BMJ Open

Built to last? The sustainability of healthcare system improvements, programmes and interventions: a systematic integrative review

Jeffrey Braithwaite 1,2, Kristiana Ludlow 1,2, Luke Testa, Jessica Herkes, Hanna Augustsson 1, Gina Lamprell, Elise McPherson, Yvonne Zurynski 1,2

ABSTRACT

Introduction The sustainability of healthcare delivery systems is challenged by ageing populations, complex systems, increasing rates of chronic disease, increasing costs associated with new medical technologies and growing expectations by healthcare consumers. Healthcare programmes, innovations and interventions are increasingly implemented at the front lines of care to increase effectiveness and efficiency; however, little is known about how sustainability is conceptualised and measured in programme evaluations.

Objectives We aimed to describe theoretical frameworks, definitions and measures of sustainability, as applied in published evaluations of healthcare improvement programmes and interventions.

Design Systematic integrative review.

Methods We searched six academic databases, CINAHL, Embase, Ovid MEDLINE, Emerald Management, Scopus and Web of Science, for peer-reviewed English journal articles (July 2011–March 2018). Articles were included if they assessed programme sustainability or sustained outcomes of a programme at the healthcare system level. Six reviewers conducted the abstract and full-text review. Data were extracted on study characteristics, definitions, terminology, theoretical frameworks, methods and tools. Hawker’s Quality Assessment Tool was applied to included studies.

Results Of the 92 included studies, 75.0% were classified as high quality. Twenty-seven (29.3%) studies provided 32 different definitions of sustainability. Terms used interchangeably for sustainability included continuation, maintenance, follow-up or long term. Eighty studies (87.0%) clearly reported the timepoints at which sustainability was evaluated: 43.0% at 1–2 years and 11.3% at <12 months. Eighteen studies (19.6%) used a theoretical framework to conceptualise or assess programme sustainability, including frameworks that were not specifically designed to assess sustainability.

Conclusions The body of literature is limited by the use of inconsistent definitions and measures of programme sustainability. Evaluations of service improvement programmes and interventions seldom used theoretical frameworks. Embedding implementation science and healthcare service researchers into the healthcare system is a promising strategy to improve the rigour of programme sustainability evaluations.

INTRODUCTION

Background

Healthcare systems across the world strive to provide safe, high-quality care and deliver the best possible health outcomes for the populations they serve. At the same time, fiscal constraints necessitate the delivery of healthcare in an efficient and cost-effective way.1 This creates a challenge to the sustainability of healthcare systems globally.2 3 Lead international agencies, including the World Health Organization, the Organisation for Economic Cooperation and Development and the World Economic Forum, have recently highlighted significant threats to the sustainability of healthcare system performance.1,4 13 Ageing populations and the rapidly increasing burden of chronic conditions also pose challenges to healthcare system sustainability.5 17 The introduction of new medical technologies, including new diagnostic tests, new drugs, medical equipment and digital healthcare services,8 as well as a growing ‘consumer
culture’, have led to demands for higher standards of patient safety and quality of care and lower costs. At the same time, the level of wasteful spending on low-value care has remained static, at approximately 30%, while high-value care which aligns with level I evidence or best-practice consensus guidelines accounts for approximately 60% of delivered care and has done for two decades.10–15

Sweeping reforms are relatively rare, mainly because they require enormous efforts to mobilise multiple stakeholders who work within entrenched cultures, structures and approaches that make up complex healthcare systems.16–19 Much of the change implementation to improve healthcare system sustainability occurs closer to the front lines of care, through innovative projects, improvement programmes and interventions, referred to as ‘programmes’ from this point forward.

