



Review article

Clinical prognostic factors for older people: A systematic review and meta-analysis

Nicola Veronese^{a,1,*}, Anna Fazzari^{a,1}, Maria Armata^a, Angela Parisi^a, Alessandra Parrinello^a,
Valentina Petralia^a, Carlo Saccaro^a, Miriam Vesco^a, Federica Tagliaferri^a,
Veronica Andrea Fittipaldo^b, Jacopo Demurtas^c, Lee Smith^d, Ligia J. Dominguez^{a,e},
Alberto Pilotto^{f,g}, Mario Barbagallo^a

^a Department of Internal Medicine, Geriatrics Section, University of Palermo, Palermo, Italy

^b IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy

^c Clinical and Experimental Medicine PhD Program, Università di Modena e Reggio Emilia, Modena - Azienda USL Sud Est Toscana, Grosseto, Italy

^d Centre for Health Performance and Wellbeing, Anglia Ruskin University, Cambridge, UK

^e School of Medicine, University Kore, Enna, Italy

^f Department of Interdisciplinary Medicine, University of Bari "Aldo Moro", Bari, Italy

^g Geriatrics Unit, Department Geriatric Care, Orthogeriatrics and Rehabilitation, E.O. Galliera Hospital, Genoa, Italy, University of Bari "Aldo Moro", Bari, Italy

ARTICLE INFO

Keywords:

Prognosis

Accuracy

Precision

Systematic review

Meta-analysis

Older people

ABSTRACT

Objective: To explore the accuracy and precision of prognostic tools used in older people in predicting mortality, hospitalization, and nursing home admission across different settings and timings.

Design: Systematic review and meta-analysis of prospective and retrospective studies.

Data sources: A systematic search from database inception until 01st February 2023 was run in Medline, Embase, Cinhal, Cochrane Library.

Eligibility criteria: Studies were eligible if they reported accuracy (area under the curve [AUC]) and/or precision (C-index) for the prognostic index in relation to any of the following outcomes: mortality, hospitalization, and nursing home admission.

Data extraction and synthesis: Two independent reviewers extracted data. Data were pooled using a random effects model. The risk of bias was assessed with the Quality in Prognosis Studies (QUIPS) tool. If more than three studies for the same setting and time were available, a meta-analysis was performed and evaluated using the GRADE tool; other data were reported descriptively.

Results: Among 16,082 studies initially considered, 159 studies with a total of 2398856 older people (mean age: 78 years) were included. The majority of the studies was carried out in hospital or medical wards. In the community setting, only two tools (Health Assessment Tool and the Multidimensional Prognostic Index, MPI) had good precision for long-term mortality. In emergency department setting, Barthel Index had an excellent accuracy in predicting short-term mortality. In medical wards, the MPI had a moderate certainty of the evidence in predicting short-term mortality (13 studies; 11,787 patients; AUC=0.79 and 4 studies; 3915 patients; C-index=0.82). Similar findings were available for MPI when considering longer follow-up periods. When considering nursing home and surgical wards, the literature was limited. The risk of bias was generally acceptable; observed bias was mainly owing to attrition and confounding.

Conclusions: Several tools are used to predict poor prognosis in geriatric patients, but only those derived from a multidimensional evaluation have the characteristics of precision and accuracy.

* Correspondence to: Geriatric Unit, Department of Internal Medicine and Geriatrics, University of Palermo, Via del Vespro, 141, Palermo 90127, Italy
E-mail address: nicola.veronese@unipa.it (N. Veronese).

¹ These authors equally contributed to this work

<https://doi.org/10.1016/j.arr.2024.102345>

Received 12 March 2024; Received in revised form 6 May 2024; Accepted 18 May 2024

Available online 21 May 2024

1568-1637/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Prognosis is an important determinant of clinical decision-making. To include prognosis in clinical decision-making may improve the utilization of health care systems: it is, for example, known that hospice and palliative care are underused in non-malignant life-threatening conditions (Kangtanyagan and Vatcharavongvan, 2023) and, similarly, older adults having advanced dementia screened for slow-growing cancer may receive invasive workups and treatments without a real benefit in terms of survival or quality of life. (Ashley et al., 2022) Likewise, substantially healthy older adults, do not receive appropriate screening for cancer that, on the contrary, could be beneficial (Chapman et al., 2023).

Based on these considerations, guidelines are starting to incorporate life expectancy and prognosis as central factors in weighing the benefits and the burdens of tests and treatments. (Gill, 2012) Prognostic indices may offer a new perspective to move away from age-based cutoffs that do not consider the impact of other factors potentially associated with a poor prognosis such as multimorbidity, social isolation, functional and cognitive decline. (Gill, 2012) Therefore, the ability to predict negative outcomes in older people (e.g., mortality) could be important for management, treatment and prevention (Pilotto et al., 2015).

