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 <u>systematic</u> 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
<u>Medline</u>
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

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJ Open

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.

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJ Open

	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, <u>list</u> 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	