Much research, but little learned to date: A scoping review on mental health of health workers in low- and lower-middle income countries and roadmap for future research

Julia Lohmann (✉ julia.lohmann@lshtm.ac.uk)

London School of Hygiene & Tropical Medicine <https://orcid.org/0000-0003-4136-9296>

Denny John

Amrita Institute of Medical Sciences and Research Centre: Amrita Institute of Medical Sciences

Aso Dzay

London School of Hygiene & Tropical Medicine

Research

Keywords: Mental health, psychological wellbeing, health workers, low- and lower-middle-income countries, scoping review, methodological quality

Posted Date: February 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-215108/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

SARS-CoV-2 has resulted in unprecedented research efforts on health workers' work realities and their potential mental health impacts. To understand the latter, high-quality evidence on the baseline situation is paramount. With the aim of providing a comprehensive overview of existing evidence and to inform future research, we undertook a scoping review of the quantitative literature on mental health and psychological wellbeing of clinical skilled healthcare personnel working in all settings of care in low- and lower-middle income countries (LLMIC).

Methods

We performed a systematic search of the literature up to the end of 2019, in English or French language, in MEDLINE, EMBASE, PsychINFO, Global Health, and CAIRN. We included both studies estimating levels of mental health and studies investigating associations with other factors. We extracted data on study characteristics and methodology, and assessed the methodological quality of the included studies along nine criteria.

Results

We found 143 relevant studies, 135 including data on mental health levels and 126 including data on associations with other factors. The studies covered 26 of the world's 78 LLMICs, with most studies conducted in India, Nigeria, Pakistan, or Egypt, in urban secondary- and tertiary-care settings. 67% of studies assessed burnout, 25% general psychological wellbeing, and 20% other mental health outcomes. Only 19% of studies were of high quality due to shortcomings particularly in regards to sample representativeness, context-specific measurement tool validity, and reporting of methodological detail. We found much heterogeneity in investigated associated factors. Studies focused almost exclusively on potential determinants of mental health, while none linked mental health to objectively measured performance outcomes.

Conclusion

We conclude that despite its impressive size, we can learn comparatively little from the current body of literature. Based on our findings, we outline areas for expansion, methodological improvement, and standardization of reporting in future research on mental health of health workers.

Systematic review registration

PROSPERO no. CRD42019140036

Background

The SARS-CoV-2 pandemic has painfully reminded the global health community of – and alerted the general public to – the difficult and often precarious conditions among which health care professionals all over the world work to prevent and avert illness and deaths, not only in times of crisis, but also in “normal times”. Emerging studies show the impact on health workers’ mental health and psychological wellbeing [e.g. 1–6], mirroring what was also observed following earlier epi- and pandemics [7–9]. Mental health and psychological wellbeing of health workers have consequently risen high on the agenda of policy makers and researchers [10, 11].

Researchers aiming to establish to what extent the pandemic has affected health worker mental health necessitate robust and representative data on the pre-pandemic baseline situation. Pre-pandemic research is widely available for high-income countries (HIC) [e.g. 12–15], demonstrating often alarming levels of burnout and other mental health conditions, identifying a large number of determinants, and linking poor mental health to adverse consequences for patients and the health system at large.

For non-HIC settings, however, four recent systematic reviews on the mental health of health workers in low- and middle-income countries indicate a general paucity of data, regarding both the magnitude of the “problem” as well as risk and protective factors [16–19]. However, the reviews’ focus on burnout, on specific geographical areas, and partly also on only specific care settings do not allow for a comprehensive picture of the research landscape regarding mental health of health workers in non-HIC countries. Three out of four reviews performed a quality assessment of the reviewed research, implying substantial room for improvement. However, the quality assessments were done at a rather high level of abstraction as is the standard for the purpose of a systematic review, but do not allow to distil concrete areas and action points for methodological improvement.

Consequently, the four existing reviews are not sufficient in providing researchers and policy makers with a comprehensive overview over the data and literature available. With a view towards current and future research activities, they do not allow to draw sufficiently robust and comprehensive conclusions to inform a strategic research agenda aimed at providing tangible, high-quality evidence to researchers, funders, and policy makers wishing to work towards improving the psychological wellbeing of their healthcare workforce.

This review therefore aims to provide a comprehensive overview of the quantitative literature on mental health and psychological wellbeing of clinical skilled healthcare personnel working in all settings of care in low- and lower-middle income countries (LLMIC) worldwide. Specifically, we aim to highlight gaps in geographical scope, target population scope, and mental outcomes, as well as methodological strengths and weaknesses in the current body of literature. Based on our findings, we outline areas for expansion, methodological improvement, and standardization of reporting in future research on mental health of health workers. For researchers engaged in assessing the impact of the SARS-CoV-2 pandemic on health workers’ mental health, we hope that this review provides a useful resource.

Methods

Study design

We performed a systematic search of the literature on mental health and psychological wellbeing of clinical skilled healthcare personnel working in all care settings in low- and lower-middle income countries worldwide, up to the end of 2019. The review had initially been intended as a systematic review and meta-analysis of the available evidence; the protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO) (registration number: CRD42019140036), and has been published [20]. However, following the search and screening steps and as the high number of relevant studies and the methodological issues highlighted in this article became apparent, we concluded that it would be preferable to redefine our research question also as a precursor to future research and systematic reviews [21, 22].

Specifically, the review answers to the following questions: Which quantitative evidence is available on the prevalence as well as on associated factors of mental health and psychological wellbeing of clinical skilled healthcare personnel working in all settings of care in low- and lower-middle income countries worldwide? Which gaps in geographical scope, target population scope, and mental outcomes exist, consequently? What are methodological strengths and weaknesses in the current body of literature?

We therefore adopted a scoping review approach in line with the methodology proposed by Arksey and O'Malley [23] and modified by Levac, Colquhoun and O'Brien [24], which defines as key step following the definition of the research question: the identification of the relevant studies; the selection of the studies; the data charting; and the collating, summarizing and reporting of the results. These steps are described in detail in the following.

Important to note is that the change in research question from the initially intended systematic review and meta-analysis did not affect search or in- or exclusion criteria in any way, and that the identification and selection steps for scoping reviews correspond exactly to the search and selection steps in a systematic review.

In writing the manuscript, we followed the essential reporting items as proposed by the PRISMA extension for scoping reviews [25]. The respective checklist is provided in Additional Files 1.

Study identification

Eligibility criteria

In regards to study participants, we considered studies referring to:

- Formally and fully trained health professionals and health associate professionals [26, 27], specifically medical doctors, nursing and midwifery professionals, and nursing and midwifery associate professionals;
- Working in formal health care facilities (public, private not-for-profit, private for-profit);

- Working in low- and lower-middle income countries as per the World Bank's 2019 classification [28].

We focused on LLMIC rather than all low-and middle-income settings in line with recent developments in income classification at international level in response to the realization of dramatic variation within the middle-income category.

We excluded studies focusing exclusively on non-clinical or not formally or not yet fully trained personnel as well as on exclusively community-based personnel (e.g., pure management or administrative personnel, traditional or lay health workers, community health workers, students and other health workers in training), studies with an exclusive focus on non-LLMIC, and studies on migrant health workers from LLMIC to HIC.

In regards to mental health outcomes, we considered studies on burnout, depression, anxiety, trauma, general psychological wellbeing and/or distress, as well as other specific mental health/distress diagnoses or concepts if work-related and explicitly framed as a mental health issue. We considered all studies labelled by the authors as investigating the above, irrespective of whether they referred to an international disease classification system (such as ICD-10/11 or DSM-VI/V) and irrespective of the measurement tool used. We did not consider studies on stress, job or life satisfaction without specific reference to mental health.

In regards to types of studies, we included all relevant observational study designs as well as control arms of randomized controlled intervention trials published in English or French language. Where multiple papers were generated from the same data looking at the same outcome, only the most relevant/recent paper was included. However, if multiple papers were generated from the same data with different outcomes or on different subpopulations, all papers were included. We did not include qualitative studies, previously published systematic reviews (although we cross-checked included articles), opinion pieces, commentaries, policy briefs, and conference abstracts without identifiable full-text.

Information sources and search strategy

We searched for eligible studies published up to December 2019, in English or French language, in MEDLINE, EMBASE, PsychINFO, Global Health, and CAIRN. Specifically, we performed a first search in June 2019 in MEDLINE, EMBASE, PsychINFO, and Global Health, and in February 2020 updated this search to the end of December 2019 and further searched CAIRN for additional French-language articles published up to the end of December 2019.