To maximise the benefits of programme innovation in healthcare, we need the ability to rigorously assess whether programmes are adaptable to real-world settings, and sustainable beyond the programme trial period.20 Stirman et al’s 2012 systematic review of the sustainability of implemented healthcare programmes reported that few of the included studies that were published before June 2011 provided a definition of sustainability.21 The authors considered articles in which studies assessed the continuation of programmes after initial implementation efforts, staff training periods or funding had ended.22 They found that when defining sustainability, the majority of included studies referred to Scheirer’s definition22 which is based on the work of Shediac-Rizkallah and Bone.23 Scheirer proposed three levels of analysis for programme sustainability: (1) Individual level: continuing to deliver beneficial services (outcomes) to consumers; (2) Organisational level: maintaining the programme and/or its activities, even if modified and (3) Community level: maintaining the capacity of a community to deliver programme activities.22 In a subsequent paper, Scheirer and Dearing defined sustainability as ‘the continued use of program components and activities for the continued achievement of desirable program and population outcomes’.22

More recently, Moore et al24 proposed five constructs for the assessment of programme sustainability: (1) following a defined period of time, (2) a continuation of a programme and/or (3) the maintenance of individual behavioural change, (4) evolution or adaption of the programme, and individual behavioural change may occur while (5) continuing to produce benefits for individuals/systems. In a review published the following year, Lennox et al25 found continuation of programme activities and continued health benefits as the most commonly reported sustainability constructs.26 Guided by Stirman et al21 and previously established definitions, for this current review, sustainability was conceptualised as the continuation of programme or programme components, or the continuation of outcomes, after initial implementation efforts, staff training or funding has ended. In terms of outcomes, our review was concerned with healthcare system outcomes (Scheirer’s organisation level of analysis), rather than patient or community outcomes.22

In addition to the limited use of operationalised definitions, Stirman et al21 also found that included studies often lacked methodological rigour and seldom used theoretical frameworks or defined measures to evaluate programme sustainability.21 Although theories and frameworks abound, with new ones continually proposed,26 there is limited recent information about the application of theories and frameworks in the healthcare system to underpin the assessment of system-level sustainability of implemented programmes.

Objectives
With an increasing emphasis on the potential threats to healthcare system sustainability as an impetus, we aimed to describe to what extent studies of healthcare improvement programmes, as implemented in the healthcare delivery system, report on programme sustainability. We also aimed to determine which theoretical frameworks have been applied, and how sustainability is defined, conceptualised and measured.

The current systematic integrative review builds on the work of Stirman et al21 and is part of a larger project investigating healthcare system sustainability.27 A detailed account of barriers and facilitators to the sustainability of implemented healthcare programmes will be reported separately.

METHODS
Protocol and registration
The published protocol for this review27 can be found at the following web address: https://bmjopen.bmj.com/content/7/11/e018568. Since publishing the protocol, we realigned our focus with Schreier’s organisational level analysis as we were interested in system and organisational level outcomes for implemented programmes in the healthcare delivery system. Modifications to our protocol are explained and justified in the corresponding sections.

Search strategy
This review was carried out in line with the Preferred Reporting Items for Systematic Review and Meta-Analyses statement (PRISMA).28 The search strategy was developed by KL, JH, LT, EM and a clinical librarian (Mary Simons). Six academic databases were searched: CINAHL, Embase, Ovid MEDLINE, Emerald Management, Scopus and Web of Science (see box 1 for search strategy example). Emerald Management was added after the publication of the study protocol to find publications related to management or operations. To capture relevant articles not discovered by database searches, we used a snowballing approach to manually search reference lists of systematic reviews.
Box 1  Search strategy example: Embase

1. (Sustainab* OR continuation OR continual OR institutional* OR re-silien* or durab* OR viab* OR stability OR stable OR persist* OR maintenance OR routin*)ti,ab.
2. exp programme sustainability/.
3. (improve* OR innovation OR reform* OR intervention OR programme* OR strateg* OR project OR plan OR change management),ti,ab.
4. health programme/ or health promotion/ or organisation/.
5. healthcare delivery/ or integrated healthcare system/.
6. 1 OR 2.
7. 3 OR 4.
8. 5 AND 6 AND 7.
9. limit eight to (human and English language and yr= “2011 -Current” AND (article or article in press OR “review”) AND journal).
10. remove duplicates from 9.* Indicates truncation.

Study selection

Data were downloaded into EndNote V8 and duplicates removed. Table 1 outlines the selection criteria applied when reviewing abstracts and full-text publications. To establish inter-rater reliability, six reviewers (KL, LT, HA, JH, GL and EM) completed a blinded review of a random 5% sample of abstracts. Any discrepancies between reviewers’ decisions were discussed by the author group, with JB and YZ acting as arbitrators. The remaining publications were randomly allocated between the reviewers who reviewed study abstracts. Rayyan, a web and mobile app for systematic reviews, was used for the blinded and full abstract review. Publications that met the inclusion criteria were subject to a full-text review using the selection criteria.

