An Introduction to Systematic Reviews of Prognosis

Katrina Williams and Carl Moons

for the Cochrane Prognosis Review Methods Group
(Co-convenors: Doug Altman, Katrina Williams, Jill Hayden, Sue Woolfenden, Richard Riley, Carl Moons)

An Introduction to Systematic Reviews of Prognosis

- Prognosis studies (background)
  - Definitions & approaches
- Discuss three types of prognosis questions
  - Average/overall prognosis: 'What is the most likely course of this health condition?‘
  - Prognostic factors: 'What factors are associated with, or determine outcome?‘
  - Prognostic (risk prediction) models: 'Are there risk groups who are likely to have different outcomes?‘
- Introduce methods for SR of prognosis studies

Definitions

- **Prognosis:** Probable course or prediction of specific outcome of a health condition over time
- **Prognosis studies:** Aim to understand the course, determinants, or probability of outcome in a cohort
- Use of prognostic information:
  - To provide information to patients
  - Identify target groups for treatment
  - To target specific prognostic factors for modification through treatment

From *Hayden et al., . Journal of Clinical Epidemiology. 2008;61(6):552-60.*

Conducting Prognostic Reviews

- Benefits of systematic reviews (SRs):
  - Summaries for evidence-based practice
  - Summaries of state-of-the-art in a field
- Primary studies
  - Quality of primary studies poor
  - Poor reporting
- SRs may improve primary studies

Challenges for Prognosis Reviews

- Inconsistent terminology
- Observational study designs
- Different types of studies

Types of Prognosis Questions

1. Average (overall) prognosis
2. Prognostic factor studies
3. Risk prediction modeling studies
4. Treatment effect modification studies

Not IPD MA; all on aggregate level

From *Hemmingway et al., in preparation*

Average Prognosis

- Estimate overall course of a health condition
- Baseline risk
- May be defined according to particular demographic and clinical context

The prognosis of depression in the elderly

Martin G. Cole, MD, FRCPC

Controversy prompted a systematic and critical review of original research articles to determine the prognosis of depression in the elderly. A search of three computer databases for articles published from January 1980 to May 1989 and of the bibliographies of articles located revealed 10 reports, involving 990 subjects, that met the following inclusion criteria: original research, published in English or French since 1950, sample of at least 25 patients, inclusion of only patients over 60 years, mean follow-up period of 1 year or more and description of the patients' mental state during follow-up. The methods and study designs were then assessed with the use of the criteria for prognostic studies established by McMaster University Health Sciences Centre, Hamilton, Ont. All of the studies had serious, multiple flaws. Nevertheless, when the results were combined it appeared that at least 60% of the patients remained well or had relapses with recovery (mean follow-up 31.9 months); up to 25% remained continuously ill. Physical illness, cognitive impairment and severe depressive symptoms were frequently related to poor prognosis; most of the social factors studied were not related to prognosis. Future studies must pay attention to methods and design, particularly the composition of study populations, the assessment of outcomes and the control of extraneous prognostic factors.

La controverse a donné lieu à un exame
une recherche originale, afin de détern

CAN MED ASSOC J 1990; 143:633

Table 1. Mortality in Mechanically Ventilated Stroke Patients

Source	Enrollment Dates	Mortality, No./Total (%)					
		Inpatient	30 d	3 mo	6 mo	1 y	≥2 y
Ischemic Stroke							
El Ad et al[21]	1980-1990	19/21 (90)					
Wijdicks and Scott[34]	1980-1995 (Basilar occlusion)	22/25 (88)					
Gujjar et al[22]	1994-1997	41/74 (55)					
Grotta et al[23]	NA (Carotid only)	14/20 (70)				16/20 (80)	
Leker and Ben Hur[24]	1992-1998		11/16 (69)				
Mayer et al[25]	1993-1996		10/20 (50)				
Bushnell et al[26]	1994-1997		19/41 (46)				
Berroushcot et al[27]	1994-1998 (Carotid only)		39/52 (75)	42/52 (81)			
Wijdicks and Scott[28]	1976-1994 (Carotid only)			17/24 (71)			
Foerch et al[29]	1998-2001 (>65 y old)				28/46 (61)		
Burtin et al[30]	1984-1989					69/79 (87)	
Santoli et al[31]	1990-1995					42/58 (72)	
Steiner et al[32]	1992-1993					58/84 (69)	
Schielke et al[33]	1996-1999					60/101 (59)	77/101 (76)
Intracerebral Hemorrhage							
El Ad et al[21]	1980-1990	28/32 (88)					
Gujjar et al[22]	1994-1997	90/156 (58)					
Roch et al[35]	1997-1999	58/120 (48)					82/120 (68)
Mayer et al[25]	1993-1996		17/24 (71)				
Bushnell et al[26]	1994-1997		45/90 (50)				
Foerch et al[29]	1998-2001 (>65 y old)				11/19 (42)		
Lessire et al[37]	1990-1994					19/26 (73)	
Burtin et al[30]	1984-1989					97/100 (97)	
Steiner et al[32]	1992-1993					21/30 (70)	
Mean mortality, total No. (%)*		203/370 (55)	141/243 (58)	98/141 (69)		375/549 (68)	