The goal of estimating prognosis is to improve clinical decision making and, ultimately, patient outcomes. (Gill, 2012) Despite the proliferation of various prognostic instruments, currently, there is poor evidence that their routine use improves patient outcomes. The limited use of these tools in clinical practice is mainly based on the fact that several tools have a limited accuracy (i.e., quantifying the classifier's ability to distinguish between the positive and negative classes across different threshold values) and precision (i.e., to identify the discriminatory power of a predictive model for the time-to-event outcome). (Rector et al., 2012) Previous works, based on systematic approaches of the literature, have found that most tools designed to predict mortality have only modest accuracy, and there is large variability across various diseases and populations. (Siontis et al., 2011; Yourman et al., 2012) Finally, these seminal works were published more than 10 years ago. However, newer tools that are now available are not included in these previously published works and their inclusion may yield differing findings.

Based on these considerations, with this systematic review and meta-analysis, we aim to explore the accuracy and precision of prognostic tools used in geriatric medicine in predicting mortality, hospitalization, and nursing home admission across different settings and timings.

2. Materials and methods

This systematic review and meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and to the indications specific for systematic reviews about prognosis. (Debray et al., 2017; Page et al., 2021) The protocol was a priori registered in <https://osf.io/3wq7u/>.

2.1. PICOTS question and inclusion and exclusion criteria

The PICOTS question was defined as follows:

P (participants): older people, defined as having at least a mean age of 60 years in the studies included;

I (index): all the clinical factors that were explored as prognostic, both combined or single;

C (comparator): none

O (outcomes): hospitalization (re-hospitalization in people already hospitalized), mortality, and nursing home admission. Data on these outcomes were required to be reported in terms of accuracy (area under the curve, AUC) or precision (C-index, Brier Index, pseudoR2) with their 95% confidence intervals (CIs). The AUC is the most commonly used metric for assessing the accuracy of predictive tools and can be

compared to relative risks or similar metrics, across different tools (Siontis et al., 2011), whilst C-index is the most commonly used metric for precision.

T (timing): we included all the timings proposed, but they were divided in less than one month (short-period), between 1 and 6 months, between 6 and 12 months, and more than 12 months;

S (setting): we included all the studies, independently from the setting, but they were divided into community, hospital (sub-divided into medical and surgical wards), emergency department, and nursing home admission.

We excluded studies: (i) with data that could not be meta-analyzable (e.g., no 95% CI were reported); (ii) with a mean age less than 60 years or not reported; (iii) with prognostic factors in which only radiological or bio-humoral factors were included; (iv) written in any language other than English or in form of conference abstracts; (v) and including other outcomes of interest, such as intensive care unit admission.

2.2. Search strategy

From the database inception until 01 February 2023, Medline, Embase, Cinhal, Cochrane Library were searched independently by eight investigators (AF, MA, AP, AP, VP, CS, MV, FT), in couple. The detailed search strategy, for each database, is shown in [Supplementary Table 1](#).

2.3. Data extraction

The data extraction followed the indications of the CHARMS (Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies) checklist. (Moons et al., 2014) The eight investigators (AF, MA, AP, AP, VP, CS, MV, FT) extracted data independently which included author of study, year of study, type of study (retrospective or prospective), total sample size, mean age with the standard deviation (SD), percentage of females, setting, outcomes, and prognostic factors investigated. Disagreements between authors were resolved through discussion with two senior authors (NV, LJD).

2.4. Risk of bias

The eight investigators (AF, MA, AP, AP, VP, CS, MV, FT) carried out evaluation of the risk of bias using the Quality In Prognosis Studies (QUIPS) tool. (Hayden et al., 2013) This tool covers six important areas to consider when evaluating validity and bias in studies of prognostic factors, i.e., participation, attrition, prognostic factor measurement, confounding measurement and account, outcome measurement, and analysis and reporting. (Hayden et al., 2013) The data were consequently reported by single study and as a total using the robvis tool. (McGuinness and Higgins, 2020)

2.5. Statistical analysis and synthesis of the data

We planned to run a meta-analysis if the same prognostic factor was used in the same setting with the same follow-up category, having at least three studies for an analysis. (Debray et al., 2014) Findings with less than three studies were reported descriptively. The analyses are proposed by estimates of accuracy (AUC) and for precision (C-index), in agreement with the NICE guidelines. (Farmer et al., 2016) While no definitive thresholds exist, values of AUC/C-index of 0.50 indicate accuracy or precision no better than chance, between 0.50 and 0.60 very poor accuracy/precision, between 0.60 and 0.70 poor, between 0.70 and 0.80 good, between 0.80 and 0.90 very good, and more than 0.90 an excellent accuracy/discrimination. (Hanley and McNeil, 1982; Siontis et al., 2011)

