Healthcare: A Systematic Review



Nicole Rendell^{1*}, Alexander Rosewell², Kamalini Lokuge¹, Emma Field^{1,3}

Abstract

Background: Health system performance indicators are widely used to assess primary healthcare (PHC) performance. Despite the numerous tools and some convergence on indicator criteria, there is not a clear understanding of the common features of indicator selection processes. We aimed to review the literature to identify papers that document indicator selection processes for health system performance indicators in PHC.

Methods: We searched the online databases Scopus, Medline, and CINAHL, as well as the grey literature, without time restrictions, initially on July 31, 2019 followed by an update November 13, 2020. Empirical studies or reports were included if they described the selection of health system performance indicators or frameworks, that included PHC indicators. A combination of the process focussed research question and qualitative analysis meant a quality appraisal tool or assessment of bias could not meaningfully be applied to assess individual studies. We undertook an inductive analysis based on potential indicator selection processes criteria, drawn from health system performance indicator appraisal tools reported in the literature.

Results: We identified 16503 records of which 28 were included in the review. Most studies used a descriptive case study design. We found no consistent variations between indicator selection processes of health systems of high income and low- or lower-middle income countries. Identified common features of selection processes for indicators in PHC include literature review or adaption of an existing framework as an initial step; a consensus building process with stakeholders; structuring indicators into categories; and indicator criteria focusing on validity and feasibility. The evidence around field testing with utility and consideration of reporting burden was less clear.

Conclusion: Our findings highlight several characteristics of health system indicator selection processes. These features provide the groundwork to better understand how to value indicator selection processes in PHC.

Keywords: Performance Indicators, Primary Healthcare, Selection, Development

Copyright: © 2022 The Author(s); Published by Kerman University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Rendell N, Rosewell A, Lokuge K, Field E. Common features of selection processes of health system performance indicators in primary healthcare: a systematic review. *Int J Health Policy Manag.* 2022;x(x):x–x. doi:10.34172/ijhpm.2022.6239

Article History:

Received: 15 April 2021
Accepted: 6 March 2022
ePublished: 17 March 2022

*Correspondence to:

Nicole Rendell
Email:
nicole.rendell@anu.edu.au

Introduction

Health system performance indicators are used to understand how well a health system is functioning and to what level it is meeting the needs of the population it has been designed to serve. This is often specific to the country context, which uses the country expectations of their health system, to develop such indicators. Regardless of context, indicators must be developed in a way that ensures any monitoring, evaluation or review activities, accurately reflect the health priorities, local health needs and expectations regarding equitable quality of care. The coronavirus disease 2019 (COVID-19) pandemic has further highlighted the value of timely access to health system performance intelligence and the central role of primary healthcare (PHC). However, striking the balance between detailed information from numerous indicators and a smaller number of clearly communicated indicators that can be reasonably collected, is an ongoing challenge.^{1,2} Both extremes can lead to a situation where there is a weak correlation between the performance indicators and the

performance itself, negating the value of indicators and performance measurement.¹

There are many studies in the literature on health system performance assessment. Some studies have tried to align health system performance indicators between countries to draw comparisons.^{3–8} However often international comparability is only possible across core health system components, such as those defined by the World Health Organization (WHO) to align with their six building blocks framework.^{3,5,7–9} There are also instruments available in the literature that have been developed to analyse a given set of quality indicators in the healthcare system after they have been created, such as QUALIFY and the Appraisal of Indicators through Research and Evaluation (AIRE).^{10,11} However, in the context of selecting health system performance indicators, there is little consensus on the approach to ensure the indicators are fit for use. In this context, we apply the concept described by Barbazza et al where fit for purpose and fit for use are both key constructs of actionable indicators. Specifically,

fit for use is defined as “*getting the right information into the right hands at the right time*” (p. 2).¹² We extend this definition to also consider reporting burden and other implementation factors.

While there is some literature available that describes the selection of health system performance indicators or a broader performance framework in a given context,^{4,13-18} there appears to be no comprehensive review of the processes used to select them, and how they compare.^{19,20} One recent systematic review considered content validity of indicator sets specifically across the full spectrum of healthcare settings. Procedural criteria formed only part of the review’s findings and included consideration of assessment purpose, develop/use conceptual framework, stakeholder involvement and transparency of the development process.²¹ We ascertain that a set of health system performance indicators could be considered successful if they were fit for use. However, in the first instance, we must understand what the current approaches to selection of health system performance indicators are and in what ways they vary. We aimed to review the literature to identify papers that document indicator selection processes for health system performance indicators in PHC.

For the purposes of this research, we will focus on health system performance indicators used to measure performance in PHC. Indicators at this level of the health system focus on local service delivery ranging from preventive services such as vaccinations, to ongoing management of non-communicable diseases.²² PHC has long been considered integral to health system functioning²³⁻²⁷ and there are many health system frameworks and indicators available that have been devised to measure and assess PHC specifically.²⁸⁻³⁰ These include Primary Care Assessment Tools³⁰ and the Primary Health Care Performance Initiative²⁹ in addition to broader health system frameworks such as the WHO Health System Building Blocks⁹ and the Sustainable Development Goals.³¹ Despite a range of standardised PHC frameworks and indicators to choose from, each with tools and resources available to support countries in implementing them, evidence suggests they are not consistently implemented.^{8,20,32-34}

Methods

This systematic review was completed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.³⁵

Search Strategy and Selection Criteria

We searched three online databases (Scopus, Medline and CINAHL), without time restrictions, and the grey literature (using the global search function on the Google platform in privacy mode, first 300 citations) on July 31, 2019. An updated database search followed on November 13, 2020. The searches for the online databases used the following syntax (see [Supplementary file 1](#) for full details):

(“health system?” OR “health care” OR “primary health*” OR “primary care”) AND (“performance indicator?” OR “quality indicator?” OR “framework?”) AND (development OR prioritization OR selection) AND NOT (acute OR hospital).

The grey literature search simply included a global search of the phrase: “development of health system performance indicators in primary care.” The databases were selected based on their reputation for content on health systems. The approach to the grey literature was adopted after testing a range of terms and search conditions, which found that a simplified global search was most effective in returning relevant results. The reference lists of all included studies were also searched for further eligible studies.

Studies were eligible for inclusion if the full text was available in English and they were empirical studies or reports in the grey literature that described the selection of health system performance indicators or frameworks, that included PHC indicators. These included clinical indicator series covering more than one disease and are used with the goal of understanding PHC performance. Indicator selection processes included any indicator or indicator set, that was identified for implementation and ongoing PHC management in a real world setting. The care setting was considered in scope, if it aligned with the definition of PHC outlined by WHO²² and no referral was required by an individual to seek the services. Although we did not record exclusions from failing the PHC criterion when completing the title and abstract screening. It was also necessary for inclusion, that the indicators were field tested, piloted or implemented (field testing). This was interpreted to include revisions of an existing indicator set and those of well-established organisations known to the authors with a clear trajectory for implementation, of indicators yet to be clearly implemented. This criterion was to ensure practical considerations of implementation were captured by the selected studies.

Empirical studies that reported only on indicators related to hospitals or acute settings; and developed and/or applied only a survey design without consideration for selection of PHC indicators were excluded. Studies were also excluded if they were secondary sources (for example, narrative reviews and systematic reviews); the indicators were specific to a single condition, due to our focus on health system performance; or based on a theoretical discussion on health system performance assessment including proposed indicators or frameworks that had not been field tested.

Two reviewers independently screened all titles, abstracts and full text articles according to the inclusion and exclusion criteria. Differences were resolved by consensus. Where the selection process of the same framework was reported across two papers, only the most comprehensive or more recent publication was selected for full text review and inclusion.

