BMJ Open

Frailty assessment tools for use by surgeons when evaluating older adults prior to surgery: a scoping review protocol

Tyler R Chesney,[1,2] Camilla Wong,[2,3] Andrea C Tricco ,[4,5] Duminda N Wijeysundera,[6,7] Karim Shiraz Ladha,[6,7] Teruko Kishibe,[2] Samuel Dubé,[8] Martine T E Puts,[9] Shabbir M H Alibhai,[10,11] Julian F Daza [1,12]

To cite: Chesney TR, Wong C, Tricco AC, *et al*. Frailty assessment tools for use by surgeons when evaluating older adults prior to surgery: a scoping review protocol. *BMJ Open* 2022;**12**:e061951. doi:10.1136/bmjopen-2022-061951

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-061951).

Received 15 February 2022
Accepted 06 July 2022

For numbered affiliations see end of article.

Correspondence to
Dr Tyler R Chesney;
tyler.chesney@unityhealth.to

ABSTRACT

Introduction Despite growing evidence, uncertainty persists about which frailty assessment tools are best suited for routine perioperative care. We aim to understand which frailty assessment tools perform well and are feasible to implement.

Methods and analysis Using a registered protocol following Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA), we will conduct a scoping review informed by the Joanna Briggs Institute Guide for Scoping Reviews and reported using PRISMA extension for Scoping Reviews recommendations. We will develop a comprehensive search strategy with information specialists using the Peer Review of Electronic Search Strategies checklist, and implement this across relevant databases from 2005 to 13 October 2021 and updated prior to final review publication. We will include all studies evaluating a frailty assessment tool preoperatively in patients 65 years or older undergoing intracavitary, non-cardiac surgery. We will exclude tools not assessed in clinical practice, or using laboratory or radiologic values alone. After pilot testing, two reviewers will independently assess information sources for eligibility first by titles and abstracts, then by full-text review. Two reviewers will independently chart data from included full texts using a piloted standardised electronic data charting. In this scoping review process, we will (1) index frailty assessment tools evaluated in the preoperative clinical setting; (2) describe the level of investigation supporting each tool; (3) describe useability of each tool and (4) describe direct comparisons between tools. The results will inform ready application of frailty assessment tools in routine clinical practice by surgeons and other perioperative clinicians.

Ethics and dissemination Ethic approval is not required for this secondary data analysis. This scoping review will be published in a peer-review journal. Results will be used to inform an ongoing implementation study focused on geriatric surgery to overcome the current lack of uptake of older adult-oriented care recommendations and ensure broad impact of research findings.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The proposed review has been designed and will be conducted and reported in accordance with best practices in evidence synthesis methodology.
⇒ By focusing solely on tools studied for clinical application before surgery, we may exclude potentially useful tools that have not yet been investigated in a surgical population.
⇒ There may be non-frailty assessment tools that are of relevance in the preoperative setting that will not be captured in this review; however, this review focuses specifically on frailty assessment tools.

INTRODUCTION

Despite exponential growth in frailty research in surgery, effective guidance for surgeons when evaluating older patients prior to surgery is lacking.[1–5] Frailty is present in 10%–30% of older adults, and is associated with inferior postoperative outcomes including major complications, death and functional decline.[6–11] A high prevalence of frailty has major health systems implications as the population ages and older adults comprise over half of inpatient surgical procedures.[12 13] Identification of frailty improves risk stratification, shared decision-making and enables targeted multidisciplinary intervention (eg, prehabilitation, nutritional supplementation, shared care).[3 14 15] Frailty assessment tools can be used in the clinical setting as measurement tools to diagnose or screen for frailty, to make a prognostic assessment of expected outcomes, or to estimate treatment effect to guide clinical decisions (ie, differential treatment effect).[16 17]

Myriad frailty assessment tools are described, with varying degrees of development and validation rigour; yet, it remains unclear which of these tools should be

applied in routine surgical care.[3–5 18] Prior efforts at evidence synthesis have summarised the association of frailty with various postoperative outcomes, but most have focused on effects summarised across various tools or methods for evaluating frailty.[11 19–23] Without focusing on the properties of individual frailty assessment tools, it is challenging to select specific tools for routine clinical application. A recent review has synthesised the psychometric properties of frailty assessment tools in the non-surgical setting.[24] A single review has examined the association of individual frailty tools with postoperative outcomes, but these have been synthesised across surgical types and diagnostic or treatment effect properties were not examined.[25] Acknowledging the lack of clarity about which tools to apply, while many specialty societies recommend frailty screening, none strongly recommend specific tools.[3 26–32]

Given the exponential growth in publications on frailty in surgery, we aim to understand the available knowledge related to frailty assessment tools applied in routine surgical care including purpose, level of investigation, usability and comparisons. We will conduct a scoping review as this methodology is designed to address broad questions and examine the extent, range and characteristics of the published literature as well as summarise findings from a heterogeneous body of knowledge.[33 34] Based on the results of this scoping review, future systematic reviews with quantitative meta-analysis may be conducted for frailty assessment tools with sufficient available evidence.

In this scoping review, we will systematically identify the published literature assessing frailty assessment tools in the preoperative clinical setting, and

1. Index the frailty assessment tools that have been developed or evaluated in the preoperative clinical setting.
2. Describe the level of investigation supporting each tool for diagnosis, prognosis or treatment effect estimation.
3. Describe useability in clinical practice of each tool.
4. Describe direct comparisons between tools.

We will use this evidence synthesis to index the level of investigation evaluating frailty assessment tools that can be applied in routine clinical practice by surgeons and other perioperative physicians.

METHODS AND ANALYSIS

We report this scoping review protocol in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols (PRISMA-P) recommendations along with updated scoping review methodologic guidance: the conduct of the review is informed by the formally Joanna Briggs Institute methodology manual for scoping reviews, and the final manuscript will be reported in accordance with the PRISMA Extension for Scoping Reviews (PRISMA-ScR) recommendations.[33–36] The completed PRISMA-P checklist can be found in online supplemental appendix A.

Patient and public involvement

Patient and public will not be involved in the design, conduct or parting of the study. Results of this review will be disseminated to relevant conferences and peer-reviewed journals, and by including them in subsequent implementation research. The results of this scoping review will be compared with any recommendations provided in current specialty society recommendations.[3 26–32]

Review question

The review question was generated in consultation with leading experts in evidence synthesis, knowledge translation, perioperative risk stratification and geriatric perioperative medicine. The research question is: What frailty assessment tools exist for use when evaluating older adults in the preoperative clinical setting, and what level of investigation is available assessing measurement, diagnostic, prognostic and useability properties?

