BMJ Open

Recommendations for the use of biomarkers for the management of adults with sepsis: a scoping review and critical appraisal

Miriam Mateos-Haro ,[1,2] Ana Garcia-Santa-Vinuela ,[3] Daniel Molano-Franco,[4] Ivan Solà ,[5,6] Federico Gordo-Vidal,[7,8] María Cruz Martín-Delgado ,[9,10] Jesus Lopez-Alcalde,[1,6,10,11] Javier Zamora [1,6]

To cite: Mateos-Haro M, Garcia-Santa-Vinuela A, Molano-Franco D, *et al*. Recommendations for the use of biomarkers for the management of adults with sepsis: a scoping review and critical appraisal. *BMJ Open* 2025;**15**:e090922. doi:10.1136/bmjopen-2024-090922

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2024-090922).

JL-A and JZ contributed equally.

Received 07 July 2024
Accepted 10 June 2025

For numbered affiliations see end of article.

Correspondence to
Professor Javier Zamora;
jzamora@salud.madrid.org

ABSTRACT

Objective A synthesis and appraisal of the recommendations for biomarkers in practice guidelines concerning sepsis is required to consolidate evidence-based practice. We generated an evidence gap map (EGM) on the use of biomarkers for managing adults with sepsis.

Design Scoping review.

Data sources MEDLINE, Guidelines International Network, Pan American Health Organization, Trip Database and UpToDate were searched from 2016 to March 2025.

Eligibility criteria Guidance documents (GD) that searched at least one literature source and provided clinical recommendations for the use of biomarkers for the management (diagnosis and prognosis, including treatment response) of adults with sepsis.

Data extraction and synthesis Two reviewers independently applied the eligibility criteria and extracted data. We used the AGREE-II (Appraisal of Guidelines for Research and Evaluation) tool to assess the GD quality. GDs that scored ≥50% on the AGREE-II 'Rigour of development' domain were considered robust. We also applied the GRADE (Grading of Recommendations Assessment, Development and Evaluation) system to evaluate if the recommendations were strong or conditional.

Results We found 10 GDs, with only half (4/8) having a robust methodology. There were 31 recommendations concerning biomarkers. Among these, 24 (77.4%) recommendations were about single biomarkers, with lactate (23; 74.2%) and procalcitonin (8; 25.8%) most frequently recommended. Biomarker testing focused on prognosis in 28 (90.3%) recommendations. Overall, 16 (51.6%) recommendations were graded strong and 13 (42.0%) were conditional, which we displayed in an EGM.

Conclusions The methodology of GDs concerning adult sepsis was poor. Our review calls for more prudent use of biomarkers in specific prognostic scenarios and in combination with standard clinical assessments. Enhancing the methodological quality of future GDs is essential to generate more valid and robust recommendations for optimising patient care.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ We deployed a systematic literature search covering the last decade, ensuring a uniform sepsis definition across the guideline documents included.
⇒ The guidance documents covered sepsis in hospital inpatients, generally and in subgroups.
⇒ There was an absence of a predefined, validated threshold for methodological quality assessment in the Appraisal of Guidelines for Research and Evaluation-II tool.
⇒ We present a state-of-the-art interactive evidence gap map concerning biomarkers in adults with sepsis.

INTRODUCTION

Biomarkers help in predicting prognosis, selecting suitable candidates for tailored treatments, optimising drug dosages and detecting therapeutic and adverse responses. By performing patient stratification, they can permit personalised medicine.[1] A broad range of biomarkers can be consulted in the Marker DB database.[2] Incorporating biomarkers into clinical practice requires careful evaluation of their performance with respect to validity, utility, cost-effectiveness and impact on patient outcomes.[3]

Over 250 biomarkers have been evaluated in the last decade.[4] They are mainly host–response inflammatory biomarkers, which can be used for diagnostic, prognostic or monitoring purposes. The Surviving Sepsis Campaign guidelines[5] state that sepsis biomarkers can complement the clinical evaluation. However, there is controversy. For example, concerning sepsis diagnosis, the Sepsis-3 consensus claims that the role of biomarkers remains undefined.[6] The most frequently studied biomarkers in sepsis include lactate, procalcitonin (PCT) and C-reactive protein (CRP). There are other

recognised biomarkers such as interleukin-6 (IL-6) and presepsin (P-SEP). Their utility in clinical practice is unclear.[6]

Guidance documents (GD), such as clinical practice guidelines (CPGs), consensus statements (CS) or position statements (PS), are tools for the translation of biomarker research evidence into practice.[7] CPGs are systematically developed statements that provide recommendations for clinical practice and should be informed by a systematic review of evidence and an assessment of the benefits and costs of alternative care options intended to optimise patient care. The CPG recommendations should be formulated with a scientific, rigorous and transparent methodology and should be periodically updated.[8] CSs and PSs tend to be less methodologically rigorous compared with CPGs, and they may serve as an initial attempt to generate recommendations based on collective opinion or consensus of experts.[9] They have been criticised for their potential bias and lack of transparency.[10] To maximise objectivity in recommendations, some sepsis GDs have been formulated with the GRADE (Grading of Recommendations Assessment, Development and Evaluation) system.[11] Frequently, however, GDs of different organisations have disparities in topic coverage and even have conflicting recommendations.[12] In this situation, evidence synthesis using an evidence gap map (EGM), a comprehensive and visually intuitive representation of the existing research, is considered helpful.[13]

In this background, we undertook a scoping literature review to identify, describe and assess the quality of the GDs making recommendations about the use of biomarkers in the management (diagnosis and prognosis, including treatment response) of adult patients with sepsis. We identified, described and summarised the clinical recommendations for biomarker use in the included GDs, creating an EGM with a user-friendly interface and searchable functionality.