We used the GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) assessment, adapted for prognostic studies. (Iorio et al., 2015) Briefly, we considered the risk of bias according to the QUIPS; statistical heterogeneity (inconsistency) between

different studies was assessed using the I^2 , with a low heterogeneity that was based on I^2 from 0% to 49%, moderately heterogenous from 50% to 74%, and highly heterogeneous from 75% and above (Higgins et al., 2003); indirectness was based on the assumption that the analyses should reflect the PICOTS question and in particular if only a specific population was included (e.g., patients with pneumonia for all medical wards); imprecision: the evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, AUC or C-index 0.5–0.8 and 0.8–1) and by 2 increments if the individual studies varied across 3 areas (for example, AUC or C-index 0–0.5, 0.5–0.8 and 0.8–1); and publication bias was assessed using the Egger bias test. (Egger et al., 1997) The GRADE gives four different degrees of certainty of evidence from very low (we have very little confidence in the estimate: the true prognosis [probability of future events] is likely to be substantially different from the estimate) to high (we are very confident that the true prognosis [probability of future events] lies close to that of the estimate) (Iorio et al., 2015).

All analyses were performed with MedCalc, version 22.04 for Windows. A p value of <0.05 was considered statistically significant.

3. Results

3.1. Literature search

As shown in Fig. 1, we initially considered 16,082 title/abstracts. Of

them, 510 full-texts were examined: the works were mainly not included since data were not meta-analysable (e.g., no data about accuracy or precision were reported or without estimates of variance) or since the works included patients with a mean age less than 60 years. The full list of the references excluded is reported in Supplementary Table 2. Finally, we included 159 articles in this systematic review (see the list in Supplementary Table 3).

3.2. Descriptive findings

The most important descriptive characteristics of the 159 studies included are summarised in Supplementary Table 4. Overall, the studies included a total of 2398856 older participants with a mean age of 78 years who were mainly female (51%). The majority of the studies were carried out among older people hospitalized in medical wards (n=108), 22 studies were conducted in the community, 14 in surgical hospital wards, 13 in emergency departments, and only two in nursing homes. The majority of the studies included overall mortality as an outcome (n=147), whilst more limited evidence was available for nursing home admission (n=1) or hospital admission/re-admission (n=5), whilst the other six studies reported mixed outcomes. Finally, as fully detailed in Supplementary Table 4, several prognostic factors were explored from clinical scores (e.g., Charlson Comorbidity Index) to physical performance parameters (e.g., gait speed) to multidimensional scores (e.g., multidimensional prognostic index, MPI).