Data collection processes and data items

Data were extracted by reviewers into a purpose-designed Microsoft Excel Spreadsheet. Data items are summarised in box 2. These categories were derived from an initial review of key papers on the topic of healthcare sustainability. During regular team meetings, the categories were further refined, and descriptions were amended as the team progressed with the full-text review and data extraction. Data were extracted on study characteristics; definitions and terminology; programme evaluations, funding and evolution; theoretical frameworks; and methods and tools used to assess programme sustainability (box 2).

We recorded the timepoints at which the continuation of the programme or systems outcomes were assessed. We included evaluations of programmes which were deemed to have continued and reported system or organisational outcomes after the initial implementation phase, staff training or programme funding had ended. The reporting of multiple evaluation timepoints in many of the publications was ambiguous and therefore not reported in our review. For articles reporting on

### Table 1  Selection criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Inclusion</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Language</td>
<td>English language</td>
<td>Languages other than English</td>
</tr>
<tr>
<td>2. Types of publications</td>
<td>Peer-reviewed journal articles</td>
<td>Posters, conference proceedings, thesis dissertations</td>
</tr>
<tr>
<td>3. Types of research</td>
<td>Primary empirical research including secondary analyses of data</td>
<td>Systematic reviews, protocols, grey literature, and ‘lessons learned’ documents (unless presenting empirical data analysis)</td>
</tr>
<tr>
<td>4. Publication data</td>
<td>Published July 2011–March 2018</td>
<td>Published before July 2011 or after March 2018</td>
</tr>
<tr>
<td>5. Setting</td>
<td>Healthcare settings including hospitals, primary care, residential aged care, mental health and community health</td>
<td>Settings other than healthcare, such as environmental sustainability and primary/high school education</td>
</tr>
<tr>
<td>6. Evaluation</td>
<td>Evaluated programmes, interventions or change strategies, including studies of multiple projects</td>
<td>Models of care, evaluations of new centres, and government reforms or policies, for example, health insurance</td>
</tr>
<tr>
<td>7. Sustainability</td>
<td>Assesses sustainability of a programme from a systems or organisational view point: (A) Evaluation of a programme after funding has ended or after the initial training or implementation phase (B) Explicitly assesses sustainability, for example, stakeholders’ views of sustainability even if a programme is in its implementation phase (C) Longitudinal studies consisting of follow-up assessments or evaluations conducted over multiple timepoints</td>
<td>Studies that reported on outcomes for patients or clients only, broad public health programmes or community initiatives that did not report on system-based or organisational outcomes or impacts, pilots and studies of early implementation</td>
</tr>
</tbody>
</table>

Continued
more than one programme, the longest time frame was recorded. ‘Evolution of programmes’ referred to whether the programme had been changed, modified or adapted from its initial design.

**Data analysis and synthesis**

Our analysis and synthesis was guided by Miles and Huberman and Whittemore and Knaf. The data reduction stage involved extracting data using the purpose-designed Excel spreadsheet and frequency counting techniques. Data were displayed using matrices to aid comparisons and synthesis across studies. Data were compared and synthesised to summarise study characteristics, definitions of sustainability and terminology, healthcare programme features (e.g., funding), the use of theoretical frameworks and assessments of sustainability. Verification of the accuracy and meanings of the extracted data was undertaken by KL and LT, with YZ and JB arbitrating when questions arose.

**Quality assessment**

The quality of individual studies was assessed by KL, LT, HA, JH and GL using Hawker’s Quality Assessment Tool. This tool comprises the following domains: abstract and title; introduction and aims; method and data; sampling; data analysis; ethics and bias; results; transferability (generalisable); and implications and usefulness. An overall quality rating of low, medium or high was assigned to each study based on Lorenc et al. The reviewers first completed a blinded quality assessment of 6% of studies before each assessing a proportion of the remainder. Although in our protocol we planned to use the Cochrane Collaboration’s tool to assess risk of bias, Hawker’s Quality Assessment Tool was deemed more appropriate as it is specifically designed for assessing quality across different study methodologies.