Definitions

Frailty is a state of vulnerability to stressors due to multisystem decline in physiological reserve and function, thereby increasing the risk of adverse health outcomes.[37–41] Several evidence-based models have been developed to operationalise frailty including (1) the cumulative deficit model reflecting a cumulative effect of deficits acquired across many domains including medical, social and functional; and (2) the phenotype model reflecting a biological syndrome of decline across multiple physiological systems.[37 38 42]

For this review, frailty assessment tools will be considered health measurement tools (either based on formative or reflective models), with a specific and reproducible set of variables used to assess frailty in older adults as reflected by the definition above; these tools typically are multicomponent tools developed to assess frailty based on an established model of frailty.[16]

Eligibility criteria

The eligibility criteria are summarised here and elaborated in the tables provided in online supplemental appendix B.

Population

We will include studies of older adults undergoing major intracavitary, non-cardiac surgery. This will be considered those aged 65 or older, or a study population with a median age of 65 or older, or where the majority of included individuals are 65 years or older.[43 44] We will exclude noncavitary (eg, soft tissue, extremity and neurosurgery) and cardiac surgery as these have largely different perioperative considerations and postoperative outcomes.

Concept

We will include studies reporting on the development or evaluation of characteristics of a frailty assessment tool. Eligible studies will report on frailty assessment tool characteristics used for diagnosis, prognosis or estimating

treatment effect. Characteristics may include reliability, validity and useability (see elaboration tables below for further detail).[45–51] We will exclude tools not assessed in clinical practice (ie, assessed using administrative data alone without use in the clinic setting), assessing a single domain (eg, physical performance alone, malnutrition alone), or using laboratory/radiologic values alone, as these do not adequately represent the multidimensionality of frailty assessment tools. We will exclude studies that do not employ a formal frailty assessment tool; studies in which the 'frailty' assessment tools is in fact used to assess a different construct (eg, disability, sarcopenia); and studies that only include frailty as an adjustment covariate in a multivariable prognostic model without further reporting on individual characteristics of the frailty assessment (with respect to diagnosis, prognosis, measurement properties or other relevant outcomes).

Context

We will include studies that report on frailty assessment tools for use in the clinical setting prior to surgery.

Types of evidence sources

We will include randomised trials (primary and post hoc analyses), prospective or retrospective cohort studies, diagnostic test accuracy studies, measurement properties studies, prediction studies, useability studies and systematic reviews of any of the study types above reported in any language. We will include qualitative studies only if they include evaluation of the impact or clinical useability of frailty assessment tools. We will exclude studies that do not have full text publications.

Information sources and search strategy

We will develop a comprehensive search strategy in collaboration with an expert information specialist, and this will be peer-reviewed using the Peer Review of Electronic Search Strategies checklist.[52]

We will translate and implement the search across all relevant databases (eg, MEDLINE, EMBASE, CENTRAL, CINAHL) from 2005 to 13 October 2021 (the first use of frailty as a title word in a surgical population was in 2006), humans-only studies and without language restrictions.[38 53] The search strategy takes the basic format of: Frailty Assessment Tools AND ((Post-Operative Care) OR (Prediction/Prognosis AND Postoperative Outcomes)). Sample search strategy is available in online supplemental appendix C. Case reports, comments, editorials and letters will be removed. We will report each database and register the date of search, and the search strategy for all databases and registers. We will not include a search of the grey literature as it is unlikely that informative clinical studies of frailty assessment tools in surgical populations will be available in these sources. The search will be updated prior to publication. We will supplement these sources by scanning references lists of included studies for additional sources of evidence. The grey literature was not searched given the objectives of this review are to identify the level of evidence examining various performance measure of frailty assessment tools in the preoperative setting.

Review team calibration and consistency

We anticipate a large number of citations, so a review team will be used for selection of sources of evidence, and data charting of individual sources of evidence. For each step, a standardised electronic form and explanation and elaboration document will be developed by the study team based on eligibility criteria and objectives. The review leads will pilot test the forms on a sample of eligible and non-eligible papers to assess appropriateness and comprehensiveness and make revisions as needed. The full review team will be trained to use these forms sequentially in each review step. After training for each step, calibration exercises will be conducted with the full review team by pilot testing the forms on 50 randomly selected citations for screening, and five citations for data charting. We will review discrepancies in group discussions and refinements will be made to the forms as needed and reported. Additional calibration exercises may be done if sufficient agreement across reviewers is not reached or if reviewers express the need for more training. Sufficient agreement in level 1 screening will be accepted if no more than 20% of studies included by the review leads are excluded by one or more reviewer. Sufficient agreement in level 2 screening and data charting will be at least 80% agreement across all reviewers. Pilot forms will be set up in Excel; an example template is in online supplemental appendix D.

Records and selection of sources of evidence

We will employ the Covidence web-based platform for systematic review management.[54] After deduplication, all citations will be loaded into Covidence. We will conduct two-level citation screening for eligibility: level 1 screening titles and abstracts and level 2 screening full texts. Two independent reviewers will screen each citation at both screening stages using the piloted selection form and accompanying elaboration and explanation document reflecting the eligibility criteria. Calibration exercises will be conducted as described previously. Citations selected for inclusion by at least one reviewer in stage 1 will be included in stage 2 screening. The reason for exclusion in full-text screening will be recorded. The results of the search and selection will be reported and presented in a PRISMA flow diagram. We will request additional information from study authors by email if needed to decide on eligibility. Disagreement will be resolved through discussion involving a third reviewer if needed.

Data charting process

Once all records for final inclusion have been selected, we will chart data from all full text records.

Process

Two independent reviewers will chart data from each eligible record. When more than one record exists for

the same study, the record with the most complete or most recent information will be used. We will develop and pilot electronic data charting forms in Google Forms and a detailed explanation and elaboration manual for additional details and definitions on data items to be charted prior to data charting (this will incorporate the below data items and definitions). Calibration exercises will be conducted as described previously. Disagreement will be resolved through discussion, and involving a third reviewer if agreement is not reached. We will request additional information from study authors if needed.

Data items

We will chart data on study and population (eg, publication type, country, age, proportion with frailty, setting); frailty assessment tool characteristics (eg, name, type, geriatric domains, scoring, purpose, access); measurement properties of frailty assessment tools (eg, conceptual framework, validity, reliability); information on tools for diagnosis (eg, type, reference standard, diagnostic accuracy measures); information on tools for prognosis (eg, analysis type, predicted outcomes, purpose, accuracy measures); and useability (eg, feasibility, acceptability, time, equipment).