METHODS

This study[14] was prospectively registered in the Open Science Framework repository on 6 September 2023, and it was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews (PRISMA-ScR)[15] (online supplemental Appendix 1).

Information sources and search strategy

The search strategy was designed by an experienced librarian (IS). The searches were conducted between 2016 and March 2025 (online supplemental Appendix 2). It covered the following search engines: Medline (via Pubmed), Guides International Network (GIN) (https://g-i-n.net/), Pan American Health Organisation (PAHO) (www.paho.org/en), Trip Medical Database (www.tripdatabase.com/) and UpToDate (www.uptodate.com/contents/search). We restricted the time limit to 2016 to ensure that the GDs adhered to the most recent recommended definition of sepsis and septic shock.[6] We did not restrict to any language or geographic area. We used EndNote X9 software[16] for the management of the search results, and the study selection process was detailed using a PRISMA flow diagram.[17]

Eligibility criteria

We included all GDs, whether CPGs, CS or PS, as long as they deployed a documented search strategy (at least one search source) and made recommendations for the clinical use of biomarkers for any aspect of the management (diagnosis and prognosis, including treatment response) of adults (>18 years of age) with sepsis in general hospital setting, emergency departments or intensive care units (ICU). The structure of the clinical question and the exclusion/inclusion criteria can be accessed in online supplemental Appendix 3.

Document selection and data extraction

The search results were exported to Rayyan software[18] and de-duplicated. Two independent reviewers (AG-S-V and MM-H) screened studies by title and abstract and full text, following prespecified eligibility criteria. Discrepancies were resolved by discussion. Two reviewers (AG-S-V and MM-H) developed and tested a charting form before use. Finally, the same two independent reviewers collected data on the EPPI-Reviewer web.[19] Extracted data included details related to the characteristics of the documents and the recommendations: study methods, participants, concept, context and key findings relevant to the review question. Disagreements in data extraction were resolved through discussion or with an additional reviewer (JL-A). If appropriate, authors of papers were contacted to request missing or additional data.

The authors of some GDs used the GRADE system to generate recommendations, classifying the certainty of the evidence in one of four levels: high: further research is very unlikely to change our confidence in the estimate of effect; moderate: further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate; low: further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate; and very low: any estimate of effect is very uncertain. The authors of the GDs, using the GRADE system,[20] also categorised the recommendations by their strength in two levels: strong: the majority, if not all individuals, would opt for the suggested intervention; and conditional: there is likely to be significant variation in the decisions made by informed individuals. When analysing the wording of recommendations, where the strength was not explicitly mentioned among guidelines that did not use GRADE, we adopted a specific interpretation strategy as follows: If the authors utilised the term ‘must’, we inferred the recommendation as being strong. Conversely, if the terms ‘should’ or ‘suggest’ were employed, we interpreted the recommendation’s strength as conditional.

Critical appraisal of the guidance documents

The methodological quality of the included GDs was assessed using the Appraisal of Guidelines for Research and Evaluation (AGREE)-II tool,[21] which evaluates rigour and transparency across six domains: scope and purpose, stakeholder involvement, rigour of development, clarity of presentation, applicability and editorial independence, which contained various items. Detailed information about the items is presented in online supplemental Appendix 4. Two independent reviewers graded each item from 1 (not reported) to 7 (reported with high quality). The disagreements were resolved by consensus. We obtained a final score for each of the six domains, calculated by adding each reviewer's rating and converting it into a percentage of the maximum possible score, presenting from 0% to 100%. Moreover, the overall score for each rating was classified into three categories: recommended, recommended with modifications and not recommended, based on their potential clinical application.[21] Each reviewer assigned an overall score ranging from 1 (lowest) to 7 (highest), defined by our team's consensus opinion, as follows: recommended (7.0), recommended with modifications (4.0–5.9) and not recommended (≤3.9).[22] Finally, the scores per domain (%) and overall per document (1–7) were summarised in a table. Without a specific cutoff point to determine the quality of the CPGs, we considered, based on our team's consensus opinion, that documents with a score of ≥50% in domain 3 ('rigour of development') had a robust methodology, addressing the systematic search, study evaluation, recommendation formulation processes and expert external review.[22 23]

Data synthesis

We performed a descriptive analysis of the characteristics of the included GDs and their findings. We examined the recommendations' methodological framework to determine if they aligned with the clinical question structures of PIRD (population, index test, reference test, diagnostic outcome) and PICOTS (population, intervention, comparator, outcome, timing, setting) for diagnosis and prognosis, respectively. We reported the results of this review using tables and figures.

