Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Metrics and indicators used to assess health system resilience in response to shocks to health systems in high income countries—A systematic review

Pádraic Fleming, Catherine O’Donoghue, Arianna Almirall-Sanchez, David Mockler, Conor Keegan, Jon Cylus, Anna Sagan, Steve Thomas

Abstract

Health system resilience has never been more important than with the COVID-19 pandemic. There is need to identify feasible measures of resilience, potential strategies to build resilience and weaknesses of health systems experiencing shocks. The purpose of this systematic review is to examine how the resilience of health systems has been measured across various health system shocks. Following PRISMA guidelines, with double screening at each stage, the review identified 3175 studies of which 68 studies were finally included for analysis. Almost half (46%) were focused on COVID-19, followed by the economic crises, disasters and previous pandemics. Over 80% of studies included quantitative metrics. The most common WHO health system functions studied were resources and service delivery. In relation to the shock cycle, most studies reported metrics related to the management stage (79%) with the fewest addressing recovery and learning (22%). Common metrics related to staff headcount, staff wellbeing, bed number and type, impact on utilisation and quality, public and private health spending, access and coverage, and information systems. Limited progress has been made with developing standardised qualitative metrics particularly around governance. Quantitative metrics need to be analysed in relation to change and the impact of the shock. The review notes problems with measuring preparedness and the fact that few studies have really assessed the legacy or enduring impact of shocks.

1. Introduction

Health system resilience has become a concept of growing interest to academics, policy and decision makers, practitioners and health service managers, over the past decade, given the notable increase in catastrophic events that directly impact health systems. Such catastrophic events, or shocks, have included pandemics, natural disasters, wars, terrorism and financial crises. The focus on health system resilience has intensified with the rapidly emergent COVID-19 pandemic. Understanding health system resilience has therefore never been more essential. Learning from past shocks is important for health system resilience, building back better rather than reverting to the status quo, thereby poising health systems for better preparedness in the future [1,2].

While health system resilience is key to coping with catastrophic events, there is some confusion about what resilience means, how to assess it and how to strengthen it. There have been recent literature reviews [3–5] and concept papers on resilience [6–8], but there is not always consensus, with ongoing methodological explorations to develop complimentary resilience strategies for both known challenges and unpredictable shocks [9]. Most definitions focus on the health system response to a shock and how the system can absorb, adapt and transform to cope with sudden changes.

The purpose of this systematic review is to examine how the resilience of health systems was measured and assessed across various health system shocks over the past two decades. The reason for focusing on metrics is threefold:

- It helps identify feasible measures of resilience that have been used.
- It helps focus on potential strategies to improve these metrics and build health system resilience.
- It highlights potential weaknesses of health systems experiencing shocks allowing preventive action.

For this review we define health system resilience as the ability to prepare for, manage (absorb, adapt and transform) and learn from shocks [6]. This broader definition takes into account the concepts of absorption, adaptation and transformation but also sets it in the context of a dynamic shock cycle [6]. Given the complexity of health systems,
with many components, interactions and feedback [10], two pre-defined frameworks were utilized to guide and categorize the results of the review following Steve Thomas et al. [6]. First, the WHO Health System Functions outline four components of a health system (governance, resourcing, financing and service delivery). Second, the shock cycle outlines four stages of a shock and the associated system response (preparedness, onset and alert, management, and learning and legacy) [6].

This review will therefore advance our understanding of health system resilience and introduce a common language and set of metrics when discussing future health system resilience. It will also identify potential gaps in the literature and attendant data. The review question and sub-questions were as follows:

• What types of metrics and indicators have been used to assess and measure ‘Health System Resilience’ in relation to a shock, crisis or sudden change which has occurred to a health system?
  - What type of measures were used to assess weaknesses and points of vulnerability in healthcare systems?
  - What strategies should be a key focus for health system resilience?

2. Materials and methods

The aim of this systematic review was to categorize how health system resilience was assessed and reported, and based on these metrics, to discuss strategies for strengthening health system resilience. The search was completed in line with the published protocol [11] and PRISMA guidelines (Appendix 1), a summary of which will follow.

2.1. Developing the search strategy

A shock was defined as ‘a sudden and extreme change’ that impacts a health system, for example economic shocks, pandemics, sudden climate event and natural disasters. Studies that only examined ‘everyday resilience’, defined as predictable and enduring health stresses were excluded, for example general staff shortages or increasing costs of healthcare; population ageing; or seasonal flu [3,12,4].

There was no specific limit on study designs, with both quantitative and qualitative studies included, however it was essential that data collection and/or data analysis was undertaken to assess health system resilience. Studies could be prospective (for example, preparedness/forecasting) or retrospective in nature (for example, assessing impact). Cross-sectional and longitudinal designs were eligible for inclusion. Only studies written in English and Spanish were included, based on the fluency of the research team. There was no limitation in terms of the population of interest, including patients, healthcare professionals, managerial and administrative staff, policy makers and decision makers, government, and wider communities.

2.2. The search

Four databases were selected to search literature based on relevant subject matter (EMBASE, CINAHL, MEDLINE, Web of Science), in addition to a grey literature search using Google Scholar and expert advice through a stakeholder engagement process, as well as forward citation searching. The search was conducted in February 2021, with no limit based on year of study. Database specific search strings were developed and can be seen in supplementary materials (Appendix 2). An example of a search string is presented in Box 1.

