EMA Emergency Medicine Australasia

Emergency Medicine Australasia (2021) **33**, 592–600 doi: 10.1111/1742-6723.13743

REVIEW ARTICLE

Review article: Emergency department crowding measures associations with quality of care: A systematic review

Peter G JONES [1,2] David MOUNTAIN [3] and Roberto FORERO [4]

[1]Department of Surgery, School of Medicine, Faculty of Medicine and Health Sciences, The University of Auckland, Auckland, New Zealand, [2]Adult Emergency Department, Auckland City Hospital, Auckland, New Zealand, [3]Emergency Department, Sir Charles Gardner Hospital, Perth, Western Australia, Australia, and [4]Simpson Centre for Health Services Research, The University of New South Wales, Sydney, New South Wales, Australia

Abstract

ED crowding has been reported to reduce the quality of care. There are many proposed crowding metrics, but the metric most strongly associated with care quality remains unknown. The present study aims to determine the crowding metric with the strongest links with processes and outcomes of care linked to the Institute of Medicine quality domains. Systematic searches in healthcare databases were conducted using terms for 'crowding', 'metrics' and 'performance', supplemented by grey literature and citation searches. The level of evidence for each association was assessed using an explicit tool. The body of evidence was assessed using the Grading of Recommendations Assessment, Development and Evaluation approach. Evidence was synthesised using harvest plots. Titles and abstracts of 2052 studies were screened, 452 selected for full-text review and 183 included. Inter-observer agreement was moderate $\kappa = 0.54$ (95% confidence interval 0.50–0.59). Two thirds were from urban tertiary hospitals in North America (65%), Australasia (13%), Europe (12%) and Asia (8%). One third provided Level 3 or higher evidence. Metrics were based on occupancy (38%), time (31%), workload (19%) or combinations (9%). Data were synthesised from 25 607 375 patients, 2368 staff, 9089 hospitals and 101 177 sampling times. Almost all crowding metrics were patient-centred and reflect timeliness and efficiency. ED length of stay, boarding time and total occupancy had the strongest association with safety and effectiveness of care. ED length of stay was also associated with equity. The certainty of evidence for associations between crowding measures varied across domains of quality, from very low to moderate certainty.

Key words: *crowding, emergency service, hospital, length of stay, quality of healthcare, systematic review.*

Key findings

- Crowding measures with the strongest associations with quality of care were total ED occupancy and total EDLOS (measured as a departmental average rather than at individual level, and in higher acuity settings).
- Occupancy varies from hour to hour and is a useful measure to trigger real-time interventions to mitigate immediate harm to patients. Whereas EDLOS at departmental level may be a more reliable measure of overall system performance over time and be more suited as a quality indicator at state or national level to drive long term system changes.

Introduction

ED crowding has long been considered a major cause of preventable harm,[1] suggesting it may be an appropriate indicator of the quality of care within health systems. However, there remains no consensus around how best to measure crowding, with many measures suggested previously. Establishing the association between crowding metrics and processes and outcomes of care is an essential step towards ascertaining which metrics should be used as quality indicators. If there is no association between the metric and either a process or outcome of care then the metric should not be used.

Previous systematic reviews of ED crowding were unable to draw conclusions about which metrics to recommend, because of the complexity of the source literature and the

Correspondence: Associate Professor Peter G Jones, Auckland District Health Board, Private Bag 92024, Victoria Street West, Auckland 1142, New Zealand. Email: peterj@adhb.govt.nz

Peter G Jones, PhD, Director of Emergency Research, Emergency Physician; David Mountain, MBBS, Emergency Physician; Roberto Forero, PhD, Senior Research Fellow.

Preliminary data from this review was presented at the Australasian College for Emergency Medicine Annual Scientific Meeting in Perth, Western Australia, Australia on November 22, 2018.

Accepted 21 January 2021

imprecise nature of the review findings.[2–13] Despite including 478 unique study references, with over 700 different metrics identified, only one previous review concluded which metrics were associated with quality of care (Appendix S1). The suggested measures were: 'the number of patients in the waiting room, ED occupancy (percentage of overall ED beds filled), and the number of admitted patients in the ED awaiting inpatient beds'.[13] However, this conclusion was based solely on a count of how often links were made between the crowding measure and the Institute of Medicine (IOM) quality domains[14] in 32 included studies,[13] representing only a fraction of the available evidence. Furthermore, the strength and direction of the associations and the quality of evidence behind the associations were not factored into this conclusion.

The question of which measures of ED crowding are most strongly associated with quality of care remains unanswered. The aim of the current study is to determine which measures were most strongly associated with care quality.

Methods

Protocol

The protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) CRD42016053307 on 14 December 2016.

Search strategy

Structured searches in the Cochrane Library, MEDLINE, Embase, CINAHL and the World of Science CORE collection (for theses) using MeSH and free-text terms for 'emergency department' 'crowding' 'metrics' and 'performance' were undertaken for all available years until October 2017 (Appendix S1). Citation searches were undertaken, and abstract authors contacted where possible for other unpublished work. There was no restriction by language.