The extracted data from EPPI-Reviewer[18] was compiled into a JSON file and exported into an EPPI-Mapper wizard,[24] which created the interactive EGM. The EGM presented the characteristics of the recommendations. Some of them mentioned multiple biomarkers, so finally, we created an entry for each biomarker separately. The rows containing the frequency of each biomarker in the recommendations were mapped onto the columns containing the purpose (diagnosis or prognosis) and direction of the recommendations (against use, for use, no recommendation). The squares comprised the strength and had different colours: green (strong), yellow (conditional) or pink (no recommendation) and different sizes, depending on the frequency (n) of the strength category.

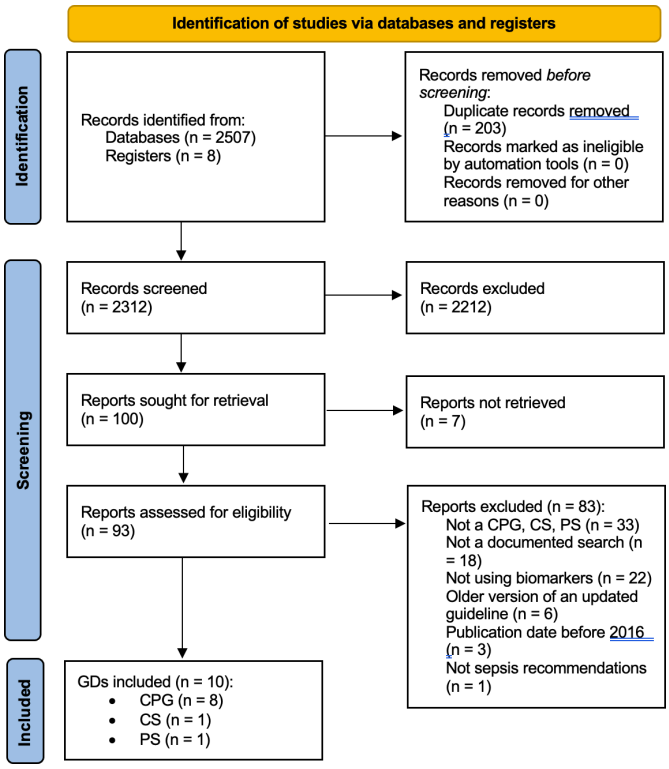

Figure 1 Flow diagram of guidance document selection in the scoping review of the use of biomarkers in adult sepsis. CPG, clinical practice guidelines; CS, consensus statements; PS, position statements.

Patient and public involvement

None.

RESULTS

Search and selection

The searches identified 2515 records (2507 from databases and eight from registers). Once duplicates had been removed, we had a total of 2312 records for the initial screening. After the initial screening of titles and abstracts, 93 documents were eligible for assessment in full text. From them, 83 were excluded for the following reasons: not CPGs, CS, PS (n=33), not supported on a documented search (n=18), not using biomarkers (n=22), older version of an updated guideline (n=6), publication date before 2016 (n=3) and not sepsis recommendations (n=1). Finally, 10 GDs (8 CPGs, 1 CS and 1 PS) were included (figure 1).

Characteristics and quality of the included guidance documents

This scoping review included ten GDs[5 25–33] outlined in online supplemental Appendix 5.

The searches supporting each GD were performed from 2017 to 2023, and in four cases, the date was not reported (40.0%).[27 30–32] All GDs were published from 2017 to 2025. The documents exhibit a varied geographical distribution, with 40.0% originating from Europe (n=4), and the remaining 60.0% representing a diverse

range of regions, including East Asia and the Pacific (n=2), Europe and North America (n=1), Latin America and the Caribbean (n=1), North America (n=1), and Oceania (n=1). The documents were mainly led by scientific societies (60.0%), followed by government organisations (30.0%) and non-governmental organisations (10.0%). All documents focused on adult populations. They targeted adults only in five GDs (50.0%) and adults plus children (≥12 years) with sepsis in two GDs (20.0%) and three GDs (30.0%) targeted subgroups: neutropenic adults (n=1)[25] and pregnant women (n=2).[30 33]

The domain-standardised scores and the overall scores for the evaluated documents with the AGREE-II tool are presented in online supplemental Appendix 5. Full assessments are detailed in the supplementary materials. The quality of the included documents showed a median overall score of 50.0% (IQR 33–75%). In relation to the evaluation per domain, the ones which obtained the highest scores were 'Scope and purpose' (median 72.0%; IQR 43–78%) and 'Editorial independence' (71.0%; 58–79%), followed by 'Clarity of presentation' (56.0%; 50–58%), all scoring ≥50%. The domains that scored <50% were 'rigour of development' (44.0%, 20–75%), 'Stakeholder involvement' (43.0%; 22–50%) and 'Applicability' (21.0%; 8–31%). Five documents (5/10; 50.0%) showed scores ≥50% for at least three AGREE II domains.[5 26–29] The rigour of the development domain, which evaluates the reliability of the documents based on its methodology, revealed that only four (4/10, 40.0%) achieved a score ≥50%. According to the six domain scores, NICE[28] had the best overall score (83.0%), followed by J-SSCG[27] and SSC[5] (75.0%), NCEC[29] and IETSI[26] (67.0%). The following four with low scores were DGHO,[25] KSCCM,[32] RCOG[33] and SOMANZ[30] (33.0%). Notably, AAEM,[31] the unique CS, had the lowest score (8.0%).