2.3. Data collection and analysis

Once the search strategy was executed, the results were initially uploaded into Endnote reference manager (version 9), where duplicates were removed. The final set of articles were subsequently transferred into the screening software tool COVIDENCE to assist with both title/abstract and full-text screening, where a further five duplicates were identified and removed. Three reviewers in total conducted title/abstract and full-text screening, with the third reviewer mediating when a disagreement arose between the first two reviewers. The inclusion criteria were intentionally broad, including (1) any study design, (2) utilizing qualitative or quantitative methods, (3) that collected, analyzed and reported metrics or indicators to assess health system resilience in relation to a shock, (4) within any type of health system (universal, national insurance, private, multi-tiered). Studies were excluded based on the following criteria:

• No methods outlined/no formal data collection.
• Everyday resilience/not a sudden shock.
• Not related to health systems (for example, impact on specific disease or condition).
• Potential metrics/indicators of resilience only discussed, suggested, or recommended but not collected/used.
• A metric/indicator was planned in detail but not applied to a shock.
• Only an abstract available.
• No abstract or full text available.
• Low-and-middle income country.
• Not available in English or Spanish.

Data were extracted by one reviewer, with consultation with two other reviewers. This was initially based on descriptive study data, including: authors, year, location, healthcare setting, study population (s), type of shock, along with information related to study design and methods were initially extracted, into tabular format in Microsoft Word and Excel. A formal meta-analysis was not possible due to high heterogeneity between study type, design and data reported. Therefore, data related to health system resilience were synthesized according to (1) health system function: governance, finance, resources, and service delivery and (2) stage of shock: (1) preparedness, (2) shock onset and alert, (3) shock impact and management; and (4) recovery and learning. Full details of extracted data can be seen in the supplementary materials (Appendix 3 and Appendix 4).

Quality assessments of 193 studies were carried out on each study by two independent reviewers, including any articles identified through grey literature search or citation chasing. As the review included both quantitative and qualitative study designs, the quality assessments were based on an amalgamation of published checklists including:

• CASP Qualitative Checklist (Critical Appraisal Skills Programme);
• JBI (Joanna Briggs Institute) Critical Appraisal Checklist for Analytical Cross-Sectional Studies;
• NIH (National Institute of Health) checklists
  o Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies, and
  o Quality Assessment Tool for Before-After (Pre-Post) Studies with No Control Group; and
• MMAT (Mixed Methods Appraisal Tool).

The eight questions included in the quality assessment were purposefully focused on design and methodology since the aim of the review was to determine how resilience is measured within health systems literature, and not to determine the reliability or validity of such measures. As a result the threshold for passing the quality assessment was set high. Any study that scored less than 75% was excluded. The questions used to assess quality, along with a composite score for each study, can be seen in the supplementary materials (Appendix 5). There was disagreement on 10% of the quality assessments which were discussed until consensus was reached. Forty-five studies were excluded based on quality.
2.4. Methodological changes to the study protocol

Due to the high volume of studies after full-text screen and quality assessment, a total of 80 studies were excluded where the shock occurred in a low or middle-income country, a departure from the published protocol (Appendix 6). Countries were included if they were considered high-income at the time the review was conducted. This decision was made in conjunction with the wider research team, with broad expertise in global health systems. This was primarily taken on the basis of feasibility related to the number of excluded studies. In addition, it became apparent from the full text review that some low-and-middle income countries appear to experience and handle shocks very differently to high-income countries. This phenomenon appears to be twofold: (1) many low-and-middle income countries tended to be preoccupied with measuring the impact on infectious diseases, eradicated or under control in most high-income countries, such as measles, polio or malaria [13–19] or (2) the health system was frequently not resilient given the nature of the shock, becoming dysfunctional due to displaced populations [20–22] or political chaos [23] requiring a complete rebuilding of the health system [24], while often relying on international aid [25, 26].

Furthermore, as described in the quality assessment section, rather than using specific tools outlined in the protocol, an amalgamation of several published checklists were used, due to the heterogeneous nature of the study design and methodologies adopted. Finally, where feasible the protocol sub-questions intended to assess resilience and strategies. However, the authors narrowed the scope of the review to facilitate feasibility focusing on the measurement of resilience across health systems. Nevertheless, interpretation of these metrics may shed light on the aforementioned sub-questions.

2.5. Study limitations

Despite a robust rationale for excluding low-and-middle income countries from the review, there are associated limitations. It is possible that certain resilience metrics reported in low-and-middle income countries are relevant and informative for high-income countries. While this is true, 55% of excluded studies related to pandemics, either COVID-19 or pre-COVID pandemics – a similar proportion of included studies, reducing the possibility of missed metrics for these shock types (Appendix 6). Nonetheless, the excluded studies from low-and-middle income countries may be particularly insightful regarding under-represented shocks of growing importance. Natural disasters, for example, represented 15% of excluded studies, while countries experiencing conflict (23%) have become increasingly pertinent given the ongoing invasion of Ukraine by Russia and the associated impact globally and particularly for neighboring high-income countries.

Another limitation regards the timing of the search strategy, which concluded in February 2021 – likely excluding much literature related to health system resilience during COVID-19, particularly given the growing interest amongst academics. While covering a 20-year period and many incidences and types of shock, it must be recognized that the results of this review only relate to the early stages of the ongoing COVID-19 pandemic. Despite this, the majority of included studies related to COVID-19 and therefore represents many of the common metrics used to measure health system resilience during this time.

