

Patient-reported outcome measures in community-acquired pneumonia: a systematic review of application and content validity

Melanie Lloyd,^{1,2} Emily Callander,³ Amalia Karahalios,⁴ Lucy Desmond,⁵ Harin Karunajeewa^{1,5}

To cite: Lloyd M, Callander E, Karahalios A, *et al*. Patient-reported outcome measures in community-acquired pneumonia: a systematic review of application and content validity. *BMJ Open Res* 2019;**6**:e000398. doi:10.1136/bmjresp-2018-000398

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjresp-2018-000398>).

Received 20 December 2018
Revised 3 April 2019
Accepted 26 April 2019

ABSTRACT

Introduction Patient-reported outcome measures (PROMs) are a vital component of patient-centred care. Community-acquired pneumonia (CAP) is a significant contributor to morbidity, mortality and health service costs globally, but there is a lack of consensus regarding PROMs for this condition.

Methods We searched MEDLINE, EMBASE and Cochrane Collaboration for studies, both interventional and observational, of adult recovery from CAP that applied at least one validated PROM instrument and were published before 31 December 2017. The full text of included studies was examined and data collected on study design, PROM instruments applied, constructs examined and the demographic characteristics of the populations measured. For all CAP-specific PROM instruments identified, content validity was assessed using the COnsensus based Standards for selection of health Measurement INstruments guidelines (COSMIN).

Results Forty-two articles met the inclusion criteria and applied a total of 17 different PROM instruments including five (30%) classified as CAP specific, six (35%) as generic and six (35%) that measured functional performance or were specific to another disease. The 36-Item Short Form Survey (SF-36) was the most commonly used instrument (15 articles). Only one of 11 (9%) patient cohorts assessed using a CAP-specific instrument had a mean age ≥ 70 years. The CAP-Sym and CAP-BIQ questionnaires had sufficient content validity, though the quality of evidence for all CAP-specific instruments was rated as very low to low.

Discussion PROM instruments used to measure recovery from CAP are inconsistent in constructs measured and have frequently been developed and validated in highly selective patient samples that are not fully representative of the hospitalised CAP population. The overall content validity of all available CAP-specific instruments is unclear, particularly in the context of elderly hospitalised populations. Based on current evidence, generic health instruments are likely to be of greater value for measuring recovery from CAP in this group.

INTRODUCTION

Patient-reported outcome measures (PROMs) are critical endpoints for assessing

Key messages

- What patient-reported outcome measures (PROMs) have been used in studies of community-acquired pneumonia (CAP), what constructs do they measure and do they have adequate content validity?
- PROMs are a vital valuation tool in the current era of unsustainable growth in health costs; however, there is little consistency in the instruments used and constructs measured in studies of CAP, and no CAP-specific instrument has high-quality evidence to support its content validity.
- The limited evidence supporting existing instruments developed to measure patient-reported recovery from CAP must be addressed if we are to reduce the considerable health burden associated with this disease.

the effectiveness of patient care, and their purpose is a more sensitive and meaningful evaluation of the patient illness experience,¹ thereby quantifying and identifying changes in health.² This defining characteristic is of crucial importance as it offers a departure from system or process metrics that have historically been relied on for performance measurement.³ PROMs are therefore integral to determining value in the allocation of health resources as they measure the actual health outcomes produced.⁴ Their role is particularly relevant for conditions that are generating large and increasing health service costs, as PROMs help quantify the actual health outcomes that are achieved relative to monetary investment. The momentum surrounding PROMs usage has grown in recent years with the establishment of international collaborations for standardisation of outcome measurement^{5,6} and increasing integration of PROMs into routine clinical care within national health systems.⁷

Community-acquired pneumonia (CAP) is a common and complex disease with high



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Melanie Lloyd;
melanie.lloyd@wh.org.au

mortality, morbidity and health service expenditure.^{8 9} The incidence of CAP is strongly linked to ageing, and the costs attributable to this condition will increase into the future as elderly populations grow in most high-income countries.¹⁰ In Australia, over 40% of the half-million annual hospital bed days attributable to adult CAP occur in those aged over 80 years,¹¹ a trend which is reflected across the USA and Europe.^{12 13} The potential of PROMs to drive innovation and improve patient-centred care should be realised for this high burden condition¹⁴; however, there is ongoing debate around which instrument should be used, the relevance of alternative constructs measured and methods of application.¹⁵

Measurement of outcomes in CAP is complicated by the complexity of the illness. By manifesting in elderly populations, CAP often occurs in the context of multiple overlapping comorbidities and multisystem complications are common.^{16 17} While recent progress has been made to define patient-reported end-points for registrational drug trials in CAP,^{18 19} consideration must be given to the demographics of the sample used in psychometric evaluation of PROM instruments to ensure results are generalisable to the target population. Additionally, disease-specific tools may be of less use in pragmatic studies and routine clinical practice due to complex interactions between conditions, and this may be reflected in the decision of many CAP researchers to apply alternative generic measures of patient-reported health status. A review of PROM instruments for CAP is timely to determine what constructs have been measured in studies of recovery to date and the generalisability and relative quality of different instruments. Without access to suitably relevant and robust measurement tools, identifying optimal patient-centred management strategies for this significant and common illness is difficult.²⁰

The objectives of this systematic review are to identify, appraise and synthesise the available literature regarding the use of psychometrically validated PROMs in studies of recovery from adult CAP. Specifically, this review aims to determine which PROM instruments have been used in evaluation studies relating to this illness, the constructs measured and the settings and characteristics of CAP populations where each instrument has been applied. The secondary aim is to determine, for CAP-specific instruments identified, the settings and characteristics of populations in which these PROMs have been developed and psychometrically tested and their content validity using an established evaluation framework.²¹

METHODS

Protocol and registration

This systematic review was prospectively registered on The International Prospective Register of Systematic Reviews (PROSPERO (CRD42018099739)). It was completed under the guidance of the CONsensus based Standards for selection of health Measurement INSTRUMENTS (COSMIN) guidelines for systematic reviews of

PROMs,²² and the manuscript was prepared in accordance with the Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA) checklist.²³

Patient and public involvement

This research was done without patient involvement. Patients were not invited to participate in study design and interpretation of results or asked to contribute to the writing or editing of this paper for readability and accuracy.