The search strategy included a combination of subject terms and free-text terms from three categories based on the inclusion and exclusion criteria: (1) geographic focus: all LLMIC as well as overarching regional terms; (2) population: generic terms for healthcare professionals as well as terms for specific health worker cadres; and (3) outcomes: specific terms for burnout, depression, and psychological wellbeing, generic terms for mental health/illness and work-related psychological stress/distress/trauma, and terms for specific common measurement instruments. In regards to the latter, we also included "motivation" and "satisfaction" as search terms, based on our experience that studies labelled as such

sometimes contain mental health measures as part of the motivation or satisfaction measurement tool. The search strategy for MEDLINE is provided in the protocol [20]. Search strategies for the other databases were adapted according to their specific requirements.

Study selection

Given the purposively broad search strategy, the search resulted in a total of 8,687 unique studies. AD and JL independently examined titles and abstracts of a subset of the studies against the eligibility criteria, comparing and discussing results until convergence. The remainder of the study titles and abstracts were screened by a single researcher. We retained and retrieved 435 studies for full text screening. The high number of initial search results compared to retained studies was largely due to the search algorithm picking up studies on mental health of non-health workers, which however made mention of treatment and thereby the health workforce in the abstract. Full-text screening was performed in full double screening by AD and JL. Discrepancies were minimal and resolved in discussion. 114 studies were retained following this stage, whereas 321 were excluded. Of the latter, the vast majority were studies on motivation and satisfaction, which we had retained for the full text screening to ensure they do not include “hidden” psychological wellbeing measures. The remaining studies were excluded because they measured non-specific stress or substance abuse (usually not clearly marked as mental health problem and including eg. smoking), were qualitative studies, were exclusively conducted in an upper-middle or high-income country, or included only community health workers, trainees or students, or other health worker groups outside our inclusion criteria. Despite the broad search strategy, 7 relevant articles included in one or several of the three prior systematic reviews [16–18] (Kesarwani, Husaain and George [19] was published after the search and screening was completed) were not picked up by our search and therefore manually included. A screening of the reference lists of the 121 resulting articles led to the inclusion of a further 22 studies. Figure 1 outlines the study search and selection process.

Data charting

Data extraction

From each included study, we extracted the information summarized in Table 1 into an Excel-based data extraction form developed ex-ante. JL and AD independently extracted data from a subset of studies and compared and discussed results until reaching convergence. Data extraction was then continued by a single researcher, as full double extraction would have not been feasible within the scope of the project given the large number of studies.

Table 1
Extracted data elements

Category	Data elements
Basic study information	Journal, year of publication, year of data collection, study country(ies), mental health outcome(s) measured, study type (levels vs associated factors assessed)
Study setting and population	Multiple vs single region/city/facility, number of health facilities, level of care, sector (private, public), urban vs rural setting, health worker cadres, further information on inclusion and exclusion criteria; adequacy of the description of study setting and population
Study sample	Sampling entry point (health care facilities vs other), sampling strategy, reasons for sample size, response rate, actual sample size (overall, by cadre, by gender), basic sample characteristics
Mental health outcome(s)	Measurement tool name, tool details (number of items, response scale incl numeric representation of anchors) and references provided, translation, pretest, reliability, validity, reporting of findings (proportions, continuous), calculation of composite scores, categorization into good/poor mental health groups, relevant analytical details (eg adjustment for sample composition), findings (overall, by cadre, by gender)*, clarity of results tables and description
Associated factors	Factors investigated, including whether the paper made clear which factors were considered and how these were measured; analytical techniques (bivariate vs multivariate); reporting of all associations vs significant ones only; statistics used for reporting of associations (e.g. correlation coefficients, odds ratios, beta weights, etc.)
* We extracted data regarding mental health levels found by the studies, but decided not to report them as explained above in "Study design" ; for associated factors, we did not extract data on findings as we realized even before the data extraction stage that this would exceed the capacity of the study team	

Methodological quality assessment

In relation to the study aim of highlighting methodological strengths and weaknesses, based on the extracted information, we then evaluated the included studies assessing levels of mental health in regards to nine quality criteria pertaining to the availability of information on and adequacy of basic study characteristics, study population and sample, outcome measurement, and reporting of results.

The quality criteria are described in detail in Additional Files 2 and explained in the respective results sections. In the absence of specific quality checklists for mental health prevalence studies, the quality criteria were developed by adapting and expanding commonly used quality checklists for systematic reviews, notably the JBI Checklists for Analytical Cross-Sectional Studies and for Prevalence Studies [29], the CASP Case Control Study Checklist [30], and the NIH Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies [31].

While we borrowed heavily from these checklists, our quality criteria contain more detail on measurement and validity aspects in line with the specific importance of these aspects in the measurement of mental health constructs, and focus less on study results in line with the early change in study focus away from a systematic review and meta-analysis and towards a scoping review outlined above.

JL and AD independently evaluated a subset of studies on the quality criteria and compared and discussed results until reaching common understanding indicated by convergence in assessment. The resulting studies were assessed by only one researcher, as full double assessment was beyond the capacity of the team given the large number of studies.

We did not assess the quality of studies investigating factors associated with mental health outcomes for reasons of capacity. As reported in the results, interpretability and comparability of findings varied not only from study to study, but also within studies by the specific factors assessed and analytical techniques used. This would have necessitated a quality assessment by factor assessed rather than by study, exceeding the scope of the project.

Collating, summarizing and reporting of the results

In a final step, we summarized and organized the extracted data by themes, prioritizing the results based on relevance to the research questions as per Arksey & O'Malley's [23] framework. We summarized and described data separately for studies having assessed levels of good or poor mental health, as well as for studies having assessed factors associated with mental health. In regards to the former, we further described data separately for studies with an explicit study aim to assess mental health prevalence as explained in more detail below. All analysis was descriptive in line with the descriptive study aim to describe scope and methodological strengths and weaknesses of the literature.

Results

The search and selection process (Fig. 1) resulted in a total 143 articles mental health or psychological wellbeing among clinical, skilled healthcare personnel in low- or lower-middle income countries published up to December 2019. Of those 143 articles, 118 included data on both levels of mental health as well as on one or several factors associated with these levels; 17 studies reported only overall mental health levels in their samples, without any breakdown by sample subgroup or further associated factors; and 8 studies reported only associations with other factors, without presenting any absolute data on mental health levels.

There is a clear increase in availability of studies on levels of mental health of health workers over time, with half of the included articles being published in 2015 and later and less than 5% before 2005 (Fig. 2). Of note, 40% of studies did not report the year in which data collection took place (*quality criterion 1*).

88% of studies assessed only one mental health outcome, 11% two outcomes, and 1% three outcomes. As described in more detail below, 67% of studies assessed burnout, 25% general psychological wellbeing, 10% depression, and 10% other mental health outcomes. Given this relative homogeneity, particularly in light of an on-going debate around whether it is clinically meaningful to differentiate burnout and depression, especially when using self-reported measures [32], as well as the large heterogeneity in measurement tools used for general psychological wellbeing, we decided against organizing results by mental health outcome as might have been most intuitive.

Rather, in the following, we first present data related to availability and characteristics of studies on mental health levels, by study setting, study population, and study methodology and outcomes, along with findings related to the methodological quality of the included studies. The second part of the results section presents studies on correlates of mental health levels, with a focus on the investigated factors and on methodological aspects.

A comprehensive overview of all included studies including key characteristics can be found in Additional Files 3.

Research on levels of good or poor mental health amongst health workers in LLMIC

Our broad search strategy allowed us to include both studies with an explicit aim to assess levels of mental health as well as studies with a different primary study aim, but having reported data on mental health levels in the process. 84 of the 135 included studies (62%) explicitly aimed to assess mental health levels. For simplicity and in line with how the overwhelming majority of authors have called their studies, we will call these studies “prevalence studies” in the following.

The remaining 49% included studies with a focus on assessing factors associated with mental health, studies where mental health was one predictor variable among many for another substantive outcome variable, studies testing specific hypotheses including mental health variables (usually from the psychology literature), or validation studies of tools to measure the mental health outcome or a different construct.

This section provides an overview over findings from studies investigating levels of mental health. Key results pertaining to study characteristics are summarized in Table 2, overall as well as for the subset of explicit prevalence studies. Results regarding study quality are summarized in Table 3.