Finally, the concept of resilience is broad, transcending public health and health policy and crossing into many academic disciplines, however the search strategy intentionally focused on health systems, which likely resulted in the absence of metrics related to broader considerations that impact resilience and the component parts. Rather than capture all possible metrics of health system resilience, this review outlines how it has been purposefully measured to date, with a view to informing debate on appropriate strategies to measure health system resilience in future.

3. Results

3.1. Summary of search results

The search strategy returned a total of 3175 articles, including 300 from the grey literature and 12 from forward citation searching. After duplicates were removed, 1908 articles were screened for relevance based on title and abstract, where 1466 articles were excluded, leaving 442 articles for full-text screening. A further 249 article were excluded after full text screen, the majority (32%) were excluded as there were no methods described or no formal data collection. This was followed by studies examining ‘everyday resilience’ (21%), while 38 studies (15%) were not related to health systems. A further 28 studies (11%) included metrics that were only discussed but not utilised, which was similar to the 15 studies where preparedness plans were put in place but not applied to a shock. Finally, other reasons (n = 35), as outlined in Fig. 1, related to abstracts only or articles that were not found or not available in English or Spanish.

3.2. Descriptive data

There were 68 studies that met the inclusion criteria and passed the quality assessment, of which three were Spanish (Appendix 3). The majority of studies were quantitative by design (59%), followed by 16 studies (24%) that used mixed methods and finally 12 qualitative studies (18%). Of the included studies, almost half (46%) were focused on COVID-19, followed by economic crisis (26%), natural disasters (18%), pre-Covid-19 pandemic (9%) and finally one study related to health system resilience during the Croatian war (1991–1992) (Table 1).

In terms of geographical focus, almost half of the studies took place in Europe (n = 32), followed by another quarter in North America (n = 18). Six studies were conducted in Asia, with four in Australasia, two in the Middle East and one in South America. Six studies were conducted across multiple global regions.

Appendix 7 graphically depicts the distribution of studies by shock from the year 2000 to 2021. It is perhaps unsurprising that half of the studies included were published in 2020 or early 2021, with 31 of the 34 studies during this period related to COVID-19 (Appendix 7).

3.3. Data synthesis

As outlined in the methods, the key frameworks for analyzing the metrics in the included studies were by health system function and by stage of the shock cycle. The analyses are presented below.

3.3.1. Health system function

The most common health system function area studied was Resources (n = 46), followed by Service Delivery (n = 44), Governance (n = 36), with the least on Finance (n = 26). Categories of common metrics are presented under each health function in Table 2, with more detailed examples presented in Appendix 4.

3.3.1.1. Resources. ‘Resources’, including human and physical resources, was the health system function where most measures of resilience were captured. Half of these studies reported quantitative metrics (n = 23), ten reported qualitative metrics, while 13 reported both. The vast majority of studies reported a metric related to human resources, whether that was (1) an absolute quantitative measure, for example, headcount of staff (usually doctors and nurses) as a proportion of the general population; or (2) qualitative research exploring the impact of the shock on the wellbeing of the workforce. With regard to the former, rates of clinical personnel were often compared regionally or nationally within and across countries, or to EU or OECD averages. These metrics were often used to evaluate relative preparedness of a system for a disaster. Fukuma et al. [27] used these comparative metrics, for
example, to show how the Fukushima region was medically underserved, compared to national average in Japan, when accounting for physicians, nurses, ambulance calls and clinics per 100,000 residents.

With regard to qualitative metrics, these tended to focus on staff well-being, for the most part. For economic crises, these studies often examined the impact of worsening work conditions, whether that was capturing changes to wages or contractual arrangements, or the intensification of workload and the knock-on effect this had on staff motivation, self-esteem, burn-out as well as the quality of care available for patients [28–31]. For the more recent COVID-19 shock, studies also focused on well-being capturing fear and anxiety amongst staff, heightened by external factors such as poor communication from decision makers [32]. Collado-Boira et al. [33], for example, interviewed nursing and medical students in Spain about their experience and concerns about entering the workforce early to support the pandemic effort. The students described fear about infection, transmission to family, lack of PPE, lack of knowledge and concerns about decision making. These types of studies were often complemented with quantitative metrics, such as rates of absenteeism or standardized mental well-being scales [34]. Interestingly, sometimes these quantitative measures revealed unexpected results, such as better coping skills than would have been anticipated [35]. In addition to capturing staff experiences of the shocks, studies were also concerned with metrics related to available supports, for example, childminding options or flexible working conditions during COVID-19, as well as occupational health initiatives such as psychological supports or telephone support lines [36,37].

Table 1
Study design by type of shock.

|                              | COVID-19 | Economic Crisis | Natural/Man-made Disasters | Pre-Covid-19 Pandemic: H1N1 (09/10) or SARS (03) | War/Conflict | Total (%) |
|------------------------------|----------|-----------------|---------------------------|-----------------------------------------------|--------------|-----------|
| Quantitative                 | 17       | 9               | 8                         | 5                                             |              | 40 (59%)  |
| Mixed Methods                | 10       | 5               | 1                         | –                                             |              | 16 (24%)  |
| Qualitative                  | 4        | 4               | 3                         | 1                                             |              | 12 (18%)  |
| Total (%)                    | 31 (46%) | 18 (26%)        | 12 (18%)                  | 6 (9%)                                        | 1 (1%)       | 68        |

Fig. 1. PRISMA flow diagram.
Table 2
Common metrics by health system function.