Data Extraction

For included studies or reports, we extracted information on the country context and key features of indicator selection including scale (international, national or subnational level indicators); type of indicator(s) (ie, subject); whether it was an original framework or revision; the key steps taken to develop the relevant indicator(s); consideration of existing frameworks and/other global reporting requirements; consideration of causal chain (for example, use of a logic model); stakeholders consulted; data quality and validity; data availability,

reliability and coherence with existing systems; reporting burden/resources relative to the context; and ongoing review of indicators. These criteria for data extraction were informed by health system indicator appraisal tools reported in the literature^{10,11,36} and then agreed by all authors.

Quality Assessment

The premise of this work is to summarise common approaches to understand possible factors in reported health system performance indicator selection processes that could potentially be used to assess the value of one indicator selection process over another. This is laying the groundwork for development of a quality assessment tool of indicator selection processes. To determine relevant criteria for data extraction, we used published indicator appraisal tools that assessed the quality of health system performance indicators themselves, ie, *outcomes*, to inform the structure of our dataset on *processes*.^{10,11,36} This infers that such criteria applied to indicators, is a valid starting point for understanding different aspects of processes used to select them. The criteria are outlined in the previous section, under *Data Extraction*, and were agreed to by consensus among authors. Therefore, the quality assessment of included papers formed part of the data extraction process used to answer our research question and was informed by categories for assessment of indicator quality which were agreed by all authors. For this reason,

application of a quality appraisal tool or assessment of the risk of bias were not appropriate.

Data Analysis

Following data extraction, indicator selection processes were compared and contrasted against the criteria drawn from indicator appraisal tools,^{10,11,36} their country context, emerging themes and patterns of interest. The analysis was inductive by design and allowed for key themes to emerge from the qualitative data extracted.

Results

Search Results

Our search identified 16 503 records of which 16 404 were excluded after screening titles and abstracts. After assessing the full text of 99 articles, we excluded 71 because there was insufficient information for data extraction (n=20), not an empirical study (n=12), a more comprehensive paper was available for inclusion (n=10), the care setting was not PHC (n=9), the indicators were specific for a single condition (n=8), the testing criteria was not met (n=7) or the full text was not available in English (n=5). In total 28 studies and reports were included for analysis (see Figure).

Study Characteristics

The characteristics of the included studies are summarised

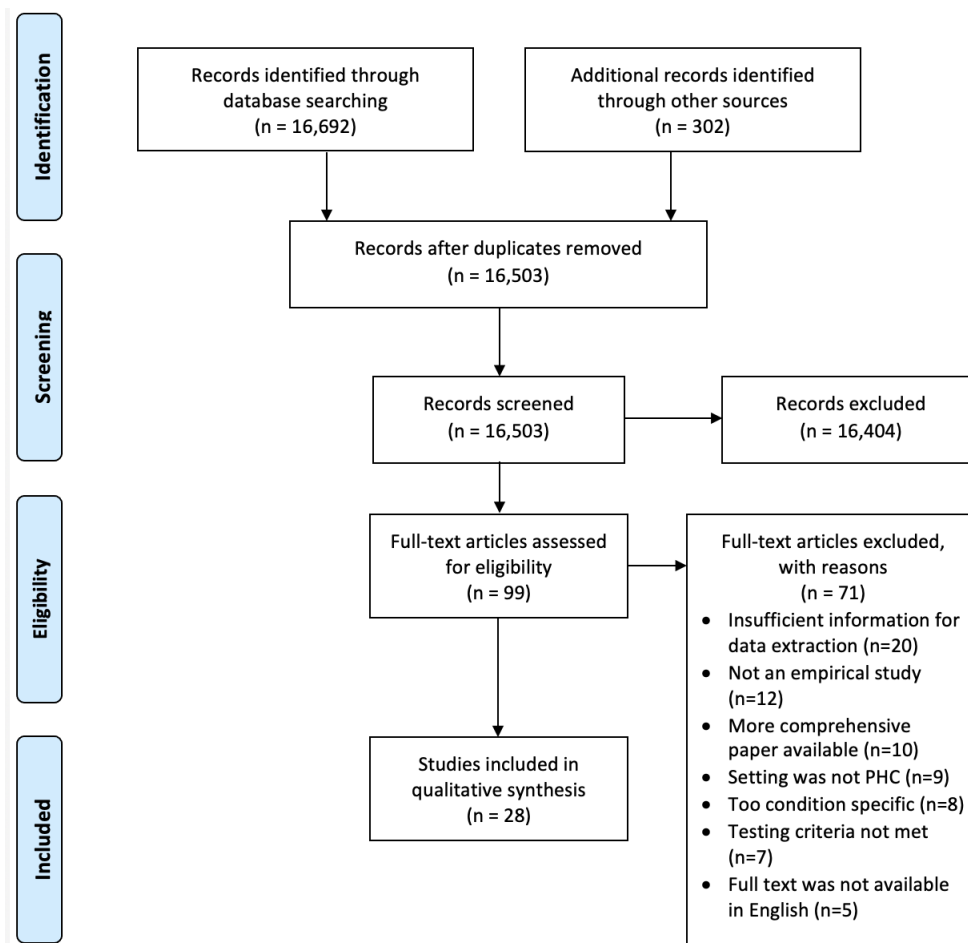


Figure 1. Flow Diagram of Screening Process.

in Table 1. The included studies were classified according to the World Bank Country and Lending Groups³⁷ at the time of their publication. They were predominantly from high income countries (n=23) of which 15 were Europe centric,³⁸⁻⁵² three each from the United States⁵³⁻⁵⁵ and Canada⁵⁶⁻⁵⁸ and one each from Australia⁵⁹ and New Zealand.⁶⁰ The few papers from upper-middle (n=1)⁶¹ or low- and middle-income countries (n=4), were set in Asia and Africa including China, Nepal, Kenya and India.⁶²⁻⁶⁵ The selection criteria meant that most of the included papers centred around a descriptive case study methodology (n=23).^{39-44,46,48-56,58-63,65} The other papers were an exploratory case study (n=1)⁵⁷ or cross sectional studies

(n=4).^{38,45,47,64} Half of the included papers were published in last six years from 2015 (n=14),^{38-40,42,45,49-52,54,56,62,63,65} with the oldest included paper published in 1998.⁵³ All themes described below form the results from the analysis.

Country Context

All indicator selection processes that were set in low- and low-middle income countries (n=4)⁶²⁻⁶⁵ were published relatively recently, with the earliest published in 2009⁶⁴ and the other three papers published within the last four years.^{62,63,65} There were no other consistent characteristics noted among papers from low- and low-middle income countries. However, all