Table 2
Studies assessing levels of mental health - key results by study type

	All studies	Prevalence study subset
Total number of studies	135	84
Study region		
WHO Africa Region	56 (41.5 %)	34 (40.5 %)
WHO Eastern Mediterranean Region	39 (28.9 %)	23 (27.4 %)
WHO South-East Asia Region	35 (25.9 %)	23 (27.4 %)
WHO Western Pacific Region	5 (3.7 %)	4 (4.7 %)
Study setting*		
Urban	118 (87.4 %)	77 (91.7 %)
Rural	39 (28.9 %)	22 (26.2 %)
Multi-site	90 (66.7 %)	52 (61.9 %)
Study population*		
Primary level of care	35 (26.3 %)	20 (23.8 %)
Second and/or tertiary level of care	126 (94.7 %)	80 (95.2 %)
Public sector	123 (91.1 %)	75 (89.3 %)
Private sector	45 (33.3 %)	28 (33.3 %)
Medical doctors	72 (53.3 %)	50 (59.5 %)
Nurses	96 (71.1 %)	55 (65.5 %)
Other clinical staff	38 (28.2 %)	26 (31.0 %)
Other managerial staff	3 (2.2 %)	3 (3.6 %)
Sampling		
Census	50 (37.0 %)	40 (47.6 %)
Random / stratified random sample	24 (17.8 %)	15 (17.9 %)
Convenience sample	37 (27.4 %)	17 (20.2 %)
Unclear	24 (17.8 %)	12 (14.3 %)
Study outcomes*		

*Proportions do not add up to 100% as many studies include mixed settings/study populations and/or several outcomes

	All studies	Prevalence study subset
Burnout	90 (66.7 %)	57 (67.9 %)
Depression	14 (10.4 %)	10 (11.9 %)
Anxiety	10 (7.4 %)	7 (8.3 %)
Trauma	5 (3.7 %)	3 (3.6 %)
General psychological wellbeing	33 (24.4 %)	20 (23.8 %)
Outcome measurement*		
Continuous outcome	73 (54.1 %)	36 (42.9 %)
Proportions / categorical outcome	90 (66.7 %)	71 (84.5 %)
Results reporting		
By cadre: yes, by design	87 (64.4 %)	52 (61.9 %)
By cadre: yes	22 (16.3 %)	17 (20.2 %)
By cadre: no	26 (19.3 %)	15 (17.9 %)
By gender: yes, by design	14 (10.4 %)	6 (7.1 %)
By gender: yes	47 (34.8 %)	36 (42.9 %)
By gender: no	74 (54.8 %)	42 (50.0 %)
*Proportions do not add up to 100% as many studies include mixed settings/study populations and/or several outcomes		

Table 3
Studies assessing levels of mental health - study quality by study type

	All studies	Prevalence study subset
Qual 1: Is the year of data collection mentioned?		
1 Mentioned	82 (60.7 %)	56 (66.7 %)
0 Not mentioned	53 (39.3 %)	28 (33.3 %)
Qual 2: Is the study setting adequately described?		
2 Fully adequate	116 (85.9 %)	75 (89.3 %)
1 Mostly adequate, with only some ambiguity	10 (7.4 %)	4 (4.8 %)
0 Insufficient	9 (6.7 %)	5 (5.9 %)
Qual 3: Is the study population clearly specified and defined?		
2 Fully clear	129 (95.6 %)	82 (97.6 %)
1 Mostly clear, with only some ambiguity	5 (3.7 %)	2 (2.4 %)
0 Unclear	1 (0.7 %)	0 (0 %)
Qual 4: Is the sample likely to be representative for the intended study population?		
2 Highly likely	30 (22.2 %)	19 (22.6 %)
1 Somewhat likely	46 (34.1 %)	36 (42.9 %)
0 Unlikely	59 (43.7 %)	29 (34.5 %)
Qual 5: Are the study participant characteristics described in sufficient detail?		

	All studies	Prevalence study subset
2 Fully adequate	112 (83.0 %)	72 (85.7 %)
1 Mostly adequate, with only some ambiguity	15 (11.1 %)	9 (19.7 %)
0 Insufficient	8 (5.9 %)	3 (3.6 %)
Qual 6: Is the tool(s) used to measure the outcome(s) adequately reported?		
2 Fully adequate	115 (85.2 %)	74 (88.1 %)
1 Mostly adequate, with only some ambiguity	10 (7.4 %)	7 (8.3 %)
0 Insufficient	10 (7.4 %)	3 (3.6 %)
Qual 7: Is convincing information on the validity of the tool(s) used to measure the outcome(s) reported?		
2 Convincing	11 (8.2 %)	5 (6.0 %)
1 Partially convincing	36 (26.7 %)	15 (17.8%)
0 Unconvincing or no information provided	88 (65.2 %)	64 (76.2 %)
Qual 8: Is all necessary background information to interpret numeric representations of measurements provided?		
2 Convincing	11 (8.2 %)	5 (6.0 %)
1 Partially convincing	36 (26.7 %)	15 (17.8%)
0 Unconvincing or no information provided	88 (65.2 %)	64 (76.2 %)
Qual 9: Are results adequately displayed?		

	All studies	Prevalence study subset
1 Adequate	124 (91.8 %)	76 (90.5 %)
0 Unclear or containing obvious errors	11 (8.2 %)	8 (9.5 %)
Combined quality classification		
High quality	25 (18.5 %)	18 (21.4 %)
Moderate quality	81 (60.0 %)	52 (61.9 %)
Low quality	29 (21.5 %)	14 (16.7 %)

Study countries and settings

Study country. Of the 135 studies reporting on levels of poor or good mental health, all but two reported on one country only, while two reported on three countries each. The total number of country samples therefore amounts to 139, and covers a total of 26 unique countries (Fig. 3).

Geographically and in alignment with where most LLMIC are located, most studies (41%) were conducted in countries of the WHO Africa Region, with 56 studies across 12 countries, therefore covering about one third of the region's 39 LLMIC. 39 studies (29%) were conducted in the WHO Eastern Mediterranean Region, in 6 out of the region's 10 LLMIC (60%). 35 studies (26%) were conducted in the WHO South-East Asia Region, in 5 out of the region's 9 LLMIC (56%). Finally, 5 studies (4%) were conducted in the WHO Western Pacific Region, in 3 out of the regions 10 LLMIC (33%). None of the 6 LLMIC in the WHO European Region and of the 4 LLMIC in the WHO Region for the Americas is represented in the included studies.

Close to two thirds of studies were conducted in four countries, namely India (29 studies), Nigeria (25 studies), Pakistan (15 studies), and Egypt (10 studies). The number of available studies in each of the remaining 22 countries ranged between 1 and 7.

Study setting. 65% of studies were conducted in urban settings, 7% in rural settings, 22% in both urban and rural settings, and for 6%, it was not possible to judge based on the reported information. 80 studies (59%) were conducted within only one city, of which 45 studies (33% of all included studies) within only one healthcare facility, most of which university or other tertiary care hospitals. For multi-site studies, the

number of healthcare facilities from which respondents were sampled ranged from 2 to 89 (mean = 12.8, sd = 18.4), with 13% of studies not reporting and 14% of studies having sampled through other channels.

Reporting of the study setting (*quality criterion 2*) was largely satisfactory, with 86% of studies reporting adequate information. Among explicit prevalence studies, reporting was even slightly better.

Study populations and samples

Study population. 93% of study samples included health workers working at secondary- and/or tertiary-level health facilities, of which 42% exclusively tertiary-level staff. Only 26% of study samples, from a total of 13 countries, included health workers at the primary level of care. 2 studies did not provide information.

64% of studies focused exclusively on the public sector, 6% exclusively on the private (including faith-based) sector, 27% included a mix of both, and for 3% it was not possible to judge based on the information provided. Studies including private-sector health workers were conducted across a total of 14 countries.

We classified study populations into medical doctors, nurses, other clinical staff, and other managerial staff. The “other clinical staff” category contains dentists, pharmacists, auxiliary staff (i.e., “untrained” clinical staff according to the WHO/ILO definition used as inclusion criterion), laboratory and imaging personnel, physiotherapists, environmental health staff, and other paramedical staff. It must be noted that within the medical doctor and nursing categories, many studies focused on staff with specific specialization.

64% of studies investigated only one staff category, whereas the remaining 36% two or more. 53% of studies included medical doctors, 71% of studies included nurses, 28% of studies included other clinical staff, and 2% of studies included other managerial staff.

96% of studies reported fully adequate information on the study population (*quality criterion 3*; 98% in prevalence studies).

Sampling and resulting samples. 85% of studies sampled respondents using multi-step procedures, where they first selected health facilities, some then specific departments within health facilities, and then respondents. Some studies had further explicit pragmatic inclusion criteria such as only staff who had been working at the facility for a specific time frame, or inclusion criteria related to the main study aim, such as only respondents who had witnessed death recently. Only 14% of studies sampled respondents directly, for instance by making use of other channels such as mailing lists or meetings of professional associations or snowballing in the researchers’ own networks.

37% of studies described their sampling strategy as a census, 18% as a random or stratified random sample, and 27% as a convenience sample. For 18% of studies, the sampling strategy was unclear (14% among prevalence studies). The proportion of studies with a census or random/stratified random sample

was substantially higher among explicit prevalence studies than among studies with a different primary aim (65% vs. 37%).