Eligibility criteria

Eligibility of articles

The eligibility criteria for inclusion of an article in the review were based on published recommendations.^{21 23} Articles were included if they reported on a study that was: (1) conducted in an adult (≥ 18 years) population with a diagnosis of CAP and (2) included an evaluative application at least one validated PROM that satisfied the PROM eligibility criteria listed below. Evaluative application refers to the measurement of outcomes after the commencement of treatment of a particular disease.²⁴ Articles reporting on the development or validation of PROMs in the target population were also included. Exclusion criteria were as follows: (1) case studies or papers published in publications that are not peer reviewed, (2) conference proceedings, (3) articles with no available English language version, (4) articles exclusively investigating hospital associated, ventilator associated or aspiration pneumonias, (5) articles without independent reporting of CAP cohort data if multiple disease processes are examined together and (6) articles with diagnostic, prognostic and prophylactic objectives, where outcomes are not measured after the commencement of treatment for CAP.

Eligibility of PROM instruments

PROM instruments were eligible for inclusion in the review if they measured a construct related to: (1) symptoms; (2) general quality of life or (3) function, disability or mobility. PROMs are defined as 'a direct subjective assessment by the patient of elements of their health'²⁵ and both disease specific and generic instruments were included provided one of the above constructs is measured. The PROM in question must have undergone at least partial formal assessment of its measurement properties²² and its use adequately described in the literature to enable it to be reproduced in subsequent studies. Therefore, articles that did not use PROMs that satisfied these preliminary conditions were excluded from the systematic review. The elements of recovery from acute illness from the perspective of the patient should be considered separately from patient-reported satisfaction measures which focus on the humanity of care, rather than the outcome of treatment.⁴ Instruments that measured patient-reported satisfaction with care alone were therefore excluded. For the secondary objective, a PROM was

defined as CAP specific if the instrument was developed explicitly for use in individuals with a diagnosis of CAP.

Study identification and selection

Two databases (MEDLINE and EMBASE) were searched from their inception until 31 December 2017. Additionally, the Cochrane Library of Clinical Reviews (www.cochranelibrary.com) was searched for any systematic reviews listed under the keyword 'pneumonia' that provided a relevant source of reference studies. The search strategy for identification of relevant publications was developed in conjunction with a biomedical librarian and is included in the online supplementary material 1.

A three-step search process was used to identify relevant articles. First, duplicates were removed and then two authors (ML and LD) independently reviewed the title and abstracts of the articles, excluding studies that did not meet the inclusion criteria. Next, full-text versions of the remaining articles were obtained and reviewed to confirm eligibility. Any discrepancy was resolved through discussion with a third author (HK). Care was taken to identify situations where multiple reports were compiled from a single study to avoid repeated reporting of the same data. Where multiple articles reported data from the same participant group, we refer to the 'cohort' and only include that participant data once in the analysis. A list of included articles and associated cohorts is included in the online supplementary material 2. Finally, reference lists of included articles and any literature reviews or summary papers identified in the full-text screen were hand searched to identify potentially relevant articles that might have been missed by the search strategy (online supplementary material 3). A list of excluded full-text studies, with reasons for exclusion, is provided in the online supplementary material 4.

DATA COLLECTION

Data were extracted through the use of customised electronic forms. The following study characteristics were recorded for all articles meeting the inclusion criteria: participant demographics (ie, number, age, sex and other demographics if described), disease severity, conditions other than CAP included in study population (eg, heart failure, influenza, chronic obstructive pulmonary disease; note that CAP data must be presented separately), setting (hospitalised/managed in outpatient setting/both/other), geographical location of study (country), study design, study date and duration, name of PROM(s) utilised, citation of development and validation studies and timing of outcome measurement. In addition to the above, the following data were extracted from studies validating CAP-specific PROM instruments: mode of administration and recall period.

The following information was obtained for all included PROM instruments via references given in the included articles: target population, concepts measured, number of domains and items, scales and scoring and language.

This was to inform the understanding of the different patient-reported constructs that have been measured in studies of CAP.

Measurement framework for assessment of content validity of CAP-specific PROM instruments

CAP-specific PROMs identified through the inclusion criteria (above) were subject to a formal assessment of content validity in accordance with the COSMIN guidelines.²² This framework has been established to promote consistency in the assessment and reporting of measurement properties and has been published in conjunction with comprehensive guidance.²⁶ Under the COSMIN framework, each development and validation study for a particular PROM instrument is assessed individually for methodological quality, and findings are pooled to determine the sufficiency of the measurement property. The COSMIN framework separately examines PROMs development studies (concept elicitation, cognitive interview and pilot studies involving participants from the target population) from validity studies (where both patients and professionals may be involved in testing).

As relevancy, comprehensiveness and comprehensibility are key elements for patient-reported instruments; content validity is considered the most important measurement property and is assessed first.²⁴ The quality of the evidence supporting each measurement property is also assessed separately, using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) criteria.²⁷ Two authors (ML and LD) independently completed the quality scoring framework (outlined below) for CAP-specific PROMs identified through the search process. Where disagreement was identified between scores, a third author (EC or HK) reviewed the assessments and provided consensus.