Table 1. Characteristics of Included Studies

Reference	Care Setting	Study Type*	Scale, Type of Indicator(s)	Name of Framework Developed or Project
Aller et al, 2015 ³⁸	Three healthcare areas of the Catalan health system, Spain	Cross-sectional	Sub-national, coordination of care	NA
AIHW, 2009 ⁵⁹	Healthcare services, Australia	Government report	National level, safety and quality	National Indicators of Safety and Quality in Healthcare
Barbaza et al, 2019 ³⁹	Primary care practices, Europe	Case study - descriptive	International, PHC performance	PHC-IMPACT
Barrett et al, 1998 ⁵³	Two community mental health agencies in Colorado, United States	Case study - descriptive	Sub-national, mental health services performance	NA
Blozik et al, 2018 ⁴⁰	Swiss mandatory basic health insurance (Helsana group), Switzerland	Case study - descriptive	National, quality of care of ambulatory services	NA
Campbell et al, 2011 ⁴¹	Health services, United Kingdom	Case study - descriptive	National, clinical and organisational quality	QOF
Carinci et al, 2015 ⁴²	OECD countries	Case study - descriptive	International, healthcare quality	Healthcare Quality Indicators project
Claessen et al, 2011 ⁴³	Palliative care services, the Netherlands	Case study - descriptive	National, palliative care	NA
Coma et al, 2013 ⁴⁴	PHC professionals in Catalonia, Spain	Case-study - descriptive	Sub-national, PHC quality	EQA
Cookson et al, 2016 ⁴⁵	Small-area level, England	Time-series cross-sectional	Sub-national, health equity	NA
De Bie et al, 2011 ⁴⁶	Community pharmacies, Netherlands	Case study - descriptive	National, community pharmacy care	NA
Engels et al, 2006 ⁴⁷	General practices in selected European countries - Austria, Belgium, France, Germany, Israel, The Netherlands, Slovenia, Switzerland and the United Kingdom	Cross-sectional	International, management of primary care practices in Europe	EPA project
Gribben et al, 2002 ⁶⁰	First Health network of general practices, New Zealand	Case study - descriptive	Sub-national, PHC quality	NA
Herndon et al, 2015 ⁵⁴	Pediatric oral healthcare in the United States	Case study - descriptive	National, pediatric oral care	NA
Hutchison et al, 2020 ⁵⁶	PHC in Ontario, Canada	Case study - descriptive	Sub-national, PHC performance	PCPM
Katz et al, 2006 ⁵⁷	Family practices in Manitoba, Canada	Case study - exploratory	Sub-national, PHC quality	NA
Leemans et al, 2013 ⁴⁸	Palliative care in Flanders, Belgium	Case study (protocol) - descriptive	Sub-national, Palliative care	Q-PAC
Nambiar et al, 2020 ⁶²	PHC facilities in Kerala, India	Case study - descriptive	Subnational, PHC performance	NA
Parker et al, 2015 ⁴⁹	Patient safety in primary care organisations, Europe	Case study - descriptive	International, patient safety in PHC	LINNEAUS collaboration
Prytherch et al, 2017 ⁶³	Antenatal Care, Postnatal, Family Planning and Maternity services, Kenya	Case study - descriptive	National, PHC quality	KQMH

Table 1. Continued

Reference	Care Setting	Study Type*	Scale, Type of Indicator(s)	Name of Framework Developed or Project
Reedy et al, 2005 ⁵⁵	Santa Clara County public health services, United States	Case study - descriptive	Sub-national, patient safety in PHC	NA
Riain et al, 2015 ⁵⁰	General Practices in Ireland	Case study - descriptive	National, PHC quality	GP-IQ
Rushforth et al, 2015 ⁵¹	Primary care practices, United Kingdom	Case study - descriptive	National, PHC quality	NA
Sarriot et al, 2009 ⁶⁴	Health districts, Nepal	Cross-sectional	Sub-national, health sector aid investments	Nepal Specific Sustainability Framework
Stanciu et al, 2020 ⁵²	Primary care clusters, Wales	Non-experimental mixed methods	Subnational, PHC performance	PCCMA
Terner et al, 2013 ⁵⁸	PHC, Canada	Case study - descriptive	National, PHC performance	NA
Veillard et al, 2017 ⁶⁵	PHC systems in 135 low- and middle-income countries	Case study - descriptive	International, PHC quality	PHCPI
Wong et al, 2010 ⁶¹	PHC, China	Case study - descriptive	Sub-national, PHC performance	China results- based Logic Model for CHS

Abbreviations: NA, Not Applicable; PHC-IMPACT, Primary Healthcare Impact, Performance and Capacity Tool; AIHW, Australian Institute of Health and Welfare; QOF, Quality and Outcomes Framework; OECD, Organisation for Economic Co-operation and Development; EQA, the Catalan acronym for Estandard de Qualitat Assistencial; EPA, European Practice Assessment; PCPM, Primary Care Performance Measurement; Q-PAC, Quality Indicators for Palliative Care; QQMH, Kenya Quality Assurance Model for Health; GP-IQ, General Practice Indicators of Quality; PCCMA, Primary Care Clusters Multidimensional Assessment; PHCPI, Primary Healthcare Performance Initiative; CHS, Community Health Facilities and Stations.

papers that were included as revisions (n = 4) were set in high income countries,^{41,42,56,58} two of which were set in Canada, one focusing on subnational indicators in Ontario.⁵⁶

Key Steps

The included papers highlighted common steps taken to develop health system performance indicators in PHC (see Table 2). Most of the indicator selection processes began with a review of the literature and other policy documents. Some of these adopted a stricter methodology opting for a rigorous systematic review (n = 5)^{38,39,43,48,52} as an initial step, while the others adopted a more flexible approach to reviewing the literature and other publicly available documents (n = 13).^{42,46,49,50,54,55,57-60,62,63,65} All those that did not explicitly use a literature review as part of their indicator selection process, relied heavily on adapting an existing framework into the relevant context (n = 10).^{40,41,44,45,47,51,53,56,61,64}

All of the included indicator selection processes incorporated a consensus process with context specific stakeholders to reach the final set of indicators, with two exceptions.^{55,60} In just under half of the selected papers, consensus building was achieved using either a (modified) Delphi process (n = 7)^{42,44,47,49,50,62,65} or RAND Corporation/University of California Los Angeles (UCLA) Appropriateness Method (n = 5).^{41,48,51,54,63} Outside of field testing (see below), there were no identified consistencies reported among these papers that used known consensus building methodologies – their contexts ranged from low- and middle-income countries to high income countries; sub-national, national and international indicator sets were all represented; and the dates of publication ranged from 2006 to 2020, similar to the papers that did not use the known consensus methodologies. In the cases that neither Delphi nor RAND/UCLA Appropriateness methods were explicitly applied, there were consultations reported with a

carefully selected expert group or stakeholders relevant to the context.^{38-40,43,45,46,52,53,56-59,61,64} These accounted for half of the papers included (n = 14). Around one third of all included papers cited at least one patient representative as part of their consultations (n = 9).^{40,43,45,47,48,50,51,54,56}

While the methodology of this review explicitly states that field testing the derived indicators in practice is part of the inclusion criteria, this concept of testing in practice was not applied uniformly across each of the included indicator selection processes. Just over half of the included papers (n = 16) reported field testing where by adjustments could be made to the indicator set in response to the findings ie, there was utility in testing the indicators for the purpose of improving them and their application.^{39-41,43,46-52,54,60,62,63,65} All but two^{42,44} of the papers that reported using a known consensus methodology (Delphi or RAND/UCLA Appropriateness methods), conducted field testing in this way (n = 10).^{41,47-51,54,62,63,65} For those papers that did not adopt this approach to field testing, the indicator selection processes tested the indicators in practice more as a proof of concept and the opportunity to respond to feedback was not clear.