Study selection

We included studies that were original research articles meeting the inclusion criteria: focused on ED crowding; specified a variable that measures crowding (the metric); specified an effect measure (process or outcome) influenced by the metric; and tested the association between the metric and the effect measure. Studies making general comments without specifying which measure of crowding were associated with which outcomes were excluded, as were review articles, secondary reports, duplicates, news articles, opinion pieces and editorials.

Abstracts were screened by two reviewers independently. If selection criteria appeared to be met, or it was unclear that they should be excluded, studies were considered possibly relevant. Agreement between reviewers was assessed using exact agreement and the kappa statistic. Consensus was used to resolve disagreements. Excluded studies (with reasons) are listed in Appendix S2.

Description of metrics

Metrics reported in each study were categorised as occupancy, time, workload, compound or other (see Appendix S1 for definitions).[15]

Effect measures

Effect measures were what authors used to quantify associations between crowding metrics and quality of care (Appendix S1).

Reference standards were categorised into Processes and Outcomes of care as suggested by Donabedian[16,17] and associated with the domains of quality according to those described by the IOM:[14]

- Patient-centred: Patient experience;
- Timely: Timeliness/Time to Assessment;
- Efficient: ED length of stay (LOS) components/did not wait for assessment (DNW)/Diversion;
- Effective: Receiving Care/Representation;
- Safe: Mortality/Adverse Events/Hand Hygiene/Violence in ED; and
- Equitable: Effect measure explored with respect to healthcare inequities.

Clinically important differences for times to treatment are condition-specific and discussed in the evidence summaries where relevant, based on prior studies (Appendix S1).[18–20]

Data extraction

Data were extracted and tabulated using a standard electronic form.

Strength of evidence

Included studies were critically appraised using the Graphic Appraisal Tool for Epidemiology to assign a Level of Evidence based on the 'Oxford 2011 Levels of Evidence' (Appendix S1). We used the Grading of Recommendations Assessment, Development and Evaluation process to appraise the certainty of the body of evidence for the association of each metric with effect measures.[21] The certainty of evidence from Level 4 studies were downgraded one level, and Level 5 evidence was excluded because of high risk of bias (highly uncertain).

Summarising the evidence

We used harvest plots to show the multiple dimensions of the body of evidence.[22] Clinical importance rather than degree of statistical significance determined the strength of association (see Appendix S1 for an explanation of harvest plots and how strength of association was determined). Publication bias was also assessed (Appendix S1).

Results

Search results

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses diagram (Fig. 1) shows the search results. The database search found 2095 titles and 189 were sourced by citation searchers and other sources. After removal of duplicates, 1777 titles and abstracts were screened independently by two reviewers and 1219 were excluded. There was moderate agreement between the reviewers on selection for full-text review $\kappa = 0.54$ (95% confidence interval 0.50–0.59)[23,24] and

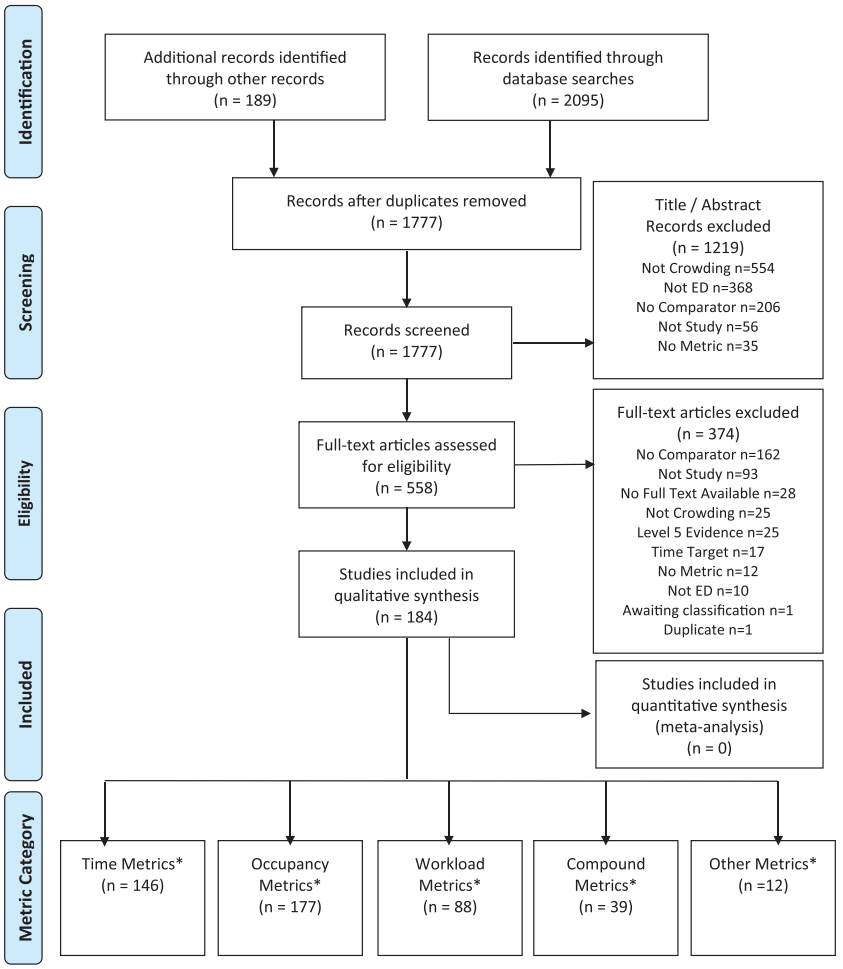

Figure 1. *The Preferred Reporting Items for Systematic Reviews and Meta-Analyses diagram. *Numerator is greater than the number of studies as more than one metric and association often reported in the source studies.*

differences were resolved by consensus. After full-text review of 558 studies, another 374 studies were excluded (Appendix S1). The 184 included studies with detailed results and their level of evidence are shown in Appendix S3.