RESULTS

PROM instruments used in evaluation studies of CAP populations

Forty-two articles were included in this review, and these articles reported results from 17 different PROM instruments meeting the inclusion criteria (figure 1). The most commonly used instrument was the 36-Item Short Form Survey (SF-36)²⁸ (15 articles, 11 cohorts), although a variety of generic and function-specific instruments were applied including instruments that were developed for other respiratory conditions (table 1). Five validated CAP-specific instruments were identified.^{19 29-32}

Characteristics of CAP populations examined using PROMs

The CAP-specific instruments have been applied in relatively young populations; only one of the 11 (9%) cohorts reported a mean age of ≥ 70 years (table 1). In contrast, each of the four functional status instruments has been applied in at least one CAP cohort with a mean age ≥ 70 years. PROMs had predominantly been applied

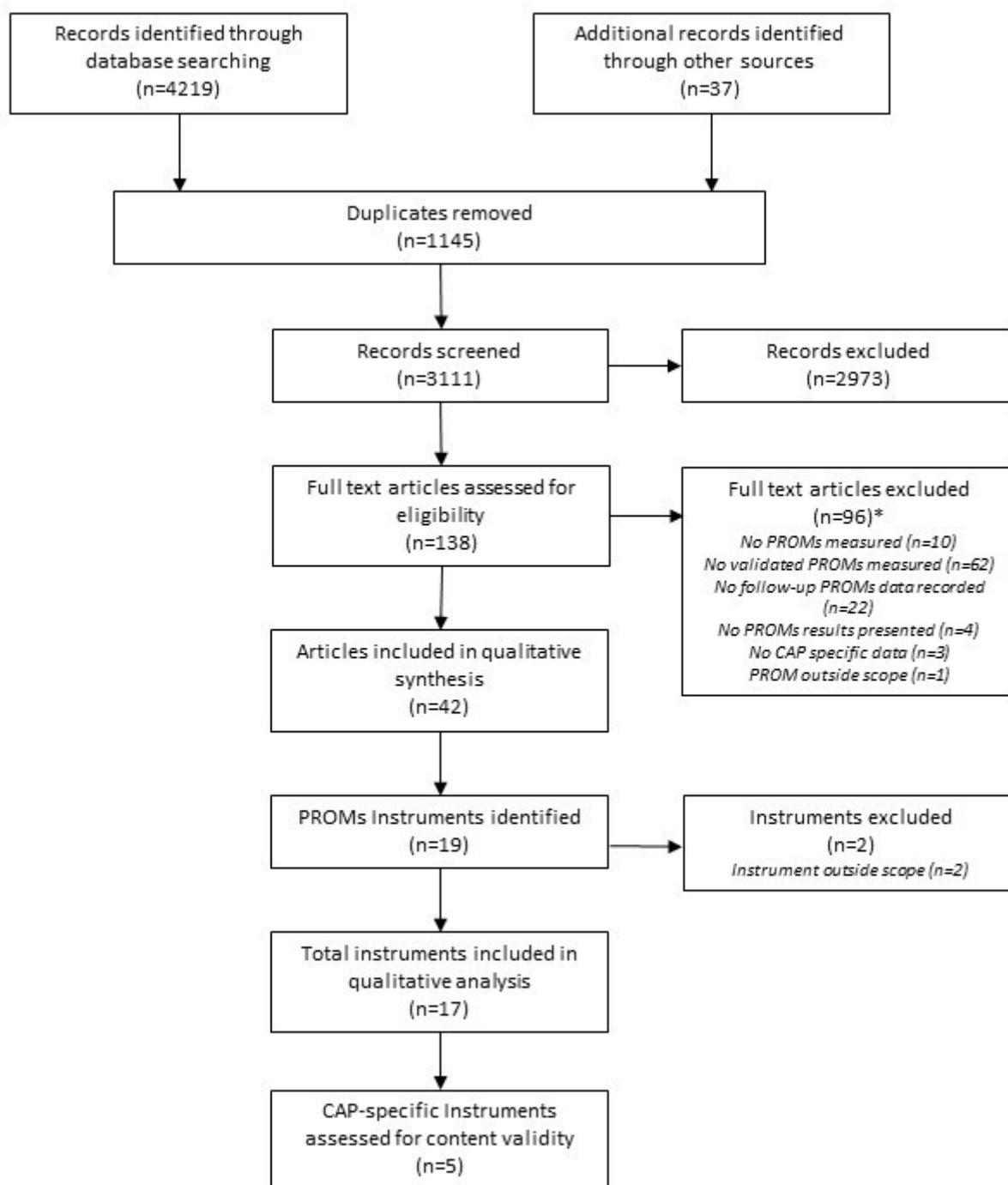


Figure 1 Preferred Reporting Items for Systematic Review and Meta-Analysis flow chart. *More than one reason for exclusion may apply to each article.

in majority hospitalised cohorts, and less than half of the study cohorts were measured beyond 1 month of diagnosis.

Constructs measured

All of the CAP-specific instruments had a primary focus on patient-reported symptoms, although two instruments (CAP-BIQ and CAP Score) also included items relating to general well-being, or psychosocial impacts and physical functioning (table 2). The generic instruments

examined constructs relating to performance of activities of daily living and overall health-related quality of life. The relative length of the different questionnaires varied greatly.