Structure of Frameworks and Indicator Criteria

Of all included indicator selection processes, only one produced a single synthetic indicator⁴⁴ and the remaining indicator sets were organised into frameworks (see Table 2). The frameworks were predominantly structured around PHC categories (n = 22).^{38,40-43,45-51,53-60,63,64} While there was some overlap of defined categories across included papers, these were limited. Of 123 categories of indicators identified, overlaps could be identified across nine categories: access, efficiency, effectiveness, safety, clinical care, quality, people, preventive health activities and chronic disease management. In contrast, a small selection of indicator selection processes

Table 2. Number of Included Papers by Indicator Selection Feature

	High-Income	Upper-Middle Income	Low- and Lower-Middle Income	
Country context	2 ^{38,60}	1 ⁶¹	4 ^{62,65}	
	Systematic Review	Literature Review	No Review – Relied on Existing Framework	
Review of existing publications	5 ^{38,39,43,48,52}	13 ^{42,46,49,50,54,55,57-60,62,63,65}	10 ^{40,41,44,45,47,51,53,56,61,64}	
	Delphi	RAND/UCLA Appropriateness	No Formal Methodology Applied	None Reported
Consensus process	7 ^{42,44,47,49,50,62,65}	5 ^{41,48,51,54,63}	14 ^{38-40,43,45,46,52,53,56-59,61,64}	2 ^{55,60}
Engagement with patients	2 ^{47,50}	3 ^{48,51,54}	4 ^{40,43,45,56}	
Field Testing	5 ^{47,49,50,62,65}	5 ^{41,48,51,54,63}	5 ^{39,40,43,46,52}	1 ⁶⁰
	Categories/Dimensions	Donabedian	Logic Model	Single Composite Indicator
Structure of indicators	22 ^{38,40-43,45-51,53-60,63,64}	2 ^{39,52}	3 ^{61,62,65}	1 ⁴⁴
	Validity	Feasibility	Burden of Reporting	Ongoing Review
Indicator criteria	All	All	13 ^{39,41-44,48,50,51,53,56,59,62,64}	6 ^{41,42,47,56,59,62}

Total Included papers = 28.

used a causal chain structure, either Donabedian's structure-process-outcomes classification ($n=2$)^{39,52} or standard logic model ($n=3$).^{61,62,65} Interestingly, these five papers did not include national level indicators (ie, they were either international or sub-national indicators), and four of them were published within the last four years.^{39,52,62,65}

In terms of indicator criteria, every included paper had considered the validity of each proposed indicator as part of their selection process. Feasibility issues, including consideration of data availability, reliability, integration with existing systems and/or whether a modification of existing practices would be required, also featured in all indicator selection processes but to varying degrees. In six of the included papers, this formed a central tenet of the selection process.^{43,45,51,57,62,64} Around half of papers factored in the resulting burden of reporting relative to context that can be generated by large and complex indicator frameworks ($n=13$).^{39,41-44,48,50,51,53,56,59,62,64} There were no other consistent characteristics identified among these studies that considered reporting burden. The need for ongoing review was also reported inconsistently with just six of 28 included papers earmarking to revisit application of the indicator set in the future.^{41,42,47,56,59,62} Not all papers that reported ongoing review were included on the basis of a revision (see methodology), although half of them did ($n=3$).^{41,42,56}

Quality Assessment

We did not undertake a separate quality appraisal for individual studies as quality assessment of the processes used in each of the included papers formed the basis of our data extraction process and analyses. While our approach did not directly assess different aspects of indicator process against predefined benchmarks, as this is the aim beyond the findings of this paper, we did ascertain meaningful information on different aspects of indicator selection processes. For example, extracting data around formal consensus processes provided insights to the rigour applied to the study design. Likewise, extracting data around consulted stakeholders, including patient representatives, provided insights into the level of stakeholder engagement incorporated into the study design.

Discussion

Our review provides an overview of the key features of an indicator selection process (the process) used in the selection of PHC indicators. After a comprehensive search of databases and grey literature sources, 28 indicator selection processes met the inclusion criteria, where the indicators were subsequently field tested. We found no consistent variations between selection processes of health systems of high income and low- or lower-middle income countries.

A literature review was the most common initial step, with adaptation of an existing framework also prevalent, but less common. A consensus building process with a range of stakeholders featured in nearly all of the included processes, with only around a third reporting inclusion of patient perspectives. A structured methodology for the consensus building process was not universally applied. Field testing was also only integrated into the process for just over half the papers. The resultant indicators were predominantly structured into PHC categories with limited overlap of categories across the different processes. All processes considered validity and feasibility issues while the reporting burden relative to resources was considered in around half of included papers and few papers reporting the need for ongoing review.

Fit for Use

The term fit for use has been used to reflect that health system performance indicators, including those for PHC, are highly context specific and are needed by different parts of the health system, at different times. The selection of indicators should therefore aim to be well adapted to the context in which they are implemented while allowing assessment of a benchmarked standard of PHC. Currently there is no agreed criteria for assessment of fit for use for health system performance or PHC indicators. Recent work by Barbazza et al explored the complementary concepts of fit for purpose, fit for use and how these apply to the overall actionability of indicators.¹² They identified three clusters for consideration when it comes to fit for use – methodological, contextual and managerial considerations.¹² Determinations on whether an

indicator selection process was successfully implemented in terms of meeting fit for use criteria was beyond the scope of this paper. However, future work could merge our findings with these categories to consider such an assessment.

We sought to review the literature on selection processes of indicator sets that had already been implemented in practice and examining them through the lens of key criteria drawn from indicator appraisal frameworks.^{10,11,36} These appraisal frameworks were developed as assessment tools for health system performance indicators, with one of them designed specifically for assessing performance indicators for PHC.³⁶

A similar approach of using an indicator appraisal framework was adopted by de Bruin-Kooistra et al for selecting quality indicators for midwifery care in the Netherlands.⁶⁷ The authors used the AIRE instrument as a manual and subsequent checklist for developing the indicators in their project.⁶⁷ Likewise, Perara et al also incorporated a checklist into their development of the Systematic Indicator Development Method.⁶⁸ However, they use a definition of “fitness for purpose” that gives little consideration to use beyond technical capacity.⁶⁸

One of the first papers to raise the notion of a ‘best practice’ selection process for health performance indicators was published in 2003 by Mainz.⁶⁹ More recently, a systematic review by Kötter et al was undertaken analysing guideline based approaches to indicator development (including selection).⁷⁰ The authors advocate for a ‘gold standard’ process of indicator selection to foster transparency and efficiency of resources and conclude that “It remains unclear which method leads to the best [*quality indicators*], since no randomized controlled or other comparative studies investigating this issue exist” (p. 20).⁷⁰ Our systematic review adds to this groundwork and goes further to propose that any indicator selection process considered ‘gold standard’ will need to be sufficiently nimble to accommodate different contexts and therefore produce indicators that could be fit for use.

Consensus Among Stakeholders

One of the emerging themes present in nearly all the processes was a procedure for reaching consensus among stakeholders. Not only were relevant stakeholders consulted but they were engaged in a deliberate way to reach consensus on the indicator set being developed. Two methodologies that feature strongly in the literature did not emerge as strong themes in our analysis: the Delphi technique and the RAND/UCLA appropriateness method. By definition, the Delphi technique involves repetitive administration of anonymous questionnaires, usually two or three rounds, with each round building towards a consensus usually without a face to face meeting.⁷¹ The RAND/UCLA appropriateness method, a derivative of the Delphi technique, similarly involves a series of rounds to reach consensus but adopts a more comprehensive approach by expressly combining expert opinion (from questionnaires and in person panel meetings) with evidence, usually drawn from a systematic literature review.⁷¹ Both of these methods have been used widely in the literature.^{14-17,72,73}

It is not clear why these or an alternative method did not feature more strongly among the selected papers in our review.

Most papers that did use one of the consensus generating methodologies did also include field testing with utility (see below). As such, potentially it is an issue of resources. A recent systematic review by Jandhyala on consensus generating methods suggests that they have been modified over time and no longer reflect their original principles.⁷² In addition the inclusion of patients as part of stakeholder consultations was identified in only a third of the selected papers. This finding does not align well with the literature for best practice indicator selection as there is a large body of evidence that advocates for the inclusion of patients among stakeholder consultations.^{21,74-79}

Utility of Field Testing

One of the unique aspects of our review’s methodology is the inclusion of criterion around field testing. It was included to ensure implementation issues were adequately considered in line with the concept of selecting fit for use PHC indicators. There are many proposed indicator sets in the literature that are formed on the basis of a literature review, followed up with a consensus process among relevant experts. The goal of these papers is often to arrive at a final set indicators rather than ensure the indicator set is meeting the requirement that determined the need for the indicator(s) in the first place.^{14-17,73,80} In the absence of field testing, any proposed indicator(s) would still be theoretical.