Description of studies

Table 1 summarises the included studies, which were all observational.

Two thirds of associations were supported by Level 4 evidence at higher risk of bias. Occupancy (33%) and time (32%) metrics were most often reported followed by workload (19%) and compound metrics (8%). DNW was rarely used as a crowding metric (3%).

Patients were the unit of analysis in 141 studies (75%). The median (interquartile range [IQR]) number of included patients was 8717 (778–51 669), range 17–14 551 553. Staff were the unit of analysis in six studies (3%). The median (IQR) number of staff involved was 215 (92–347), range 76–1419. Sampling times were used as the unit of analysis in 32 studies (17%), with median (IQR) 225 (129–1179) times, range 20–52 263. Hospitals were the unit of analysis in nine studies (5%), median (IQR) 405 (67–467) hospitals, range 18–4810. As not all studies stated the number of participants, times or hospitals, respectively, the exact number involved remains unknown. Based on the studies that reported this, the results of this review are founded on 25 607 375 patients; 2368 staff; 9089 hospitals and 101 177 sampling times. Four studies used two units of analysis, so the denominator for the proportions of the units of analysis was 188.[25–28]

Associations between crowding measures and quality of care

Summary of findings, tables and harvest plots in Appendix S4 show the strength and direction of the associations, and the detailed certainty of evidence for each metric, with individual study references for each effect measure.

The measures with the strongest associations with quality of care were total ED occupancy and total EDLOS (when measured as a departmental average rather than at individual level, and in higher acuity settings). Harvest plots summarising the evidence for these metrics are shown in Figures 2 and 3, respectively.

Patient-centred

Nearly every metric studied was associated with a worse patient experience, although not in all studies. The certainty of evidence for these associations was very low for most metrics. The strongest associations were for time to assessment, waiting room and boarder occupancy (low certainty evidence). The other time metrics, DNW and number of arrivals were weakly associated (very low or low certainty evidence). Indirect occupancy was not associated with patient experience (one study) and hospital occupancy only very weakly associated in one study (very low certainty evidence).

Timely

Timeliness of care was most strongly associated with total EDLOS in most studies although there was some inconsistency, with some very weak associations in the direction opposite to expected (low certainty evidence). ED occupancy, especially occupancy by boarders and in the treatment area were moderately associated with delays to care in most studies (low certainty evidence). However, for boarder occupancy there were strong associations with more timely care in a small minority of conditions (opposite to expected), most notably for

TABLE 1. *Description of included studies*

Parameter	*n* = 184	Percent
Country		
North America	120	65
Australia	24	13
Europe	23	13
Asia	14	8
UK	3	2
Study type†, *n* = 187		
Observational retrospective	149	79
Observational prospective	39	21
Type of hospital		
Level 4 (urban tertiary academic)	111	60
Level 3 (urban secondary)	21	11
Level 2 (rural or regional)	1	1
Multiple types	39	26
Unclear	4	2
Level of evidence†, *n* = 189		
1	0	0
2	8	4
3	53	28
4	131	68
Metric category†, *n* = 459		
Time	146	32
Occupancy	177	38
Workload	88	19
Compound	39	8
DNW	12	3
Effect measures, *n* = 668†		
Processes		
Timeliness	109	16
EDLOS	106	16
Hospital LOS	24	4
Staff experience	76	11
Outcomes		
Mortality	77	12
Receiving care	46	7
DNW	59	9
Re-presentation	30	5
Other	92	14
Patient experience	49	7

Key: Level of Evidence based on critical appraisal using Graphic Appraisal Tool for Epidemiology, 1 is lowest risk of bias, 4 is highest. †Numerator is greater than the number of studies as some source studies nested more than one study within a published report, reported multiple metrics or effect measures, or provided different levels of evidence for different effect measures within the same study. DNW, did not wait for assessment; LOS, length of stay.

repeat cardiac markers. This was also the case for boarding time (Appendix S4). Indirect occupancy measures were strongly associated with timeliness of care (low certainty evidence). There was moderate certainty evidence that waiting room occupancy was weakly associated with timeliness of care as was the number of arrivals (low certainty evidence). Compound metrics varied in their association with timeliness and all evidence was very low certainty. DNW was weakly associated with timeliness, although not all associations were in the direction expected (very low certainty evidence). Hospital occupancy was mostly not associated with timeliness of care but occasionally a moderate association was observed (very low certainty evidence).