Development and content validity of CAP-specific PROM instruments

Development and validation populations

The CAP-Sym was developed for the most inclusive population (any disease severity, any age >18 years) (table 3),

Table 1 PROM instruments used in studies of recovery in CAP populations

Instrument	Number of articles	Number of cohorts*	Average number of participants per cohort	Age, years (weighted mean†, (range of study means))	Cohorts including majority hospitalised n (%)	Cohorts including high-severity illness n (%)	Cohorts measured >1 month after admission n (%)
CAP specific							
CAP-Sym	5	4	262	58.2 (51, 68)	3 (75)	4 (100)	1 (25)
CAP-BIQ	1	1	500	64.2	0 (0)	1 (100)	1 (100)
CAP Score	5	2	453	71.7 (56.6, 74)	2 (100)	1 (50)	1 (50)
Metlay Score 1	3	3	322	<60‡ (50, 64)	1 (33)	1 (33)	3 (100)
Metlay Score 2	1	1	126	52.7	0 (0)	0 (0)	0 (0)
Respiratory specific							
LCQ	1	1	44	77.0	1 (100)	0§ (0)	0 (0)
MRC scale	1	1	51	55.0	1 (100)	0 (0)	0 (0)
Functional performance							
Barthel Index	5	5	366	70.4§ (68, 79.1)	5 (100)	4§ (80)	2 (40)
Lawton ADL Index	1	1	195	72.8	0 (0)	0 (0)	0 (0)
Katz Index	3	3	698	78.6 (56.7, 86)	3 (100)	3 (100)	2 (66)
ECOG	1	1	968	72.7	1 (100)	1 (100)	0 (0)
Generic							
SF-36	15	11	455	60.2 (44.2, 76)	9 (82)	6 (55)	6 (55)
SF-12	4	4	101	56.0 (50, 68.5)	2 (50)	1 (25)	2 (50)
SF-8	1	1	475	77.0	0 (0)	0 (0)	0 (0)
EQ-5D	5	4	469	69.7 (58, 76)	4 (80)	4 (80)	3 (60)
Sickness impact profile	1	1	140	60.0	1 (100)	0§ (0)	1 (100)
WHOQOL-BREF	1	1	240	71.0	1 (100)	1 (100)	0 (0)
Total	42¶	33¶			74%	61%	48%

*A cohort is defined as a group of study participants who may have contributed data to more than one article.

†The weighted mean was calculated by multiplying the mean age of a cohort by the proportion of participants that cohort contributed to the aggregate number of participants for a given instrument and then adding across all cohorts.

‡One study reported only the proportion of participants in different age categories, not overall mean or median, although 78% were aged <60 years.

§Not reported for one cohort (excluded from aggregate calculation for that variable).

¶Each article/cohort may apply more than one PROM, and hence column does not sum to total.

ADL, activities of daily living; CAP, community-acquired pneumonia; CAP-BIQ, CAP Burden of Illness Questionnaire; CAP-Sym, CAP Symptom Questionnaire; ECOG, Eastern Cooperative Oncology Group; EQ-5D, EuroQoL-five dimension; LCQ, London Chest Questionnaire; MRC, Medical Research Council; PROM, patient-reported outcome measure; SF, short form; WHOQOL-BREF, WHO quality of life questionnaire short form.

although a relatively young cohort was used in the concept elicitation study (mean age 52 years).²⁹ The CAP-BIQ target population was aged over 50 years, and the concept elicitation cohort was recruited primarily (60%) from the outpatient setting.¹⁹ The CAP Score underwent psychometric evaluation in a hospitalised cohort (mean age 56 years),³⁰ while the Metlay symptom scores were developed for use in low-risk outpatients only.^{31 32}

Development and validation methodologies and quality of evidence

In accordance with the COSMIN framework, both the CAP Score and Metlay Score 1 received an inadequate

rating for concept elicitation as the target population were not involved in the selection of some or all items, respectively (table 4). While patients were involved in concept elicitation studies for the CAP-Sym and Metlay Score 2, these were of doubtful quality owing to lack of detail provided for the qualitative methodologies used and the questionable representativeness of the study participants. Only the CAP-BIQ concept elicitation study was judged as adequate. All CAP-specific instruments were given an inadequate rating for overall quality of the PROM development study, except for the CAP-BIQ which was rated as doubtful quality. A cognitive interview

Table 2 Design characteristics of PROM instruments used in studies of recovery from CAP

Instrument	Reference for development	Target population	Concepts measured	Items, scores and scales	Language*
CAP specific					
CAP-Sym	Lamping <i>et al</i> ²⁹	Diagnosed with CAP	Symptoms/problems associated with CAP.	Eighteen items, each six-point scale. Total score range 0–90, high values indicate poorer outcomes.	English (extensively translated)
CAP-BIQ	Wyrwich <i>et al</i> ¹⁹	Diagnosed with CAP, aged >50 years	Presence of key symptoms, comorbidities affected, psychosocial impacts, productivity and health resource use.	Twelve domains for symptoms and four other domains. Aggregate scoring methods unclear.	English
CAP-Score	El Moussaoui <i>et al</i> ³⁰	Diagnosed with CAP and hospitalised	Symptoms and general state of well-being associated with CAP.	Nine items, two domains, multiple scales. Composite score with 0–100 scale, low values indicate poorer outcomes.	Dutch (English)
Metlay Score 1	Metlay <i>et al</i> ³¹	Diagnosed with mild-CAP	Symptoms associated with CAP.	Five items, various scoring scales. Composite score 0–100 scale, low values indicate poorer outcomes.	English
Metlay Score 2	Metlay <i>et al</i> ³²	Diagnosed with mild-CAP	Symptoms associated with CAP.	Five items, each six-point scale. Composite score 0–100 scale, low values indicate poorer outcomes.	English
Respiratory specific					
LCQ	Garrod <i>et al</i> ⁴³	Severe COPD	Limitation in daily activity.	Fifteen items, four domains, each six-point scale. Composite score 0–90, high values indicate greater disability.	English (multiple translation)
MRC dyspnoea scale	Fletcher ⁴⁴	Chronic respiratory disease	Disability associated with breathlessness.	One item, graded 1–5. Higher grade indicates greater incapacity.	English (multiple translations)
Functional performance					
Barthel Index	Mahoney and Barthel ⁴⁵	Chronic neuromuscular or musculoskeletal disorders	Level of independence with ADLs.	Ten items, three-level rating with associated score. Composite score 0–100, high values indicate greater independence (modified versions exist).	English (multiple translations)
Lawton IADL Scale	Lawton and Brody ⁴⁶	Older adults (not institutionalised)	Ability to perform tasks necessary to live independently in the community.	Eight items, each with 3–5 corresponding statements from which participant selects one. Each item scored 0 or 1. Composite score 0–8, low values indicate greater incapacity.	English (multiple translations)