Tables elaborating data items that will be charted are included in supplementary appedix E, providing definitions, prespecified assumptions and simplifications, as well as information on how items should be interpreted.

Synthesis of results

Synthesis will be guided by the main study objectives. The outputs will be stratified across settings (emergency, benign, oncology) and surgery types (eg, colorectal, gynaecology, urology). We will collate all studies and relevant information relating to each frailty assessment tool. The expected results of the scoping review include an index of the frailty assessment tools developed or evaluated in the preoperative setting with older adults undergoing surgery, along with the level of investigation across diagnosis, prognosis, treatment effect estimation and other measurement properties.

The synthesis will including the following items:

Characteristics of frailty assessment tools

We will list all frailty assessment tools described in the literature. For each tool we will list characteristics including the number of studies evaluating each tool, scale, scoring method, language versions, development population, purpose of tool and access.

Frailty assessment tool measurement properties

We will summarise which measurement properties have been assessed for each tool as listed in the data charting items.

Level of investigation

We will summarise the level of investigation for each frailty assessment tool stratified by purpose (eg, diagnosis, prognosis). Specifically, level of investigation will be ordered as development study, validation study, replication study, impact study and systematic review. Studies will be included if they only provide unadjusted or adjusted associations between preoperative frailty (measured using a frailty assessment tool) and a postoperative outcome, but will be ranked as the lowest level of investigation unless formal diagnostic test, prognostic, treatment effect estimate, impact analysis or measurement properties study methodology is used.

Useability

For tools with higher levels of investigation, we will summarise assessments of useability descriptively.

Comparisons

We will summarise the outcomes of any direct comparisons between tools.

Summary of results

We will create a visual summary (eg, bubble plot) to visually summarise the number of studies and level of investigation supporting each tool, stratified by setting (overall, oncology, benign, emergency) and by surgery type.

Author affiliations
[1]Department of Surgery, Division of General Surgery, University of Toronto, Toronto, Ontario, Canada
[2]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health Toronto, Toronto, Ontario, Canada
[3]Division of Geriatric Medicine, St. Michael's Hospital, Unity Health Toronto, Toronto, Ontario, Canada
[4]Knowledge Translation Program, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health Toronto, Toronto, Ontario, Canada
[5]Queen's Collaboration for Health Care Quality JBI Centre of Excellence, School of Nursing, Queen's University, Kingston, Ontario, Canada
[6]Department of Anesthesiology and Pain Medicine, St. Michael's Hospital, Unity Health Toronto, Toronto, Ontario, Canada
[7]Department of Anesthesiology and Pain Medicine, and Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada
[8]Division of Gynecologic Oncology, Université de Montréal, Montreal, Quebec, Canada
[9]Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, Toronto, Ontario, Canada
[10]Department of Medicine, and Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada
[11]Department of Medicine, University Health Network, Toronto, Ontario, Canada
[12]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada

Contributors All authors (TRC, CW, ACT, DNW, KSL, TK, SD, MTEP, SMHA and JFD) contributed substantially to the design and conception of the study. TK developed the search criteria with input from TRC and JFD. TRC and JFD drafted the initial manuscript. All authors (TRC, CW, ACT, DNW, KSL, TK, SD, MTEP, SMHA and JFD) revised the manuscript critically for important intellectual content, approved the final submission and agreed to be held accountable for all aspects of the work.

Funding This work was supported by MDFM Ontario Association of General Surgeons Resident Research Grant 2020 grant number N/A and the St. Michael's Hospital AFP Innovation Fund grant number SMH-21-020. ACT is funded by a tier 2 Canada Research Chair in Knowledge Synthesis number N/A. DNW is supported in part by a Merit Award from the Department of Anesthesiology and Pain Medicine at the University of Toronto number N/A and by the Endowed Chair in Translational Anesthesiology Research at St. Michael's Hospital and the University of Toronto number N/A. MTEP is funded by a tier 2 Canada Research Chair in Care for Frail Older Adults number N/A.

Disclaimer Funders had no role in developing or reviewing the protocol.

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.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Andrea C Tricco http://orcid.org/0000-0002-4114-8971
Julian F Daza http://orcid.org/0000-0002-4486-9571