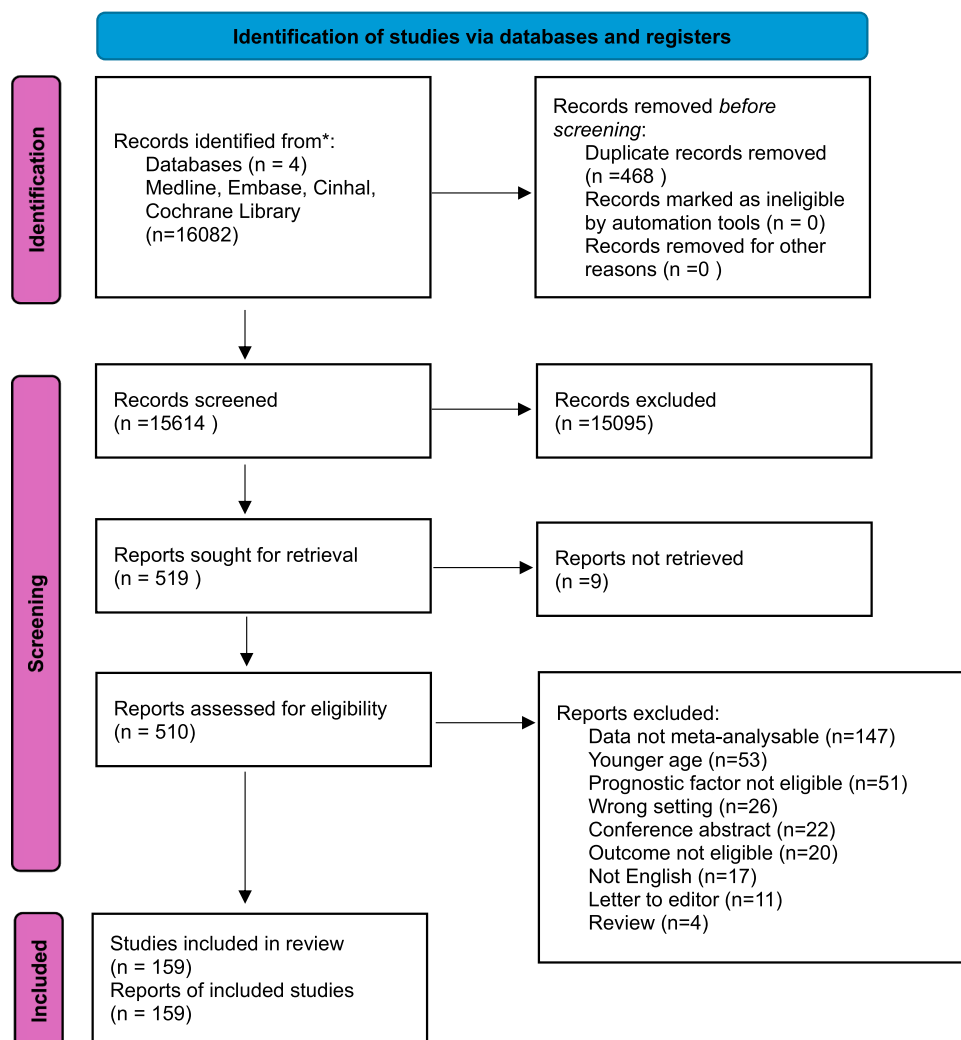


Fig. 1. PRISMA flow-chart.

3.3. Main findings

The main findings of our systematic review are reported in [Tables 1 and 2](#) and in [Supplementary Tables 5–6–7–8–9](#), divided by setting.

3.4. Community

In the community setting, as detailed in [Supplementary Table 5](#), we found data from 86 different cohorts in 22 independent studies. Overall, in this setting, considering hospitalization as an outcome, only one study including 16,280 participants reported that the combination of age and gender had a very good accuracy in predicting hospitalization with a follow-up less than one month (AUC=0.83; 95%CI: 0.81–0.85). For the studies considering hospitalization and a longer follow-up, no one reported a very good accuracy having an AUC less than 0.80 ([Supplementary Table 5](#)). Similarly, in one study including 7033 participants, having a follow-up of 60 months and comparing Frailty Index, Lee Index, and Schonberg Index, the overall precision was poor (C-index less than 0.70) ([Supplementary Table 5](#)).

Considering mortality as the outcome, it is noteworthy that the tools used in the different studies were of poor accuracy or precision (AUC or C-index less than 0.70) in follow-up between 1 and 6 months and between 6 and 12 months. When assessing the studies having a follow-up period more than 12 months, the best accuracy was reached by a Health Assessment Tool (AUC=0.87; 95%CI: 0.85–0.88) in an Italian study including 3363 participants over three years of follow-up, whilst the highest precision was reached by the MPI in the InChianti study of 1453 older participants over 15 years of follow-up (C-index= 0.821; 95%CI: 0.806–0.835) ([Supplementary Table 5](#)).

3.5. Emergency department (ED)

[Supplementary Table 6](#) reports the main findings of the studies made in ED settings. It is noteworthy that no study reported data about precision, such as C-index.

When taking hospitalization as the outcome, no studies reported data about precision. As shown in [Supplementary Table 6](#), the accuracy of the tools examined was generally modest in predicting hospitalization across different follow-up cut-offs.

The most consistent literature was about mortality. In studies having a follow-up time less than one month, the most accurate tool was the Barthel Index that had an excellent accuracy (AUC=0.97; 95%CI: 0.95–0.99) as shown in a Korean study of 488 older patients. Similarly, in studies with a follow-up time between 1 and 6 months, the Urgent Surgical Elderly Mortality risk score had the highest accuracy (AUC= 0.83; 95%CI: 0.78–0.87) in a cohort of 500 older patients over two months of follow-up, whilst the Trauma and Injury Severity Score had an excellent accuracy among studies having a follow-up between 6 and 12 months (AUC= 0.94; 95%CI: 0.90–0.98).

Finally, in one study including 889 older patients attending the ED and with a follow-up of 3 months, the acutely presenting older patient screening program 1 had a higher accuracy in predicting nursing home admission than the International Resident Assessment Instrument Emergency Department ([Supplementary Table 6](#)).

3.6. Hospital, medical wards

Hospital medical wards included the largest number of the studies available, as shown in [Supplementary Table 7](#). Taking re-hospitalization as the outcome, only a few factors had a good accuracy, since almost all of the prognostic factors included had an AUC less than 0.70, across different follow-up times.

For this setting, we were able to do a meta-analysis since at least three studies were available for the outcome mortality having a similar follow-up period. [Table 1](#) shows these data. In studies having a follow-up period less than one month and therefore considering also in-hospital

Table 1
Main findings of the meta-analysis in hospital setting using mortality as outcome: data about accuracy.