Continued

Table 2 Continued

Instrument	Reference for development	Target population	Concepts measured	Items, scores and scales	Language*
Katz Index	Katz <i>et al</i> ⁴⁷	Older adults	Ability to perform ADLs independently.	Six items, each rated 0 or 1. Composite score 0–6, where low scores indicate greater dependence.	English (multiple translations)
ECOG	Oken <i>et al</i> ⁴⁸	Cancer	Impact of disease on the daily living abilities of the individual.	One item, graded 1–5. Higher grade indicates greater incapacity.	English (multiple translations)
Generic					
SF-36	Ware and Sherbourne ²⁸	Any illness	Health-related quality of life.	Two component scores, eight-dimensions, 36-items (SF-36), 12-items (SF-12) and 8-items (SF-8). Composite score 0–100 for each component/dimension, higher scores indicate higher quality of life.	English (extensively translated)
SF-12	Ware <i>et al</i> ⁴⁹				
SF-8	Ware <i>et al</i> ⁶⁰				
EQ-5D	The EuroQoL Group ⁵¹	Any illness, cross-cultural evaluations	Health-related quality of life.	Five items representing five health domains, plus one global item rated on a visual analogue scale 0–100 points, higher scores indicate good health.	Extensively translated
Sickness impact profile	Bergner <i>et al</i> ⁶²	Any illness	Behavioural measure of sickness-related dysfunction.	Three dimensions, 12 categories, 136 items rated on binary scale. Dimension and category subscores may be calculated. Higher scaled scores indicate greater dysfunction.	English (multiple translations)
WHOQOL-BREF	WHOQOL Group ⁵³	Any illness, cross-cultural evaluations	Health status based on the individual's perceptions in the context of their culture and value systems.	Four domains, 26 items.	Extensively translated

*Language used in development with alternative translations in parentheses.

ADL, activities of daily living; CAP, community-acquired pneumonia; CAP-BIQ, CAP Burden of Illness Questionnaire; CAP-Sym, CAP Symptom Questionnaire; COPD, chronic obstructive pulmonary disease; ECOG, Eastern Cooperative Oncology Group; EQ-5D, EuroQol-five dimensions; IADL, instrumental activities of daily living; LCQ, London Chest Questionnaire; MRC, Medical Research Council; SF, Short Form Survey; WHOQOL-BREF, WHO quality of life questionnaire short form.

Table 3 Settings and populations used in content validity evaluation of CAP-specific PROM instruments

Instrument	Study reference	Setting	Population	Disease	Mode of instrument administration	Recall period
CAP-Sym	Lamping <i>et al</i> ²⁹	62% hospitalised. Participants from 13 countries.	n=556 Mean (SD) age 50.4 (18.7) years. 13 language groups.	Any severity	Unspecified	On day of completion
CAP-BIQ	Wyrwich <i>et al</i> ¹⁹	60% outpatients. Participants from USA.	n=500 Mean (SD) age 62.4 (8.6) years. 51% college educated.	Any severity	Web-based self-administration	Up to 120 days after diagnosis
CAP Score	El Moussaoui <i>et al</i> ³⁰	All hospitalised. Participants from the Netherlands.	n=67 Mean (SD) age 56 (17.8) years.	Severe excluded	Face-to face or telephone interview	Up to 1 month
Metlay Score 1	Metlay <i>et al</i> ³¹	65% outpatients. Participants from USA and Canada.	n=576 78% aged <60 years. 33% college educated.	Low-risk of mortality only	Face-to face or telephone interview or self-complete.	Up to 1 month
Metlay Score 2	Metlay <i>et al</i> ³²	56% outpatients. Participants from USA.	n=126 Mean age 52.7 years.	Low-risk of mortality only	Administered daily in self-complete diary format	On day of completion

CAP, community-acquired pneumonia; CAP-BIQ, CAP Burden of Illness Questionnaire; CAP-Sym, CAP Symptom Questionnaire; PROM, patient-reported outcome measure.

or other pilot test was conducted for the CAP Score and CAP-BIQ, though both were assessed to be of doubtful quality (online supplementary material 5).

While formal validation studies were conducted with patients for all instruments, these were, in all cases, inadequate for determining the content validity of the instrument. Using the GRADE criteria,²⁷ the quality of evidence for measurement of content validity was judged to be of very low quality for the CAP Score and Metlay Scores 1 and 2 (table 5). For these instruments, under the COSMIN framework, the ratings for relevance, comprehensiveness and comprehensibility are therefore

based entirely on the opinion of the authors of this review. The marginally higher ratings for evidence quality for the CAP-Sym and CAP-BIQ are attributable to completion of a content validity study in professionals (CAP-Sym) and concept elicitation or pilot study of adequate quality (CAP-BIQ).