Clinical recommendations in the guidance documents

The 10 included documents provide 31 recommendation statements (online supplemental Appendix 6). The biomarker most frequently assessed was lactate (addressed in 23 recommendations), followed by PCT (n=8), CRP (n=8), IL-6 (n=3) and P-SEP (n=2). The recommendations predominantly focused on a single biomarker (n=24, 77.4%). Moreover, seven recommendations addressed combinations of diverse biomarkers: CRP, IL-6, PCT, and P-SEP (n=2[27]), CRP, IL6 and PCT (n=1,[25]) and CRP plus lactate (n=4[28 33]). The recommendations had a prognostic (n=28, 90.3%) or diagnostic (n=3, 9.6%) purpose.

Regarding the diagnostic purpose (n=3), all the recommendations endorse the use of biomarkers and none specified the population under consideration. The index test is described in all cases, and all recommendations (n=3) focused on various biomarkers (recommendation #1: 'biomarkers in general', recommendation #8, #9: 'CRP, IL-6, P-SEP and PCT'). Two recommendations described the reference test (2/3; 75.0%) (recommendation #8 and #9: 'observation of general conditions') and also the outcome (recommendation #8 and #9: 'sensitivity and specificity')

Within the prognostic category, mostly all recommendations support the use of biomarkers, mainly in combination with other actions (24/31, 77.4%) as laboratory parameters, as part of the resuscitation protocol and as support of other interventions. Three recommendations suggested the use of biomarkers alone (recommendation #19, lactate: 'in all manners measured in the ED', recommendation #25, lactate: 'measuring blood lactate', recommendation #10, PCT: 'as an indicator for stopping antimicrobial therapy' and recommendation #30, lactate: 'serum lactate should be measured urgently'). On the contrary, one recommendation was against their use alone (recommendation #28, PCT 'to decide when to start antimicrobials' (online supplemental Appendix 6)). In the two remaining cases, there was insufficient evidence to provide a recommendation. 19 recommendations were developed according to a formal method, the GRADE system (18/31; 58.0%) and the remaining recommendations (13/31; 42.0%) did not report a development method.

We ascertained if the recommendations were structured in a methodological framework, such as PIRD and PICOTS. For the prognostic purpose (n=28), the population element of the framework was present in mainly all recommendations (26/28; 92.8%). Furthermore, the description of the prognostic factor was addressed across all recommendations. The primary prognostic factor, the biomarker, was available also in the Open Science as explicitly mentioned in 89.2% of cases, being not specified in three cases (recommendations #3, #4, #5). The comparator was not explicitly mentioned in all cases. The outcome was mainly not included except once (recommendation #19: 'increasing mortality and poor outcomes'). Moreover, the timing of the outcome was not described in any case, and the setting was mentioned only on 3/28 (referred to as 'acute hospital setting' in recommendation #17) and as 'emergency department' in recommendation #19, #20).

Regarding the direction of the recommendations (online supplemental Appendix 6), most recommendations pointed to using the biomarker (28/31; 90.3%) and one case guided against its use (3.2%) (recommendation #16). The remaining two did not emit any recommendations because no evidence was found for the question of interest.

Regarding the certainty of the evidence supporting the recommendations, stated by the authors of the GDs that deployed GRADE, the majority were classified as low or very low (15/19; 79.0%), while a few recommendations had moderate certainty (3/19; 15.8%) or were unclear (1/19; 5.2%). The system for evaluating the remaining GDs was not reported (11/28, 39.3%).

Only 10 out of the 31 recommendations (32.2%) made the strength of the recommendation explicit. However, assuming the strength based on the use of language, the strength was defined by our team as conditional in 13