**Patient and public involvement**

The NHMRC Partnership Centre in Health System Sustainability (PCHSS) includes among its members the Consumers Health Forum of Australia (CHF). Members of the CHF were present at meetings of the PCHSS and had opportunity to comment on this study.

**RESULTS**

**Study selection**

The search of academic databases identified 5679 records, with an additional 118 records obtained through snow-balling. The agreement rate between the six reviewers of the blinded 5% abstract review was 92%. This high rate, along with a high proportion of exclusion decisions, had the potential to reduce the value of Fleiss’ kappa, resulting in a misrepresentative kappa score. To account for this, Brenann-Prediger’s kappa was calculated at 0.84 (95% CI: 0.78 to 0.90). The results of the review strategy are detailed in figure 1. After removal of duplicates, the abstracts and titles of 4973 records were screened using the selection criteria. Four hundred and forty records were retained for full-text review, yielding 92 articles for inclusion in data synthesis.

**Study characteristics and quality assessment**

Most included studies used quantitative methods and had longitudinal study designs (table 2). Sixty-nine studies (75.0%) were classified as high quality, 20 (21.7%) were medium quality and 3 (3.3%) were rated as low quality (online supplementary file 1). Studies came from 33 countries. The studies covered high-income (n=10, 30.3%), upper-middle-income (n=5, 15.1%),

---

**Table 1** Continued

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Inclusion</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Systems outcomes</td>
<td>Focus is on changes or improvements to the healthcare system</td>
<td>Public health or prevention programmes, for example, physical activity, immunisation, smoking, contraceptive use, screening; patient-based outcomes only; community-based outcomes only; and studies of cost-effectiveness if only projected or hypothetical savings—not actual cost-savings</td>
</tr>
</tbody>
</table>

---

**Box 2** Data items extracted from included publications

<table>
<thead>
<tr>
<th>Study characteristics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Study design</td>
<td>Method</td>
</tr>
<tr>
<td>Setting type</td>
<td>Country</td>
</tr>
<tr>
<td>Description of the programme</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Defining sustainability</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitions</td>
<td>Terminology</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conceptualising sustainability</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation of sustainability: whether the focus of the evaluation was (A) the sustainability of the programme, (B) the continuation of systems-based outcomes or (C) both A and B</td>
<td>Funding of the programme</td>
</tr>
<tr>
<td>Evaluation of timepoints</td>
<td>Evolution of the programme</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Theoretical frameworks</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of framework</td>
<td>Framework details</td>
</tr>
<tr>
<td>How the framework was used</td>
<td>Stage of framework use (programme design, implementation or evaluation)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assessing sustainability</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods used</td>
<td>Tools used</td>
</tr>
</tbody>
</table>
lower-middle-income (n=8, 24.2%) and low-income (n=10, 30.3%) countries as classified by the World Bank (table 2). Almost half of the studies (n=44, 47.8%) originated from North America, and of these, 36 (81.8%) were conducted in the USA. The second-most common setting was the UK (n=9, 9.8%), followed by the Netherlands (n=8, 8.7%) and Canada (n=7, 7.6%).

Defining sustainability

Definitions

Over half of studies (n=53, 57.6%) explicitly referred to sustainability as part of the study aim. Only 27 studies (29.3%) defined sustainability, whether this was in reference to an established definition, a composite of established definitions or authors' own definition (see table 3). Thirty-two definitions were identified across the included studies. Nine pre-existing definitions were cited by multiple studies. In four of the studies, the authors provided their own definitions (table 3). Collectively, the two most frequently cited definitions from Shedia-Rizkallah and Bone’s (23) and Scheirer and Dearing’s (24) work were cited by 15 of the 27 studies (55.6%) that defined sustainability. There were 19 additional previously published definitions identified, each cited by single studies.

Terminology

The terminology used to describe sustainability varied greatly (table 3). The most commonly used terms were a variation of ‘sustainability’ or ‘sustained’, followed by ‘continuation’, ‘maintenance’, ‘follow-up’ and ‘long term’.