Content validity

The items contained in all instruments were found to be relevant to the construct of interest for this review: symptoms, physical function and quality of life of adults

Table 4 Instrument development and concept elicitation study quality ratings for CAP-specific PROM instruments

Instrument	Reference	Quality of concept elicitation study		Total quality of PROM development study
		COSMIN quality rating*	Were patients involved?	COSMIN quality rating*
CAP-Sym	Lamping <i>et al</i> ²⁹	Doubtful	Yes	Inadequate
CAP-BIQ	Wyrwich <i>et al</i> ¹⁹	Adequate	Yes	Doubtful
CAP Score	El Moussaoui <i>et al</i> ³⁰	Inadequate	No	Inadequate
Metlay score 1	Metlay <i>et al</i> ³¹	Inadequate	No	Inadequate
Metlay score 2	Metlay <i>et al</i> ³²	Inadequate	Yes	Inadequate

*The COSMIN quality rating applies a four-point scale: very good, adequate, doubtful, inadequate.

CAP, community-acquired pneumonia; CAP-BIQ, CAP Burden of Illness Questionnaire; CAP-Sym, CAP Symptom Questionnaire; COSMIN, COnsensus based Standards for selection of health Measurement Instruments; PROM, patient-reported outcome measure.

Table 5 Results of overall content validity evaluation of CAP-specific PROM instruments

Instrument and reference	Relevance		Comprehensiveness		Comprehensibility	
	Rating of results	Quality of evidence	Rating of results	Quality of evidence	Rating of results	Quality of evidence
CAP-Sym ²⁹	+	Moderate	+	Moderate	+	Very low
CAP-BIQ ¹⁹	+	Low	+	Low	+	Moderate
CAP Score ³⁰	+	Very low	±	Very low	+	Low
Metlay Score 1 ³¹	±	Very low	–	Very low	?	Very low
Metlay Score 2 ³²	±	Very low	±	Very low	?	Very low

'+' indicates sufficient; '–' indicates insufficient; '±' indicates inconsistent (as defined by the COnsensus based Standards for selection of health Measurement Instruments guidelines²⁴); '?' indicates assessment unable to be completed due to sample instrument unavailable. Quality of evidence is rated on a four-point scale: high, moderate, low, very low.

CAP, community-acquired pneumonia; CAP-BIQ, CAP Burden of Illness Questionnaire; CAP-Sym, CAP Symptom Questionnaire.

recovering from CAP. The inconsistent rating for relevance given to both Metlay Scores is attributable to a lack of clear description of the construct to be measured by the instrument. Both the CAP-Sym and CAP-BIQ were found to be sufficiently comprehensive, while comprehensiveness of the other instruments was judged by the reviewers to be insufficient or inconsistent. The CAP-BIQ, CAP-Score and CAP-Sym were found to have sufficient comprehensibility, though in the case of the latter, this is based on the opinion of the reviewers, not patients themselves. The comprehensibility of the Metlay Scores could not be assessed due to the unavailability of sample data collection instruments.

DISCUSSION

This study is the first, to the authors' knowledge, to systematically review and document the availability, quality and use of PROM instruments in CAP. The significance of this illness, in terms of both morbidity and health service cost, promotes a degree of urgency for the establishment of a robust outcome framework to both understand the health gains produced and the value associated with care. Consensus on the outcomes of most importance in CAP is lacking, and ongoing debate has failed to resolve the question of the appropriateness of various clinical and patient-centred measures.^{33 34} PROMs are often overlooked in favour of objective clinical measures such as mortality or time to clinical stability which do not necessarily reflect the outcomes of most value to the patient.⁴ Recent work relating to development of standardised outcome frameworks for antimicrobial drug trials in CAP, including design of the CAP-BIQ, is certainly a move in the right direction but is yet to be completed or extrapolated to settings outside pharmaceutical registrational trials.¹⁸

The results of this review have identified that the patient-reported constructs measured in studies of recovery from CAP to date are inconsistent, and the quality of evidence to support the CAP-specific instruments applied is suboptimal. The relatively recent CAP-BIQ concept elicitation study found that CAP

survivors reported a wide range of symptoms and problems, including a significant need for caregiver assistance during recovery.¹⁹ The latter construct was not measured by any of the other CAP-specific instruments but may be captured by alternative measures such as the Barthel Index or generic quality of life instruments. In the absence of an adequately developed and validated CAP-specific instrument, clinicians have relied on these generic instruments, or those designed for other respiratory illnesses. The consequence of inconsistency in the constructs measured and instruments applied is that the value of new interventions, clinical tools and models of care to patients themselves cannot be determined or compared.

Despite the existence of five validated CAP-specific instruments, none were supported by high-quality evidence of their content validity. Four of the instruments were developed more than a decade ago, prior to establishment of advanced methodologies for concept elicitation and instrument evaluation, and none have been further validated since their development.^{29–32} Only the newer CAP-BIQ questionnaire underwent adequate instrument development methodologies, but has yet to undergo (to the authors' knowledge) formal evaluation of content validity.¹⁹ Additionally, other important measurement properties, such as structural and cross-cultural validity, responsiveness, reliability and measurement error,²² have not been evaluated for this instrument.

The other key issue with the existing CAP-specific instruments relates to the fundamental disconnect between the population used in development and testing and the population generating the bulk of the CAP health burden—those of advanced age and with complex multimorbidity.^{8 9} None of the CAP-specific instruments identified in the review were explicitly designed for use in an elderly inpatient cohort. Each instrument has been applied in only one validation study, all of which were conducted in populations with a mean age under 65 years. The only instrument that has been designed specifically for hospitalised populations (CAP Score) did not include patients in concept elicitation.