Efficient

ED occupancy, notably occupancy in the main treatment area of ED was strongly associated with efficiency, as were indirect occupancy measures (low certainty evidence). Boarder occupancy had a very weak association (low certainty evidence). There were also rare but moderate associations opposite to expected for these occupancy measures (low and very low certainty evidence). Conversely, nearly all associations reported for EDLOS and time to assessment were moderately associated with efficiency (moderate certainty evidence). Boarding time was weakly associated and treatment time very weakly associated with efficiency (low certainty evidence). Workload metrics were very weakly associated with better and worse efficiency, or not associated (very low or low certainty evidence). In contrast, the compound metrics NEDOCS and the work score were strongly associated with efficiency (moderate and low certainty evidence respectively). The EDWIN score was weakly associated with efficiency and the READI-DV was very weakly associated (both low certainty evidence). Hospital occupancy was weakly associated with efficiency in most studies (low certainty evidence).

Effective

Time metrics were found to have more associations with the

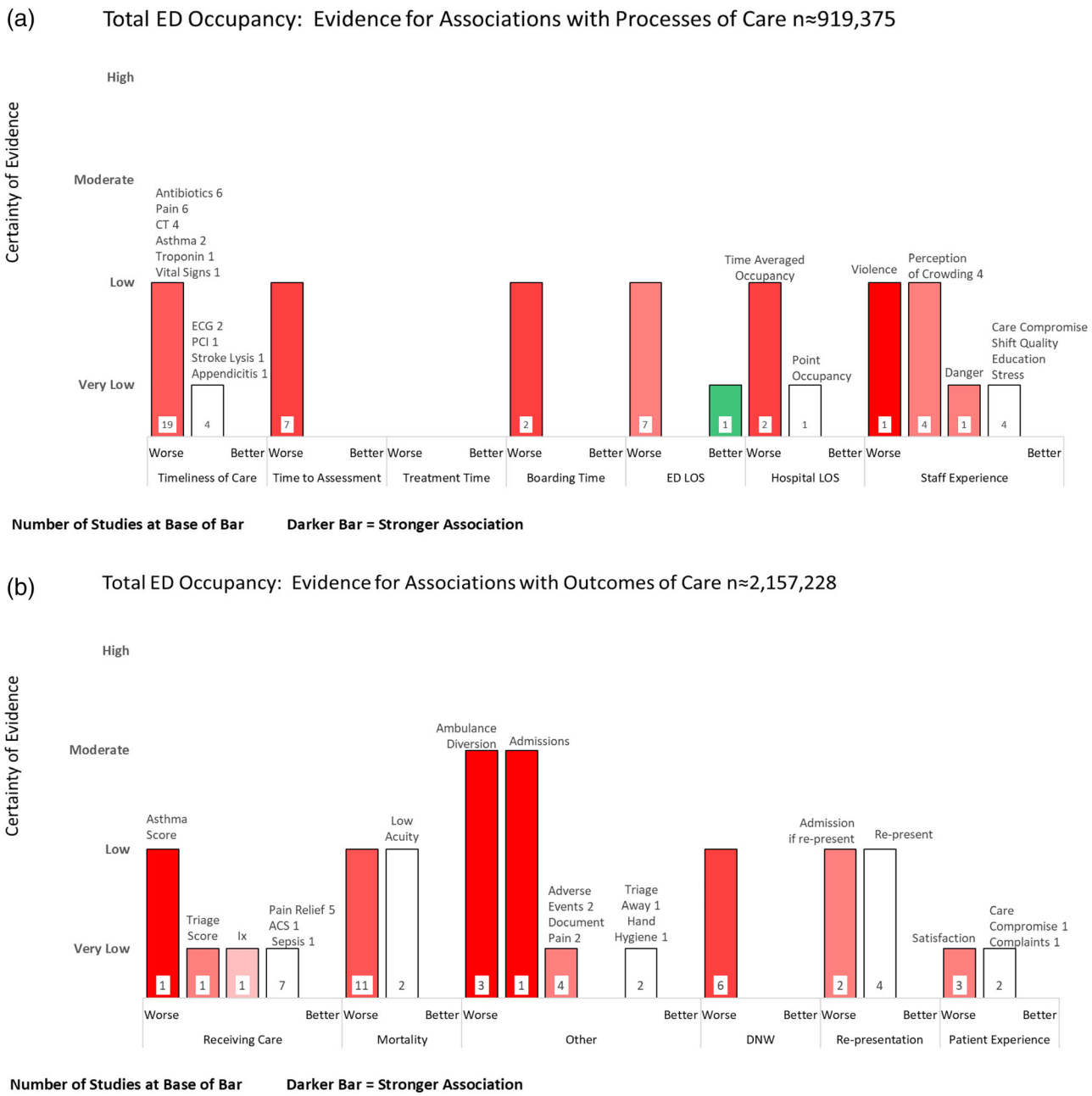

Figure 2. *(a) Processes and staff experience. (b) Outcomes and patient experience. (■), Worse with crowding; (□), no association; (■), better with crowding. ACS, acute coronary syndrome; CAP, community-acquired pneumonia; CT, computed tomography; DNW, did not wait for assessment; ECG, electrocardiogram; Ix, investigations; LOS, length of stay; PCI, percutaneous coronary intervention; Rx, treatment.*

effectiveness of care than occupancy metrics, although the associations varied from very weak to strong and there were also very weak associations in the direction opposite to expected, except for treatment time which had a strong association with effectiveness, but this moderate certainty evidence came from a single study. Similarly, the evidence of a moderate or strong association between the workload metrics number of arrivals and number waiting to be seen came from one or two studies only and was low or very low certainty. Other workload metrics, DNW and compound metrics were not associated with efficiency.