Conceptualising sustainability

Evaluation of sustainability

Over a third of studies (n=33, 35.9%) focused on the sustainability of a programme or its components. Thirty-seven studies (40.2%) looked at the continuation or improvement of healthcare systems outcomes, such as length of stay, hospital costs, quality of care and hand hygiene compliance. Twenty-two studies (23.9%) examined both the sustainability of programmes and systems outcomes.

Funding of programmes

A quarter of studies (n=22, 23.9%) specified whether funding had ended (n=12, 13.0%) or was ongoing (n=10, 100.9%), and two studies (2.2%) indicated that funding
Table 3  Definitions of sustainability and terminology use

<table>
<thead>
<tr>
<th>Defined sustainability</th>
<th>No of studies</th>
<th>%*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>27</td>
<td>29.3</td>
</tr>
<tr>
<td>No</td>
<td>65</td>
<td>70.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Definition</th>
<th>No of studies†</th>
<th>%††</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shediac-Rizkallah and Bone23</td>
<td>10</td>
<td>37.0</td>
</tr>
<tr>
<td>Scheirer and Dearing24</td>
<td>5</td>
<td>18.5</td>
</tr>
<tr>
<td>Pluye et al17</td>
<td>4</td>
<td>14.8</td>
</tr>
<tr>
<td>Scheirer22</td>
<td>3</td>
<td>11.1</td>
</tr>
<tr>
<td>Buchanen et a78</td>
<td>2</td>
<td>7.4</td>
</tr>
<tr>
<td>Stirman et al21</td>
<td>2</td>
<td>7.4</td>
</tr>
<tr>
<td>Slaghuis et al24</td>
<td>2</td>
<td>7.4</td>
</tr>
<tr>
<td>Procter et al79</td>
<td>2</td>
<td>7.4</td>
</tr>
<tr>
<td>Gruen et al52</td>
<td>2</td>
<td>7.4</td>
</tr>
<tr>
<td>Other pre-existing definitions§</td>
<td>19</td>
<td>70.4</td>
</tr>
<tr>
<td>Authors’ own definition</td>
<td>4</td>
<td>14.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Terminology</th>
<th>No of studies</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sustainability/sustainable/sustainably</td>
<td>70</td>
<td>76.1</td>
</tr>
<tr>
<td>Sustain/sustained/sustainment/sustainers</td>
<td>69</td>
<td>75.0</td>
</tr>
<tr>
<td>Continuation/continues/continued/continuance/continue</td>
<td>47</td>
<td>51.1</td>
</tr>
<tr>
<td>Maintenance/maintained</td>
<td>46</td>
<td>50.0</td>
</tr>
<tr>
<td>Follow-up/followed up</td>
<td>43</td>
<td>46.7</td>
</tr>
<tr>
<td>Long term/longer term</td>
<td>42</td>
<td>45.7</td>
</tr>
<tr>
<td>Adoption/adopted/adopt/adopters</td>
<td>36</td>
<td>39.1</td>
</tr>
<tr>
<td>Post/after/following:</td>
<td>31</td>
<td>33.7</td>
</tr>
<tr>
<td>trial/intervention/phase/programme/training/design/inception/project/initiation/establishment/competition/assessment/test/funding/enrolment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-implementation/after implementation/following implementation</td>
<td>25</td>
<td>27.2</td>
</tr>
<tr>
<td>Routine/routinisation/routinely</td>
<td>22</td>
<td>23.9</td>
</tr>
<tr>
<td>Institutionalised/institutional/ institutionalisation/institutionalising</td>
<td>13</td>
<td>14.1</td>
</tr>
<tr>
<td>Discontinuation/discontinued/discontinuity/not continued</td>
<td>13</td>
<td>14.1</td>
</tr>
<tr>
<td>Durability/durable</td>
<td>3</td>
<td>3.3</td>
</tr>
<tr>
<td>Scalability/scale-up</td>
<td>2</td>
<td>2.2</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>3.3</td>
</tr>
</tbody>
</table>

*As a proportion of total studies.
†The number of studies and associated percentage exceeds the total number of included studies and 100%, respectively, as some studies referred to more than one definition or term.
‡As a proportion of studies defining sustainability.
§Definitions each cited by single studies.

was not applicable. Sixteen studies (17.4%) did not report funding. The remaining 52 studies (56.5%) referred to funding or specified funding organisation(s); however, it was not clear what the length of funding was or whether funding had ended or was ongoing.