Factor	Number of cohorts	Sample size	Follow-up (months)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	AUC	95% low	95% high	p-value	Overall quality
CURB-65	3	899	<1	not serious	Serious (I ² =65%)	serious	not serious	no	0.75	0.70	0.79	<0.0001	Low
MPI	13	11787	<1	not serious	Serious (I ² =70%)	not serious	not serious	no	0.79	0.76	0.81	<0.0001	Moderate
PSI	3	773	<1	not serious	not serious (I ² =0%)	serious	not serious	no	0.79	0.78	0.80	<0.0001	Moderate
qSOFA	3	1452	<1	not serious	very serious (I ² =83%)	serious	not serious	no	0.74	0.69	0.78	<0.0001	Very low
SIRS	3	1452	<1	not serious	Serious (I ² =63%)	serious	not serious	no	0.76	0.73	0.78	<0.0001	Low
MPI	12	9751	6	not serious	Serious (I ² =70%)	not serious	not serious	no	0.74	0.71	0.76	<0.0001	Moderate
MPI	14	13997	>12	not serious	Serious (I ² =71%)	not serious	not serious	no	0.72	0.69	0.76	0.02	Moderate
qSOFA	3	1518	>12	not serious	Serious (I ² =69%)	serious	not serious	no	0.65	0.60	0.70	0.03	Low

Abbreviations: AUC: area under the curve; CURB-65: confusion, respiratory rate, BP, age ≥ 65 years; MPI: multidimensional prognostic index; PSI: pneumonia severity index; qSOFA: quick sequential organ failure assessment; SIRS: systemic inflammatory response syndrome

Table 2
Main findings of the meta-analysis in hospital setting using mortality as outcome: data about precision.

Factor	Number of cohorts	Sample size	Follow-up (months)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	C	95% low	95% high	p-value	Overall quality
MPI	4	3915	<1	not serious	not serious ($I^2=3\%$)	not serious	not serious	no	0.82	0.78	0.85	<0.0001	Moderate
Age	3	15834	>12	Not serious	Not serious ($I^2=28\%$)	Not serious	Not serious	yes	0.88	0.84	0.91	<0.0001	Moderate

Abbreviations: MPI: multidimensional prognostic index

mortality, only two tools had a moderate certainty of evidence according to the GRADE evaluation, i.e., the MPI and PSI (Pneumonia Severity Index). The MPI analysis included 11,787 older hospitalized patients across 13 different studies and had a good accuracy in determining short-period mortality (AUC=0.79; 95%CI: 0.76–0.81); the PSI included three studies for less than 1000 older hospitalized patients, having an AUC of 0.80 (95%CI: 0.79–0.81). On the contrary, SIRS, CURB-65 and qSOFA were graded, for the studies with a follow-up less than one month, as low and very low certainty of evidence, respectively. As reported in [Table 2](#), the MPI had a very good precision in predicting mortality in the short-period, since in four different studies including 3915 patients the C-index was 0.82 (95%CI: 0.78–0.85). Considering other follow-up periods, MPI had a good accuracy in predicting mortality in 12 studies having a follow-up of six months (AUC=0.74; 95%CI: 0.71–0.76) and in 14 studies, including 13,997 patients, in one year mortality (AUC=0.72; 95%CI: 0.69–0.76). The evidence for six and twelve months was graded as moderate, according to the GRADE ([Table 1](#)).

Evidence in relation to the other tools investigating mortality as the outcome is fully detailed in [Supplementary Table 7](#).

Finally, a few studies explored nursing home admission as the outcome. Among the four cohorts included the best in accuracy was the Geriatric Trauma Outcome Score II (AUC=0.82; 95%CI: 0.75–0.88) in one study including 144 patients followed up for three years ([Supplementary Table 7](#)).

3.7. Hospital, surgical wards

[Supplementary Table 8](#) reports the data for the tools used for predicting mortality in surgical wards, since no study reported data about hospitalization or nursing home admission.

In studies having a follow-up time less than one month, the Acute Physiology and Chronic Health Evaluation IV score had a good accuracy in predicting mortality in 443 older patients attending surgery. In studies having a follow-up time between 6 and 12 months, a modified-Krishnan’s frailty index had the highest accuracy in predicting mortality in surgical wards (AUC= 0.856; 95%CI: 0.767–0.945). Finally, among studies with a follow-up period more than 12 months, a frailty index among 239 older patients followed-up for 43.5 months had an excellent accuracy (AUC=0.90; 95%CI: 0.85–0.95) ([Supplementary Table 8](#)). On the contrary, a combination of the American Society of Anaesthesiologist (ASA) score and the Tumor Nodes Metastasis (TNM) cancer stage reached the highest precision over 24 months of follow-up in 405 older patients (C-index= 0.80; 95%CI: 0.78–0.87).