REFERENCES

1 Soto-Perez-de-Celis E, Li D, Yuan Y, *et al*. Functional versus chronological age: geriatric assessments to guide decision making in older patients with cancer. *Lancet Oncol* 2018;19:e305–16.
2 Suskind AM, Finlayson E. A call for frailty screening in the preoperative setting. *JAMA Surg* 2017;152:240–1.
3 Alvarez-Nebreda ML, Bentov N, Urman RD, *et al*. Recommendations for preoperative management of frailty from the Society for perioperative assessment and quality improvement (SPAQI). *J Clin Anesth* 2018;47:33–42.
4 Chesney TR, Pang G, Ahmed N. Caring for older surgical patients: contemporary attitudes, knowledge, practices, and needs of general surgeons and residents. *Ann Surg* 2017;268:77–85.
5 Ghignone F, van Leeuwen BL, Montroni I, *et al*. The assessment and management of older cancer patients: a SIOG surgical task force survey on surgeons' attitudes. *Eur J Surg Oncol* 2016;42:297–302.
6 Collard RM, Boter H, Schoevers RA, *et al*. Prevalence of frailty in community-dwelling older persons: a systematic review. *J Am Geriatr Soc* 2012;60:1487–92.
7 Song X, Mitnitski A, Rockwood K. Prevalence and 10-year outcomes of frailty in older adults in relation to deficit accumulation. *J Am Geriatr Soc* 2010;58:681–7.
8 Handforth C, Clegg A, Young C, *et al*. The prevalence and outcomes of frailty in older cancer patients: a systematic review. *Ann Oncol* 2015;26:1091–101.
9 Makary MA, Segev DL, Pronovost PJ, *et al*. Frailty as a predictor of surgical outcomes in older patients. *J Am Coll Surg* 2010;210:901–8.
10 Li Y, Pederson JL, Churchill TA, *et al*. Impact of frailty on outcomes after discharge in older surgical patients: a prospective cohort study. *CMAJ* 2018;190:E184–90.
11 Hewitt J, Long S, Carter B, *et al*. The prevalence of frailty and its association with clinical outcomes in general surgery: a systematic review and meta-analysis. *Age Ageing* 2018;47:793–800.
12 Etzioni DA, Liu JH, Maggard MA, *et al*. The aging population and its impact on the surgery workforce. *Ann Surg* 2003;238:170–7.
13 Statistics Canada. *Population projections for Canada, provinces and territories: 2009 to 2036*. Statistics Canada, 2015. https://www150.statcan.gc.ca/n1/pub/91-520-x/2010001/aftertoc-aprestdm1-eng.htm
14 Hall DE, Arya S, Schmid KK, *et al*. Association of a frailty screening initiative with postoperative survival at 30, 180, and 365 days. *JAMA Surg* 2017;152:233–40.
15 McDonald SR, Heflin MT, Whitson HE, *et al*. Association of integrated care coordination with postsurgical outcomes in high-risk older adults: the perioperative optimization of senior health (POSH) initiative. *JAMA Surg* 2018;153:454–62.
16 Buta BJ, Walston JD, Godino JG, *et al*. Frailty assessment instruments: systematic characterization of the uses and contexts of highly-cited instruments. *Ageing Res Rev* 2016;26:53–61.
17 Rockwood K, Theou O, Mitnitski A. What are frailty instruments for? *Age Ageing* 2015;44:545–7.
18 McIsaac DI, MacDonald DB, Aucoin SD. Frailty for perioperative clinicians: a narrative review. *Anesth Analg* 2020;130:1450–60.
19 Ward MAR, Alenazi A, Delisle M, *et al*. The impact of frailty on acute care general surgery patients: a systematic review. *J Trauma Acute Care Surg* 2019;86:148–54.
20 Buigues C, Juarros-Folgado P, Fernández-Garrido J, *et al*. Frailty syndrome and pre-operative risk evaluation: a systematic review. *Arch Gerontol Geriatr* 2015;61:309–21.
21 Beggs T, Sepehri A, Szwajcer A, *et al*. Frailty and perioperative outcomes: a narrative review. *Can J Anesth/J Can Anesth* 2015;62:143–57.
22 Rostoft S, van Leeuwen B. Frailty assessment tools and geriatric assessment in older patients with hepatobiliary and pancreatic malignancies. *Eur J Surg Oncol* 2021;47:514–8.
23 Lin H-S, Watts JN, Peel NM, *et al*. Frailty and post-operative outcomes in older surgical patients: a systematic review. *BMC Geriatr* 2016;16:157.
24 Sutton JL, Gould RL, Daley S, *et al*. Psychometric properties of multicomponent tools designed to assess frailty in older adults: a systematic review. *BMC Geriatr* 2016;16:1–20.
25 Aucoin SD, Hao M, Sohi R, *et al*. Accuracy and feasibility of clinically applied frailty instruments before surgery: a systematic review and meta-analysis. *Anesthesiology* 2020;133:78–95.
26 Chow WB, Rosenthal RA, Merkow RP, *et al*. Optimal preoperative assessment of the geriatric surgical patient: a best practices guideline from the American College of surgeons national surgical quality improvement program and the American geriatrics Society. *J Am Coll Surg* 2012;215:453–66.
27 Griffiths R, Beech F, Brown A, *et al*. Peri-Operative care of the elderly 2014: association of anaesthetists of great britain and ireland. *Anaesthesia* 2014;69 Suppl 1:81–98.
28 Montroni I, Ugolini G, Saur NM, *et al*. Personalized management of elderly patients with rectal cancer: expert recommendations of the European Society of surgical oncology, European Society of Coloproctology, International Society of geriatric oncology, and American College of surgeons Commission on cancer. *Eur J Surg Oncol* 2018;44:1685–702.
29 Frailty Guideline Working Group. Guideline for perioperative care for people living with frailty undergoing elective and emergency surgery. Available: https://cpoc.org.uk/sites/cpoc/files/documents/2021-09/CPOC-BGS-Frailty-Guideline-2021.pdf [Accessed 1 Jul 2022].
30 Saur NM, Davis BR, Montroni I, *et al*. The American Society of colon and rectal surgeons clinical practice guidelines for the perioperative evaluation and management of frailty among older adults undergoing colorectal surgery. *Dis Colon Rectum* 2022;65:473–88.
31 Aceto P, Antonelli Incalzi R, Bettelli G, *et al*. Perioperative management of elderly patients (prime): recommendations from an Italian intersociety consensus. *Aging Clin Exp Res* 2020;32:1647–73.
32 Geriatric Surgery Verification Program. Optimal resources for geriatric surgery. Available: https://www.facs.org/media/f10eka54/geriatricsv_standards.pdf [Accessed 1 Jul 2022].
33 Peters M, Godfrey C, McInerney P. Chapter 11: Scoping Reviews. In: Aromataris E, Munn Z, eds. *JBI Manual for Evidence Synthesis*. JBI, 2020.
34 Peters MDJ, Marnie C, Tricco AC, *et al*. Updated methodological guidance for the conduct of scoping reviews. *JBI Evid Synth* 2020;18:2119–26.
35 Shamseer L, Moher D, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647.
36 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.
37 Clegg A, Young J, Iliffe S, *et al*. Frailty in elderly people. *The Lancet* 2013;381:752–62.
38 Robinson TN, Walston JD, Brummel NE, *et al*. Frailty for surgeons: review of a national Institute on aging conference on frailty for specialists. *J Am Coll Surg* 2015;221:1083–92.
39 Anaya DA, Johanning J, Spector SA, *et al*. Summary of the panel session at the 38th annual surgical symposium of the association of Va surgeons: what is the big deal about frailty? *JAMA Surg* 2014;149:1191–7.
40 Ethun CG, Bilen MA, Jani AB, *et al*. Frailty and cancer: implications for oncology surgery, medical oncology, and radiation oncology. *CA Cancer J Clin* 2017;67:362–77.
41 Morley JE, Vellas B, van Kan GA, *et al*. Frailty consensus: a call to action. *J Am Med Dir Assoc* 2013;14:392–7.
42 Rodríguez-Mañas L, Féart C, Mann G, *et al*. Searching for an operational definition of frailty: a Delphi method based consensus