Safe

Of the time metrics, EDLOS (moderate certainty evidence) and boarding time (low certainty evidence) had the strongest associations with safety, although for both there were also very weak associations in the direction opposite to expected. Total ED occupancy was

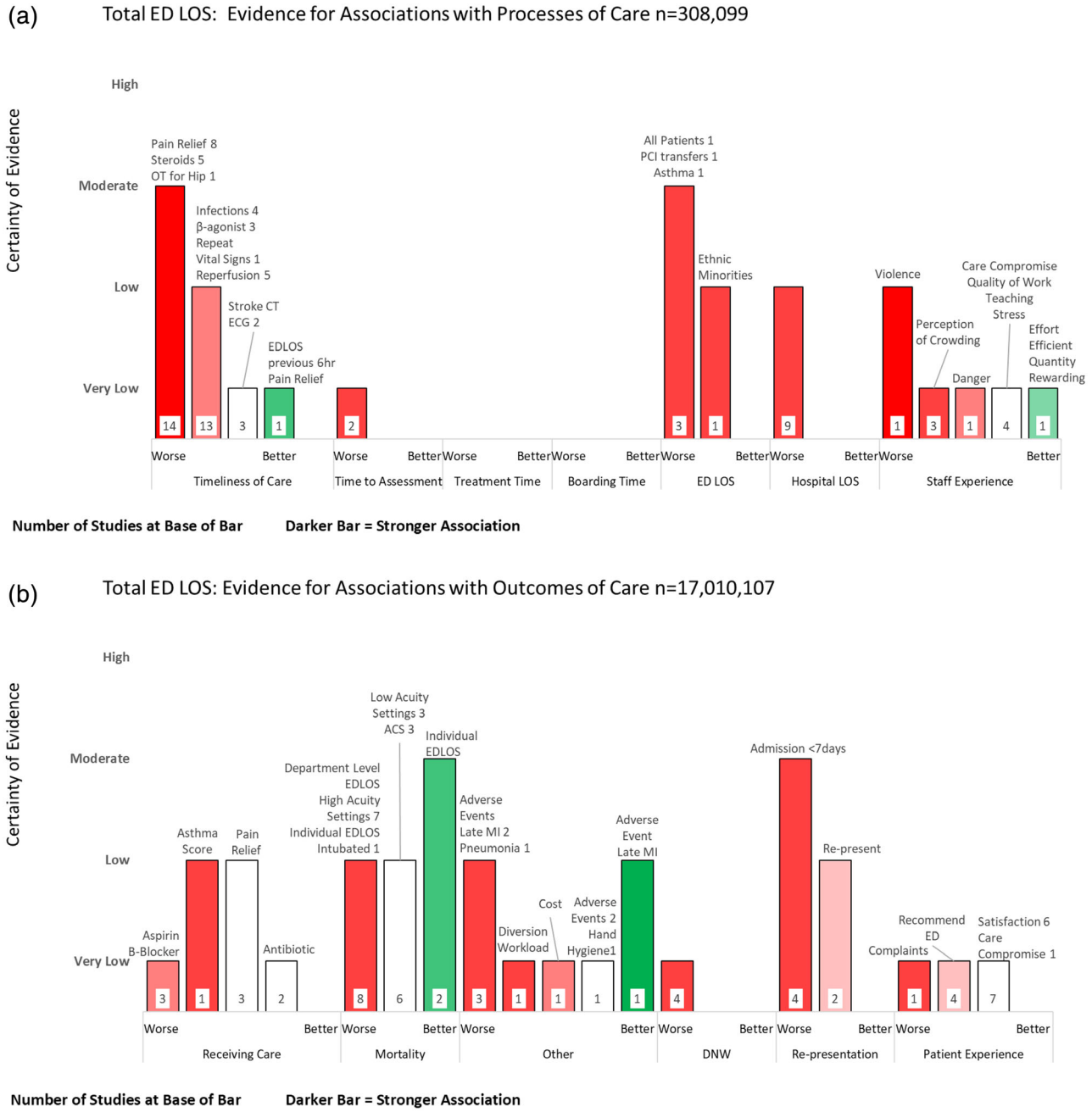

Figure 3. *(a) All processes and staff experience. (b) Outcomes and patient experience. (■), Worse with crowding; (□), no association; (■), better with crowding. ACS, acute coronary syndrome; CAP, community-acquired pneumonia; CT, computed tomography; DNW, did not wait for assessment; ECG, electrocardiogram; Ix, investigations; LOS, length of stay; PCI, percutaneous coronary intervention; Rx, treatment.*

also strongly associated with safety in about a third of studies (low certainty evidence). The occupancy metric with the strongest association was waiting room occupancy, although this evidence came from only two studies and was low certainty. Indirect occupancy measures were more strongly associated with improved safety than worse safety, although mostly there was no association (very low certainty evidence). Similarly, DNW was more often associated with improved safety although where found, all associations were very weak (very low certainty evidence). There was no evidence that workload metrics were associated with safety. Of the three compound metrics studied with respect to safety, the overcrowding hazard scale was strongly associated with safety in a single study (moderate certainty evidence),

NEDOCS was weakly associated in one study (low certainty evidence) and the EDWIN score was mostly not associate with safety (very low certainty evidence).