Among the 24 studies with a declared random or stratified random sample, 42% provided a rationale for the envisioned sample size (40% of prevalence studies). 19 studies (79%) provided a response rate (87% of prevalence studies). 15 of the 19 studies reported a response rate above 70% (10 of the 13 prevalence studies). It is important to note, however, that 6 studies reported response rates close to or of 100%, calling into question whether the studies actually used a fully random sample as opposed to some elements of replacement and/or convenience.

Of the 37 studies with a declared convenience sample, 8 provided a rationale for the envisioned sample size, 12 did not, and 17 did not state which sample size they attempted to reach at all.

Among the 50 studies with a declared (attempted) census, 38 (76%) provided a response rate (80% of explicit prevalence studies). 23 of the 38 studies reported a response rate above 70% (18 of the 32 prevalence studies).

Resulting sample sizes ranged from 29 to 2245 respondents (mean = 274.4, sd = 273.9). The difference in sample size between prevalence studies and studies with a different primary aim was small (mean 281.3 vs. 262.8, not statistically significant).

Based on the description of the sampling strategy and resulting sample (*quality criterion 4*), we judged only 22% of studies to be based on a sample highly likely to be representative of the intended study population (23% of prevalence studies). 34% of studies presented a convincing description of a census or random sample, but with a response rate below 70% or not provided, or a very convincing description of a convenience sample (43% of prevalence studies). 44% of studies presented an unconvincing description or not enough information to judge (35% of prevalence studies).

Of note, as most studies sampled health workers through health facilities and relied on health workers present at work, they by design did not capture health workers ill enough not to be able to work, making them prone to a systematic underestimation of severe cases of mental illness. This was discussed and acknowledged as a limitation by only a handful of studies.

Reporting of key respondent characteristics (*quality criterion 5*), defined by us as at minimum sex, age and/or seniority in health care, and health worker type or cadre, was fully adequate in 83% of studies (85% among prevalence studies), mostly adequate with only some ambiguity or omission in 11% of studies, and insufficient in 6% of studies.

Study outcomes and measures

Study outcomes. 88% of studies assessed only one mental health outcome, 11% two outcomes, and 1% three outcomes. 67% of studies assessed burnout, 10% depression, 7% anxieties, 4% traumata, and 25% general psychological wellbeing.

Outcome measures. All burnout studies used self-reported measures rather than diagnostic interviews. Among the 90 burnout studies, the Maslach Burnout Inventory - Health Services Survey (MBI-22 HSS) was by far the most common measurement tool, used by 55 studies (61%). A further 15 studies used adaptations of the MBI (e.g., only one subscale, only selected items) or unspecified MBI versions. 7 studies (7%) used the Copenhagen Burnout Inventory, either in full or one of the subscales. The remaining 13 studies used other established or self-developed tools including the Freudenberger Burnout Scale; the Oldenburg Burnout Inventory; the Shirom Melamed Burnout Inventory; a two-item measure developed by Mbindyo et al [33] as part of a motivation inventory; or a single-item direct question ("Do you feel burned out?").

Among the 14 studies having assessed depression, only one used a clinical interview (depression component of the Structured Clinical Interview for DSM-IV), while the remaining 13 studies used self-reported measures, including the Depression, Anxiety and Stress Scale (DASS-21; 4 studies); the Aga Khan University Anxiety and Depression Scale (2 studies); the Patient Health Questionnaire (PHQ-9; 2 studies); the Zung Depression Scale (2 studies); the Beck Depression Inventory (1 study); the Standardized Hospital Anxiety and Depression Scale (1 study); and the Death Distress Scale (1 study).

Similar to depression, only 1 of the 10 studies assessing anxiety used a clinical interview (anxiety component of the Structured Clinical Interview for DSM-IV), while the remaining 9 studies used self-reported measures. The latter included the DASS-21 (4 studies); the Aga Khan University Anxiety and Depression Scale (1 study); the Spielberger State-Trait Anxiety Inventory (1 study); the Zung Anxiety Scale (1 study); the Standardized Hospital Anxiety and Depression Scale (1 study); and the Death Distress Scale (1 study).

All 5 studies assessing trauma used self-reported measures, including the Impact Event Scale – Revised (3 studies); the PTSD Checklist (1 study); and PROQoL (1 study).

Finally, all 33 studies assessing general (ie diagnosis-unspecific) psychological wellbeing used self-reported tools, including the General Health Questionnaire (GHQ) in the 12-item (11 studies), the 28-item (3 studies), the 30-item (5 studies), or an unspecified (1 study) version; the Warwick Edinburgh Mental Wellbeing Scale (4 studies); the WHOQOL-BREF (4 studies); the WHO-5 Wellbeing Index (1 study); the SF-36 mental health subscale (1 study); the SRQ-20 (1 study); the Reker Wong Perceived Wellbeing Scale (1 study); and an unspecified tool (1 study).

We considered reporting of the tool(s) to measure outcome(s) adequate (*quality criterion 6*) if name, version, language, and any potential modifications were clearly reported or referenced. In the case of non-established tools, we expected a clear description including the item list and response modalities. 85% of studies met our criteria (88% of prevalence studies), whereas 7% of studies each either reported with some ambiguity or insufficiently.

Of note, only few articles demonstrate awareness of the limitations and implications associated with using a self-reported tool rather than a clinical interview to measure the mental health outcome.

Validity considerations. Given the culture-sensitive nature of mental health and the predominant use of self-reported measures, we further assessed the extent to which studies provided convincing information of the validity of the tool used to measure the intended mental health constructs (*quality criterion 7*). We considered validity information as convincing if the study provided self-generated content and criterion validity (e.g. a convincing combination of expert judgement/qualitative pre-study, Confirmatory Factor Analysis, and assessment of relationships with related constructs) which based on the description was achieved following standard psychometric quality criteria and yielded adequate psychometric results, or if the study referred to an external validation paper which was accessible, provided similar high-quality evidence, and was carried out in a similar population (ie at minimum same country or cultural context, even if different population). For studies reporting their measurements in categorical fashion (see below), we further required context-appropriate validity evidence of the threshold used to classify respondents into different mental health categories.

Only 8% of studies provided information which we considered convincing. 27% of studies provided some, but incomprehensive or not fully convincing validity information. 65% of studies provided no or unconvincing validity information. The proportion of studies providing convincing validity information was even lower among explicit prevalence studies (convincing: 6%; somewhat convincing: 18%; insufficient: 76%), which is due to the higher proportion of studies reporting categorical outcomes and failing to provide validity evidence regarding the used thresholds to categorize respondents.

Of note, only 25% of studies reported having performed a pretest (both overall and among explicit prevalence studies). Irrespective of the quality or appropriateness of the information, 62% of studies provided some information on reliability (usually Cronbach's alpha) and 49% of studies provided some information on validity (usually references to the tool manual and/or validation studies conducted in high-income settings). Among explicit prevalence studies, some information on reliability and validity (irrespective of quality of the information) was provided by 54% and 50%, respectively.

Measurement. Beyond the measured outcomes and tools themselves, studies differed in how they reported the outcome measurements. All utilized measurement tools employed either Likert response scales or symptom counts and therefore, in a first analytical step, resulted in a quasi-continuous numeric measurement. 33% of the included studies reported outcome measurements in this "crude" metric, i.e. as means of sum scores or scale means (16% of prevalence studies). 46% of studies divided respondents into different categories along this quasi-continuous raw score, and reported proportions of participants in each category (57% of prevalence studies). 21% of studies reported data both in quasi-continuous and in proportional form (27% of prevalence studies).

In this context, we also assessed the extent to which the authors provided all necessary background information to interpret numeric representations of measurements (*quality criterion 8*), including the numeric codes used for response options, information on aggregation, and for studies reporting proportions, information on thresholds for categorization. 58% of studies reported sufficient information to allow interpretation and comparison to other studies having used the same measurement tool (56% of

prevalence studies). For 28% of studies, there was some ambiguity, about which reasonable assumptions can be made however (33% of prevalence studies). For 14% of studies, information was insufficient (11% of prevalence studies).