Our review identified that around half of the included papers incorporated a field testing component in a way that added utility to the indicators’ selection process.^{41,43,46-52,57,62-65} Other condition specific indicator selection processes in the literature have also emphasised this approach or the absence of it.^{19,80-82} Of note, Hilarion et al articulate that “...indicator development and their application should not be separated” (p. 99).⁸⁰

Structure of Indicators

Only few studies selected in our review structured their indicators according to a causal chain such as Donabedian’s structure-process-outcomes⁸³ or a more standard logic model. Those that did not were structured around categories relevant to the context and there was limited overlap between the selected papers on PHC aspects. It is not clear why there was a preference for a non-linear categorisation of indicators over a sequential causal chain. Potentially a categorisation approach allows for easier comparison across different contexts and more flexibility to align within existing established frameworks.

It is not clear which approach is favoured by the literature.^{2,18,84} For example, one widely used assessment tool for PHC, the Primary Care Assessment Tools, is organised according to principles of PHC – first contact, person-focused care over time, comprehensiveness and coordination.³⁰ Further, on behalf of the World Organization of Family Doctors’ executive committee, Kidd et al argue that if PHC indicator(s) are too focussed on clinical conditions, they risk subsequent action favouring vertical oriented approaches.⁸⁵ They suggest standard aspects of PHC be integrated into indicator frameworks such as comprehensiveness, coordination,

continuity of care, safety and quality and workforce selection.⁸⁵ However, grouping indicators according to Donabedian's structure-process-outcome categorisation has also featured in the literature.^{5,18} A recent 2019 umbrella review of PHC quality indicators uses this grouping as a framework for analysing the indicators identified through its review.¹⁸

In addition, our review only identified one selection process for a single synthetic indicator. This finding aligns with recent literature advising against composite indicators, in favour of multidimensional frameworks. This is because they can mask what is happening in reality and even if they do indicate an issue, it is difficult to unpack the system and other related factors that would have led to that finding.^{5,86}

Strengths and Limitations

Our review has several strengths. It was conducted in line with PRISMA guidelines and used a simplified search strategy to ensure it would comprehensively capture the range of contexts and terminology used when developing PHC indicators. This syntax was decided upon because it allowed for a large variation in terminology within the same topic, even though it increased screening burden and subsequent researcher bias (discussed below). The review sets the scene for further work on criteria that could be used to assess a 'best practice' in selecting indicators for PHC.

Limitations of this review relate to the homogeneity of the types of studies that were included. Most of the included papers had a descriptive case study design. When a qualitative appraisal tool was applied (ie, Mixed Methods Appraisal Tool⁸⁷), it yielded the same results for each included study because there was an overlap between the inclusion criteria and that of the appraisal tool. This meant appraising the quality of individual studies or assessing the risk of bias using an existing quality appraisal tool did not produce results that could meaningfully distinguish the quality of evidence from each of the different included papers. Other systematic reviews with process research questions have also not applied quality assessment tools.^{21,88} As explained by Carroll and Booth, quality appraisal in qualitative research is still vulnerable to subjectivity and any tool applied may only evaluate the reporting of the study rather than its actual conduct, thereby questioning its value.⁸⁹ Further, the outcomes of this review are more closely aligned with methodologies for developing quality assessment tools. In the continuum proposed by Whiting et al, this review could be considered within the second stage of 'tool development'.⁹⁰ Thereby, attempting to apply an appraisal tool to this kind of tool development process research is duplicative and does not contribute to the credibility of the studies selected, as intend by quality appraisal.

Process research is an emerging methodology in health systems research and is more commonly applied in psychotherapy and business management research.^{88,91} In the context of psychotherapy literature, process research is used to identify, describe, explain and predict the effects of processes that lead to therapeutic change and understand the mechanisms of action for a given result.⁸⁸ There is more variation in how process research is applied in the field of

business management. One application is in the context of New Venture Creation and focuses on the process of non-existence to existence of economic activities.⁹¹ Across these disciplines it is clear that there is value in understanding the process to achieve a given outcome, yet challenges remain in ensuring the rigour and absence of bias in qualitative process research. Berends and Deken argue that the challenge lies in demonstrating the link between process data and process theory.⁹² This is especially challenging when considering novel research questions where a clear theory is yet to be established, as is the case with this systematic review, which relied on pre-existing indicator appraisal tools as a foundation to understanding indicator selection processes. An exploratory or comparative design would offer more definitive insights about what features of an indicator selection process are conducive to fit for use PHC indicators. For example, an exploratory case study,⁶⁶ qualitative comparative analysis⁹³ or quasi experimental field trial⁹⁴ with an emphasis on qualitative data collection, although these designs are more difficult and resource intensive. Also, most of the included papers were set in high-income countries which may affect the translation of the findings into other country contexts. However, the underlying theme of this analysis is to identify process features that transcend context so the impact of such differences on the findings may be limited. Further, the criteria applied to restrict included papers to PHC settings limits the generalisability of the findings, even though some characteristics may resonate at multiple levels of the health system.

In addition, the base assumption for our findings is that comparisons across criteria drawn from a selection of health system indicator appraisal tools, among indicator sets that had been implemented in practice, leads to knowledge about the criteria for assessing the selection of indicators that are fit for use. In reality, there could be several factors in a given indicator selection process that could lead to indicators that are fit for use and were not captured by criteria used in our dataset. It is also possible that by restricting our criteria to indicators that were implemented in practice, comprehensive indicator selection processes in scope may have been excluded if such a process was developed and reported across more than one paper. The implementation criteria led to the exclusion of a number of papers during the title and abstract screening. Inclusion of these papers may have broadened the range of study designs included but this is unlikely as the research question would still focus on processes which are commonly reported through case study designs.

Researcher Bias

Key aspects of the inclusion criteria were inherently vulnerable to researcher bias. For example, by including only papers that had sufficient information to extract for data collection, it is possible that there were papers that reported on indicator selection and were excluded due to the subjective level of detail in which they had described their processes. Further, by excluding less comprehensive or earlier papers when the same framework was reported across more than one paper, it is possible that different or more comprehensive processes

were excluded, and may have contributed to the homogenous set of results.

Our results may have also been affected by the variation of key words and naming conventions in this field which meant the selection of papers for inclusion were subject to researcher bias. The categorisation process undertaken to do data extraction and the subsequent analysis are other examples of unavoidable researcher bias.

Lastly, a systematic review protocol was not registered for this study, limiting the transparency and opportunity for peer feedback on the methodology. The in-house protocol is available upon reasonable request from the corresponding author.

Conclusion

We identified several characteristics of health system indicator selection processes in the literature. These include use of a literature review as an initial step, more so than adapting an existing framework; stakeholder engagement with a known methodology to consensus building; structuring the indicator framework according to context specific PHC domains; and indicator criteria focusing on validity and feasibility (including reliability). The evidence around field testing with utility and consideration of reporting burden was not as strong despite being critical to implementation success. The evidence presented here provides some key principles to guide future work on assessing PHC indicator selection processes for health program staff, policy officials, donors and researchers. Future research using an explorative or comparative designs will strengthen these findings.

Ethical issues

Not applicable.

Competing interests

Authors declare that they have no competing interests.

Authors' contributions

NR, AR, KL, and EF designed the study. NR completed the database and grey literature searches. Both NR and EF undertook the screening process and quality review. NR analysed the data and wrote the paper with input from all authors.

Funding

This research is supported by an Australian Government Research Training Program Fee Offset Scholarship and the Australian Government Research Training Program Domestic Scholarship.