**Timepoints at which sustainability was assessed**

The majority of studies (n=80, 87.0%) provided a clear time frame between the end of initial implementation, staff training or funding and the final evaluation timepoint. An additional 10 studies (10.9%) specified that evaluations occurred post-implementation; however, a final clear evaluation timepoint was not provided. Two programmes (2.2%) were still in their implementation phase but were included in this review as stakeholders were interviewed about the future sustainability of the programme. The evaluation periods from the 80 studies providing clear final evaluation
programmes. The mean evaluation period was 40.7 months. For most studies (n=34, 42.5%), the final evaluation timepoint was between 1 and 2 years post-implementation, post-training or post-funding. Nine studies (11.3%) used an evaluation time of less than a year.

Only 11 of the 92 studies (12.0%) reported that evaluation occurred after initial programme funding had ended. Of these 11, 9 studies (81.8%) evaluated programme sustainability, 1 (9.1%) evaluated the continuation of systems-level outcomes and 1 (9.1%) evaluated the sustainability of both programme and outcomes. Eight of the 11 studies (72.7%) clearly specified the evaluation period after funding had ended, ranging from 8 months to 6 years (mean=35 months). For the other three studies (27.3%), it was not clear when the funding ended in relation to the evaluation.40–42

Evolution of programme
Thirty-four studies (34.0%) described evolution, adaptation or modification of programmes, which included for example, flexibility of the programme,43 adjustments to suit local context,44–46 incorporating feedback from front line staff,47 48 evolution of the programme over time49 50 and establishing a dedicated team responsible for continuous monitoring and making adaptations to programmes.51

Theoretical frameworks
Eighteen of the 92 included studies (19.6%) used a sustainability-related theoretical framework (online supplementary file 2). Fourteen of the 18 studies using a framework (77.8%) made explicit reference to sustainability in the study aims (online supplementary file 2). The other four studies (22.2%) reported on sustainability, despite not referring to the concept in the aims. Some frameworks were purpose-designed to evaluate sustainability (eg, Gruen et al’s dynamic model of health programme sustainability).52 Other frameworks were originally designed for other purposes and not for the assessment of sustainability, for example, the Consolidated Framework for Implementation Research53; Atun et al’s conceptual framework for analysing integration of targeted health interventions into health systems54; or Greenhalgh et al’s Conceptual Model for Considering the Determinants of Diffusion, Dissemination, and Implementation of Innovations in Health Service Delivery and Organization.55 The Reach, Effectiveness, Adoption, Implementation, Maintenance framework56 was designed to examine success across the life-cycle of a programme, including its sustainability (maintenance), and was used by two studies (online supplementary file 2).

Reasons for using frameworks
Seventeen of the 18 studies reported using a framework (94.4%) to underpin programme evaluation, with three key purposes: (1) to assess quantitative data related to outcomes; (2) to frame interview guides and (3) to inform, structure, map or verify qualitative findings. Six studies (33.3%) used frameworks for multiple purposes (online supplementary file 2).

Only 1 of the 18 studies (5.6%) used a framework to support implementation. Licskai et al used the Canadian Institutes of Health Research ‘knowledge-to-action’ framework to implement an asthma guidelines programme.57 As part of this cycle, a key action phase was ‘sustained knowledge use’. No studies reported using frameworks to support the design of programmes.

Assessing sustainability
Methods used
The most common research method used (n=46, 50.0%) to measure sustainability was the analysis of routinely collected data by healthcare organisations (eg, patient length of stay, admissions/readmission, financial data). Interviews were used in 41 studies (44.6%) and surveys in 28 (30.4%) studies. Other methods included checklists, observations, cost-effectiveness evaluations and focus group discussions. Forty-six of the studies (50.0%) used more than one method, including mixed-methods approaches.