statement. The frailty operative Definition-Consensus conference project. *J Gerontol A Biol Sci Med Sci* 2013;68:62–7.
43 Hurria A, Levit LA, Dale W, *et al*. Improving the evidence base for treating older adults with cancer: American Society of clinical oncology statement. *J Clin Oncol* 2015;33:3826–33.
44 Lundebjerg NE, Trucil DE, Hammond EC, *et al*. When it comes to older adults, language matters: Journal of the American geriatrics Society adopts modified American Medical association style. *J Am Geriatr Soc* 2017;65:1386–8.
45 Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis* 1985;38:27–36.
46 Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
47 Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
48 Terwee CB, Mokkink LB, Knol DL, *et al*. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651–7.
49 Moons KGM, Altman DG, Reitsma JB, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1–73.
50 Steyerberg EW, Pencina MJ, Lingsma HF, *et al*. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. *Eur J Clin Invest* 2012;42:216–28.
51 Cohen JF, Korevaar DA, Altman DG, *et al*. Stard 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6:e012799.
52 McGowan J, Sampson M, Salzwedel DM, *et al*. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol* 2016;75:40–6.
53 Ferrucci L, Maggio M, Ceda GP, *et al*. Acute postoperative frailty. *J Am Coll Surg* 2006;203:134–5.
54 Covidence - Better systematic review management. Covidence. Available: https://www.covidence.org/ [Accessed 31 Jan 2021].

APPENDIX A

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Checklist item	Page Number
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	Identify the report as a protocol of a systematic review	1
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	Na
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	7
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	1
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	14
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	7
Support:			
Sources	5a	Indicate sources of financial or other support for the review	14
Sponsor	5b	Provide name for the review funder and/or sponsor	14
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	14
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	6
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	7
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	8
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	9
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	23
Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	11

Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	11
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	11
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	12
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	12
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	Na
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	Na
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)	Na
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	Na
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	Na
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	Na
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	Na

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.

APPENDIX B

Eligibility criteria elaboration tables

POPULATION – Does this study include older adults undergoing major intracavitary, noncardiac surgery?	
INCLUDE	EXCLUDE
Human participants Older adults (age 65 years or older) ▪ Study includes only those aged 65 years or older ▪ Mean or median age of 65 or older ▪ Majority of individuals are 65 years or older ▪ Results are reported for a subgroup of patients 65 years or older Major intracavitary, noncardiac surgery ▪ Surgery conducted with either open, laparoscopic, or robotic technique within the abdomen, pelvis, or thorax ▪ Studies with a mixed surgery types should include >80% intracavitary surgery, or present results for this subgroup ▪ Elective or emergency surgery ▪ Major surgery will be defined based on the definition in the study, or including surgery types that have an expected length of stay greater than 2 days.	▪ Nonhuman subjects ▪ Not surgical patients ▪ Patients younger than 65 years old ▪ Cardiac surgery ▪ Extremity, soft tissue, or breast surgery; orthopedic surgery, plastic surgery ▪ Endovascular and extremity vascular surgery ▪ Head and neck surgery ▪ Neurosurgery or spine surgery; including anterior approach spine surgery ▪ Trauma care (traumatic injury care and trauma surgery) ▪ Studies with mixed surgery types and less that 20% intracavitary surgery, or not reporting results for intracavitary surgery as a subgroup ▪ Minor surgery or surgery types with an expected length of stay less than 2 days. ▪ Mixed surgical and nonsurgical populations should be excluded; unless frailty tool assessment has been conducted and reported on the surgical subgroup separately. NOTE* Studies that retrospectively derive their population from a large database of routinely collected data should be excluded unless they explicitly state that a clinically applied frailty tool was used. A common example of the is the "American College of Surgeons National Surgical Quality Improvement Program" or "NSQIP". Studies that use this database should be excluded, unless they clearly use a clinical frailty tool, and use the NSQIP data only for outcomes assessments. This can be confusing because they talk of using the "mFI" or "Modified frailty Index", but this is actually retrospectively computed based on data in the database, rather than an assessment done with the patient. If a study explicitly describes using the mFI prospectively, this could be included.

CONCEPT – Does this study assess a frailty assessment tool?	
INCLUDE	EXCLUDE
▪ Health measurement tools assessing frailty. Frailty is a state of vulnerability to stressors increasing the risk of adverse health outcomes due to multisystem decline in physiologic reserve and function. These tools typically are multicomponent tools developed to assess frailty based on an established model of frailty such as the phenotype model or the cumulative deficit model ▪ Other terms that may be used for frailty are: vulnerability, physiologic age, functional age, accumulation of deficits	▪ Health measurement tools assessing constructs other than frailty (assessing other constructs such as disability, sarcopenia, comorbidity, function, chronologic age) ▪ Tools assessing single domains such as physical performance alone, malnutrition alone, cognition alone, daily function alone, etc. ▪ Single laboratory or radiologic values ▪ Studies that base frailty assessment only on Comprehensive Geriatric Assessment or assessments that require an expert geriatrics team; that is, studies that do not assess frailty using an assessment tool that can be applied by a non-geriatrics expert. NOTE* that studies that compare another frailty assessment tool to Comprehensive Geriatric Assessment should be included. Also, sometimes "GA" or "geriatric assessment" is a term used, this can be included if conducted by non-geriatrician/geriatrics expert team ▪ Studies using surgical risk scores that do not measure, assess, screen for, or diagnose frailty NOTE* Disease-specific risk calculators (like RCRI, Revised Cardia Risk Index), the ASA (Anesthesia Society of America Physical Status assessment), and comorbidity indexes (like Charlson Comorbidity Index, CCI) should not be considered frailty assessment tools and should not be included.

CONCEPT – Does this study report on the characteristics of a frailty assessment tool?	
INCLUDE	EXCLUDE
Studies that report on any of the following characteristics of a frailty assessment tool: ▪ Validity (content/face validity, construct validity, structural validity, criterion validity)	▪ Studies that only include frailty as a variable in a multivariable model without reporting further characteristics of the frailty assessment on diagnosis, prognosis, measurement properties, or other outcomes

▪ Reliability (internal consistency, inter-rater reliability, test-retest reliability) ▪ Measurement properties ▪ Diagnostic test accuracy ▪ Diagnostic performance ▪ Sensitivity ▪ Specificity ▪ Positive predictive value/Negative predictive value ▪ Accuracy ▪ Likelihood ratio ▪ Predictive performance ▪ Prognostic performance ▪ Overall performance (R^2, Brier score) ▪ Discrimination (Receiver operating curve, ROC, AUC, c statistic, concordance index, Integrated Discrimination Improvement) ▪ Calibration (calibration plot, Hosmer-Lemeshow) ▪ Reclassification (Net Reclassification Index) ▪ Clinical usefulness, clinical impact, impact analysis, clinical decision rule, Difference in net benefit (NB), decision curve analysis (DCA) ▪ Decision curve analysis (DCA) ▪ Utility ▪ Useability (feasibility, acceptability, satisfaction, implementation, availability, language versions, practicality or training)[41] ▪ For systematic reviews, these will be only included if a relevant summary estimate of any of these outcomes is reported for individual frailty assessment tool(s).	▪ Studies developing novel prognostic or diagnostic models ▪ Systematic reviews that report pooled outcomes across frailty assessment tools without outcomes for individual frailty assessment tools Note. if frailty is used only in a multivariable model to adjust an analysis of another exposure (like surgical type, surgery vs no surgery, old vs young), this should be excluded. In contrast, if other variables are used to adjust for the effect of frailty, like what is the impact of frailty on mortality adjusted for age and comorbidity, this study could be included.