Authors' affiliations

¹National Centre for Epidemiology and Population Health, Australian National University, Canberra, ACT, Australia. ²School of Public Health and Community Medicine, University of New South Wales, Sydney, NSW, Australia. ³Menzies School of Health Research, Brisbane, QLD, Australia.

Supplementary file

Supplementary file 1. Detailed Search Strategy.

References

- European Union. Expert Panel on effective ways of investing in Health (EXPH): Preliminary report on Tools and Methodologies for Assessing the Performance of Primary Care. https://ec.europa.eu/health/sites/health/files/expert_panel/docs/017_assessing_performance_primarycare_en.pdf. Published 2017.
- Levesque JF, Sutherland K. Combining patient, clinical and system perspectives in assessing performance in healthcare: an integrated measurement framework. *BMC Health Serv Res.* 2020;20(1):23. doi:10.1186/s12913-019-4807-5
- Hirschhorn LR, Baynes C, Sherr K, et al. Approaches to ensuring and improving quality in the context of health system strengthening: a cross-site analysis of the five African Health Initiative Partnership programs. *BMC Health Serv Res.* 2013;13 Suppl 2:S8. doi:10.1186/1472-6963-13-s2-s8
- Perić N, Hofmarcher MM, Simon J. Headline indicators for monitoring the performance of health systems: findings from the european Health Systems_Indicator (euHS_I) survey. *Arch Public Health.* 2018;76:32. doi:10.1186/s13690-018-0278-0
- Braithwaite J, Hibbert P, Blakely B, et al. Health system frameworks and performance indicators in eight countries: a comparative international analysis. *SAGE Open Med.* 2017;5:2050312116686516. doi:10.1177/2050312116686516
- Marshall MN, Shekelle PG, McGlynn EA, Campbell S, Brook RH, Roland MO. Can health care quality indicators be transferred between countries? *Qual Saf Health Care.* 2003;12(1):8-12. doi:10.1136/qhc.12.1.8
- Noto G, Corazza I, Kļaviņa K, Lepiksone J, Nuti S. Health system performance assessment in small countries: the case study of Latvia. *Int J Health Plann Manage.* 2019;34(4):1408-1422. doi:10.1002/hpm.2803
- Fekri O, Macarayan ER, Klazinga N. Health System Performance Assessment in the WHO European Region: Which Domains and Indicators have been Used by Member States for its Measurement? Health Evidence Network synthesis report 55. <https://www.euro.who.int/en/publications/abstracts/health-system-performance-assessment-in-the-who-european-region-which-domains-and-indicators-have-been-used-by-member-states-for-its-measurement-2018>. Published 2018.
- World Health Organization. Monitoring the building blocks of health systems: a handbook of indicators and their measurement strategies. https://www.who.int/healthinfo/systems/WHO_MBHSS_2010_full_web.pdf. Published 2010.
- de Koning J, Smulders A, Klazinga N. *Appraisal of Indicators through Research and Evaluation (AIRE)*. Amsterdam: Academic Medical Center-University of Amsterdam; 2007.
- Reiter A, Fischer B, Kötting J, et al. QUALIFY: Instrument for the Assessment of Quality Indicators. https://www.researchgate.net/publication/267256474_QUALIFY_Instrument_for_the_Assessment_of_Quality_Indicators. Published 2007.
- Barbaza E, Klazinga NS, Kringos DS. Exploring the actionability of healthcare performance indicators for quality of care: a qualitative analysis of the literature, expert opinion and user experience. *BMJ Qual Saf.* 2021;30(12):1010-1020. doi:10.1136/bmjqs-2020-011247
- Ahuja S, Gronholm PC, Shidhaye R, Jordans M, Thornicroft G. Development of mental health indicators at the district level in Madhya Pradesh, India: mixed methods study. *BMC Health Serv Res.* 2018; 18(1):867. doi:10.1186/s12913-018-3695-4
- Ebert ST, Pittet V, Cornuz J, Senn N. Development of a monitoring instrument to assess the performance of the Swiss primary care system. *BMC Health Serv Res.* 2017;17(1):789. doi:10.1186/s12913-017-2696-z
- Fukuma S, Shimizu S, Niihata K, et al. Development of quality indicators for care of chronic kidney disease in the primary care setting using electronic health data: a RAND-modified Delphi method. *Clin Exp Nephrol.* 2017;21(2):247-256. doi:10.1007/s10157-016-1274-8
- Lee B, Park SY. Developing key performance indicators for guaranteeing right to health and access to medical service for persons with disabilities in Korea: using a modified Delphi. *PLoS One.* 2018;13(12):e0208651. doi:10.1371/journal.pone.0208651
- O'Donnell S, Doyle G, O'Malley G, et al. Establishing consensus on key public health indicators for the monitoring and evaluating childhood obesity interventions: a Delphi panel study. *BMC Public Health.* 2020;20(1):1733. doi:10.1186/s12889-020-09814-y
- Ramalho A, Castro P, Gonçalves-Pinho M, et al. Primary health care quality indicators: an umbrella review. *PLoS One.* 2019;14(8):e0220888. doi:10.1371/journal.pone.0220888
- Blozik E, Nothacker M, Bunk T, Szecsenyi J, Ollenschläger G, Scherer M. Simultaneous development of guidelines and quality indicators -- how do guideline groups act? A worldwide survey. *Int J Health Care Qual Assur.* 2012;25(8):712-729. doi:10.1108/09526861211270659
- Klazinga N, Fischer C, ten Asbroek A. Health services research related to performance indicators and benchmarking in Europe. *J Health Serv Res*