| Resources (n=46) | Service delivery (n=44) | Governance (n=36) | Finance (n=26) |
|------------------|-------------------------|-------------------|----------------|
| Staff motivation, support and well-being (n=17) | Patient demographics (n=18) | Preparedness plans and enacted (n=16) | Healthcare expenditure as % of GDP (n=14) |
| Increasing capacity - physical infrastructure (n=11) | Impact on normal service delivery (n=16) | Coordination (and communication) (n=15) | Health spending per capita (n=7) |
| Nurse and doctor density per capita (n=8) | Patient activity data (n=16) | Surveillance (n=13) | Out of pocket payments (n=7) |
| Whole population hospital beds, ICU beds density, testing capacity (n=7) | Impact on timeliness/waiting times (n=8) | Information systems (n=11) | Public spend as a % of total (n=6) |
| Forecasting resources (n=7) | Telemedicine (n=7) | Reorganisation and establishment of new organisations/units (n=7) | Coverage of population/access (n=6) |
| Increasing capacity - staff (n=5) | Patient outcomes and experiences/Quality of care (n=6) | Involvement in decision making (n=5) |
| Local level staff numbers and variations (n=3) | New services created (n=4) | Transparency (n=5) |
| | Training and guidelines (n=5) | Reforms (n=4) |
| | Timeliness of response (n=3) | |

Box 1
Search string utilized for MEDLINE database (Appendix 2).

"Delivery of Health Care"/og, an OR ('Delivery of Health Care'/ AND Evaluation Studies as Topic/)
((health system* OR healthcare system* OR health care system* OR health service* OR healthcare service* OR health care service* OR health organisation* OR healthcare organisation* OR health care organisation*) adj4 (Stress test* OR metric* OR indicator* OR assess* OR measure* OR quantit* OR scale OR framework* OR analyt* OR evaluate* OR monitor* OR resilience OR resilient OR resiliency OR attribute* OR surveillance)).ab,or/1–2
((health system* OR healthcare system* OR health care system* OR health service* OR healthcare service* OR health care service* OR health organisation* OR healthcare organisation* OR health care organisation* OR health surveillance system*) adj8 (shock* OR vulnerable OR vulnerabilit* OR fragile* OR fragiliti* OR pressure* OR insecurity OR insecurity OR crisis OR crisis OR disaster* OR outbreak* OR threat* OR overwhelm* OR disrupt* OR disruption* OR disturb* OR unpredictable OR pandemic* OR epidemic* OR outbreak* OR instabiliti* OR war* OR conflict?)).ti,ab.

Finally, several COVID-19 studies were interested in capturing strategies to increase capacity. As well as numbers of new staff recruited, measures of surge capacity included, for example:

- medical and nursing students who were incorporated early into the workforce,
- retired healthcare workers returned,
- part-time staff moved to full-time hours,
- participation of the private sector,
- telehealth to extend access to remote areas,
- elective procedures delayed.

Of course, not all resources related to staff. Physical infrastructure was also measured in terms of availability of beds (general or ICU) or backup generators, for example, or innovative changes in response to the shock [38,39]. In terms of the aforementioned surge capacity for COVID-19, this took the form of ambulatory clinics converted into wards or operating theatres to intensive care units, anesthesia machines converted to ventilators, or indeed changes to standard practice for example, using ventilators for more than one patient [40,41]. Zarzylo-vych et al. [37] described a similar process in the Czech Republic, where public fundraising initiatives, in response to COVID-19, provided research and preparation for mass production of ventilators.

3.3.1.2. Service delivery. Measures of service delivery accounted for the second highest number of metrics related to a health system function. Of the 44 studies, 23 reported quantitative metrics, 9 qualitative and 12 reported both. In line with the quantitative or mixed method design of included studies, there were several examples of absolute measures both in terms of activity and quality of care. Activity metrics included number of people attending various services, for example, emergency departments, which were often disaggregated into type and origin of the referrals. Other more specific activity metrics related to:

- testing numbers,
- vaccinations,
- investigations,
- procedures (scheduled, attended, cancelled),
- treatment,
- changes in diagnoses numbers.

Many of the studies detailed the impact of shocks on normal service provision. Rios et al. [39] highlighted the impact that Hurricane Maria had on chronically ill patients. Smith et al. [42] retrospectively reviewed Emergency Medical Services activity and call types within New York City’s 911 computer assisted dispatch database to document increased patient loads on surrounding hospitals after Bellevue Hospital closure due to Hurricane Sandy.

Multiple economic crisis studies documented the negative effect on services and the factors contributing to poorer patient experiences and outcomes [28,29]. Karanikolos et al. [43] looked at a specific metrics of unmet medical need and issues related to access in the Baltic states from 2005 to 2012. The main drivers of increased unmet need were inability to afford care in Latvia and long waiting lists in Estonia, while experiences from Lithuania indicated that health policies prioritizing the maintenance and availability of existing services, can curb the deterioration of access.

Metrics to assess quality of care were often focused on health outcomes such as survival rates [44] or condition specific metrics such as cardiology [45], but they also took a staged approach to measuring service delivery, such as ‘time to treatment’ and changes to these metrics as a result of a shock. Several COVID-19 studies examined the impact on non-Covid-19 admissions to emergency departments, for example Lee at al. [46] and Mulholland et al. [47]. In order to assess impact, chronological data related to the shock were often compared to baseline data that were derived from the same time period during the previous year or an average of several years before the shock.

Other studies compared a discrete period of time, for example a defined number of weeks, before the shock followed by a similar number of weeks following the shock onset. Interestingly, these chronological analyses often revealed which metrics showed an immediate impact (hospital activity), and which required more time to reveal the full impact (quality of care, health outcomes). In the case of COVID-19, some data were more sensitive to the peaks and troughs, often directly related to the progression of the shock and associated public health measures, as was seen with the various waves of COVID-19.