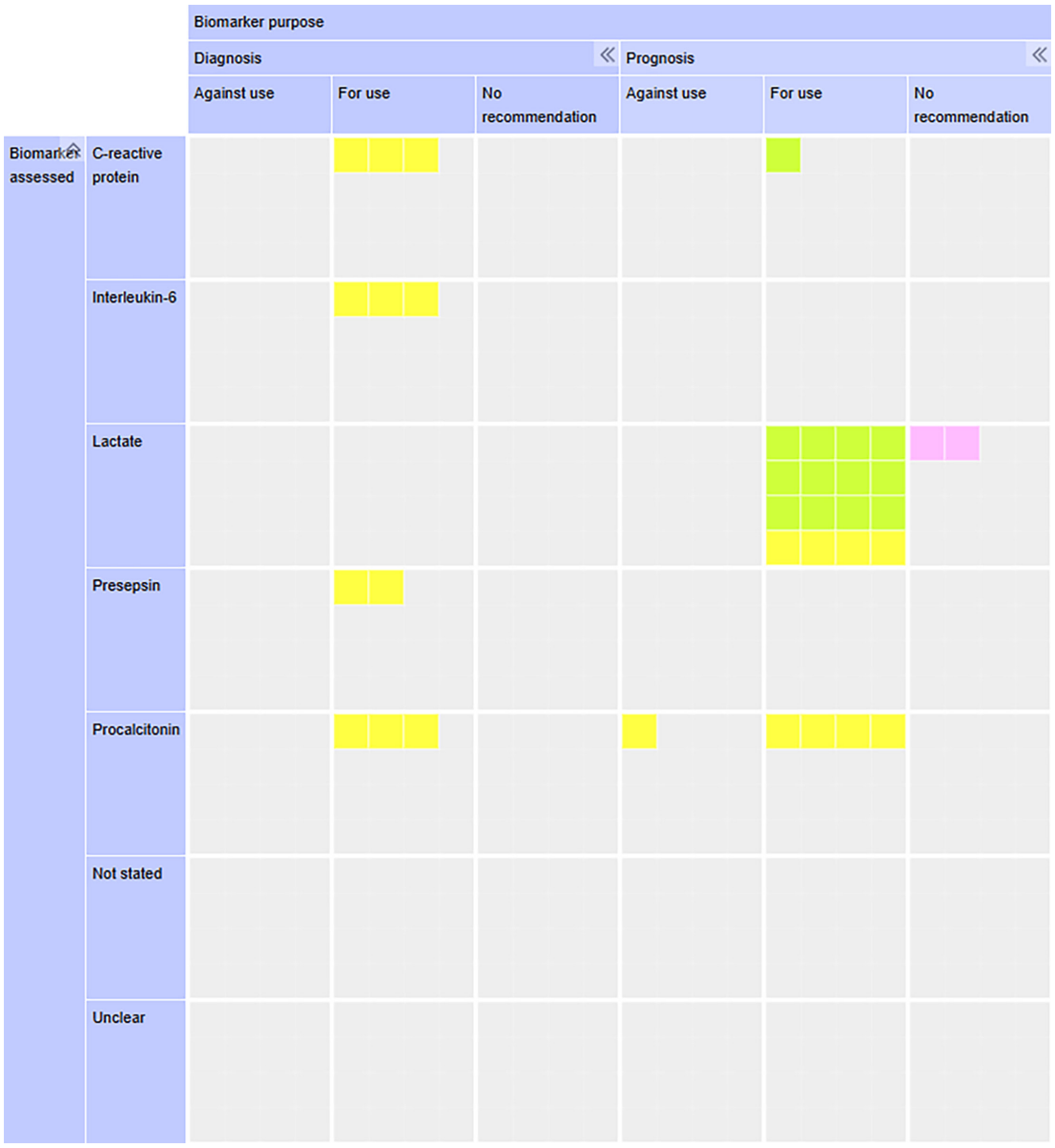

Figure 2 Static image of the EGM generated with EPPI-Mapper in the scoping review of the use of biomarkers in adult sepsis (DOI 10.17605/OSF.IO/N9RW6). Green: strong, yellow: conditional, pink: no recommendation. EGM, evidence gap map.

out of 31 cases (42.0%) and strong in 16 out of 31 cases (51.6%), being not applicable in two cases (recommendation #6 and #21). Among the 16 strong recommendations (#2, #3, #4, #6, #7, #11, #15, #16, #17, #20, #22, #24, #30, #31), all supported the utilisation of the biomarker.

All the characteristics regarding the recommendations are displayed in an EGM (figure 2, online supplemental Appendix 4), available also in the Open Science Framework repository, DOI: [10.17605/OSF.IO/N9RW6]).

DISCUSSION

Summary of findings

Our scoping review on the use of biomarkers to manage adult patients with sepsis captured a small number of GDs in the last decade. Despite the global impact of sepsis, the documents come mainly from Europe, produced by healthcare organisations, government agencies and scientific societies. The methodological standard for guideline development showed considerable variability, with an average score across the guidelines indicating moderate performance, pointing towards deficiencies in the rigour of development and stakeholder involvement.[34] Notably, only two GDs involved stakeholders adequately or provided tools for integrating recommendations into practice,[28 29] something that is crucial for ensuring implementation in real-world settings.

The recommendations covered CRP, IL-6, lactate, PCT and P-SEP, delineating the roles of these biomarkers in clinical practice. Most recommendations were related to lactate, recognising its role as a prognostic indicator combined with clinical measurements. Few recommendations focused on a combination of biomarkers.

Combining a panel of biomarkers with clinical data could be especially beneficial for diagnosing or stratifying patients' risks.[35]

The GDs mostly focused on prognostic recommendations. These often involved monitoring the treatment response, such as changes in vasopressor use or inflammation control, crucial for managing the progression of sepsis. However, there was a lack of consensus on the type of response monitored. Only a few recommendations were focused on diagnostic purposes, reflecting the uncertainty about this indication. The practical utility of biomarkers, as described in the GD, appeared to be limited with underpinning evidence unclear. For example, PCT and CRP were frequently recommended, although their roles in diagnosis or prognosis have not been fully established.

Strengths and weaknesses

A key strength of our work is its novel approach to evaluating and summarising GDs, facilitated by a comprehensive literature search from 2016 onwards to maintain a consistent sepsis definition over the last decade. We used the AGREE II instrument, the most suitable instrument for critically evaluating CPGs. Additionally, an EGM was created to identify areas needing further research and to help prioritise future efforts.