- Policy*. 2011;16 Suppl 2:38-47. doi:10.1258/jhsrp.2011.011042
21. Schang L, Blotenberg I, Boywitt D. What makes a good quality indicator set? A systematic review of criteria. *Int J Qual Health Care*. 2021; 33(3):mzab107. doi:10.1093/intqhc/mzab107
 22. World Health Organization. Primary Health Care (PHC). <https://www.who.int/primary-health/en/>. Published 2019.
 23. World Health Organization. Declaration on Primary Health Care: Astana, 2018. <https://www.who.int/primary-health/conference-phc/declaration>. Published 2019.
 24. World Health Organization. The world health report 2008: primary health care now more than ever. https://www.who.int/whr/2008/whr08_en.pdf. Published 2008.
 25. Pettigrew LM, De Maeseeneer J, Anderson MI, Essuman A, Kidd MR, Haines A. Primary health care and the Sustainable Development Goals. *Lancet*. 2015;386(10009):2119-2121. doi:10.1016/s0140-6736(15)00949-6
 26. Starfield B. Primary care: an increasingly important contributor to effectiveness, equity, and efficiency of health services. SESPAS report 2012. *Gac Sanit*. 2012;26 Suppl 1:20-26. doi:10.1016/j.gaceta.2011.10.009
 27. Kringos DS, Boerma WG, Hutchinson A, van der Zee J, Groenewegen PP. The breadth of primary care: a systematic literature review of its core dimensions. *BMC Health Serv Res*. 2010;10:65. doi:10.1186/1472-6963-10-65
 28. Kringos DS, Boerma WG, Bourgueil Y, et al. The European primary care monitor: structure, process and outcome indicators. *BMC Fam Pract*. 2010;11:81. doi:10.1186/1471-2296-11-81
 29. Primary Health Care Performance Initiative. Measuring Primary Health Care Performance. <https://improvingphc.org/measuring-primary-health-care-performance>. Published 2018.
 30. Shi L, Masís DP, Guanais FC. Measurement of Primary Care: The Johns Hopkins Primary Care Assessment Tool 2012. https://www.jhsph.edu/research/centers-and-institutes/johns-hopkins-primary-care-policy-center/pca_tools.html.
 31. United Nations. Sustainable Development Goals Knowledge Platform. <https://sustainabledevelopment.un.org/sdgs>.
 32. Fraccolli LA, Gomes MF, Nabão FR, Santos MS, Cappellini VK, de Almeida AC. Primary health care assessment tools: a literature review and metasynthesis. *Cien Saude Colet*. 2014;19(12):4851-4860. doi:10.1590/1413-812320141912.00572014
 33. Bangalore Sathyananda R, de Rijk A, Manjunath U, Krumeich A, van Schayck CP. Primary health Centres' performance assessment measures in developing countries: review of the empirical literature. *BMC Health Serv Res*. 2018;18(1):627. doi:10.1186/s12913-018-3423-0
 34. European Union. A New Drive for Primary Care in Europe: Rethinking the Assessment Tools and Methodologies. Report of the Expert Group on Health Systems Performance Assessment. https://ec.europa.eu/health/sites/health/files/systems_performance_assessment/docs/2018_primarycare_eg_en.pdf. Published 2018.
 35. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097
 36. Perera R, Dowell T, Crampton P, Kearns R. Panning for gold: an evidence-based tool for assessment of performance indicators in primary health care. *Health Policy*. 2007;80(2):314-327. doi:10.1016/j.healthpol.2006.03.011
 37. The World Bank. World Bank Country and Lending Groups. <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>. Published 2020.
 38. Aller MB, Vargas I, Coderch J, et al. Development and testing of indicators to measure coordination of clinical information and management across levels of care. *BMC Health Serv Res*. 2015;15:323. doi:10.1186/s12913-015-0968-z
 39. Barbazza E, Kringos D, Kruse I, Klazinga NS, Tello JE. Creating performance intelligence for primary health care strengthening in Europe. *BMC Health Serv Res*. 2019;19(1):1006. doi:10.1186/s12913-019-4853-z
 40. Blozik E, Reich O, Rapold R, Scherer M. Evidence-based indicators for the measurement of quality of primary care using health insurance claims data in Switzerland: results of a pragmatic consensus process. *BMC Health Serv Res*. 2018;18(1):743. doi:10.1186/s12913-018-3477-z
 41. Campbell SM, Kontopantelis E, Hannon K, Burke M, Barber A, Lester HE. Framework and indicator testing protocol for developing and piloting quality indicators for the UK quality and outcomes framework. *BMC Fam Pract*. 2011;12:85. doi:10.1186/1471-2296-12-85
 42. Carinci F, Van Gool K, Mainz J, et al. Towards actionable international comparisons of health system performance: expert revision of the OECD framework and quality indicators. *Int J Qual Health Care*. 2015;27(2):137-146. doi:10.1093/intqhc/mzv004
 43. Claessen SJ, Francke AL, Belarbi HE, Pasman HR, van der Putten MJ, Deliens L. A new set of quality indicators for palliative care: process and results of the development trajectory. *J Pain Symptom Manage*. 2011; 42(2):169-182. doi:10.1016/j.jpainsymman.2010.10.267
 44. Coma E, Ferran M, Méndez L, Iglesias B, Fina F, Medina M. Creation of a synthetic indicator of quality of care as a clinical management standard in primary care. *Springerplus*. 2013;2(1):51. doi:10.1186/2193-1801-2-51
 45. Cookson R, Asaria M, Ali S. Health Equity Indicators for the English NHS: a longitudinal whole-population study at the small-area level. Southampton (UK): NIHR Journals Library; 2016. <https://www.ncbi.nlm.nih.gov/books/NBK385236/>.
 46. De Bie J, Kijlstra NB, Daemen BJ, Bouvy ML. The development of quality indicators for community pharmacy care. *BMJ Qual Saf*. 2011;20(8):666-671. doi:10.1136/bmjqs.2010.045237
 47. Engels Y, Dautzenberg M, Campbell S, et al. Testing a European set of indicators for the evaluation of the management of primary care practices. *Fam Pract*. 2006;23(1):137-147. doi:10.1093/fampra/cmi091
 48. Leemans K, Cohen J, Francke AL, et al. Towards a standardized method of developing quality indicators for palliative care: protocol of the quality indicators for palliative care (Q-PAC) study. *BMC Palliat Care*. 2013;12:6. doi:10.1186/1472-684x-12-6
 49. Parker D, Wensing M, Esmail A, Valderas JM. Measurement tools and process indicators of patient safety culture in primary care. A mixed methods study by the LINNEAUS collaboration on patient safety in primary care. *Eur J Gen Pract*. 2015;21 Suppl 1:26-30. doi:10.3109/13814788.2015.1043732
 50. ni Riain A, Vahey C, Kennedy C, Campbell S, Collins C. Roadmap for developing a national quality indicator set for general practice. *Int J Health Care Qual Assur*. 2015;28(4):382-393. doi:10.1108/ijhcqa-09-2014-0091
 51. Rushforth B, Stokes T, Andrews E, et al. Developing 'high impact' guideline-based quality indicators for UK primary care: a multi-stage consensus process. *BMC Fam Pract*. 2015;16:156. doi:10.1186/s12875-015-0350-6
 52. Stanciu MA, Law RJ, Myres P, et al. The development of the Primary Care Clusters Multidimensional Assessment (PCCMA): a mixed-methods study. *Health Policy*. 2020;124(2):152-163. doi:10.1016/j.healthpol.2019.12.004
 53. Barrett TJ, Bartsch DA, Zahniser JH, Belanger S. Implementing and evaluating outcome indicators of performance for mental health agencies. *J Healthc Qual*. 1998;20(3):6-13. doi:10.1111/j.1945-1474.1998.tb00254.x
 54. Herndon JB, Crall JJ, Aravamudhan K, et al. Developing and testing pediatric oral healthcare quality measures. *J Public Health Dent*. 2015;75(3):191-201. doi:10.1111/jphd.12087
 55. Reedy AM, Luna RG, Olivas GS, Sujeer A. Local public health performance measurement: implementation strategies and lessons learned from aligning program evaluation indicators with the 10 essential public health services. *J Public Health Manag Pract*. 2005;11(4):317-325. doi:10.1097/00124784-200507000-00010
 56. Hutchison B, Haj-Ali W, Dobell G, Yeritsyan N, Degani N, Gushue S. Prioritizing and implementing primary care performance measures for Ontario. *Healthc Policy*. 2020;16(1):43-57. doi:10.12927/hcpol.2020.26291
 57. Katz A, Soodeen RA, Bogdanovic B, De Coster C, Chateau D. Can the quality of care in family practice be measured using administrative data? *Health Serv Res*. 2006;41(6):2238-2254. doi:10.1111/j.1475-6773.2006.00589.x
 58. Turner M, D'Silva J, Tipper B, Krylova O, Webster G. Assessing primary healthcare using pan-Canadian indicators of health and health system performance. *Healthc Q*. 2013;16(2):9-12.
 59. Australian Institute of Health and Welfare (AIHW). Towards national indicators of safety and quality in health care. https://www.aihw.gov.au/getmedia/a143a228-0e9e-4098-bea6-a84053446bbc/hse-75-10792_c02.pdf.aspx. Published 2009.
 60. Gribben B, Coster G, Pringle M, Simon J. Quality of care indicators for population-based primary care in New Zealand. *N Z Med J*. 2002;115(1151):163-166.
 61. Wong ST, Yin D, Bhattacharyya O, Wang B, Liu L, Chen B. Developing