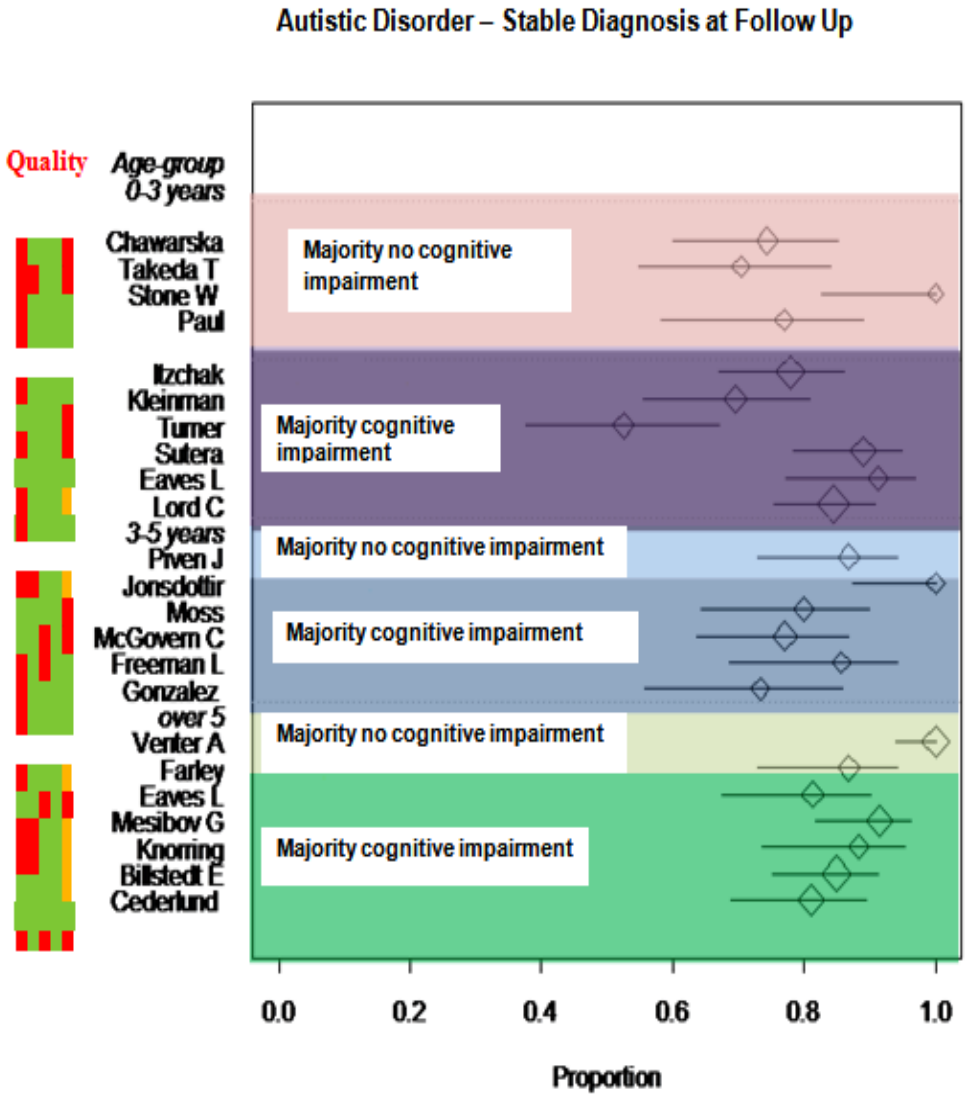
Abbreviation: NA, not available.
*Excludes basilar occlusion study[34] and studies that