3.3.1.3. Governance. Governance is of critical importance to dealing with health system shocks, but not the function most studied within this
review. This might be related to the challenge of developing suitable metrics for governance. Of the 36 studies reporting metrics related to governance, 14 reported quantitative metrics, 10 qualitative and 12 reported both. The three most common categories of measures related to:

- information systems,
- coordination across partners/coordination capacity (and communication),
- leadership/transparency/decision making.

A consistent message across the literature was the need for timely and reliable information systems and flows to enable an effective response to a shock. This became especially evident in the COVID-19 studies, showcasing how new information and surveillance systems can be set up in a relatively short time-period, facilitating timely, up-to-date and relevant information to key stakeholders, including the public [34,48–53]. During the crisis in Italy, for example, Romani et al. [40], detailed the Pagoda dashboard, created in Modena province within two weeks of the pandemic starting. The system updated every three hours, creating fully linkable demographic, hospital, emergency department, specialty, laboratory and medication data. This greatly aided regional leaders’ ability to manage the crisis. Such information systems can serve as a snapshot of the current state of the crisis but were often used to assess the efficacy of other interventions, such as government strategies and the link with the stringency of lockdown measures. Benitez et al. [41] also pointed to innovations from the pandemic, specifically with regard to some Latin-American health ministry’s centralizing information for both public and private health resources.

In terms of the 2009 H1N1 pandemic, three of the four studies evaluated surveillance and information systems. Wicker et al. [54] described a pilot scheme used where routine administrative data on antiviral medication distribution was used to track hotspots of the virus in Australia, using a low cost and adaptable way to help operational planning during an emergency. However, some limitations were also highlighted, for example, the imperfect nature of public surveillance systems during the H1N1 pandemic. Stoto et al. [55] highlighted that surveillance data is impacted (sometimes disproportionately) by a series of health-seeking and/or reporting decisions made by patients, health care providers, and public health professionals, which is influenced by a changing information environment. An information base that is less dependent on individual decision makers, requires proactive efforts to improve situational awareness by conducting prospective, representative, population-based telephone surveys and seroprevalence surveys in well-defined cohorts, to enable appropriate targeted emergency responses during shock onset [55].

Effective information systems also require leadership by decision makers, who should appropriately delegate authority – with clearly defined responsibilities of government, for example in relation to reorganization of services or licensing of medications [37]. Decisions should be evidence-based, actively involving key stakeholders and experts in decision making process, such as medical workforce, while ensuring the process is fully transparent to all stakeholders, avoiding real or perceived knee-jerk reactions [29,31,56,57].

Finally, coordination across partners is required to ensure there is sufficient capacity to respond to the shock. This requires high levels of communication and collaboration across multiple sectors. In addition to emergency management and emergency medical services, this also relies on coordinated action from law enforcement, fire services, public health, public works and the media – particularly relevant for sudden disasters where rapid information sharing is key [58,59]. The various shocks tested the capacity of the current structure to deal with the event, and often involved stakeholders working together in a way they had not done previously. Several qualitative studies reported coordination as a key component of successful outcomes [37,60,61].

3.3.1.4. Finance. The low number of studies assessing metrics around financing might indicate a gap in the literature, with most studies measuring the absolute or perceived economic impact of the shock, and to a lesser degree capturing additional funding made available post-shock. Unsurprisingly most of the included studies analyzing finances related to the economic crisis following the 2008 financial crisis, although nine also related to COVID-19 (Fig. 2). Of the 26 included studies, half reported quantitative metrics (n = 13), three qualitative and ten reported both. Across the quantitative and mixed methods studies, detailed in Appendix 3 and 4, a range of absolute measures were often captured, including:

- health expenditure as proportion of GDP,
- public health expenditure (total and%),
- private health expenditure (total and%),
- insurance cover of population,
- out of pocket payments (total and%),
- financial stability of system,
- salary costs and changes (health professionals),
- spending on pharmaceuticals.

Analyses included total funding over time, which was sometimes disaggregated into different types of funding, for example, by hospital, primary care, dental, and laboratories. Studies sometimes examined additional investments made during a shock and where these were targeted, for example additional investment in mental health or telehealth (during COVID-19) [36,62]. Finally, finance measures of resilience also examined the impact of the shock, often focusing on service users, for example out-of-pocket payments [57,63–65], loss or limitation of health coverage or entitlements [64,66,67], access, unmet need and waiting lists [31,43,66,67], stakeholder perceptions of the impact of austerity on the health system and health outcomes [29–31], as well as protective measures put in place for disadvantaged groups [68].

3.3.1.5. Health system function by shock type. As can be seen from Fig. 2, many studies examined more than one health system function. COVID-19 represents the most studied shock across all functions, with the exception of ‘Finance’, where economic crises were the main shocks under investigation. However, pre-COVID-19 pandemics or conflicts did not feature in the health system function of finance. As well as COVID-19 and economic crises, it is interesting to note that metrics related to disasters are also present across all health system functions, particularly for governance. A quarter of studies reporting governance metrics were studying a disaster related shock.

3.3.2. Shock cycle stage

Examining metrics by stage of the shock cycle also provides interesting insights. Immediately, we can see from Fig. 2 that most studies reported metrics related to the management stage (79%). Most informatively, the fewest number of metrics related to the recovery and learning stage (22%), though studies on economic crises were more prominent during this stage.