CONTEXT – Does this study assess a frailty assessment tool conducted clinically before surgery?	
INCLUDE	EXCLUDE
▪ Tool applied in the clinical setting either with patient-reported, caregiver-reported, or clinician-reported information using information derived in the clinical context including from the health record ▪ Tool used before surgery	▪ Tools applied to administrative data (such as NSQIP), simulation models, computer models, laboratory setting, nonclinical research setting ▪ Studies assessing frailty only after surgery

TYPE OF EVIDENCE SOURCE – Is this the right publication type?	
INCLUDE	EXCLUDE
Include all study designs that report of properties of a frailty assessment tool ▪ Randomized clinical trial (RCT) ▪ Cohort Study (prospective or retrospective) ▪ Prognosis study ▪ Diagnostic accuracy study ▪ Measurement properties study ▪ Prediction study ▪ Useability study (this may include feasibility, acceptability, satisfaction, implementation, or required training) ▪ Systematic reviews of these types of studies ▪ Qualitative studies only if evaluation of the impact or clinical useability of frailty assessment tools	▪ Narrative reviews ▪ Clinical overviews ▪ Literature reviews ▪ Book chapters ▪ Case reports ▪ Editorials, letters to the editor, commentaries ▪ Opinion pieces ▪ Protocols ▪ Studies without full text publications NOTE* that narrative reviews, better termed clinical overviews, clinical summaries, or literature reviews are descriptive write-ups often styled like a book chapter, meaning they are not conducted as a study either using primary quantitative data or as a systematic review of the literature. Systematic reviews use explicit methods to systematically search the literature, extract data, and synthesis this (*note that sometimes narrative synthesis is used in a systematic review, but the study uses explicit methodology). Narrative reviews/Clinical overviews are often indicated when the abstract is simply a narrative block of text without any methods, objectives, or outcomes described.

APPENDIX C

Sample Search Strategy

Medline

1 Frailty/di [Diagnosis] 1551
2 Frailty/ep [Epidemiology] 1475
3 Geriatric Assessment/ 30072
4 frailty.ti. 8301
5 ((frail or frailty or frailties) adj3 (diagnos* or assessment* or index or indices or score or scores or scale or scales or tool or tools or evaluat* or performance or instrument or instruments or analysis or analyses or questionnaire* or survey* or measure* or screen* or test or tests or testing)).tw,kf. 6846
6 geriatric assessment*.tw,kf. 4967
7 (prefrail or nonfrail).tw,kf. 902
8 physiologic age.tw,kf. 141
9 functional age.tw,kf. 173
10 "accumulation of deficit*".tw,kf. 111
11 (Balducci or Bern Scale or Columbia Scale or Essential Frailty Toolset or Frailty Phenotype or Fried or G8 or Geriatric-8 or Groningen Frailty Indicator or Risk Analysis Index or Rockwood or Triage Risk Screening Tool or Vulnerable Elderly Survey).tw,kf. 7808
12 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 45773
13 preoperative period/ 8946
14 Preoperative Care/ 64211
15 (preoperativ* or pre-operativ*).tw,kf. 350153
16 surg*.ti. 700454
17 "before surg*".kf,ab. 43153
18 13 or 14 or 15 or 16 or 17 1020021
19 "Predictive Value of Tests"/ 215260
20 Prognosis/ 550865
21 Risk Assessment/ 289432
22 risk factors/ 890626
23 "reproducibility of results"/ 427661
24 (predict* or prognos* or risk or risks).tw,kf. 4277976
25 19 or 20 or 21 or 22 or 23 or 24 5135764
26 Postoperative Complications/ 383329
27 Postoperative Period/ 54483
28 (surg* or postoperativ* or post-operativ* or postsurg*).tw,kf. 2381606
29 26 or 27 or 28 2529224
30 25 and 29 679397
31 18 or 30 1449527
32 12 and 31 3376
33 limit 32 to yr="2005 -Current" 3043
34 limit 33 to "humans only (removes records about animals)" 3043
35 limit 34 to (case reports or comment or editorial or letter) 180
36 34 not 35 2863

https://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=N&PAGE=main&SHAREDSEARCHID=4GUsmt4qUz6tGWQ5BOL5vAA7VledBItzQ73uxutK75iRPZG4ORewhbEdnJZdtCIsY

APPENDIX D

Example pilot test form

Citation	Title	Abstract	Include (0=no, 1=yes)	Reason (free text)	Comments (anything that should be clarified in eligibility criteria)
#	…	…			

APPENDIX E

Study and Population Items	
This section includes items about the study and bibliometrics, the study population, study type, and setting.	
Publication type	▪ Randomized clinical trial (RCT) (original study or post-hoc analysis) ▪ Cohort Study (prospective or retrospective) ▪ Prognosis study ▪ Diagnostic accuracy study ▪ Measurement properties study ▪ Prediction study ▪ Useability study (this may include feasibility, acceptability, satisfaction, implementation, or training) ▪ Systematic review ▪ Qualitative study
Study site	Single centre or multicentre
Country of origin	List country or multinational
Funding source	List funding source
Study size	▪ Number of participants meeting criteria for older adult ▪ If number of older adults not reported in a study including younger adults as well, list total number of participants
Age of study participants	Preferentially list in this order ▪ Age cutoff if 65 years or older ▪ Mean or median age if no age cutoff of 65 or older ▪ Percentage older adults (%) if no age cutoff or mean/median and majority of individuals are 65 years or older
Sex and gender of participants	List number or proportion (%) of each sex and gender reported.
Proportion rated as frail	List proportion (%) or participants rated as frail in the study
Follow-up	Mean/median (not applicable for diagnosis studies)
Setting	Emergency, Benign, Oncology, mixed
Surgery Type	General surgery, colorectal, gynecology, etc, mixed
Type of tool assessment	List if this is an initial development study for the tool in the surgical population, a subsequent evaluation study of the properties of a tool previously developed, or a study that directly compares two or more assessment tools. Note that comparisons should be between two frailty assessment tools, not against a gold standard such as Comprehensive Geriatric Assessment (if this is the case, list as either development or validation/evaluation as appropriate). ▪ Development ▪ Validation/Evaluation ▪ Comparison ▪ Useability (list only if this is the only purpose) ▪ Mixed (used for more than one purpose in this study)