Holloway et al. JAMA. 2005;294:725-733

Autism

- Clinicians need evidence based advice when they are counselling families and children with autism about:
  - diagnostic stability
  - future risks caused by their condition
    - epilepsy
    - mortality

Autistic Disorder – Stable Diagnosis at Follow Up

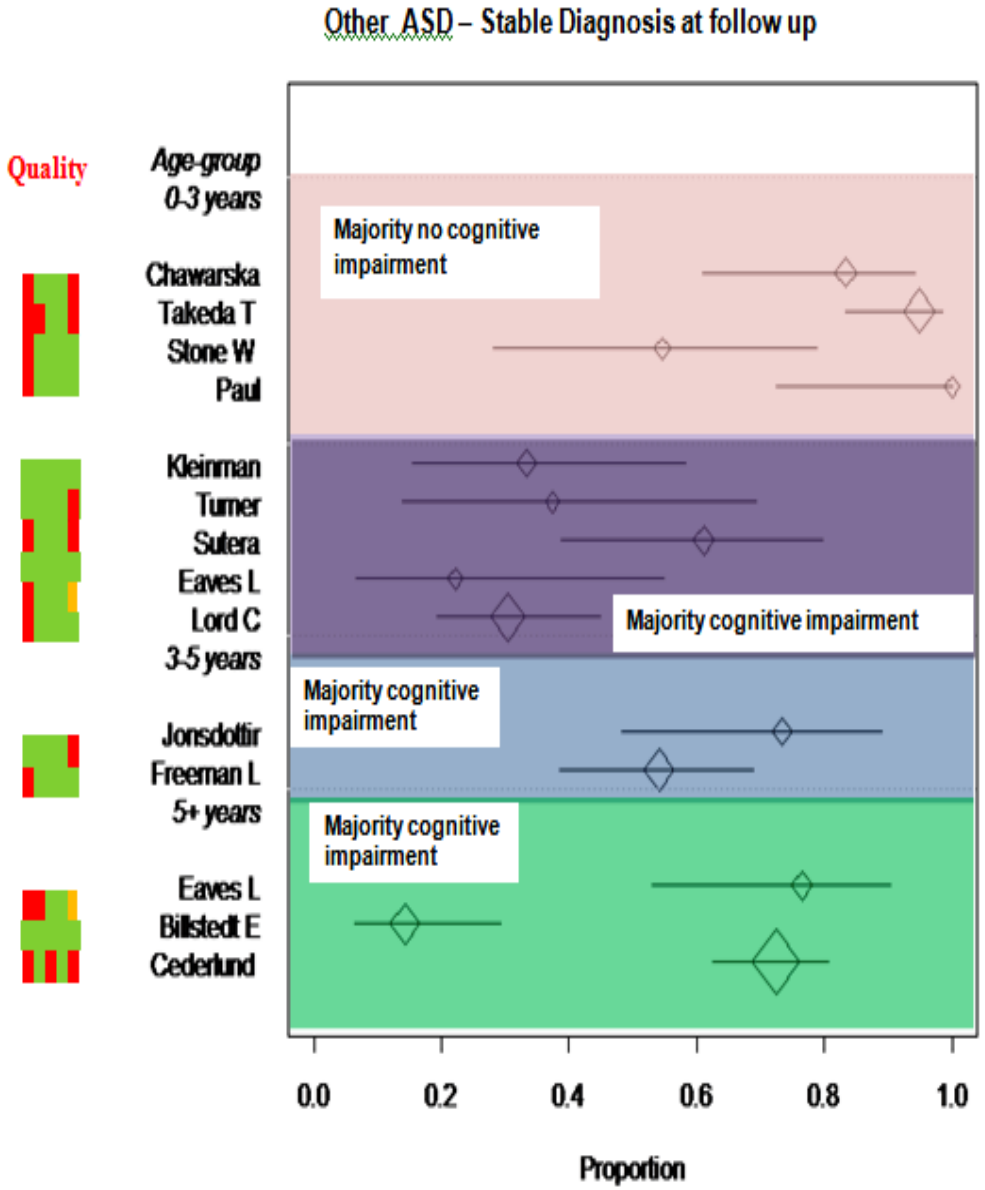


- **Baseline Mean Age 0-3-**
- **53 to 100%** of children still had an Autistic Disorder diagnosis at follow up
- **Lowest risk study - 89% still had AD diagnosis**