Studies related to COVID-19 strongly influenced the findings, with the exception of ‘recovery and learning’, since the pandemic was still an ongoing crisis at the time of conducting this review. A third of studies in the management stage were related to economic crises, representing the main focus for this shock type. The proportion of studies reporting metrics related to disasters ranged from 15% to 33%, across the four stages of the shock cycle. The majority of these metrics, however, appeared in preparedness and management stages (n = 8, respectively), followed by recovery and learning (n = 5).

3.3.2.1. Stage 1: preparedness of health systems to shocks. Twelve COVID-19 studies examined stage one of the shock cycle, in addition to five pre-COVID-19 pandemic studies. Many of the COVID-19 studies...
used composite measures of health system preparedness, based on pre-COVID-19 data, which were then correlated with COVID-19 outcomes such as confirmed case numbers or total deaths. Often these studies compared across countries to determine if these composite measures of the quality and preparedness could explain differences in COVID-19 outcomes.

Counterintuitively in some studies, countries with greater preparedness scores had worse COVID-19 outcomes. Ding et al. [49] proposed several reasons for this including detection capabilities, where more prepared countries tested more and therefore had higher numbers. Another reason proposed is that more industrial countries may have better prepared systems, but they may also have conditions for greater transmission and spread of the virus.

Several COVID-19 studies also alluded to the preparation effects that previous influenza pandemics had on the emergency response. Cheng et al. [48] describe the detailed hospital plan carried out during COVID-19, which was developed after SARS 2003. The plan included different structures of coordination between hospital agencies depending on the level of emergency declared within a hospital. The authors detailed the surveillance of patients at 43 public hospitals in Hong Kong looking at numbers tested and confirmed cases to chart the infection control strategy at the hospital and showed that hospital acquired infection rates were lower than the rates during the 2003 SARS outbreak in that region.

Metrics related to disasters were the second most commonly reported within the preparedness stage (n = 8). The presence of preparedness plans, surveillance data, early warning systems and communication strategies were key metrics within these studies [42,59,69–71] as was the availability of appropriately trained staff, medical infrastructure and surge capacity within these facilities [27,39,69,71]. Finally, patient activity was used as a metric to assess preparedness and surge capacity, although to a lesser degree [27,42,72].

3.3.2.2. Stage 2: shock onset and alert. Only 25 of the 68 studies reported metrics related to shock onset and alert (37%). Pandemics featured prominently within this stage of the shock cycle accounting for 80% of included studies. While there were some metrics related to governance, namely mitigation, containment, re-distribution of functions and staff responsibilities [37,41], the key focus tended to be on resources, service delivery and impact. In terms of resources, the metrics not only captured the availability, capacity or strategic measures around increasing staff and infrastructure [34,40,41,60,73], but also on testing, protecting and training staff [32,48,50,62,74], as well as new or innovative mechanisms for delivering care [36,45,62], and strengthening forecasting and reporting mechanisms [40,73,75].

Moving away from the pandemic studies, four studies examined shock onset and alert metrics for disasters and one economic crisis study focused on the first seven months of the economic crisis in Iceland, revealing a rapidly changing situation with decision-making under particular scrutiny [57]. For the disaster studies, finances appear to be of little concern, with metrics again focusing on information availability and flow [42,58,59,70].

3.3.2.3. Stage 3: impact and management. As previously mentioned, stage three of the shock cycle accounted for most of the reported metrics, across all shocks, with COVID-19 (n = 22) and economic crises (n = 18) dominating, although there were the same number of studies in this stage for disasters and pre-COVID pandemics as seen in the preparedness stage (n = 8 and 5, respectively).

While many of the management type metrics have been previously reported in the health functions analysis, ‘testing and tracing’ was a crucial pandemic specific metric that was key to management efforts [41,50,52,60,64]. In addition, impact data was particularly evident within this stage of the shock cycle. In terms of impact measures for pandemics, studies tended to focus on activity data [51,64] infrastructure [32] and the implications for non-pandemic related conditions and elective surgeries [46,53,76–78]. In terms of the workforce, impact metrics highlighted shortcomings in terms of staff shortages, knowledge, PPE availability, closed decision-making processes, supply chains [32,33,45,79], as well as physical and mental well-being of staff [32,35,50,62]. A small number of studies focused on the financial impact of the pandemics, mainly capturing spending [41,64] as well as funding sources [52,64,65]. Important impact metrics related to alternative modes of delivery (telemedicine or deploying medical staff to evacuation centres) [45,62,80], and the closure and gradual reopening of services [62]. These latter two points were also important metrics for conflict and disaster management studies [42,70,81,82], as was coordination and cooperation of the wider emergency response community [39,81,83].

In terms of economic crises, the metrics related to governance during management stage reveal key shortcomings, such as corruption, lack of accountability, planning, transparency and stakeholder buy-in to decision making [29,56,57,66], although there was also evidence of opportunities for reform [57,66,68]. In terms of measuring impact, metrics related to staff not only measured the number of staff [30,66,84,85] but crucially the adverse impact of the economic shock on staff motivation, workload, burnout, emigration, turnover [28,29,86–88], supplemented by metrics that allude to adverse effects, such as restrictive policies in the delivery of care, rationing, quality of care, reduced infrastructure and patient outcomes [28,29,43,57,67,68,85].