3.8. Nursing home

[Supplementary Table 9](#) shows that only two studies were identified for older residents in nursing homes. In one Italian study including 653 nursing home residents, the MPI had a good accuracy in determining mortality over 12 months of follow-up, whilst in another study of 710 residents the combination of body mass index, Mini-Mental State Examination and activities of daily living had a good precision over five years of follow-up. No studies were available about hospitalization.

3.9. Risk of bias evaluation

[Fig. 2](#) shows the risk of bias evaluated using the QUIPS tool as summary, whilst [Supplementary Table 10](#) shows the assessment of the risk of bias, study by study. Overall, the bias due to participation was rated as high in 5/159 studies (=3.1%), whilst a consistent part of the studies included could have a high risk of bias due to attrition (69/159; 43.4%). A high risk of bias could be present for confounding since 61/159 studies reported this issue (=38.4%).

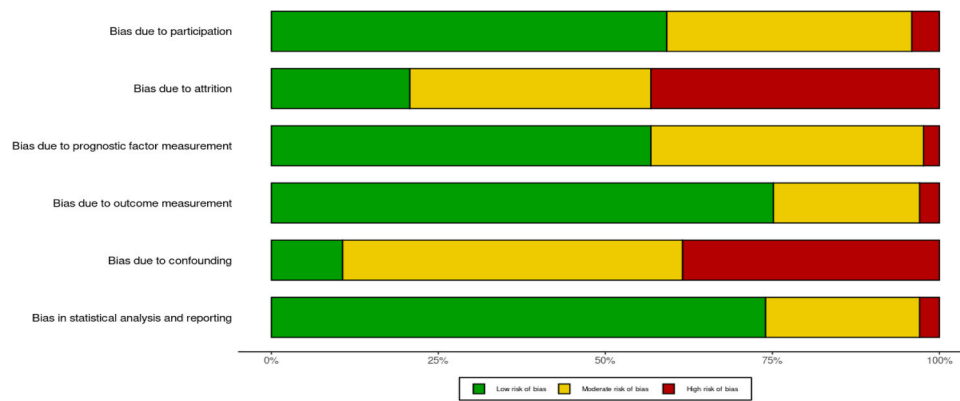


Fig. 2. Evaluation of the potential risk of bias with the QUIPS tool.

4. Discussion

In this systematic review including 159 studies for approximately 2.5 million older adults we explored the role and the use of common tools to predict hard outcomes, such as mortality, hospitalization and nursing home admission. Briefly, we found that several prognostic tools are used, from mono to multidimensional ones, but only a few reached a sufficient accuracy and precision making them ideal to use in daily clinical practice. In particular, only in hospital medical wards, we were able to run a meta-analysis showing that among the instruments explored, the MPI seems to have the best values in terms of accuracy and precision. For other settings, only sparse data are available; however, in community and in nursing home settings MPI seems to be a reliable tool.

An ideal predictive tool should have some important prerequisites. (Siontis et al., 2011) First, the tool must be validated in populations other than the one in which it was initially developed: in this sense, the tool could be reproducible. Second, an important methodological characteristic is that an ideal tool should have good accuracy and precision. Moreover, an ideal predictive tool can make accurate predictions in different settings and in different medical situations. Finally, to include in daily clinical practice and in clinical decision making, an ideal prognostic tool should be short and should include scales or evaluations commonly used in medicine. (Siontis et al., 2011) Unfortunately, following the findings of our work, hundreds of tools commonly exist to predict poor prognosis among older people. At the same time, our systematic review suggests that very few tools reached these characteristics and often, even if they have a good accuracy and/or precision, the findings are limited to less than three studies making an evaluation using a meta-analytic approach not possible.

When analysing community settings, for example, we found that the prognostic tools available had a good accuracy and precision only when considering long follow-up periods. In particular, the MPI, derived from an adaptation from the InChianti study and the Health Assessment Tool could be useful in this setting, even if the findings are limited to only one study for each. The role of prognostic tools in this setting is noteworthy. For example, several screenings proposed to detect cancer in older people are made among community-dwellers and, as observed by Schonberg et al. (Schonberg and Smith, 2016), the inclusion of validated prognostic tools might better rationalize the possible life-prolonging benefits of cancer screening. Similarly, the use of some medications that could be used in primary prevention of cardiovascular diseases, such as statins, may have favourable benefits from the use of accurate and precise prognostic tools.