Given the widespread use of the MBI-22, we wish to highlight three specific issues frequently encountered and complicating comparability of results across studies. First, how to use the MBI is rather strictly prescribed by the publishers. However, many studies did not adhere to the publisher's prescription, rather having altered certain items, used different response scales in regards to number of answer options and numeric representation of answer options, leading to different score ranges, as well as in regards to labelling of answer options. Often, not enough detail was provided to understand what was done exactly, compromising the extent to which findings can be compared across studies. Second, the publishers made small updates to the MBI-22 over time, particularly in relation to the thresholds used to categorize severity of burnout. Many studies unfortunately did not report which version they used, compromising comparability between studies even in otherwise relatively homogeneous settings. For instance, of the seven studies from India having used the MBI-22 in its recommended form and with categorical reporting of results, two used the 2nd edition threshold, four did not report or allow to somehow infer whether they used the 1st or 2nd edition thresholds, and one appears to have used a mix of both. Finally, the MBI-22 measures three sub-constructs of burnout, one of which is "reduced personal accomplishment" (RPA). The items intended to measure RPA are reversely phrased, however, so that a high raw score indicates low burnout, unlike for the other two subscales. In a large number of studies, it did not become fully clear whether the authors had reversed RPA scores/proportions so that they are interpretable "in the same direction" as the other two dimensions, or whether they report original scores. Sometimes, inference from the description of results or discussion was possible, whereas in other cases, both numeric estimates and description left doubt as to whether the authors had or had not reversed scores and/or interpreted results correctly.

Results reporting

Depending on the study aim and population, studies reported estimates of levels of mental health either overall for the entire study sample, or broken down by different sample subgroups. For simplicity, we only assessed the extent to which studies broke down results by two key sample characteristics, namely cadre and gender. However, many studies also reported estimates broken down by other demographic characteristics such as marital status or age, by work-related factors such as qualification level, work experience, or place of work, or by other factors assessed in relation to mental health, as described in more detail below.

In regards to health worker cadre, 64% of studies included respondents from only one staff category and therefore by design reported estimates by cadre (62% of prevalence studies). Of the remaining 48 studies including mixed samples, 46% reported estimates separately by cadre and 54% provided only overall estimates (among prevalence studies: 53% vs 47%, respectively).

In regards to gender, 10% of studies only estimated mental health levels among one gender (usually female nurses) or had a heavily skewed sample in terms of gender, presumably reflecting the reality in the context (usually predominantly male medical doctors with female doctors below 5%) (7% among prevalence studies). Of the remaining 121 studies, 39% reported estimates separately by male and female participants, whereas 61% only provided overall estimates (among prevalence studies: 46% vs 54%, respectively).

We further assessed the extent to which results were displayed adequately, meaning that they could be read without necessitating guesses and that they did not contain obvious errors (*quality criterion 8*). This was the case for 92% of all studies, and 91% among explicit prevalence studies.

Overall study quality

From the quality judgements in the nine individual categories presented above, we further calculated an overall quality classification for each study as outlined in detail in Additional Files 2. In order to be classified as of high quality, a study had to report results in a readable manner (quality criterion 9), provide sufficient information to allow for the measurements to be interpreted (quality criteria 6 and 8), and reach satisfactory quality scores on all other criteria combined.

As shown at the bottom of Table 3, only 19% of all studies fulfilled our criteria for high quality. We found the majority of studies as of moderate quality (60%), and 21% as of low quality. Among explicit prevalence studies, the proportion of studies in the high and moderate categories was only marginally higher than among all studies.

We did not observe any trend in study quality over time, nor any striking differences by region.

Research on factors associated with mental health amongst health workers in LLMIC

As described above, of the 143 relevant studies identified by our search, 126 included information on associations of mental health with other factors. 118 studies included both estimates of levels of good or poor mental health as well as relationships with one or multiple correlates of mental health. Only 8 studies focused exclusively on associations; these were mostly studies from the field of psychology where mental health was one of multiple constructs assessed in relation to a different focal outcome. Given this large extent of overlap as well as the relatively lower importance of study setting in the context of establishing associations, we will not present study characteristics (setting, population, sample, outcome, measurement), but rather focus on which factors have been investigated as well as on methodological and reporting considerations in the following.

Of the 126 studies reporting associations of mental health with one or more other factors, 100 (79%) reported this investigation of associations to be a key study aim, whereas the remaining 26 studies did so as a side-product of an estimation of mental health levels, as part of a validation study, or with mental health being a minor correlate to a different focal construct of interest.

Table 4 summarizes the key findings elaborated below, for all studies as well as only for the subset of studies with an explicit aim to assess associations of health worker mental health with other factors (“associations studies”).

Table 4
Studies assessing factors associated with mental health - key results by study type

	All studies	Associations study subset
Total number of studies	126	100
Assessed factors*		
Socio-demographic characteristics	94 (75.6 %)	74 (74.0 %)
Tangible work factors	96 (76.2 %)	75 (75.0 %)
Intangible work factors (perceptions, attitudes)	79 (62.7 %)	69 (69.0 %)
Health factors	29 (23.0 %)	23 (23.0 %)
Job performance	6 (4.8 %)	4 (4.0 %)
Is it clear which factors were considered?		
Clear and unambiguous from introduction/methods	64 (50.8 %)	57 (57.0 %)
Inference possible from results section	56 (44.4 %)	39 (39.0 %)
Unclear	6 (4.8 %)	4 (4.0 %)
Is it clear how the factors were measured?		
Largely clear	74 (58.7 %)	55 (55.0 %)
Partly clear, partly unclear	46 (36.5 %)	39 (39.0 %)
Mostly unclear	6 (4.8 %)	6 (6.0 %)
Is the analytical technique clearly described or inferable?		
Clear	112 (88.9 %)	89 (89.0 %)
Unclear	14 (11.1 %)	11 (11.0 %)
Was bivariate/univariate analysis performed?		
Yes	114 (90.5 %)	89 (89.0 %)
No	12 (9.5 %)	11 (11.0 %)
Was multivariate analysis performed?		
Yes	66 (52.4 %)	60 (60.0 %)
No	56 (44.4 %)	36 (36.0 %)

*Proportions do not add up to 100% as many studies assess factors from multiple categories and/or use multiple analytical techniques

	All studies	Associations study subset
Not applicable (only one associated factor investigated)	4 (3.2 %)	4 (4.0 %)
*Proportions do not add up to 100% as many studies assess factors from multiple categories and/or use multiple analytical techniques		

Of note, among the studies with an explicit focus on assessing associations of mental health with other factors, most frame their study in causal terms, in the sense of aiming to investigate determinants or consequences of mental health. However, all but a handful use cross-sectional study designs and analytical techniques that strictly speaking do not allow for causal inference.

Assessed correlates of mental health

The number of investigated correlates of mental health ranged from 1 to 30 (mean = 9.4, sd = 6.3), without notable differences between studies with and without explicit aim to investigate associations.

75% of studies assessed one or more socio-demographic characteristics; 76% tangible work factors; 63% intangible work factors; and 23% other health factors in relation to mental health, most frequently smoking, alcohol consumption, and chronic health issues. Finally, 6 studies assessed self-reported adverse behavior towards patients or other perceived job performance aspects in relation to mental health. Figure 4 provides an overview over all factors assessed by more than 5% of the studies.

It must be noted that with the exception of gender, studies differed in how they captured even basic demographic characteristics such as marital status (differences in grouping) and children (any vs number; number in continuous form vs various categories). Unsurprisingly, differences became more pronounced the more complex the investigated factors. Highly complex constructs, such as job satisfaction or work demands, were rarely measured with the same tool and reported in the same way by two or more studies.

Irrespective of the methodological challenges related to causality discussed above, there seems to be a clear focus of the currently available literature on measuring potential determinants or proximal and intangible outcomes of mental health, whereas research on relationships between mental health levels and more distal and tangible outcomes such as performance or actual turnover or absenteeism is practically non-existent. The few studies having assessed performance outcomes have done so via self-report rather than an objective performance measure.

Methodological and reporting considerations

As described above, we did not perform a study quality assessment analogous to studies reporting on levels of mental health due to the large heterogeneity of factors assessed, which would have necessitated a quality assessment by factor rather than by study. However, we did assess whether studies provided

key information necessary for understanding the results, and for potentially comparing them across studies.

Measurement. 64 studies (51%) provided a clear and unambiguous description of which factors they considered in relation to mental health, either in the introduction or methods section. For a further 56 studies (44%), inference of all considered factors was possible from the results section. Only 6 studies (5%) left doubts as to which factors the authors had considered. This was usually because a full list was absent from the introduction and/or methods sections, and only significant associations were reported. Reporting was slightly better among studies with an explicit aim of assessing associations with mental health.

For the majority of studies (74; 59%), the description of how the investigated factors were measured was largely clear in the sense that the measurement tool and response scale/options were clearly reported and allowed for interpretation and comparison of the reported measures of strength of association. For 37% of studies, the description was clear for some of the investigated factors, but not for others. 6 studies provided insufficient information on most or all investigated factors. Reporting was slightly worse among studies with an explicit aim of assessing associations with mental health. Among the 35 studies having assessed only socio-demographic characteristics or tangible work factors, so factors which are comparatively more straightforward in measurement, 89% provided clear information on how factors were measured.