Frailty Assessment Tool Characteristics	
This section includes items about the frailty assessment tool including its name, details of the tools, and the purpose of the tool.[22,23,27]	
Name	Record the name of the frailty assessment tool including abbreviation

Tool type	Frailty assessment tools can be of several types based on the type of data used as input for the tool.[42] ▪ Questionnaire: based on a list of questions. These can be answered by a clinician, the patient, or a proxy. ▪ Performance: based on a physical performance test ▪ Judgement: uses an overall assessment by a clinician based on input data ▪ Routine data: based on routine health data collected for other purposes, usually housed in administrative datasets. These are not clinically applied frailty tools, and have been excluded from this review ▪ Mixed: includes more than one type
Geriatric Domains	List which geriatrics domains are included in the frailty assessment tool.[43,44] ▪ Functional independence ▪ Physical performance ▪ Falls ▪ Comorbidity ▪ Nutrition ▪ Polypharmacy ▪ Social support ▪ Cognition ▪ Mood ▪ Other
Scale	List the type of scale of the frailty assessment tool output ▪ Categorical ▪ Ordinal ▪ Continuous Categorical scales have distinct groups, but are not in a particular order; this could include dichotomous scale with only two groups (e.g., frail, not frail). Ordinal scales have distinct groups, but the order of the groups is defined (e.g., frail, pre-frail, fit). Continuous scales do not have distinct groups, are ordered, and the difference between intervals is consistent (e.g., a scale from 0 to 100)
Scoring Method	How is the tool scored ▪ Sum ▪ Mean ▪ Weighted ▪ Threshold
Language Versions	List the languages that the tool is available in
Purpose of tool	List the purpose of the tool in this study ▪ Diagnosis ▪ Prognosis ▪ Estimating treatment effect ▪ Clinical decision rule ▪ Useability Diagnosis: the tool is used for identification of frailty at a single timepoint (preoperatively in this case). This may be either a screening test to identify patients that should have further diagnostic assessment to confirm the diagnosis of frailty, or a more definitive diagnostic test

	Prognosis: the tool is used for estimating the likelihood of a future outcome. Estimating treatment effect: the tool is used to estimate the benefit from a given treatment Clinical decision rule: tool results used to direct a treatment decision, and studies of this kind evaluate outcomes based on treatment decision guided by the tool/clinical decision rule Useability: the study may report on the useability of the frailty assessment tool with or without one of the above purposes. Useability refers to the ease of application of the tool in its intended setting given constraints such as time or money.[45] List as one of the above if present, list as useability if one of the above purposes is not present. Useability may be assessed in many ways including the following: feasibility, completion time, cost, required equipment, acceptability, satisfaction, implementation, availability, language versions, practicality or training.
Tool development	Was the tool developed in a surgical population? ▪ Surgical ▪ Non-surgical ▪ Unknown If this is not the initial development study, this may be described in the introduction of methods sections. If not described, mark as unknown
Access	List where the tool can be accessed (e.g., publication or website)

Measurement Properties of Frailty Assessment Tools	
This section includes items on measurement properties that can be assessed across all measurement tools in medicine.	
Conceptual framework	List the theoretical or operational definition of frailty that supports to the frailty assessment tool. ▪ Cumulative deficits model (see note below) ▪ Phenotype model ▪ Other ▪ None Content validity refers to the degree to which the content of a measurement tool reflects the construct being measured.[31] Frailty assessment tools will be considered to reflect a conceptual framework if they are based on an accepted theoretical or operational definition of frailty. If the operational definition used to develop the frailty assessment tool is not reported explicitly in the study report, this will be sought from an original development study for the frailty assessment tool. If this does not exist, the study team will assign an operational definition if this can be discerned by comparison with other tools. This will be reported. If this is not possible, this will also be reported. Definitions of frailty are listed in the section on definitions above.

	NOTE* There are standard procedures for creating a cumulative deficits frailty index.[48] This includes selecting at least 30-40 variables that represent deficits associated with health status, more prevalent with increasing age but not be too common at younger ages, and cover a range of systems or domains. Indexes composed of around 10 or fewer variables are unstable. When these elements are not met, the frailty assessment tool will not be considered to align with the cumulative deficits model.
Other measurement properties of health measurement tools	List any other measurement properties that are reported about the frailty assessment tool. Health measurement tools can be assessed across multiple properties to document their performance including measures of validity and reliability. Examples include Validity ▪ Construct validity: this is an evaluation of the internal structure of a measurement tool and includes, cross-cultural validity, and convergent or divergent (discriminant) validity, or extreme-groups/known-groups (discriminative) validity. ▪ Criterion validity (assessed separately in tables below for diagnosis and prognosis) Reliability ▪ Internal consistency ▪ Inter-rater reliability ▪ Test-retest reliability ▪ Measurement error Responsiveness (this will not be relevant to tools used for a single preoperative assessment) Structural validity: an assessment of the adequacy of the dimensionality of the tool for measuring the construct (e.g., frailty) and may involve factor analysis, or item response theory/Rasch analysis. Internal consistency: an assessment of the interrelatedness among scale items, and may be assessed by Cronbach's alpha. Measurement invariance: also called cross-cultural validity, is an assessment of the degree to which the performance of a new population reflects the initial measurement performance. This can be assess by differential item functioning (DIF), or multigroup confirmatory factor analysis (MGCFA) Convergent or discriminative validity: this involves hypothesis testing about how the tool compares to another reference standard of good quality, or how the tools differs between groups known to be different. Criterion validity: an assessment of the degree the tool reflects a reference or criterion standard. The elements of this are described in more detail below separately for diagnosis (concurrent criterion), and prognosis (predictive criterion)

	Internal consistency: an assessment of the interrelatedness among scale items, and may be assessed by Cronbach's alpha. Reliability: an assessment of degree to which a the measurement is the same for patients who have not changed by different raters (inter-rater), and over time (test-retest). This may be assessed by interclass correlation (ICC), or kappa. Responsiveness: generally considered an assessment of the ability to detect important changes; multiple approaches to assessing responsiveness exist. Interpretability: ability to assign qualitative meaning with clinical or commonly understood connotations for a score or change in score.