3.3.2.4. Stage 4: legacy: recovery and learning. While few studies (n = 15) focused on the final stage of the shock cycle, key metrics highlight legacies for the health system in general and specifically for future shocks. Since this review purposely focused on the first year of the ongoing COVID-19 pandemic, unlike stages one to three, COVID-19 does
not dominate the literature of the final stage of the shock cycle (n = 3). Instead historical economic shocks (n = 6) and disasters (n = 5) feature more prominently, in addition to one pre-COVID-19 pandemic (SARS).

Many of the economic metrics focused on the impact of the crisis, as outlined in stage three of the shock cycle. Yet legacy issues persisted beyond the crisis and metrics included the long term funding of the health system, with the sustainability and efficiency of health insurance funds being questioned in Romania, consistently under-resourcing population needs and requiring state top-ups to cover deficits during the crisis [86]. Financial constraints, introduced to counteract these inefficiencies, further impacted the relatively poor pay and working conditions, leading to higher emigration rates and resultant workforce supply issues [86]. Other sustainability metrics reported related to persisting access problems as a result of high direct costs, including informal payments [56].

Similarly, the disaster literature noted and evaluated metrics around legacy, highlighting a lack of integration of care and information flows between ambulance services, regional and national, hospitals, as well as noting problems with measuring back-up infrastructure (generators) and, disrupted care for those with chronic diseases during and after Hurricane Maria in Puerto Rico [39]. Metrics highlighting problems of coordination between hospitals was repeated for the COVID-19 pandemic [48], while information and analyses of lessons learned from past pandemics/economic crises were recognized as a strength, particularly for financial and governance considerations [53,65].

Keys lessons were also highlighted regarding human resources, specifically for staff protection and surveillance [50] and the need to develop HR policies to combat emigration of physicians following a shock [86]. Lorenzoni et al. [83], having studied multiple disasters, also indicated the need for greater coordination between organizations by, for example, establishing working groups and new response units and updated emergency plans.

To facilitate a coordinated response amongst agencies, Ryan et al. [70] and Economou et al. [66] highlighted the need for accurate and easily accessible data, as well as the ability to easily share data amongst agencies. In line with the cyclical nature of the shock cycle, data from disasters, for example, could be used to inform future preparedness by utilizing the data to pre-plan with medical suppliers, to plan the location and prepare supply needs for treatment hubs and to develop strategies for targeted evacuation plans, by identifying with high-risk populations [70]. These data could also be supplemented by ongoing public health surveillance [72].

4. Discussion

4.1. Breadth of resilience and scope of measurement

This systematic review has deliberately sought to identify metrics that relate to the handling of shocks on health systems. In so doing the focus has been broader than articles which formally recognize health system resilience according to one of the prominent definitions. While the review omits resilience in low-and-middle income countries, as well as COVID-19 specific literature published after February 2021, it does cast the net over a 20-year period and across a variety of shocks. This allows a more thorough analysis of the ways that health system resilience is measured in practice. In so doing the authors have also mapped out the different points of interest and impact of different kinds of shocks through the metrics used to evaluate resilience over the past two decades.

An interesting question is how to benchmark these indicators particularly where they are related to levels of resources. In some cases these are standardized measures like health spending in relation to the size of the economy or numbers of beds or staff per population. However the interpretation of these statistics is often not developed fully. Questions remain as to what the appropriate level of additional resources is and not just how have resources changed in response to the shock. Further work is needed to understand this issue more fully.

Recent literature drawing on lessons learned from the COVID-19 echoes many of the metrics identified in this review in relation to governance, for example the importance of surveillance infrastructure, data generation, analysis and transparent national and international communication [2]. It does however go beyond the somewhat limited perspective outlined in the studies included in this review, highlighting important aspects such as political leadership and their ability to rapidly and decisively respond during shock onset, in an adaptive way – informed by emerging scientific evidence, while also considering their constitutional means to implement these decisions, all while maintaining public trust and political consensus [2,89]. Pivotal to successful governance lies community engagement and participatory approaches, not least, to counteract the growing threat of misinformation - in the case of COVID-19, often targeting vaccine efforts amongst marginalized populations [89].

A rethinking of organizational resilience is also important, with novel approaches proffered, such as “real-time anticipatory response”, whereby Wells et al. [99] describe an organizational process that concurrently prepares for and responds to a shock, in this case COVID-19 but applicable to many shocks, when organizations experience high pressure and uncertainty. The lack of metrics directly related to these considerations point to a gap in the health systems resilience literature, pointing to the need careful consideration in terms of how best to prepare for, implement and evaluate good governance.

4.2. Type of measurement

This review identifies quantitative and qualitative measures of health system resilience both of which are important for measuring resilience. Nevertheless both have associated challenges. First, for quantitative metrics the challenge is one of dynamics. Frequently health system resilience is only revealed in how key metrics change in response to the shock over time. Hence a once-off measurement is not enough but must be assessed over time. Consequently, metrics of system responses and resilience may only be understood appropriately as we relate them to the severity and dynamics of the shock itself.

Yet there are some aspects of health system functioning which are difficult to assess with quantitative metrics, such as governance. However qualitative metrics are frequently harder to specify and identify. They also tend to be less standardized and in the included studies are rarely precisely defined. Hence identifying qualitative indicators has required discussion and judgement across the review team. Nevertheless, they are an essential component of measuring resilience and more work may need to be done on evaluating whether more standardization can be achieved across countries.