Analysis and reporting of results. For 89% of studies, either the methods or the results section allowed to understand which analyses were performed and how, the latter particularly in relation to multivariate analyses and whether all or only a subset of factors were included.

114 (91%) studies reported, exclusively or in a first step, the results of bivariate analyses between the measured mental health outcome and each associated factor under investigation. Of those, 55 (48%) reported the strength of association, either as correlations or as bivariate regression (beta or odds ratios) results. 66 studies (58%) rather reported group differences, resulting from either ttests/ANOVAs or from chi2 or some other non-parametric test as per the respective variables under investigation. In the vast majority of studies where group differences were reported, group means or proportions on the mental health outcome were provided along with the results of the difference test.

66 studies (52%) reported the results of multivariate analyses between the measured mental health outcome and the associated factors under investigation, of which 12 exclusively, so without prior bivariate analysis. All analysis employed multivariate regression, including either all potential associated factors under investigation or utilizing a step-wise approach where only factors significantly associated with mental health in a prior series of bivariate analyses were included. 33 studies provided estimates of strength of association for all factors included in the analysis, whereas the other 33 studies provided details only for those coefficients which emerged as statistically significant from the analysis. 55% of the studies reported beta coefficients, 32% odds ratios. In 8% of studies, we could not infer which type of

effect estimate the reported coefficient represented. The remaining 6% of studies did not report any coefficients, but only stated in the text that associations were or were not statistically significant.

Discussion

In the context of the SARS-CoV-2 pandemic, health workers' work realities and their particular risk of facing mental health issues has quickly risen in public awareness and attention, and generated unprecedented research efforts. A major challenge in establishing the impact of the pandemic on health workers' mental health is the availability – or lack thereof – of baseline data. Prior reviews [16–19] suggest that tangible research on levels of good or poor mental health among health workers in low- and middle-income countries as well as associated factors is limited. However, none allow for a comprehensive picture due to limitations in geographic scope, included outcomes, and – as became clear from this review – also due to omission of relevant articles. We therefore undertook a scoping review of the quantitative literature on mental health of health workers in low- and lower-middle income countries, with the aims of providing a comprehensive overview of existing evidence and informing a strategic research agenda.

We were surprised to find a much larger than anticipated number of studies, most published within the last 10 years. We wish to highlight that much of this research appears to be led or exclusively conducted by research teams from the respective LLMIC, which is a welcome finding in light of recent increased calls for more equitable global health research structures and partnerships [34–36].

However, as our review title implies, this impressive body of literature is rather limited in regards to what can be learned from it. Geographically, the available research is concentrated in few countries, while we did not find any studies for most LLMIC. Available research further mostly concentrates on hospital settings, often single tertiary care facilities in urban centers, while studies conducted in rural primary care settings are comparatively rare. In terms of the mental health outcomes assessed, the available literature clearly focuses on burnout, while research on other diagnoses or general psychological wellbeing, including also on the role of “suboptimal” but not clinically relevant states, is less available.

In regards to factors associated with mental health, we found a large level of heterogeneity beyond a handful of key demographic characteristics such as age, gender, marital status, cadre, and work experience. None of the studies has investigated mental health in relation to tangible and objectively measured outcomes such as performance or absenteeism.

Study quality tended to be below expectation for most included studies. In particular, we identified major issues with regards to sample representativeness, validity of measurement tools in the respective context, and provision of key information necessary to interpret the numeric figures provided by the authors. In consequence, there are major doubts in regards to the robustness, interpretability, and external validity of the vast majority of available studies. Interestingly, studies with an explicit aim to establish prevalence of mental health issues among health workers, which we would have expected to be of higher quality, did

not perform better on our quality criteria than studies which produced estimates of mental health levels as a by-product.

Some of the identified quality issues can easily be overcome by improved reporting, and we will provide recommendations to this extent below. In this context, it is interesting to note that a separate analysis of journal quality performed on a subset of the studies revealed that a significant proportion were published in journals likely to be predatory, with questionable peer review processes and editorial quality standards [37]. This highlights the need to financially enable researchers from LLMIC to publish in more reputed journals [38], which would likely increase reporting quality.

Other issues are potentially more difficult to address. Issues of representativeness likely necessitate improvements and increased effort in sampling, which however tends to be more resource-intensive. As noted, all but two studies used self-report to measure the mental health outcome, rather than the “gold standard” of a clinician-led diagnostic interview. This is less of an issue for studies investigating associations of mental health with other factors using continuous measurements rather than proportions of ill and healthy individuals. However, accuracy of self-reported measures in the estimation of mental illness prevalence has been shown a major issue even in high-income settings where self-reported measures tend to be comparatively well validated and calibrated [39, 40]. Conducting prevalence studies with clinical interviews is not only extremely resource-intensive, however, but possibly also unrealistic in settings where sufficiently trained personnel is not readily available.

The complete lack of studies linking mental health to tangible performance measures is surprising and unfortunate from both a general knowledge and political perspective. With the SARS-CoV-2 pandemic, health workers’ state of mind and health suddenly became an issue of major public concern, presumably at least in part because it threatened to compromise patient care. More general research on the relationship between health worker mental health and their performance will likely be instrumental in sustaining this new-found interest and associated investments in the field. However, doing so necessitates more complex study designs than the standard “questionnaire research”, drawing on multiple sources of data and data collection techniques, with the associated budgetary and capacity implications.

Limited geographic scope necessitates more and more extensive research – as noted above, many studies focus on one university or tertiary care hospital and appear to have been student dissertations. In order to cover not only more countries, but also a broader scope within countries, including primary care settings, larger research projects are necessary and require respective funding opportunities and research capacity.

Similarly and perhaps most importantly, the issue of lack of local validation of measurement tools urgently necessitates substantially more methodological research, as also highlighted by a 2016 systematic review [41]. Even assuming the cross-cultural validity of “Western” concepts of mental health and illness, which is debated [42], measurement tools cannot simply be assumed to measure the same constructs across cultural contexts. This issue of pertains to both the items and answer options

themselves, but even more importantly to the thresholds used to categorize respondents into severity of illness categories. Studies comparing self-reported screening tools to clinical interviews have clearly shown that appropriate thresholds vary substantially between different study populations [43–45]. Of note, while cultural differences might be most important, there might well be differences in measurement properties between different sub-population within a defined cultural context, for instance between generally well-educated health workers and the likely less educated general or patient population. Unless tools and thresholds are therefore robustly validated in the respective context, interpretation of results particularly in regards to mental health prevalence remains difficult and speculative, and comparison of study results particularly between contexts close to impossible.

Our review did not look at whether high-quality evidence regarding the culture-specific validity of the used tools actually exists in the study countries, but we focused only on what was provided by the studies. However, as less than 10% of studies provided information which we judged as convincing, we are confident in concluding that urgent investment into validation studies is necessary. Beyond specific validation studies, authors wishing to rather conduct substantive research can build validation elements into their studies, for instance by advancing the main data collection with a qualitative pre-study or expert assessment of the tool, by collecting additional data for criterion validation, or by performing psychometric analysis of the data. However, our finding that half of the studies made no mention of validity at all, combined with almost none of the studies discussing issues of validity in the study limitations, indicates that capacity building in this area is urgently necessary.

Roadmap for future research

Based on the above-presented findings, in order to develop a more comprehensive understanding of the mental health situation among health workers in LLMIC, we urge researcher and funders to consider the following in planning future research and funding opportunities:

Funders:

- Invest in research in LLMIC which have been under-researched to date, and in explicit comparative studies across countries;
- Invest in representative research at the primary level of care and utilizing robust methods of estimating mental health prevalence;
- Make funding available for research with an explicit focus on validating robust methods of estimating mental health prevalence in a variety of socio-cultural contexts;
- Strengthen comprehensive mental health research capacity, integrating psychiatric, psychological, psychometric, and epidemiological perspectives.

Researchers:

- Plan research in LLMIC which have been under-researched to date, particularly representative research at the primary level of care;
- Investigate mental health in relation to tangible outcomes such as performance, absenteeism, or turnover, using objective rather than self-reported outcome measures;
- Research psychological wellbeing and mental health conditions other than burnout
- Invest in strong sampling designs likely to lead to representative study samples and “gold standard” methods of estimating mental health prevalence;
- Invest in culture-specific validation of measurement tools, both as stand-alone projects and within substantive research, by building in psychometric elements into studies (e.g., qualitative pre-studies or expert validation, additional measures for criterion validity) and performing psychometric analysis on data sets (e.g., factor analysis, measurement invariance testing when comparing different sample groups);
- In designing research and writing up study findings, consider the elements summarized in Table 5 to facilitate identification, interpretation, and comparison;
- Aim for publication in reputed journals with high-quality editorial and peer review processes.