- **Baseline mean age of 3-5 years**
- **73% to 100%** of children still had an Autistic Disorder diagnosis
- **Lowest risk study 100% still had AD diagnosis**

- **Baseline mean age over 5 years**
- **81% to 100%** of children still had an Autistic Disorder diagnosis at follow up.
- **Lowest risk study 88 % still had AD diagnosis**

Stability of other ASD diagnosis

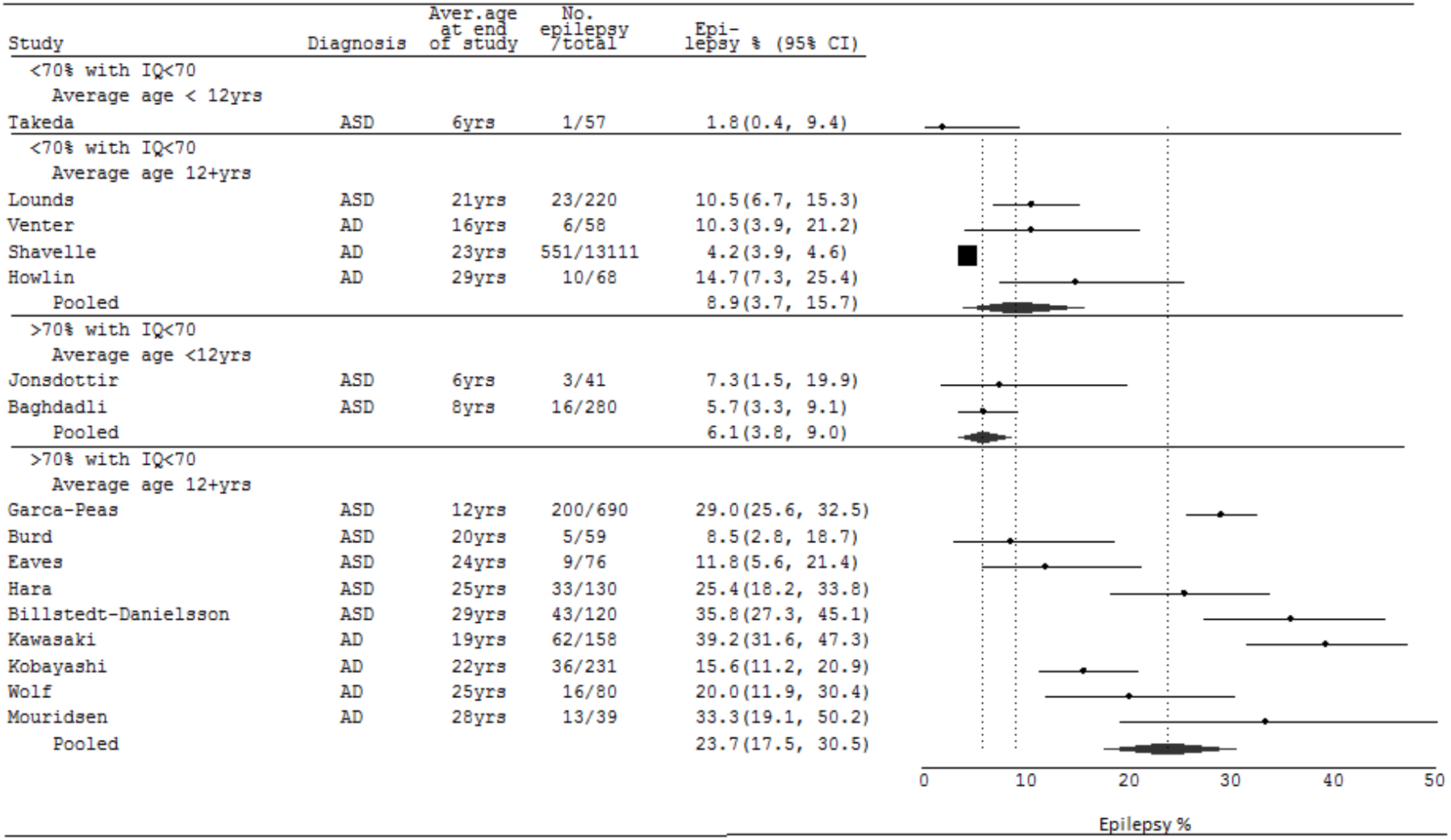


- **Baseline Mean Age 0-3-**
- **22% to 100%** of children still had an other ASD diagnosis at follow up
- **Lowest risk study 33% still had other ASD diagnosis**