Equitable

Few studies explored equity in relation to crowding metrics. Total EDLOS was moderately associated with ethnicity (longer stays for ethnic minorities) in one study (low certainty evidence), while time to assessment was also longer (weak association, very low certainty evidence). The NEDOCS score was very weakly associated with an increase in implicit (but not explicit) racism in a singly study (very low certainty evidence).

Staff experience

Several studies reported staff perception of crowding or danger and the effect of crowding on staff education or stress. While most metrics here associated with staff perceptions of crowding or danger, the compound metrics were most strongly associated (low certainty evidence). No metrics were found to be associated with the quality of education or stress (Appendix S4).

Discussion

This is the most comprehensive systematic review of ED crowding metrics to date and the first to explicitly state the certainty of evidence supporting the associations between crowding metrics and quality of care. We found that the two metrics with the highest certainty evidence for associations with worse quality of care in most settings and for most conditions were total EDLOS and total ED occupancy.

EDLOS was associated with worse quality of care across all six IOM domains. In some settings and for particular conditions, a longer individual EDLOS was associated with improved quality of care, including mortality.[29,30] This suggests that for a minority of patients a longer stay in ED was better, although this may be confounded by sicker patients being prioritised for admission or requiring definitive care outside the ED. However, there were also settings where longer individual EDLOS was associated with higher mortality,[31] and when EDLOS was measured at department level, the associations were either neutral or negative (higher mortality when on average patients have longer EDLOS).[32–34] Time to assessment was strongly associated with the timeliness and efficiency domains but less so for equity effectiveness and safety. In contrast, longer boarding times were associated with worse effectiveness and safety as well as the timeliness and efficiency domains.

ED occupancy was associated with five of the six IOM domains, with no studies reporting equity in relation to occupancy. Waiting room occupancy was strongly associated with safety in one study with respect to adverse cardiac outcomes[35] but not with mortality in another study.[36] Boarder occupancy was infrequently associated with effectiveness and not associated with safety. Indirect occupancy measures such as ambulance diversion and offload delays were strongly associated with timeliness and efficiency but not with safety and effectiveness of care.

While occupancy measures the 'crowdedness' of a system at a point in time, this is directly related to the time people spend in the system, a concept encapsulated in Little's Law ($L = \lambda W$), where L is the average number of people in the system, λ is the rate of entry and W is the average time in the system.[37] This implies that EDLOS and ED occupancy should not be regarded as different measures, rather they are different ways of measuring the same concept and each will have its role. Occupancy varies from hour to hour and is a useful measure to trigger real-time interventions to mitigate immediate harm to patients. Whereas EDLOS at departmental level may be a more reliable measure of overall system performance over time and be more suited as a quality indicator at state or national level to drive long term system changes.

Workload metrics were weakly and inconsistently associated with efficiency and timeliness and mostly not associated with effectiveness and safety. The number of arrivals was the most likely of the workload metrics to be associated with care quality, but as this is out of the control of the ED or the hospital system then it is unlikely to be a useful quality indicator, as opposed to EDLOS and ED occupancy which may be amenable to system interventions.

DNW has also been suggested as a crowding measure. We found that DNW reflects patient experience and perception of timeliness of care. At department level, DNW was not associated with safety or effectiveness[32,38] while at patient level this was associated with improved mortality[32] and less violence towards staff.[39] This means DNW has limited usefulness as a quality indicator outside of patient experience.

The compound metrics include various mixtures of other crowding measures. These were associated with all domains of care in at least one study. However, the associations were mostly weak. These metrics are complex to calculate and less useful outside the setting they were derived in. As simple and generalisable crowding measures with stronger associations with quality of care are readily available, complex compound measures of crowding are unnecessary.

Only 1 of 11 prior reviews of crowding metrics associations with quality of care were able to make a conclusion around which were the best metrics.[13] Stang *et al.* concluded that the number in the waiting room, ED occupancy and boarder occupancy were the metrics most associated with care quality based on a descriptive count of the number of studies. As their review included less than one fifth of the available studies and included studies regardless of study quality, this conclusion is itself at high risk of bias (Appendix S4). We excluded studies with only descriptive results and downgraded the certainty of evidence for or against associations if the evidence came from studies at higher risk of bias.