Older adults report vastly different patterns of symptoms^{35,36} and functional disability³⁷ than younger adults and therefore are likely to find measurement of alternative constructs of greater relevance. Population heterogeneity makes striking a balance between relevance and comprehensiveness difficult, due to the multisystem nature of the CAP illness, broad spectrum of symptoms reported, complexity of underlying comorbid disease and the range of age groups afflicted. For these reasons, the validity of existing CAP-specific instruments tested in younger patient groups cannot be generalised to the elderly population cohort generating the highest CAP-associated costs, where the role of PROMs is arguably most important.

The primary limitation of this study was that only CAP-specific instruments were assessed for their content validity, and the relative quality of the generic instruments, despite their frequency of application, was therefore not assessed. Disease-specific instruments that have been designed explicitly to measure recovery from CAP should be expected to have increased sensitivity to detect changes in health for sufferers of this illness when compared with generic instruments. This relies, of course, on the CAP-specific instruments having undergone appropriate design and validation methodologies in a sample that is representative of the general CAP population. Our review found that (1) none of the CAP-specific instruments have been adequately designed and validated and (2) the patient samples used in concept elicitation and validation do not represent the hospitalised CAP population. For these reasons, and on the basis of existing evidence, well-designed generic instruments such as the SF-36 and EuroQol-five dimension (EQ-5D) are likely to be more reliable and valid for use in elderly hospitalised CAP populations.^{38–40} Self-reported activities of daily living performance instruments, such as the Barthel and Katz Indexes, are another alternative, although prior studies are critical of these, with concerns raised regarding the impact of ceiling effects⁴¹ and changing gender roles.⁴²

There were several other limitations to this study. First, a focus on *recovery* from CAP means that instruments used in prognostic, prophylactic or diagnostic applications were not considered. Second, only English language studies were examined, meaning that good quality instruments that have been developed and validated in other languages might exist. Finally, application of the COSMIN assessment framework relies on detailed documentation of methodologies used in PROM development studies. For each of the CAP-specific instruments, the level of methodological detail provided was suboptimal, which made assessment of overall PROM quality difficult.

In summary, there is a lack of consistency and consensus regarding which constructs are important to measure when evaluating recovery from CAP from the patient's perspective. This may be due to the complexity of the underlying population and the degree to which different problems are important to various subgroups.

Additionally, the overall content validity of all available CAP-specific instruments is unclear, particularly in the context of elderly hospitalised populations who constitute a significant proportion of the overall healthcare burden from this condition. Based on current evidence, the generic instruments are likely to be of greater value in situations where representative elderly CAP populations are undergoing measurement.

Author affiliations

¹Melbourne Medical School, Western Precinct, University of Melbourne, Melbourne, Victoria, Australia

²Physiotherapy, Western Health, Melbourne, Victoria, Australia

³School of Medicine, Griffith University, Southport, Queensland, Australia

⁴Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia

⁵General Internal Medicine, Western Health, Melbourne, Victoria, Australia

Acknowledgements The authors would like to acknowledge the Western Health librarian, Evelyn Hutcheon, for her help with developing the review search strategy. The authors thank Professor Edward Janus for his ongoing support of the Western Health CAP research program.

Contributors ML conceived the study, coordinated data collection and analysis and drafted the manuscript. LD, EC, AK and HK contributed to the study design, data collection and analysis and reviewed the manuscript critically for intellectually important information.

Funding ML was supported by an Australian Government Research Training Scheme Scholarship at the University of Melbourne.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All data relevant to the study are included in the article or uploaded as supplementary information.

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

REFERENCES

1. Sloan JA, Halyard MY, Frost MH, *et al*. The Mayo Clinic manuscript series relative to the discussion, Dissemination, and operationalization of the food and drug administration guidance on patient-reported outcomes. *Value Health* 2007;10:S59–S63.
2. Alrubaiy L, Hutchings HA, Williams JG. Assessing patient reported outcome measures: a practical guide for Gastroenterologists. *United European Gastroenterol J* 2014;2:463–70.
3. Porter ME, Larsson S, Lee TH. Standardizing patient outcomes measurement. *N Engl J Med* 2016;374:504–6.
4. Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;346:f167.
5. Kelley TA. International Consortium for health outcomes measurement (ICHOM). *Trials* 2015;16.
6. Cella D, Riley W, Stone A, *et al*. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol* 2010;63:1179–94.
7. Devlin NJ, Parkin D, Browne J. Patient-reported outcome measures in the NHS: new methods for analysing and reporting EQ-5D data. *Health Econ*. 2010;19:886–905.
8. Sharma S, Sneath E, Cheng AC, *et al*. Community-acquired syndromes causing morbidity and mortality in Australia. *Commun Dis Intell Q Rep* 2017;41.
9. O'Donnell S, Costello RW, Humphreys H. Hospital resource utilisation by patients with community-acquired pneumonia. *Royal Collage of Surgeons in Ireland* 2017;110.