However, this study has some limitations. It may have missed relevant documents despite a targeted search for expert-developed guidance. GDs should have been updated regularly to reflect the latest scientific evidence, but they were not. The broad approach of the GD integrated patients from various settings, such as emergency departments and ICUs, and various groups, mixing adults with children in some cases. GDs may have potentially overlooked the unique needs of specific patient subgroups. The recommendations lacked a structured approach, e.g. they showed deficiencies in defining populations, interventions, comparators and outcomes. In the case of prognosis questions, GRADE does not provide a formal structure for evaluating the quality of evidence.[36] Moreover, the AGREE II tool's lack of a predefined threshold for quality assessment means that our evaluations are likely to be somewhat subjective. However, the 50% threshold has the advantage that it lies at the simple majority level. Furthermore, an important observation was the lack of reporting and transparency in the GD development process. For example, some documents did not report what system was used to formulate the recommendations, leading to potential quality and reliability gaps. Thus, we believe that our work merits consideration.

Implications for clinical practice

In 2020, 285 biomarkers were proposed for sepsis management, mainly assessed in fewer than five studies, indicating a lack of robust evidence supporting their use.[4] Currently, no biomarker has demonstrated sufficient utility to aid clinicians in their daily practice; however, it may serve as a complementary tool in clinical evaluation. Our review calls for more prudent use of biomarkers only in specific scenarios and in combination with standard clinical assessments to enhance the management of septic patients.

The role of biomarkers in the assessment of critically ill patients must be understood within a dynamic approach that combines clinical progression with objective decision-support tools. In the case of lactate, while its initial elevation reflects tissue hypoxia and metabolic dysfunction, its kinetics have been shown to be more relevant in guiding resuscitation and treatment response. Studies[37 38] have highlighted that progressive lactate reduction is associated with improved outcomes, supporting its inclusion in sepsis management strategies. However, in patients experiencing clinical improvement with haemodynamic stability, its usefulness is reduced, reinforcing the need for contextualised rather than systematic use.

The timing of biomarker sampling in sepsis is critical, given that the disease evolves heterogeneously and does not follow a stereotypical pattern in all patients. While the isolated measurement of a biomarker such as lactate or PCT is of limited value, serial monitoring can provide information on disease progression and response to treatment. Studies have shown that lactate reduction in the first 6 hours of resuscitation is associated with lower mortality, reinforcing the need to interpret it dynamically rather than as a single diagnostic value.[38] In this sense, the timing of sampling does not seek to establish a fixed point in the disease spectrum, but rather to allow for more precise pathophysiological monitoring over time.

Regarding the added value of biomarkers compared with established clinical parameters, their usefulness lies in their ability to improve risk stratification and guide therapeutic decisions beyond vital signs or obvious organ dysfunction. Thus, the value of a biomarker is not limited to confirming the need for obvious interventions such as vasopressors or ICU admission, but to providing additional information that refines decision-making in less clear clinical scenarios.

Implications for clinical research

In this work, the GDs obtained the lowest scores for AGREE-II in the 'Stakeholder Involvement' and 'Applicability' domains. Their 'rigour of development' was also poor. For CPGs, the AGREE reporting checklist[39] is the preferred instrument; for CS, the first reporting checklist, called ACCORD,[10] has recently been published. Furthermore, the implementation of a prognosis-structured question could aid in the formulation of the recommendations in future GDs. These elements, along with others, such as timing and setting, are critical for formulating robust clinical recommendations. With the addition of these methodological approaches for developing a GD, we could obtain more precise and reliable recommendations in the future.

Research into the use of biomarkers for sepsis should also prioritise specific patient subgroups as well as include diverse and meaningful outcomes. In prognosis, all-cause

mortality is the most studied outcome. Incorporating surrogate outcomes, such as vasopressor usage, ICU admissions, organ dysfunction and patient quality of life, could provide a more comprehensive evaluation of the biomarkers in clinical contexts. Initiatives like the COMET database[40 41] aim to standardise the outcomes that should be measured and reported in studies. This approach could contribute to establishing a universally accepted core set of outcomes for this pathology, enhancing comparability and generalisability in sepsis. Future research should also optimise biomarker sampling times and thresholds, ensuring their assessment within a standardised framework to facilitate their interpretation and application. Moreover, biomarker assessments should be integrated into panels reflecting the complex interplay of biological pathways involved in sepsis, including phenotypes associated with clinical variables, pathogens, metabolomics and proteomics. Another important aspect would be to conduct more research on specific biomarkers for sepsis, as lactate, for example, also plays a role in cardiogenic or hypovolemic shock.[42] Additionally, the documents must be updated regularly to reflect the latest scientific evidence related to clinical practices. These approaches will help to integrate biomarkers with other clinical data and standard measurements and to develop valid, practically applicable and robust recommendations.

In conclusion, this scoping review identified and evaluated the GDs and recommendations for using biomarkers to manage adults with sepsis. Only half of the GDs used a robust methodology. A high level of variability was observed among the recommendations, and only a few biomarkers were addressed, primarily for prognostic issues. The role of these biomarkers for sepsis needs to be evaluated better with improvements in the methodology to help obtain more valid and applicable recommendations, which, in turn, could contribute to the optimisation of patient care.