Limitations

Although comprehensive, our search may have missed potentially relevant articles and there was a risk of

selection bias when choosing articles to include. To mitigate against this, all studies were independently screened by two authors, with a default to include the study if consensus could not be reached between the candidate and the co-reviewer. Inter-observer agreement on selection was moderate, and similar to two prior systematic reviews on crowding which found $\kappa = 0.63$ (0.58–0.68)[7] and $\kappa = 0.47$ (0.42–0.52).[3] The present review included five times more studies than any prior review in this topic area, although it is still possible that some relevant articles were not included.

Splitting the metrics into categories based on what they measured was arbitrary and other authors have suggested other systems of categorisation,[7,12] based on Asplin's classic description of input, throughput and output measures.[40] The classification used for this review was derived through a qualitative analysis in use in Australasia[15] and embraces the recommendation of Hwang *et al.* who suggested that time and occupancy (throughput) measures may be the most promising.[12]

Categorisation of the impact of crowding using a quality framework was also arbitrary and others may not have chosen the same categories. The approach chosen is consistent with that recommended by Liu *et al.*[41] and extends the previous work by Stang *et al.* who also took this approach.[13]

The data extraction process and decisions about the strength of association between crowding metrics and processes and outcomes of care was at risk of measurement bias as these were done by a single reviewer. To reduce the risk of bias an explicit guideline was followed. No previous review authors in this field have considered the strength of association between crowding metrics and outcomes based on whether the observed differences were clinically important or not.

A further limitation relates to the outcomes of staff and patient experience. The current review only included quantitative primary research studies. This excluded qualitative research that may provide insights into the experience of care and care delivery that quantitative research may not.

The assessment of publication bias is limited in that only studies that were found in the search could be assessed. It is not possible to know if other studies were done that were not submitted as abstracts due to 'negative' results.

Finally, most evidence for crowding metrics came from larger urban academic tertiary hospitals in the high income countries, and the findings of this review may not be generalisable to other settings.

Conclusion

Total EDLOS and ED occupancy are the crowding metrics with the strongest evidence of associations with the domains of healthcare quality.

Acknowledgements

The authors would like to acknowledge the contributions of Dr Sally McCarthy, Dr Yusuf Nagree and Dr Anthony Bell for their advice and support with the concept and initial searches for this review.

Author contributions

PGJ: study concept and design, protocol registration, search strategy, study selection, data extraction, data presentation, data analysis, draft manuscript, revise manuscript. Overall responsibility for manuscript. DM: study concept and design, study selection, data checking, data analysis, revise manuscript. RF: study concept and design, study selection, data checking, data analysis, revise manuscript.

Competing interests

PGJ is a section editor for *Emergency Medicine Australasia*.

Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

1. Institute of Medicine. Committee on the Future of Emergency Care in the United States Health System. *Hospital-Based Emergency Care: At the Breaking Point*. Washington: National Acadamies Press, 2006.
2. Hwang U, Concato J. Care in the emergency department: how crowded is overcrowded? *Acad. Emerg. Med.* 2004; **11**: 1097–101.
3. Hoot NR, Aronsky D. Systematic review of emergency department crowding: causes, effects, and solutions. *Ann. Emerg. Med.* 2008; **52**: 126–36.
4. Bernstein SL, Aronsky D, Duseja R *et al.* The effect of emergency department crowding on clinically oriented outcomes. *Acad. Emerg. Med.* 2009; **16**: 1–10.
5. Forero R, Hillman KM, McCarthy S, Fatovich DM, Joseph AP, Richardson DB. Access block and ED overcrowding. *Emerg. Med. Australas.* 2010; **22**: 119–35.
6. Morris ZS, Boyle A, Beniuk K, Robinson S. Emergency department crowding: towards an agenda for evidence-based intervention. *Emerg. Med. J.* 2012; **29**: 460–6.
7. Ospina MBB, Schull M, Innes G, Blitz S, Friesen C, Rowe BH. Measuring overcrowding in emergency departments: a call for standardization (structured abstract). Health Technology Assessment Database. 2006; 47.
8. Richardson DB. Emergency Department Overcrowding and Access Block. Author Provided. 2011; Doctor of Medicine.
9. Wiler JL, Griffey RT, Olsen T. Review of modeling approaches for emergency department patient flow and crowding research. *Acad. Emerg. Med.* 2011; **18**: 1371–9.
10. Higginson I. Emergency department crowding. *Emerg. Med. J.* 2012; **29**: 437–43.
11. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: a systematic review of causes, consequences and solutions. *PLoS One* 2018; **13**: e0203316.
12. Hwang U, McCarthy ML, Aronsky D *et al.* Measures of crowding in the emergency department: a systematic