Surprisingly, in the ED setting, the tools commonly used did not report sufficient data about the risk of hospitalization, and when they did report it, the accuracy was only modest. Of importance, in one study, the Barthel Index, one of the most widely used tools to detect disability among older people had an excellent accuracy in short-period mortality

prediction, whilst in longer follow-up periods only specific scales (such as those used in trauma and surgery) were reported as having a good accuracy. Taken together, these findings suggest that more research is needed in this setting, which is often the first access for older patients to hospital and its services. For example, in one systematic review of the multidimensional tools used in the ED, the authors found that only screening tools are used, but they did not report any information about accuracy and precision about the outcomes of interest of the present work. (Graf et al., 2011)

In the hospital setting and more precisely in medical wards, we found the greatest quantity of literature. In particular, the MPI seems to be the most accurate and precise tool in predicting mortality in older people hospitalized for medical reasons based on a meta-analysis of 13 studies with almost 12,000 older people and on four studies with 3915 participants. Briefly, several multicenter studies demonstrated that the MPI, a comprehensive geriatric assessment (CGA) derived tool, had excellent accuracy and calibration in predicting the clinical outcomes typical of older people, such as hospitalization, institutionalization, need for homecare services and mortality. (Cruz-Jentoft et al., 2020; Pilotto et al., 2008) Nowadays, the MPI has been validated in over 54,000 older adults suffering from the most common chronic and acute age-related diseases associated with high mortality in over 50 international studies. (Pilotto et al., 2020) Our systematic review has only highlighted its reliability as prognostic tool in older people hospitalized in medical wards and the need of CGA in clinical decision making. (Veronese et al., 2022)

Finally, in surgical wards and in nursing home settings, we found only limited research about possible prognostic tools even if, in our opinion, these settings are of importance for physicians interested in geriatric medicine. It is noteworthy that, again, the MPI had a good accuracy in predicting mortality among nursing home residents in one year of follow-up. Future research is, however, urgently needed to better understand the role of prognostic tools in these relevant settings.

Even though this systematic review overcomes several limitations of some classical papers published approximately 10 years ago on the role of prognosis among older people (Gill, 2012; Siontis et al., 2011), several limitations of the present work must still be acknowledged. First, very few studies assessed the prognostic role of tools using hospitalization or nursing home admission as outcomes: even if these outcomes are of importance in better using healthcare resources, they are still under-represented in scientific research. Second, we considered only predictive studies that assessed accuracy or precision, also based on the NICE guidelines that clearly indicated the necessity of these estimates for prognostic tools used in people affected by multimorbidity. (Farmer et al., 2016) Finally, as observed in the risk of bias assessment, some studies are likely of a poor quality particularly due to high risk of bias due to attrition or since the role of confounding factors in analyses was only partially considered.

In conclusion, in this systematic review and meta-analysis carried out across different settings and including approximately 2.5 million older adults, we found that several tools are used to predict poor prognosis in geriatric patients, but only those derived from a multidimensional evaluation have the characteristics of precision and accuracy.

Ethics

Being a systematic review with meta-analysis of previously published literature, no informed consent was obtained and we did not ask for the approval to our local Ethical Committee.

Funding with award/grant number

None

Systematic review registration

Open Science Framework, <https://osf.io/3wq7u/>.

Authors' contribution

Nicola Veronese, Ligia J. Dominguez, Mario Barbagallo designed the protocol. Andrea Fittipaldo prepared the search strategy across different datasets. Maria Armata, Angela Parisi, Alessandra Parinello, Valentina Petralia, Carlo Saccaro, Miriam Vesco, Federica Tagliaferri screened the literature, revised, contributed to, and approved the protocol. Maria Armata, Angela Parisi, Alessandra Parinello, Valentina Petralia, Carlo Saccaro, Miriam Vesco, Federica Tagliaferri extracted the data. Nicola Veronese, Anna Fazzari, Jacopo Demurtas, Alberto Pilotto drafted the first version of the manuscript. Lee Smith, Ligia J. Dominguez, Mario Barbagallo critically revised the paper. All authors approved the final version of the manuscript.

Declaration of Competing Interest

AP declared to have been involved in the development and validation of the Multidimensional Prognostic Index (MPI), a tool analyzed in this review. The other authors declare no conflict of interest for this work.

Data availability statement

The data underlying this article will be shared on reasonable request to the corresponding author. All authors declare manuscript is an honest, accurate, and transparent account of the study being reported; no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned (and, if relevant, registered) have been explained in published protocol.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.arr.2024.102345](https://doi.org/10.1016/j.arr.2024.102345).