Tools used
A small proportion of studies (n=6, 6.5%) used purpose-designed tools to evaluate sustainability. Three studies28–30 used the National Health Service (NHS) Sustainability Model and associated index.31 Despite using the same tool, the three studies reported different names: the British NHS Sustainability Index, the NHS Sustainability Survey and the NHS Institute for Innovation and Improvement Sustainability Model self-assessment tool. Two studies32 33 used the Routinisation Instrument developed by Slaghuis et al34 and one study35 used Goodman’s Level of Institutionalisation Scales.36

DISCUSSION
We need health systems and programmes that are built to last, but studies purporting to assess such systems and programmes lack definitional consistency and rigour. Our study provides a unique summary of the current application of theoretical concepts and frameworks to assess the sustainability of implemented healthcare programmes. Surprisingly, over 40% of studies describing programmes that referred to sustainability or related concepts in the title, abstract or keywords had to be excluded at full-text review because they did not assess or report on programme sustainability. Just over 57% of studies explicitly referred to sustainability in their aims, while only 29% of studies provided an operational definition of programme sustainability. This is even lower than the 35% of studies included in Stirman’s review published in 2012,21 but higher than Moore et al’s 2017 review on definitions of sustainability, where only 11.5% of included articles provided a definition.20 Unlike Stirman and colleagues,21 the majority of the studies in our review that
provided a definition of sustainability cited a pre-existing definition in the literature suggesting that evaluators are increasingly looking to the literature for definitions of programme sustainability.

The lack of a unified definition of sustainability across the literature manifested in our review through a 16-year variation in timepoints at which the authors assessed the sustainability of their programmes or programme outcomes. Of the 80 studies reporting a clear final evaluation timepoint, the majority (n=34, 42.5%) measured sustainability of a programme or continuation of outcomes 1–2 years after initial implementation, staff training or funding ended. This was similar to the 64% reported by Stirman et al.21 Just over 11% (n=9) of the 80 studies used an evaluation time of less than a year which is higher than the 6% reported by Stirman et al.21 Furthermore, some studies assessed programme sustainability only a few months after the end of initial implementation or programme funding.43 67 68 While programmes and approaches to implementation may differ, time is an important construct of sustainability, and it must be clearly articulated and justified as part of an operational definition of sustainability before undertaking an evaluation.25 The timing of a sustainability evaluation is dependent on the individual programme, outcomes of interest and whether sustainability is viewed as an outcome or as a process.25 26 Lennox et al’s systematic review, investigating approaches to healthcare sustainability evaluation, found that the measurement over time approach was used in the majority of studies.26 Although we attempted to identify studies that assessed sustainability at multiple timepoints, this was methodologically impossible as the timepoints reported were varied and often ambiguous. Formative evaluation feedback loops are thought to be essential to support successful programme implementation processes,69 however, only about a third of our included studies reported on some aspects of programme evolution, adaptation or modification, with ongoing monitoring and evaluation running alongside the implementation.

Stirman et al argued that researchers should be guided by appropriate theoretical frameworks to advance healthcare programme sustainability research.21 Many theoretical frameworks have been published to support the implementation, monitoring and assessment of healthcare programme sustainability.26 Our review suggests that in recent years, there has been little improvement in the use of theoretical frameworks to underpin assessment of the sustainability of healthcare programmes implemented in the healthcare delivery system. Stirman et al reported that 16% of studies included in their review used a theoretical framework21 and approximately 20% of our included studies did so.21 Furthermore, studies that applied frameworks mostly reported doing so at the evaluation stage rather than at the inception of programme design or implementation. This post hoc approach limits the rigour and validity of evaluation results for these programmes.70 Robust comparisons across studies were difficult because only 6.5% of studies used published tools to assess sustainability. Although three studies used the same tool (The NHS Institute for Innovation and Improvement Sustainability Model), they published different names for this tool, further challenging the ability to compare across studies.

**Study implications**

Our review revealed a lack of consistency in the way sustainability is defined, conceptualised, assessed and reported on. Furthermore, we found little improvement since Stirman et al’s review in 2012, with ongoing limited use of sustainability-related theoretical frameworks and assessment tools. We recommend that future evaluations of programmes mobilise operational definitions of sustainability, theoretically rigorous frameworks, clear and appropriate timepoints and validated assessment tools such as the Routinisation Instrument54 to evaluate sustainability of healthcare programmes. This will build a needed evidence base to support policy and investment decisions about scaling up and spreading healthcare programmes, and for them ultimately to be longer lasting.