4.3. Links between metrics and strategies

By revealing the metrics used in the included studies the data are a helpful source for policy makers, those working in the frontline and researchers alike, revealing strategies to enhance health system resilience across different health system functions. Common focuses for metrics and strategies for building resilience are listed:

4.3.1. Workforce—surge capacity

• Students incorporated early, retired healthcare workers returned, part-time staff to full-time staff, recruitment drives, elective procedures delayed.

4.3.2. Workforce wellbeing

• Childminding, training supports, ‘support lines’ set up for healthcare workers, flexibility, psychological support.
4.3.3. Physical infrastructure—COVID-19 surge capacity

- Surgical operating rooms converted to ICU units, anesthesia machines converted into ventilators, ventilators used for more than one patient, ambulatory clinic spaces converted to inpatient wards.

4.3.4. Funding strategies

- Additional funding (mental health, telehealth, for disadvantaged groups).

4.3.5. Governance

- Improved transparency, medical workforce involvement in decision making, coordinated action, effective collaboration across sectors, high communication.

The analysis therefore confirms and builds on earlier studies identifying effective resilience building strategies [3,6,4]. It is clear, however, that metrics in this review tend to relate to the absorptive and adaptive capacity of the health system, however complex adaptive systems also need to be accompanied by transformation, particularly when adaptive capacities are exhausted - often the tipping point for system innovation [91]. It would therefore be prudent to broaden resilience strategies beyond shock-specific metrics, as reported in this review, instead looking to ensure health systems can function well regardless of shock type. As Barbash and Kahn (2021) argue, for example, hospital resilience should not be overly focused on increasing ICU bed capacity or increasing stockpiles based on COVID-19 deficiencies, instead tackling systemic changes such as building robust supply chains, streamlining systems for coordinated responses and promoting cultures of excellence and collaboration [92]. These recommendations align with a rapid synthesis of international lessons from COVID-19 recommending twenty key strategies, covering adaptive and transformational resilience, across five domains of: (1) leading and governing the response; (2) financing services; (3) mobilizing and supporting the health workforce; (4) strengthening public health interventions; and (5) transforming the delivery of health and social care services [93]. Building upon these strategies, Angeler et al. [91] also highlight the important interplay between social-ecological systems and health system resilience, promoting social transformation, for example, by helping institutions and the wider public to understand health system resilience through targeted, evidence-based outreach. Finally, the measurement of health system resilience requires a less siloed view, instead incorporating external factors that take into account different population needs and contexts, for example, older people who were particularly vulnerable and disproportionately impacted during the COVID-19 pandemic [94].

4.2. Reflections on shock cycle dynamics

The breadth of different metrics and indices and therefore strategies related to preparedness across the different kinds of shocks is clear but presents a problem for governments. To attempt to be prepared for everything might be difficult, costly and time-consuming. Alternatively, to pick and choose preparedness strategies is to risk being prepared for the wrong thing and potentially even weakening the health system toward a shock by preparing the system for another that didn’t happen. Perhaps this then explains why some of the scores around preparedness do not predict how well a health system responds to a shock. A better strategy may be to ensure that the health system is functioning well. In addition it might also be useful to carry some system inefficiency as good preparedness. This might relate to reserves for financing or spare capacity in hospitals or even additional flexible workforce. This may then allow quick deployment of these resources to meet the needs of health system shocks.

There is less of a focus on legacy, learning and recovery than other elements of the shock cycle. Policy makers attempting to deal with a shock will either consciously or inadvertently create a legacy for their health system for many years. Dealing with this more explicitly will help build future health system resilience and performance [95].

5. Conclusions

COVID-19 has highlighted the importance of health system resilience and produced a massive flow of new articles on the topic. Nevertheless, health system resilience is much broader than metrics related to COVID-19 or even pandemics and covers a range of different types of shocks, each with their own features and peculiarities. This systematic review has mapped out the measures used in assessing health system resilience over the last twenty years in high income countries. Rather than judge the appropriateness of these metrics, this review presents how health system resilience has been measured to date, highlighting both common metrics around health system performance and resilience, and metrics associated with resilience building strategies. Across the health system functions, metrics around financing receive the least attention. The review also notes the limited progress made with developing standardized qualitative metrics particularly around governance and the need for quantitative metrics to be analyzed in relation to change and the impact of the shock. The review notes also the problems with measuring preparedness as a strategy even when indicators exist and the fact that few studies have really assessed the legacy or enduring impact of shocks. Consequently, there is much research needed to understand and build resilient health systems for the known chronic challenges and unpredictable shocks ahead.

Credit author statement

Padraic Fleming: Data curation; Formal analysis; Methodology; Project administration; Validation; Visualization; Roles/Writing - original draft. Catherine O’Donoghue: Data curation; Formal analysis; Methodology; Project administration; Validation; Visualization; Writing - original draft. Arianna Almiral-Sanchez: Formal analysis; Validation; Writing - review & editing. David Mockler: Data curation; Methodology. Conor Keegan: Writing - review & editing. Jon Cylus: Writing - review & editing. Anna Sagan: Writing - review & editing. Steve Thomas: Conceptualization; Data curation; Formal analysis; Funding acquisition; Methodology; Supervision; Validation; Visualization; Writing - original draft.

Declaration of Competing Interest

There are no conflicts of interest to declare.

Acknowledgments

The authors would like to acknowledge the wider RESTORE research team and the international advisory committee, including public and patient (PPI) representatives, for their oversight and contributions to discussions throughout the review process. Finally, the authors would like to thank the Health Research Board for all their support. This research was funded by the Health Research Board [grant no. RLA-2020-001].

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.healthpol.2022.10.001.