Table 5
Recommendations for study design and reporting of results

	Design	Reporting
Basics	<ul style="list-style-type: none"> • Choose study design in relation to study aim; if focus on associations, consider study designs that allow inference of causality 	<ul style="list-style-type: none"> • Study aim and hypotheses (if applicable), including rationale for assumed associations with other factors • Year of data collection • Country, geographical region within country • Urban vs rural setting, single- vs multi-site, number of health facilities • Label your study as a mental health study and choose key words that will lead to inclusion in geographic and health workforce MeSH terms etc. in search engines
Study population	<ul style="list-style-type: none"> • Prioritize larger geographic scope and primary care • Obtain information on total study population size and characteristics 	<ul style="list-style-type: none"> • Level of care, public vs private sector • Health worker cadres • Any other inclusion or exclusion criteria, including rationale • Total study population size (if available)
Sampling	<ul style="list-style-type: none"> • Maximize representativeness • Consider absenteeism due to mental illness • Plan for intended subgroup analysis/adjustment 	<ul style="list-style-type: none"> • Sampling strategy, intended sample size, rationale for sampling strategy and size • Response rate • Any other information on sample representativeness (e.g. differences in demographic characteristics between study population and sample)
Reporting of sample characteristics		<ul style="list-style-type: none"> • Total sample size • Sample size by key sample characteristics (e.g. cadre, gender, age/seniority group) • Sample characteristics (suggested minimum set: cadre, gender, age, marital status, years of work experience, work hours per week)
Reporting of outcomes		<ul style="list-style-type: none"> • Mental health outcome, including definition where potentially ambiguous

	Design	Reporting
Outcome measure	<ul style="list-style-type: none"> • Consider feasibility of a clinical interview • Consider existing research and validity evidence available in the context when deciding on the outcome measure • Consider measuring additional mental health outcomes for criterion validation 	<ul style="list-style-type: none"> • Outcome measure, including exact name and version (if multiple), reference to tool manual and development/validation paper • Information on language/translation and any modifications • Number of items, item examples, exact response scale including number of response options and numeric codes; if non-standard tool, provide in full • Composite score calculation, including how raw scores were combined, resulting theoretical score ranges and their direction, and thresholds used to classify respondents (if applicable) • Any existing reliability and validity information from the context
Associated factors	<ul style="list-style-type: none"> • Consider existing research and validity evidence available in the context, with an eye towards comparability 	<ul style="list-style-type: none"> • Provide all information necessary to understand what was being measured, how, and how the resulting numeric measurements are to be interpreted (see mental health outcome measure)
Analysis	<ul style="list-style-type: none"> • Consider which reliability and validation analyses are possible with the data (Cronbach's alpha, factor analysis, measurement invariance testing, criterion validation) • Where robust validity evidence not available: consider performing analyses using multiple thresholds for classification of participants • Consider performing both tests of association and difference, and both bi- and multivariate analysis • Categorize in consideration of the literature which you would like to compare your results to 	<ul style="list-style-type: none"> • Information on reliability and validity analyses • Information on adjustments for sample composition • Information on which estimates are reported, and how they were derived • Information on which associated factors were assessed and how bi- and multi-variate analyses were conducted (including potential stepwise approaches and covariates for the latter)

Design	Reporting
Reporting of results	<ul style="list-style-type: none"> • Overall and by subgroup, including means/proportions (ideally both for continuous variables) • Report both significant and non-significant associations • Report both estimates and significance levels
Interpretation	<ul style="list-style-type: none"> • Report and discuss implications of methodological study limitations, in particular in regards to sample representativeness, measurement tool validity, and possibility of causal inference (for studies on associated factors) • Be careful to postulate strong prevalence estimates and with the use of the word “prevalence” in the absence of use of “gold standard” measures and well validated screening tools

Methodological considerations and recommendations for future updates and reviews

Our review must be read and interpreted in light of certain methodological considerations. First, although we believe to have generated a comprehensive overview over the available literature, we cannot exclude the possibility of inadvertently having missed a few relevant studies. In this context, we'd like to briefly comment on why we believe to have found such as substantially higher number of articles than the prior reviews have, but why we still failed to identify all relevant articles through our search algorithm. In part, this is of course due to differences in geographic scope and inclusion and exclusion criteria. All reviews focused only on burnout, Dubale et al. (2019) in Sub-Saharan Africa, Kesarwani, Husaain and George (2020) in India, and Chemali et al. (2019) in the Middle East, while Dugani et al. (2018) had a global scope, but focused only on primary care. However, we believe that suboptimal phrasing of titles, abstracts, and key words also play a major role. A substantial number of articles, for instance, did not report the country name in the title, but rather the name of a region or city, if anything. Further, rather than using generic terms for the study subjects, such as “health worker” or “healthcare professional”, many studies used the specific terminology in their respective setting, increasing the risk that studies are not picked up even by a very carefully crafted search strategy. Similarly, beyond key standardized syndromes such as burnout and depression, no univocal terminology exists to describe poor psychological wellbeing or pathological forms of stress, increasing the risk of studies not being found. This issue was likely compounded by the above-discussed journal quality issue. Quite a number of the included studies were published in journals not indexed in the major data bases and therefore not benefitting from MeSH terms and similar concepts. Beyond hoping for improved reporting in the future, we therefore urge researchers

planning future reviews to invest time and care into their search strategies in consideration of the above so as to pick up a maximum of relevant research.

Second, given the unexpected large amount of studies, we were unable to perform full double screening and data extraction. Although we took precautions by intensively testing screening procedures and data extraction tools, we cannot fully exclude certain omissions or errors. Third, we purposely did not include qualitative studies. In a future expansion of the review, this would be very interesting, although we acknowledge that the above-discussed difficulty of distinguishing mental health and psychological wellbeing from related constructs such as stress and motivation is likely even more difficult in qualitative research. Finally, we only included articles in English and French language, excluding any relevant literature in other language. In light of the above-discussed high level of “Southern-led” research as well the fact that we did not find any studies from the WHO European Region and Region for the Americas, it may well be that pertinent research published in relevant local languages is not captured by our research. It would therefore be interesting to expand the review to include further languages such as Spanish, Russian, and potentially some South-East Asian languages.

Conclusion

In the context of the SARS-CoV-2 pandemic, health workers’ work realities and their particular risk of facing mental health issues has quickly risen in public awareness and attention, and generated unprecedented research efforts. A major challenge in establishing the impact of the pandemic on health workers’ mental health is the availability – or lack thereof – of baseline data. Our review shows that the body of literature on mental health of health workers in LLMIC is rather impressive in size, but that we can learn comparatively little from it due to limitations in geographic and target population scope, in outcomes, and in methodological quality. We urge funders to enable research in settings where evidence is currently limited, as well as in validation research, and to invest in mental health research capacity building. We encourage researchers to do the same, and to further improve on methodological quality of research and on reporting of methods and findings.

Abbreviations

DASS Depression, Anxiety and Stress Scale

DSM Diagnostic and Statistical Manual of Mental Disorders

GHQ General Health Questionnaire

HIC High-income countries

ICD International Statistical Classification of Diseases and Related Health Problems

LLMIC Low- and lower-middle income countries

MBI-HSS Maslach Burnout Inventory - Health Services Survey

PHQ Patient Health Questionnaire

RPA Reduced personal accomplishment

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

References to the articles in which all data analysed during this study are published are included in the supplementary information files of this article, including the extracted data set for key indicators. Further data are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Wellcome Trust grant number 213725/Z/18/Z. The Wellcome Trust was not involved in the conception and conduct of the research in any way.

Authors' contributions

The study was conceived and designed by all three authors. JL and AD conducted the search, screening, and data extraction, with support from DJ in case of need for reconciliation. JL collated the data and drafted the manuscript, with input from DJ and AD. All authors read and approved the final manuscript.

Acknowledgements

We are grateful to Russel Burke (London School of Hygiene & Tropical Medicine library) and Vittoria Lutje (information specialist, Glasgow) for their support with defining the search strategy, to Manuela De Allegri (Heidelberg University) for her critical reading and inputs into an earlier version of the manuscript, and to Alfonso Valenzuela Hurtado (Heidelberg University) for his help with editing and formatting.