Author affiliations
[1]Clinical Biostatistics Unit, Hospital Universitario Ramón y Cajal, IRYCIS, Madrid, Spain
[2]Clinical Medicine and Public Health Programme, Universidad de Granada, Granada, Spain
[3]Preventive Medicine and Public Health Department, Hospital Universitario Ramón y Cajal, IRYCIS, Madrid, Spain
[4]Department of Critical Care, Fundación Universitaria de Ciencias de la Salud, Hospital de San José, Bogotá, Colombia
[5]Biomedical Research Institute Sant Pau (IIB Sant Pau), Iberoamerican Cochrane Centre, Barcelona, Spain
[6]CIBER Epidemiology and Public Health, CIBERESP, Madrid, Spain
[7]Intensive Medicine Service, Hospital Universitario del Henares, Madrid, Spain
[8]Critical Pathology Research Group, Faculty of Medicine, Universidad Francisco de Vitoria, Madrid, Spain
[9]Intensive Medicine Service, Hospital Universitario de Torrejón, Torrejón de Ardoz, Madrid, Spain
[10]Faculty of Medicine, Universidad Francisco de Vitoria, Madrid, Spain
[11]Cochrane Associate Centre of Madrid, Madrid, Spain

X María Cruz Martín-Delgado @MCMartinDelgado and Jesus Lopez-Alcalde @JLopezAlcalde

Acknowledgements We would like to thank Andrea Correa Perez for her contribution to the initial screening and Noelia Álvarez Díaz for her assistance with the updated searches. We would also like to thank María Nieves Plana Farras, María Dolores Corbacho Loarte, and Rosa Escudero for their help configuring the evidence gap map. We acknowledge Professor Khalid Saaed Khan for his thorough review and rewording of the manuscript.

Contributors JL-A, JZ and MM-H conceived the study. JL-A, JZ and MM-H developed the research protocol. IS developed the search strategy, executed by MM-H. AG-S-V and MM-H screened studies by title and abstract and full text. They also performed the data extraction and quality assessment. JL-A and JZ verified the study data. JL-A and MM-H configured the evidence gap map and revised by DM-F, FG-V, JZ and MCM-D. MM-H wrote the draft of the manuscript, and DM-F, FG-V, IS, JL-A, JZ and MCM-D reviewed subsequent drafts. All authors had full access to all the study data throughout the review process. All authors contributed to the final edits and agreed to submit the final manuscript. MM-H is the guarantor. JL-A and JZ directed jointly this study.

Funding This project is supported by Instituto de Salud Carlos III, Spain and European Union ('Fondo Europeo de Desarrollo Regional, Una manera de hacer Europa'), grant number (PI19/00481)

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

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

ORCID iDs
Miriam Mateos-Haro http://orcid.org/0000-0003-2967-0158
Ana Garcia-Santa-Vinuela http://orcid.org/0000-0003-4118-1170
Ivan Solà http://orcid.org/0000-0003-0078-3706
María Cruz Martín-Delgado http://orcid.org/0000-0002-7468-4594
Javier Zamora http://orcid.org/0000-0003-4901-588X

REFERENCES

1 Willis JCD, Lord GM. Immune biomarkers: the promises and pitfalls of personalized medicine. *Nat Rev Immunol* 2015;15:323–9.
2 Wishart DS, Bartok B, Oler E, *et al*. MarkerDB: an online database of molecular biomarkers. *Nucleic Acids Res* 2021;49:D1259–67.
3 Horvath AR. Are Guidelines Guiding us on How to Utilize Laboratory Tests? *EJIFCC* 2015;26:146–57.
4 Pierrakos C, Velissaris D, Bisdorff M, *et al*. Biomarkers of sepsis: time for a reappraisal. *Crit Care* 2020;24:287.
5 Evans L, Rhodes A, Alhazzani W, *et al*. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med* 2021;47:1181–247.
6 Singer M, Deutschman CS, Seymour CW, *et al*. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016;315:801–10.
7 Zhou P, Chen L, Wu Z, *et al*. The barriers and facilitators for the implementation of clinical practice guidelines in healthcare: an umbrella review of qualitative and quantitative literature. *J Clin Epidemiol* 2023;162:169–81.