References

- Ashley, L., Spencer, K., Neal, R., 2022. Cancer detection, diagnosis, treatment, and care for people with dementia: an overview of current knowledge, research, and practice recommendations. CA: A Cancer J. Clin.
- Chapman, C., Jayasekera, J., Dash, C., Sheppard, V., Mandelblatt, J., 2023. A health equity framework to support the next generation of cancer population simulation models. JNCI Monogr. 2023, 255–264.
- Cruz-Jentoft, A.J., Daragjati, J., Fratiglioni, L., Maggi, S., Mangoni, A.A., Mattace-Raso, F., Paccalin, M., Polidori, M.C., Topinkova, E., Ferrucci, L., Pilotto, A., Investigators, M.A., 2020. Using the Multidimensional Prognostic Index (MPI) to improve cost-effectiveness of interventions in multimorbid frail older persons: results and final recommendations from the MPI AGE European Project. Aging Clin. Exp. Res. 32, 861–868.
- Debray, T.P., Damen, J.A., Snell, K.L., Ensor, J., Hooft, L., Reitsma, J.B., Riley, R.D., Moons, K.G., 2017. A guide to systematic review and meta-analysis of prediction model performance. *bmj* 356.
- Debray, T.P., Koffijberg, H., Nieboer, D., Vergouwe, Y., Steyerberg, E.W., Moons, K.G., 2014. Meta-analysis and aggregation of multiple published prediction models. *Stat. Med.* 33, 2341–2362.
- Egger, M., Smith, G.D., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. *Bmj* 315, 629–634.
- Farmer, C., Fenu, E., O'Flynn, N., Guthrie, B., 2016. Clinical assessment and management of multimorbidity: summary of NICE guidance. *Bmj* 354.
- Gill, T.M., 2012. The central role of prognosis in clinical decision making. *Jama* 307, 199–200.
- Graf, C.E., Zekry, D., Giannelli, S., Michel, J.-P., Chevalley, T., 2011. Efficiency and applicability of comprehensive geriatric assessment in the emergency department: a systematic review. *Aging Clin. Exp. Res.* 23, 244–254.
- Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 29–36.
- Hayden, J.A., van der Windt, D.A., Cartwright, J.L., Côté, P., Bombardier, C., 2013. Assessing bias in studies of prognostic factors. *Ann. Intern. Med.* 158, 280–286.
- Higgins, J.P., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. *Bmj* 327, 557–560.
- Iorio, A., Spencer, F.A., Falavigna, M., Alba, C., Lang, E., Burnand, B., McGinn, T., Hayden, J., Williams, K., Shea, B., 2015. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *bmj* 350.
- Kangtanyagan, C., Vatcharavongvan, P., 2023. No terminally ill patients with non-cancer received palliative care services during hospital admission: a cross-sectional study. *Am. J. Hosp. Palliat. Med.* 40, 492–499.
- McGuinness, L.A., Higgins, J.P.T., 2020. Risk-of-bias visualization (robvis): An R package and shiny web app for visualizing risk-of-bias assessments. *Res. Synth. Methods N/a*.
- Moons, K.G., de Groot, J.A., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D.G., Reitsma, J.B., Collins, G.S., 2014. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med.* 11, e1001744.
- Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamsseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int. J. Surg.* 88, 105906.
- Pilotto, A., Custodero, C., Maggi, S., Polidori, M.C., Veronese, N., Ferrucci, L., 2020. A multidimensional approach to frailty in older people. *Ageing Res. Rev.* 60, 101047.
- Pilotto, A., Ferrucci, L., Franceschi, M., D'Ambrosio, L.P., Scarcelli, C., Cascavilla, L., Paris, F., Placentino, G., Seripa, D., Dallapiccola, B., Leandro, G., 2008. Development and validation of a multidimensional prognostic index for one-year mortality from comprehensive geriatric assessment in hospitalized older patients. *Rejuvenation Res.* 11, 151–161.
- Pilotto, A., Sancarolo, D., Daragjati, J., Panza, F., 2015. Perspective: the challenge of clinical decision-making for drug treatment in older people. The role of multidimensional assessment and prognosis. *Front. Med.* 1, 61.
- Rector, T.S., Taylor, B.C., Wilt, T.J., 2012. Chapter 12: systematic review of prognostic tests. *J. Gen. Intern. Med.* 27, 94–101.
- Schonberg, M.A., Smith, A.K., 2016. Discussing long-term prognosis in primary care: hard but necessary. *JAMA Intern. Med.* 176, 678–680.
- Siontis, G.C., Tzoulaki, I., Ioannidis, J.P., 2011. Predicting death: an empirical evaluation of predictive tools for mortality. *Arch. Intern. Med.* 171, 1721–1726.
- Veronese, N., Custodero, C., Demurtas, J., Smith, L., Barbagallo, M., Maggi, S., Cella, A., Vanacore, N., Aprile, P.L., Ferrucci, L., 2022. Comprehensive geriatric assessment in older people: an umbrella review of health outcomes. *Age Ageing* 51, afac104.
- Yourman, L.C., Lee, S.J., Schonberg, M.A., Widera, E.W., Smith, A.K., 2012. Prognostic indices for older adults: a systematic review. *Jama* 307, 182–192.