References

1. Khanal P, Devkota N, Dahal M, Paudel K, Joshi D. Mental health impacts among health workers during COVID-19 in a low resource setting: a cross-sectional survey from Nepal. *Global Health* 2020;16:89.
2. Luan R, Pu W, Dai L, Yang R, Wang P. Comparison of psychological stress levels and associated factors among healthcare workers, frontline workers, and the general public during the novel coronavirus pandemic. *Front Psychiatry* 2020;11:583971.
3. Rossi R, Socci V, Pacitti F, Mensi S, Di Marco A, Siracusano A, et al. Mental health outcomes among healthcare workers and the general population during the COVID-19 in Italy. *Front Psychology* 2020;11:608986.
4. Salari N, Khazaie H, Hosseini-Far A, Ghasemi H, Mohammadi M, Shohaimi S, et al. The prevalence of sleep disturbances among physicians and nurses facing the COVID-19 patients: a systematic review and meta-analysis. *Global Health* 2020;16:92.
5. Wang Y, Ma S, Yang C, Cai Z, Hu S, Zhang B, et al. Acute psychological effects of Coronavirus Disease 2019 outbreak among healthcare workers in China: a cross-sectional study. *Transl Psychiatry* 2020;10:348.
6. Wanigasooriya K, Palimar P, Naumann DN, Ismail K, Fellows JL, Logan P, et al. Mental health symptoms in a cohort of hospital healthcare workers following the first peak of the COVID-19 pandemic in the UK. *BJPsych Open* 2020;7:e24.
7. Kisely S, Warren N, McMahon L, Dalais C, Henry I, Siskind D. Occurrence, prevention, and management of the psychological effects of emerging virus outbreaks on healthcare workers: rapid review and meta-analysis. *BMJ* 2020;369:m1642.
8. Salazar de Pablo G, Vaquerizo-Serrano J, Catalan A, Arango C, Moreno C, Ferre F, et al. Impact of coronavirus syndromes on physical and mental health of health care workers: systematic review and meta-analysis. *J Affect Disord* 2020;275:48-57.
9. Stuijtzand S, Deforges C, Sandoz V, Sajin CT, Jaques C, Elmers J, et al. Psychological impact of an epidemic/pandemic on the mental health of healthcare professionals: a rapid review. *BMC Public Health* 2020;20:1230.
10. Maalouf FT, Mdawar B, Meho LI, Akl EA. Mental health research in response to the COVID-19, Ebola, and H1N1 outbreaks: a comparative bibliometric analysis. *J Psychiatr Res* 2021;132:198-206.
11. McCauley L, Hayes R. Taking responsibility for front-line health-care workers. *Lancet Public Health* 2020;5:e461-e462.
12. Aiken LH, Sermeus W, Van den Heede K, Sloane DM, Busse R, McKee M, et al. Patient safety, satisfaction, and quality of hospital care: cross sectional surveys of nurses and patients in 12 countries in Europe and the United States. *BMJ* 2021;344:e1717.
13. Davey MM, Cummings G, Newburn-Cook CV, Lo EA. Predictors of nurse absenteeism in hospitals: a systematic review. *J Nurs Manag* 2009;17:312–30.
14. Hall LH, Johnson J, Watt I, Tsipa A, O'Connor DB. Healthcare staff wellbeing, burnout, and patient safety: a systematic review. *PLoS One* 2016;11:e0159015.

15. Wilkinson H, Whittington R, Perry L, Eames C. Examining the relationship between burnout and empathy in healthcare professionals: a systematic review. *Burn Res* 2017;6:18–29.
16. Chemali Z, Ezzeddine FL, Gelaye B, Dossett ML, Salameh J, Bizri M, et al. Burnout among healthcare providers in the complex environment of the Middle East: a systematic review. *BMC Public Health* 2019;19:1337.
17. Dubale BW, Friedman LE, Chemali Z, Denninger JW, Mehta DH, Alem A, et al. Systematic review of burnout among healthcare providers in Sub-Saharan Africa. *BMC Public Health* 2019;19:1247.
18. Dugani S, Afari H, Hirschhorn LR, Ratcliffe H, Veillard J, Martin G, et al. Prevalence and factors associated with burnout among frontline primary health care providers in low- and middle-income countries: a systematic review. *Gates Open Res* 2018;2:4.
19. Kesarwani V, Husaain ZG, George J. Prevalence and factors associated with burnout among healthcare professionals in India: a systematic review and meta-analysis. *Indian J Psychol Med* 2020;42:108-15.
20. Lohmann J, John D, Dzay A. Prevalence and factors associated with poor mental health among healthcare professionals in low- and lower-middle-income countries: a systematic review protocol. *Syst Rev* 2019;8:294.
21. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* 2018;18:143.
22. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Baldini Soares C. Guidance for conducting systematic scoping reviews. *Int J Evid Based Healthc* 2015;13:141-6.
23. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 2005;8:19-32.
24. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci* 2010;5:69.
25. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 2018;169:467-73.
26. International Labour Organization (ILO). International standard classification of occupations: ISCO-08. Geneva: International Labour Organization; 2008.
27. World Health Organization (WHO). Classification of health workforce statistics. Geneva: World Health Organization; 2019. Accessed 25 Jun 2019.
https://www.who.int/hrh/statistics/workforce_statistics/en/
28. World Bank. World Bank country and lending groups 2019. Accessed 25 Jun 2019.
<https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>
29. Moola S, Munn Z, Tufanaru C, Aromataris E, Sears K, Sfetcu R, et al. Chapter 7: Systematic reviews of etiology and risk. In: Aromataris E, Munn Z (eds). *Joanna Briggs Institute Reviewer's Manual*. The Joanna Briggs Institute; 2017. Accessed Sep 11 2019.

https://joannabriggs.org/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Analytical_Cross_Sectional_Studies2017_0.pdf

30. Critical Appraisal Skills Programme (CASP). CASP Case Control Study Checklist. 2018. Accessed Sep 11 2019. <https://casp-uk.net/wp-content/uploads/2018/01/CASP-Case-Control-Study-Checklist-2018.pdf>
31. National Institutes of Health (NIH). Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies. 2014. Accessed Sep 11 2019. <https://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/cohort>
32. Bianchi R, Schonfeld IS, Laurent E. Burnout-depression overlap: a review. *Clin Psychol Rev* 2015;36:28-41.
33. Mbindyo PM, Blaauw D, Gilson L, English M. Developing a tool to measure health worker motivation in district hospitals in Kenya. *Hum Resour Health* 2009;7:40.
34. Abimbola S. The foreign gaze: authorship in academic global health. *BMJ Global Health* 2019;4:e002068.
35. Hedt-Gauthier BL, Jeufack HM, Neufeld NH, Alem A, Sauer S, Odhiambo J, et al. Stuck in the middle: a systematic review of authorship in collaborative health research in Africa, 2014–2016. *BMJ Global Health* 2019;4:e001853.
36. Cash-Gibson L, Rojas-Gualdrón DF, Pericàs JM, Benach J. Inequalities in global health inequalities research: a 50-year bibliometric analysis (1966-2015). *PLoS One* 2018;13:e0191901.
37. John D, Polani Chandrasekar R, Lohmann J, Dzay A. 2020. Identifying predatory journals in systematic reviews. *Advances in Evidence Synthesis: special issue. Cochrane Database Syst Rev* 2020;9(Suppl 1):147.
38. Nabyonga-Orem J, Asamani JA, Nyirenda T, Abimbola S. Article processing charges are stalling the progress of African researchers: a call for urgent reforms. *BMJ Global Health* 2020;5:e003650.
39. Levis B, Benedetti A, Ioannidis JPA, Sun Y, Negeri Z, He C, et al. Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: individual participant data meta-analysis. *J Clin Epidemiol* 2020;122:115-128.e1.
40. Thombs BD, Kwakkenbos L, Levis AW, Benedetti A. Addressing overestimation of the prevalence of depression based on self-report screening questionnaires. *CMAJ* 2018;190:E44-9.
41. Ali GC, Ryan G, De Silva MJ. Validated screening tools for common mental disorders in low and middle income countries: a systematic review. *PLoS One* 2016;11:e0156939.
42. Kirmayer LJ, Swartz L. Culture and global mental health. In: Patel V, Minas H, Cohen A, Prince MJ (eds). *Global Mental Health: Principles and Practices*. Oxford, UK: Oxford University Press, 41-62.
43. Anjara SG, Bonetto C, Van Bortel T, Brayne C. Using the GHQ-12 to screen for mental health problems among primary care patients: psychometrics and practical considerations. *Int J Ment Health Syst* 2020;14:62.

44. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ 2012;184:E191–E6.
45. Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M, Gujere O, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol Med 1997;27:191–7.

Figures

Figure 1

Flow chart outlining the study search and selection process

Figure 2

Included studies by year of publication

Figure 3

Studies assessing levels of mental health by geographic distribution. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its

authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 4

Factors frequently assessed in relation to mental health. * among studies having investigated multiple categories of health workers

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ScoprevMHHW210117AddF1.docx](#)
- [ScoprevMHHW210117AddF2.docx](#)
- [ScoprevMHHW210117AddF3.docx](#)