8 Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice. Clinical Practice Guidelines We Can Trust [Internet]. Washington (DC): National Academies Press (US), 2011. Available: https://www.ncbi.nlm.nih.gov/books/NBK209539/
9 Li T, Vedula SS, Scherer R, *et al*. What comparative effectiveness research is needed? A framework for using guidelines and systematic reviews to identify evidence gaps and research priorities. *Ann Intern Med* 2012;156:367–77.
10 Gattrell WT, Logullo P, van Zuuren EJ, *et al*. ACCORD (ACcurate COnsensus Reporting Document): A reporting guideline for consensus methods in biomedicine developed via a modified Delphi. *PLoS Med* 2024;21:e1004326.
11 Atkins D, Best D, Briss PA, *et al*. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490.
12 Hajizadeh A, Lotfi T, Falzon D, *et al*. Recommendation mapping of the World Health Organization's guidelines on tuberculosis: A new approach to digitizing and presenting recommendations. *J Clin Epidemiol* 2021;134:138–49.
13 Unicef. Evidence Gap Maps. Available: https://www.unicef-irc.org/evidence-gap-maps [Accessed 2 Apr 2025].
14 Mateos-Haro M, Lopez-Alcalde J, Garcia-Santa-Vinuela A, *et al*. Recommendations for the use of biomarkers for the management of adults with sepsis: a scoping review. 2023. Available: https://doi.org/10.17605/OSF.IO/N9RW6
15 Tricco AC, Lillie E, Zarin W, *et al*. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med* 2018;169:467–73.
16 Endnote. Philadelphia, PA: Clarivate Analytics, 2013.
17 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71n71.
18 Ouzzani M, Hammady H, Fedorowicz Z, *et al*. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016;5:210.
19 EPPI-Reviewer Web [internet]. Available: https://eppi.ioe.ac.uk/eppireviewer-web/home [Accessed 2 Apr 2025].
20 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.
21 AGREE-ii. n.d. Available: https://www.agreetrust.org/wp-content/uploads/2013/06/AGREE_II_Spanish.pdf
22 Brouwers MC, Kho ME, Browman GP, *et al*. AGREE II: advancing guideline development, reporting and evaluation in health care. *Can Med Assoc J* 2010;182:E839–42.
23 Brouwers MC, Spithoff K, Lavis J, *et al*. What to do with all the AGREEs? The AGREE portfolio of tools to support the guideline enterprise. *J Clin Epidemiol* 2020;125:191–7.
24 EPPI-Mapper Wizard. n.d. Available: https://eppimapper.digitalsolutionfoundry.co.za/
25 Kochanek M, Schalk E, von Bergwelt-Baildon M, *et al*. Management of sepsis in neutropenic cancer patients: 2018 guidelines from the Infectious Diseases Working Party (AGIHO) and Intensive Care Working Party (iCHOP) of the German Society of Hematology and Medical Oncology (DGHO). *Ann Hematol* 2019;98:1051–69.
26 Instituto de Evaluación de Tecnologías en Salud e Investigación. Guía de práctica clínica para el reconocimiento y manejo inicial de sepsis en adultos: guía en versión extensa. EsSalud Lima; 2018. Available: https://gpc-peru.com/gpcsepsis
27 Shime N, Nakada T-A, Yatabe T, *et al*. The Japanese Clinical Practice Guidelines for Management of Sepsis and Septic Shock 2024. *Acute Med Surg* 2025;12:e70037.
28 National institute for health and care excellence. Suspected sepsis: recognition, diagnosis and early management. 2024. Available: https://www.nice.org.uk/guidance/ng51
29 Department of Health. NCEC National Clinical Guideline No. 26 2021. 2021. Available: http://health.gov.ie/national-patient-safety-office/ncec/
30 Bowyer L, Cutts BA, Barrett HL, *et al*. SOMANZ position statement for the investigation and management of sepsis in pregnancy 2023. *Aust N Z J Obstet Gynaecol* 2025;65:37–46.
31 Sherwin R, Ehrman RA. Is lactate measurement in the emergency department valuable as a predictor of poor outcomes in adult patients with sepsis. 2018. Available: https://apps.aaem.org/UserFiles/file/112818BODaprvdwchngLactateSepsisfrPosting.pdf
32 Park C, Ku NS, Park DW, *et al*. Early management of adult sepsis and septic shock: Korean clinical practice guidelines. *Acute Crit Care* 2024;39:445–72.
33 Lissauer D, Morgan M, Banerjee A, *et al*. Royal college of obstetrics and gynaecology. In: *Identification and Management of Maternal Sepsis during and following Pregnancy: Green-top Guideline no.64. BJOG*. 132. 2025: e61–85.
34 Silberberg B, Aston S, Boztepe S, *et al*. Recommendations for fluid management of adults with sepsis in sub-Saharan Africa: a systematic review of guidelines. *Crit Care* 2020;24:286.
35 Mearelli F, Fiotti N, Giansante C, *et al*. Derivation and Validation of a Biomarker-Based Clinical Algorithm to Rule Out Sepsis From Noninfectious Systemic Inflammatory Response Syndrome at Emergency Department Admission: A Multicenter Prospective Study. *Crit Care Med* 2018;46:1421–9.
36 Guyatt GH, Oxman AD, Kunz R, *et al*. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011;64:395–400.
37 Hernández G, Ospina-Tascón GA, Damiani LP, *et al*. Effect of a Resuscitation Strategy Targeting Peripheral Perfusion Status vs Serum Lactate Levels on 28-Day Mortality Among Patients With Septic Shock: The ANDROMEDA-SHOCK Randomized Clinical Trial. *JAMA* 2019;321:654–64.
38 Puskarich MA, Trzeciak S, Shapiro NI, *et al*. Outcomes associated with lactate monitoring during early resuscitation of sepsis. *Crit Care Med* 2012;40:2867–72.
39 Brouwers MC. The AGREE Reporting Checklist: a tool to improve reporting of clinical practice guidelines. *BMJ* 2016;354:i4852. Available: https://www.comet-initiative.org/Studies/Details/2764
40 Development of a core outcome set for adult sepsis. n.d. Available: https://www.comet-initiative.org/Studies/Details/2764
41 Taneri PE, Kirkham JJ, Molloy EJ, *et al*. Protocol for the development of a core outcome set for neonatal sepsis (NESCOS). *PLoS One* 2023;18:e0295325.
42 Sauer CM, Gómez J, Botella MR, *et al*. Understanding critically ill sepsis patients with normal serum lactate levels: results from U.S. and European ICU cohorts. *Sci Rep* 2021;11:20076.