10. Organization for Economic Cooperation and Development. OECD Factbook, 2014. Available: <https://www.oecd.org/publications/factbook/> [Accessed 26 Apr 2018].
11. Australian Institute of Health and Welfare. Separation statistics by principal diagnosis in ICD-10-AM, Australia, 2013–14 to 2014–15. Available: <https://reporting.aihw.gov.au/Reports/openRVUurl.do> [Accessed 7 Mar 2017].
12. Welte T, Torres A, Nathwani D. Clinical and economic burden of community-acquired pneumonia among adults in Europe. *Thorax* 2012;67:71–9.
13. Jain S, Self WH, Wunderink RG, *et al.* Community-acquired pneumonia requiring hospitalization among U.S. Adults. *N Engl J Med* 2015;373:415–27.
14. Porter ME. Measuring health outcomes: the outcomes hierarchy. *N Engl J Med* 2010;363:2477–81.
15. Waterer G. Recovery from community acquired pneumonia: the view from the top of the iceberg. *Eur Respir J* 2017;49.
16. Corrales-Medina VF, Musher DM, Wells GA, *et al.* Cardiac complications in patients with community-acquired pneumonia. *Circulation* 2012;125:773–81.
17. Clermont G, Angus DC, Linde-Zwirble WT, *et al.* Does acute organ dysfunction predict patient-centered outcomes? *Chest* 2002;121:1963–71.
18. Talbot GH, Powers JH, Hoffmann SC, *et al.* Developing outcomes assessments as endpoints for Registrational clinical trials of antibacterial drugs: 2015 update from the biomarkers consortium of the foundation for the National Institutes of health: table 1. *Clin Infect Dis*. 2016;62:603–7.
19. Wyrwich KW, Yu H, Sato R, *et al.* Community-Acquired Pneumonia: Symptoms and Burden of Illness at Diagnosis among US Adults Aged 50 Years and Older. *Patient* 2013;6:125–34.
20. Reeve BB, Wyrwich KW, Wu AW, *et al.* ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013;22:1889–905.
21. Mokkink LB, Terwee CB, Patrick DL, *et al.* The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
22. Prinsen CAC, Mokkink LB, Bouter LM, *et al.* COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57.
23. Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
24. Terwee CB, Prinsen CAC, Chiarotto A, *et al.* COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159–70.
25. Rothman ML, Beltran P, Cappelleri JC, *et al.* Patient-reported outcomes: conceptual issues. *Value Health* 2007;10:S66–S75.
26. Terwee CB, Mokkink LB, Knol DL, *et al.* Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651–7.
27. Guyatt GH, Oxman AD, Vist GE, *et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
28. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
29. Lamping DL, Schroter S, Marquis P, *et al.* The community-acquired pneumonia symptom questionnaire: a new, patient-based outcome measure to evaluate symptoms in patients with community-acquired pneumonia. *Chest* 2002;122:920–9.
30. El Moussaoui Ret *et al.* Development and validation of a short questionnaire in community acquired pneumonia. *Thorax* 2004;59:591–5.
31. Metlay JP, Fine MJ, Schulz R, *et al.* Measuring symptomatic and functional recovery in patients with community-acquired pneumonia. *J Gen Int Med* 1997;12:423–30.
32. Metlay JP, Atlas SJ, Borowsky LH, *et al.* Time course of symptom resolution in patients with community-acquired pneumonia. *Resp Med* 1998;92:1137–42.
33. Barlow GD, Lamping DL, Davey PG, *et al.* Evaluation of outcomes in community-acquired pneumonia: a guide for patients, physicians, and policy-makers. *Lancet Infect Dis* 2003;3:476–88.
34. Hadfield J, Bennett L. Determining best outcomes from community-acquired pneumonia and how to achieve them. *Respirology* 2018;23.
35. Metlay JP, Schulz R, Li YH, *et al.* Influence of age on symptoms at presentation in patients with community-acquired pneumonia. *Arch Intern Med* 1997;157:1453–9.
36. Wyrwich KW, Yu H, Sato R, *et al.* Observational longitudinal study of symptom burden and time for recovery from community-acquired pneumonia reported by older adults surveyed nationwide using the cap burden of illness questionnaire. *Patient Relat Outcome Meas* 2015;6.
37. Martín-Salvador A, Torres-Sánchez I, Sáez-Roca G, *et al.* Age group analysis of psychological, physical and functional deterioration in patients hospitalized for pneumonia. *Archivos de Bronconeumología* 2015;51:496–501.
38. Brazier J, Roberts J, Tsuchiya A, *et al.* A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ* 2004;13:873–84.
39. Obradovic M, Lal A, Liedgens H. Validity and responsiveness of EuroQol-5 dimension (EQ-5D) versus short Form-6 dimension (SF-6D) questionnaire in chronic pain. *Health Qual Life Outcomes* 2013;11.
40. Johnson JA, Coons SJ. Comparison of the EQ-5D and SF-12 in an adult US sample. *Qual Life Res* 1998;7:155–66.
41. de Morton NA, Keating JL, Davidson M. Rasch analysis of the Barthel Index in the assessment of hospitalized older patients after admission for an acute medical condition. *Arch Phys Med Rehabil* 2008;89:641–7.
42. Graf C. The Lawton instrumental activities of daily living scale. *Am J Nurs* 2008;108:52–62.
43. Garrod R, Bestall JC, Paul EA, *et al.* Development and validation of a standardized measure of activity of daily living in patients with severe COPD: the London chest activity of daily living scale (LCADL). *Respiratory Medicine* 2000;94:589–96.
44. Fletcher CM. The clinical diagnosis of pulmonary emphysema; an experimental study. *Proc R Soc Med* 1952;45:577–84.
45. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Md State Med J* 1965;14.
46. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist* 1969;9:179–86.
47. Katz S, Downs TD, Cash HR, *et al.* Progress in development of the index of ADL. *Gerontologist* 1970;10:20–30.
48. Oken R, Horton D, Davis T, *et al.* ECOg performance status. *Am J Clin Oncol* 1982;5:649–55.
49. Ware Jr JE, Kosinski M, Keller SD. A 12-Item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34:220–33.
50. Ware JE, Kosinski M, Gandek B, *et al.* Development and testing of the SF-8™ health survey. *Qual Life Res* 2000;307–07.
51. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
52. Bergner M, Bobbitt RA, Carter WB, *et al.* The Sickness Impact Profile: development and final revision of a health status measure. *Medical Care* 1981;19:787–805.
53. THE WHOQOL GROUP. Development of the World Health organization WHOQOL-BREF quality of life assessment. The WHOQOL group. *Psychol Med* 1998;28:551–8.