- a performance measurement framework and indicators for community health service facilities in urban China. *BMC Fam Pract*. 2010;11:91. doi:10.1186/1471-2296-11-91
62. Nambiar D, Sankar DH, Negi J, Nair A, Sadanandan R. Monitoring universal health coverage reforms in primary health care facilities: creating a framework, selecting and field-testing indicators in Kerala, India. *PLoS One*. 2020;15(8):e0236169. doi:10.1371/journal.pone.0236169
 63. Prytherch H, Nafula M, Kandie C, et al. Quality management: where is the evidence? Developing an indicator-based approach in Kenya. *Int J Qual Health Care*. 2017;29(1):19-25. doi:10.1093/intqhc/mzw147
 64. Sarriot E, Ricca J, Ryan L, Basnet J, Arscott-Mills S. Measuring sustainability as a programming tool for health sector investments: report from a pilot sustainability assessment in five Nepalese health districts. *Int J Health Plann Manage*. 2009;24(4):326-350. doi:10.1002/hpm.1012
 65. Veillard J, Cowling K, Bitton A, et al. Better measurement for performance improvement in low- and middle-income countries: the primary health care performance initiative (PHCPI) experience of conceptual framework development and indicator selection. *Milbank Q*. 2017;95(4):836-883. doi:10.1111/1468-0009.12301
 66. Baxter P, Jack S. Qualitative case study methodology: study design and implementation for novice researchers. *Qual Rep*. 2010;13(4):544-559. doi:10.46743/2160-3715/2008.1573
 67. de Bruin-Kooistra M, Amelink-Verburg MP, Buitendijk SE, Westert GP. Finding the right indicators for assessing quality midwifery care. *Int J Qual Health Care*. 2012;24(3):301-310. doi:10.1093/intqhc/mzs006
 68. Perera R, Dowell A, Crampton P. Painting by numbers: a guide for systematically developing indicators of performance at any level of health care. *Health Policy*. 2012;108(1):49-59. doi:10.1016/j.healthpol.2012.07.008
 69. Mainz J. Developing evidence-based clinical indicators: a state of the art methods primer. *Int J Qual Health Care*. 2003;15 Suppl 1:i5-11. doi:10.1093/intqhc/mzg084
 70. Kötter T, Blozik E, Scherer M. Methods for the guideline-based development of quality indicators--a systematic review. *Implement Sci*. 2012;7:21. doi:10.1186/1748-5908-7-21
 71. Campbell SM, Braspenning J, Hutchinson A, Marshall M. Research methods used in developing and applying quality indicators in primary care. *Qual Saf Health Care*. 2002;11(4):358-364. doi:10.1136/qhc.11.4.358
 72. Jandhyala R. Delphi, non-RAND modified Delphi, RAND/UCLA appropriateness method and a novel group awareness and consensus methodology for consensus measurement: a systematic literature review. *Curr Med Res Opin*. 2020;36(11):1873-1887. doi:10.1080/03007995.2020.1816946
 73. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PLoS One*. 2011;6(6):e20476. doi:10.1371/journal.pone.0020476
 74. Barson S, Doolan-Noble F, Gray J, Gauld R. Healthcare leaders' views on successful quality improvement initiatives and context. *J Health Organ Manag*. 2017;31(1):54-63. doi:10.1108/jhom-10-2016-0191
 75. Crampton P, Perera R, Crengle S, et al. What makes a good performance indicator? Devising primary care performance indicators for New Zealand. *N Z Med J*. 2004;117(1191):U820.
 76. Freeman T. Using performance indicators to improve health care quality in the public sector: a review of the literature. *Health Serv Manage Res*. 2002;15(2):126-137. doi:10.1258/0951484021912897
 77. Klazinga N, Stronks K, Delnoij D, Verhoeff A. Indicators without a cause. Reflections on the development and use of indicators in health care from a public health perspective. *Int J Qual Health Care*. 2001;13(6):433-438. doi:10.1093/intqhc/13.6.433
 78. Kötter T, Schaefer FA, Scherer M, Blozik E. Involving patients in quality indicator development - a systematic review. *Patient Prefer Adherence*. 2013;7:259-268. doi:10.2147/ppa.s39803
 79. Santana MJ, Ahmed S, Lorenzetti D, et al. Measuring patient-centred system performance: a scoping review of patient-centred care quality indicators. *BMJ Open*. 2019;9(1):e023596. doi:10.1136/bmjopen-2018-023596
 80. Hilarion P, Suñol R, Groene O, Vallejo P, Herrera E, Saura RM. Making performance indicators work: the experience of using consensus indicators for external assessment of health and social services at regional level in Spain. *Health Policy*. 2009;90(1):94-103. doi:10.1016/j.healthpol.2008.08.002
 81. Oostra DL, Nieuwboer MS, Olde Rikkert MGM, Perry M. Development and pilot testing of quality improvement indicators for integrated primary dementia care. *BMJ Open Qual*. 2020;9(2):e000916. doi:10.1136/bmjopen-2020-000916
 82. Saturno-Hernández PJ, Fernández-Elorriaga M, Martínez-Nicolás I, Poblano-Verástegui O. Construction and pilot test of a set of indicators to assess the implementation and effectiveness of the who safe childbirth checklist. *BMC Pregnancy Childbirth*. 2018;18(1):154. doi:10.1186/s12884-018-1797-y
 83. NHS Improvement. A model for measuring quality care. <https://improvement.nhs.uk/documents/2135/measuring-quality-care-model.pdf>.
 84. Hyppönen H, Ronchi E, Adler-Milstein J. Health care performance indicators for health information systems. *Stud Health Technol Inform*. 2016;222:181-194.
 85. Kidd MR, Anderson MI, Obazee EM, Prasad PN, Pettigrew LM. The need for global primary care development indicators. *Lancet*. 2015;386(9995):737. doi:10.1016/s0140-6736(15)61532-x
 86. Ham C, Raleigh V, Foot C, Robertson R, Alderwick H. Measuring the performance of local health systems: a review for the Department of Health. https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/measuring-the-performance-of-local-health-systems-dh-review-kingsfund-oct15.pdf. Published 2015.
 87. Hong QN, Pluye P, Fabregues S, et al. Mixed Methods Appraisal Tool (MMAT) Version 2018. http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/attach/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf. Published 2018.
 88. Franklin C, Zhang A, Froerer A, Johnson S. Solution focused brief therapy: a systematic review and meta-summary of process research. *J Marital Fam Ther*. 2017;43(1):16-30. doi:10.1111/jmft.12193
 89. Carroll C, Booth A. Quality assessment of qualitative evidence for systematic review and synthesis: is it meaningful, and if so, how should it be performed? *Res Synth Methods*. 2015;6(2):149-154. doi:10.1002/jrsm.1128
 90. Whiting P, Wolff R, Mallett S, Simera I, Savović J. A proposed framework for developing quality assessment tools. *Syst Rev*. 2017;6(1):204. doi:10.1186/s13643-017-0604-6
 91. Davidsson P, Gruenhagen JH. Fulfilling the process promise: a review and agenda for new venture creation process research. *Entrep Theory Pract*. 2021;45(5):1083-1118. doi:10.1177/1042258720930991
 92. Berends H, Deken F. Composing qualitative process research. *Strateg Organ*. 2021;19(1):134-146. doi:10.1177/1476127018824838
 93. Hudson J, Kühner S. Qualitative comparative analysis and applied public policy analysis: new applications of innovative methods. *Policy Soc*. 2013;32(4):279-287. doi:10.1016/j.polsoc.2013.10.001
 94. White H, Sabarwal S. Quasi-experimental Design and Methods, Methodological Briefs: Impact Evaluation 8. https://www.unicef-irc.org/KM/IE/img/downloads/Quasi-Experimental_Design_and_Methods_ENG.pdf. Published 2014.
 95. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res*. 2005;15(9):1277-1288. doi:10.1177/1049732305276687