A concerted effort is needed to move theory into practice and to support ongoing engagement between the healthcare sector and implementation science and healthcare services researchers.71 Embedding healthcare services researchers with skills in implementation science and systems science expertise in the healthcare system is one potential solution.72 Closer links at organisational level between academic and research institutes and organisations at the front lines of the healthcare system may also be helpful.

Healthcare delivery improvement programmes are considered among the essential building blocks of sustainable healthcare systems.73 The current lack of evidence consistency about the sustainability of implemented improvement programmes will inevitably limit the understanding of broader concepts such as social, economic and environmental benefits, increasingly expected from sustainable healthcare systems.74 Although artificial intelligence (AI) is also increasingly proposed as one of the solutions to support healthcare system sustainability, our review did not pick up publications in AI, possibly because this is an emerging field. We expect that future reviews will build on currently emerging work in AI75 76 to incorporate evaluations of implemented AI solutions and their contribution to health system sustainability.

**Strengths and limitations**

The search strategy was designed in collaboration with a medical librarian to enable us to capture the diversity of healthcare programmes and the disparate nature of the study methodologies and settings. We ensured our methods were rigorous through team discussions and a double-blinded sample review to ensure consistency of study inclusion and interpretation. Regular team meetings were necessary to resolve queries and divergence of opinions about which healthcare programmes to
include and what constitutes sustainability. Our review is limited to English-language studies published in the peer-reviewed literature; however, we know that many healthcare programme evaluations are not published in the public domain or are published in the grey literature. The literature search was conducted in March 2018, and therefore, this review does not include articles published in the last 2 years. We intend to review this topic again in 2 years’ time to assess the progression of research on healthcare systems sustainability. We used a systems sustainability lens for our review and programmes that assessed patient and community outcomes only were excluded; we acknowledge that positive system outcomes do not necessarily translate to positive patient or community outcomes. Additional analyses are needed to describe the complex inter-relationships among patient, community and system aspects of healthcare programme sustainability.

CONCLUSIONS

Our review uncovered lack of conceptual clarity, poor consistency of purpose and inconsistencies in defining and assessing the sustainability of programmes implemented in healthcare systems. Many studies discussed the sustainability of healthcare programmes but failed to adequately define or measure programme sustainability. Furthermore, few studies reported using sustainability-related frameworks to support programme evaluations. There is a need therefore to upskill and build capacity in teams that design, implement and evaluate healthcare programmes to ensure conceptual clarity and rigorous methodology. Consistent and unified definitions of programme sustainability are needed to enable comparisons among evaluation studies and to generate a systematic evidence base on which to make decisions about programme sustainability. The effectiveness of embedding implementation science and healthcare services researchers into the healthcare system to form collaborative teams with decision-makers and clinicians should be trialled.

Twitter Yvonne Zurynski @YvonneZurynski

Acknowledgements The authors thank Mary Simons, clinical librarian, Macquarie University, who assisted with the development of the search strategy, and Johanna Holt for helping to refine the selection criteria. The authors also thank Huey P Ting, biostatistician at the Australian Institute of Health Innovation, Macquarie University, for calculating the kappa coefficient for the blinded 5% sample abstract review and Dr Margie Campbell, Research Fellow, Centre for Health Economics Research and Evaluation, University of Technology Sydney, for identifying additional articles to include in the review via snowballing techniques.

Contributors JB conceptualised the study and led the team’s work. KL, LT, JH and EM developed the search strategy. KL, LT, JH, HA, GL and EM conducted the abstract review, full-text review and data extraction, with JB and YZ acting as arbitrators. The quality assessment of included articles was conducted by KL, LT, JH, HA and GL. KL and LT undertook the synthesis of data. KL, LT and YZ wrote the draft manuscript, with all authors providing feedback and approving the final version.

Funding This work was supported by the NHMRC Partnership Centre in Health System Sustainability (Grant ID 9100002) and NHMRC Investigator Grant APP1766320.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not directly involved in the design, conduct, reporting or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Jeffrey Braithwaite http://orcid.org/0000-0003-0296-4957
Kristiana Ludlow http://orcid.org/0000-0001-7284-5625
Hanna Augustsson http://orcid.org/0000-0001-6203-0676
Yvonne Zurynski http://orcid.org/0000-0001-7744-8717

REFERENCES