Tools Used for Diagnosis	
For tools that are used for diagnosis, the items in this section should be charted. These items reflect important properties of diagnostic tools.	
Type of Diagnosis Study	▪ Screening or triage ▪ Diagnosis
Reference standard	List the measure that the frailty assessment tool evaluated against
Diagnostic accuracy measures (criterion/concurrent validity) *use if reports only association or correlation reported	How well does the tool identify a current health condition (frailty)? Some studies simply report on a tool-health condition association or correlation by comparing to another reference tools. This is the weakest measure of concurrent criterion validity. Was the outcome reported for the diagnostic performance an association or correlation only? ▪ Yes, univariable association (no other variables were included in the statistical model) ▪ Yes, multivariable association (other variables were included in the statistical model) ▪ Yes, correlation (e.g., Spearman's) ▪ No, other measurement properties are reported Studies reporting this will report only an effect estimate such as an odds ratio (OR), or a correlation coefficient.
Diagnostic accuracy measures (criterion/concurrent validity) *use if reports more than association or correlation reported	How well does the tool identify a current health condition (frailty)? If the study reports more than an association or correlation alone, list these measurement properties here. Examples include ▪ Sensitivity ▪ Specificity ▪ Positive predictive value/Negative predictive value ▪ Accuracy ▪ Likelihood ratio
*after initial data charting, if studies report on other measurement properties, we will create an updated data charting form to collect details of reported measurement properties	

Tools Used for Prognosis	
For tools that are used for prognosis, the items in this section should be charted. These items reflect important properties of prognostic tools.	
Multivariable analysis?	Is this tool reported as a single factor for prognosis, or added to other prognosis factors? If added to other factors, is this a previously known reference model. ▪ Single factor ▪ Added to other prognosis factors ▪ Added to reference model
Other prognosis factors	If tool is added to other factors, list them here If added to a reference model, list this here
Predicted outcome	Which outcome(s) has the frailty assessment tool reported to prognosticate/predict? Note: prognosis is the prediction of a future outcome.
Purpose of prognosis study	List the type of prognosis study. Prognosis studies of single factors that provide just effect estimates can be exploratory, confirmatory, or replication studies. Prognosis studies of a prognostic model (which are statistical models that provide estimates (proportion or percentage) for the likelihood of a given outcome), can be development, internal validation, external validation, or impact analyses. Other studies can assess a single factor against other known prognosis factors or a known reference model for incremental value.[34,35,51–53] ▪ Exploratory ▪ Confirmatory ▪ Development ▪ Internal Validation ▪ External Validation ▪ Impact analysis ▪ Incremental value If more than one type of study is reported in a single evidence source, list this as the highest type as per the order listed above. Exploratory: these studies aim to identify potential prognostic factors out of a group of collected variables. These studies usually do not focus on one specific factor and its prognostic performance. Studies of this kind may include frailty along with many other candidate variables, rather than focus on the performance of frailty specifically. Confirmatory: these studies are designed to evaluate the independent association of a specific prognostic factor in the context of other known prognostic factors. These are usually based on planned multivariable analyses including other known prognostic factors. Development & Internal Validation: these studies aim to develop a novel statistical model to predict future outcomes using a development dataset. Internal validation uses the same dataset that a model was developed on by reassessing the models performance on a component of the dataset or using resampling techniques.

	External validation: these studies assess the performance of a known model using a dataset not used during model development. These studies may also update, adjust, or recalibrate the model. Incremental value: these studies are a special type of confirmatory study that are designed to evaluate the incremental value of a specific prognostic factor by adding to a known reference model (prognostic model), or to a model generated from routinely available prognostic factors.
Prognostic or predictive accuracy (criterion/predictive validity) *use if reports association only	How well does the tool predict a future outcome? Some studies simply report on a tool-outcome association. This is the weakest measure of criterion/predictive validity. Was the outcome reported for the prognostic performance reported as an association only? ▪ Yes, univariable association (no other variables were included in the statistical model) ▪ Yes, multivariable association (other variables were included in the statistical model) ▪ No, other measurement properties are reported Studies reporting this will report only an effect estimate such as an odds ratio (OR).
Prognostic or predictive accuracy (criterion/predictive validity) *Use if reports more than association only	How well does the tool predict a future outcome? If the study reports more than an association or correlation alone with prognosis/prediction of an outcome, list these measurement properties here. Examples include ▪ Sensitivity, specificity, PPV, NPV, likelihood ratios ▪ Discrimination (Receiver operating curve, ROC, AUC, c statistic, concordance index) ▪ Calibration (calibration plot, Hosmer-Lemeshow) ▪ Overall performance (R^2, Brier score) ▪ Reclassification: Net Reclassification Index (NRI), integrated discrimination improvement (IDI) ▪ Clinical usefulness, clinical impact, impact analysis, clinical decision rule, Difference in net benefit (NB), decision curve analysis (DCA) ▪ Decision curve analysis (DCA)
*after initial data charting, if studies report on other measurement properties, we will create an updated data charting form to collect details of reported measurement properties	

Tools Used for Estimating Treatment Effect	
Type	What type of estimate is the frailty tool being used for ▪ Differential treatment effect (those with and without frailty have a different magnitude of benefit from a specific treatment) ▪ Guide a clinical decision for or against a treatment or treatment pathway
*after initial data charting, if studies report on other measurement properties, we will create an updated data charting form to collect details of reported measurement properties	

Useability Items	
Useability metric	List the useability metric(s) reported in the study. These may include: ▪ useability ▪ acceptability ▪ satisfaction (provider, patient) ▪ implementation ▪ availability ▪ language versions ▪ practicality (e.g., administration time, equipment needed) ▪ training ▪ cost List all that are reported. The definitions across studies may differ. List any additional metrics encountered even if not listed here.
*after initial data charting, if studies report on useability metrics, we will create an updated useability data charting form to collect definitions and outcomes of identified metrics	

Items for Studies Comparing Two or More Tools	
Tools included	List the frailty assessment tools
Method for comparison	List the method used for comparing the tools List the outcomes that were used to compare the tools
*after initial data charting, if studies report on comparisons, we will create an updated data charting form to collect additional items relevant to the comparison	