review. *Acad. Emerg. Med.* 2011; **18**: 527–38.
13. Stang AS, Crotts J, Johnson DW, Hartling L, Guttmann A. Crowding measures associated with the quality of emergency department care: a systematic review. *Acad. Emerg. Med.* 2015; **22**: 643–56.
14. Institute of Medicine. Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Dulles: The National Academy Press, 2001.
15. Jones P, Wells S, Ameratunga S. Towards a best measure of emergency department crowding: lessons from current Australasian practice. *Emerg. Med. Australas.* 2018; **30**: 214–21.
16. Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988; **260**: 1743–8.
17. Donabedian A. Evaluating the quality of medical care. *Milbank Mem. Fund Q.* 1966; **44**: 166–203.
18. Jones P, Wells S, Harper A *et al.* Is a national time target for emergency department stay associated with changes in the quality of care for acute asthma? A multicentre pre-intervention post-intervention study. *Emerg. Med. Australas.* 2016; **28**: 48–55.
19. Jones P, Le Fevre J, Harper A *et al.* Effect of the shorter stays in emergency departments time target policy on key indicators of quality of care. *N. Z. Med. J.* 2017; **130**: 35–44.
20. Jones P, Wells S, Harper A *et al.* Impact of a national time target for ED length of stay on patient outcomes. *N. Z. Med. J.* 2017; **130**: 15–34.
21. Schünemann H, Brożek J, Guyatt G, Oxman A. *GRADE Handbook*. 2013. [Cited 11 Nov 2018.] Available from URL: https://gdt.gradepro.org/app/handbook/handbook.html
22. Ogilvie D, Fayter D, Petticrew M *et al.* The harvest plot: a method for synthesising evidence about the differential effects of interventions. *BMC Med. Res. Methodol.* 2008; **8**: 8.
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74.
24. Seigel DG, Podgor MJ, Remaley NA. Acceptable values of kappa for comparison of two groups. *Am. J. Epidemiol.* 1992; **135**: 571–8.
25. Pines JMG, Baxt WG, Rhodes KV, Shofer FS, Hollander JE. ED crowding is associated with variable perceptions of care compromise. *Acad. Emerg. Med.* 2007; **14**: 1176–81.
26. Wang H, Robinson RD, Bunch K *et al.* The inaccuracy of determining overcrowding status by using the national ED overcrowding study tool. *Am. J. Emerg. Med.* 2014; **32**: 1230–6.
27. Truong M, Meckler G, Doan QH. Emergency department return visits within a large geographic area. *J. Emerg. Med.* 2017; **52**: 801–8.
28. Weiss SJ, Ernst AA, Derlet R, King R, Bair A, Nick TG. Relationship between the National ED Overcrowding Scale and the number of patients who leave without being seen in an academic ED. *Am. J. Emerg. Med.* 2005; **23**: 288–94.
29. Yergens DW, Ghali WA, Faris PD, Quan H, Jolley RJ, Doig CJ. Assessing the association between occupancy and outcome in critically ill hospitalized patients with sepsis. *BMC Emerg. Med.* 2015; **15**: 31.
30. Conway R, O'Riordan D, Silke B. Targets and the emergency medical system – intended and unintended consequences. *Eur. J. Emerg. Med.* 2015; **22**: 235–40.
31. Intas GD, Stergiannis PI, Vassilopoulos GS, Fildissis GA. Impact of critically ill patients in regards to the boarding time from the emergency department to the intensive care unit [Greek]. *Rostrum of Asclepius/Vima tou Asklipiou.* 2011; **10**: 1.
32. Guttmann A, Schull MJ, Vermeulen MJ, Stukel TA. Association between waiting times and short term mortality and hospital admission after departure from emergency department: population based cohort study from Ontario, Canada. *BMJ* 2011; **342**: d2983.
33. Geelhoed GC, de Klerk NH. Emergency department overcrowding, mortality and the 4-hour rule in Western Australia. *Med. J. Aust.* 2012; **196**: 122–6.
34. Mitra B, Cameron PA, Archer P *et al.* The association between time to disposition plan in the emergency department and in-hospital mortality of general medical patients. *Intern. Med. J.* 2012; **42**: 444–50.
35. Pines JM, Pollack CV Jr, Diercks DB, Chang AM, Shofer FS, Hollander JE. The association between emergency department crowding and adverse cardiovascular outcomes in patients with chest pain. *Acad. Emerg. Med.* 2009; **16**: 617–25.
36. McCusker J, Vadeboncoeur A, Levesque JF, Ciampi A, Belzile E. Increases in emergency department occupancy are associated with adverse 30-day outcomes. *Acad. Emerg. Med.* 2014; **21**: 1092–100.
37. Little JDC. Little's law as viewed on its 50th anniversary. *Oper. Res.* 2011; **59**: 536–49.
38. Mullins PM, Pines JM. National ED crowding and hospital quality: results from the 2013 hospital compare data. *Am. J. Emerg. Med.* 2014; **32**: 634–9.
39. Medley DB, Morris JE, Stone CK, Song J, Delmas T, Thakrar K. An association between occupancy rates in the emergency department and rates of violence toward staff. *J. Emerg. Med.* 2012; **43**: 736–44.
40. Asplin B, Magid D, Rhodes K, Solberg L, Lurie N, C CA. A conceptual model of emergency department crowding. *Ann. Emerg. Med.* 2003; **42**: 173–80.
41. Liu SW, Singer SJ, Sun BC, Camargo CA Jr. A conceptual model for assessing quality of care for patients boarding in the emergency department: structure-process-outcome. *Acad. Emerg. Med.* 2011; **18**: 430–5.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Appendix S1. Methods background.
Appendix S2. Excluded studies.
Appendix S3. Included studies.
Appendix S4. SOF and harvest plots for each metric.