- **Baseline mean age of 3-5 years**
- **54% to 73%** of children still had an other ASD diagnosis
- **Both studies had same risk of bias**

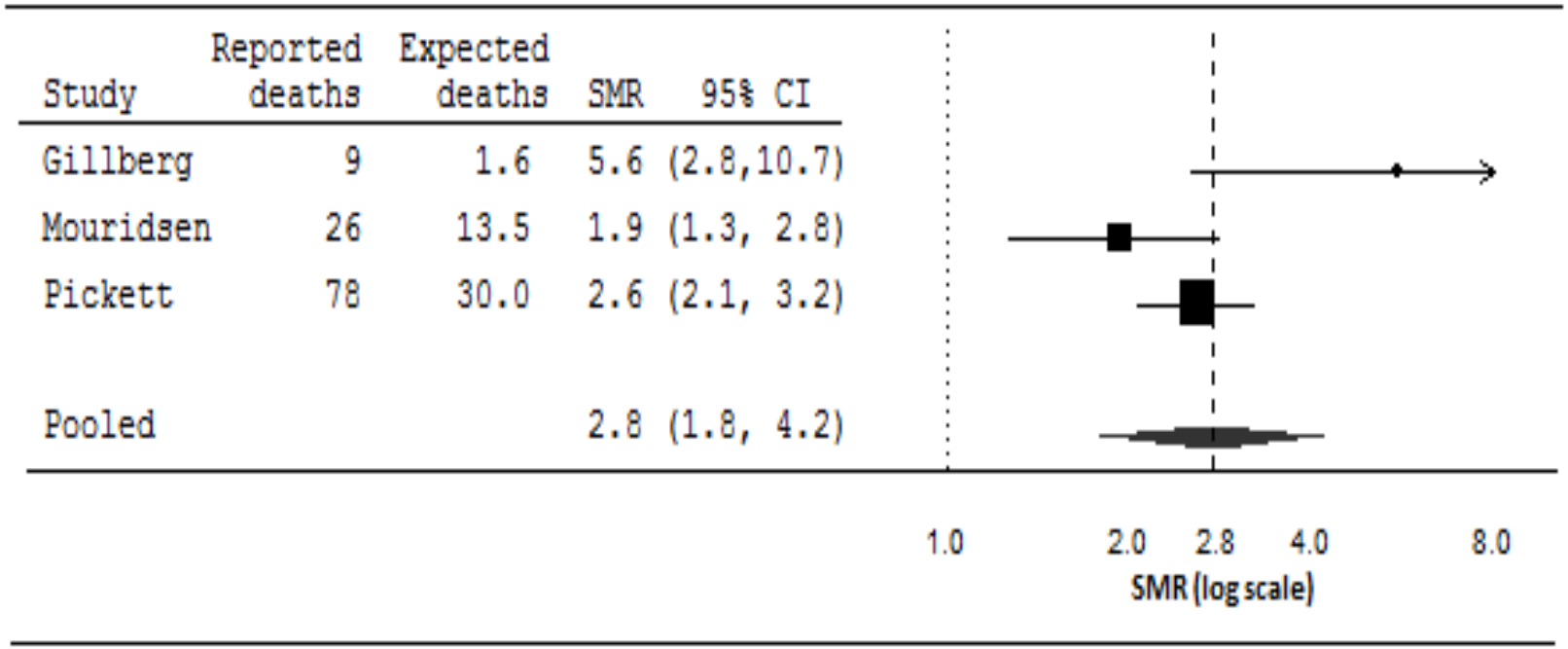
- **Baseline mean age over 5 years**
- **14% to 76%** of children still had an other ASD diagnosis
- **Lowest risk study 14 % still had other ASD diagnosis**

Pooled estimates of epilepsy percentage stratified by IQ and age and ordered by diagnosis and age



Size of data markers is proportional to sample size. The pooled diamond data markers indicate DerSimmonian-Laird pooled estimates of the epilepsy percentage for particular subgroups.

Pooled estimates of Standardised Mortality Ratio (SMR)



I^2=69.1%. Cochran's Q=6.5, degrees-of-freedom=2, p=0.04

Prognostic Factor Studies

- Focus on prognostic factors
- Aim:
  - To identify specific factors associated with subsequent outcome of a health condition
  - Not necessarily sick (patients)
- Prognostic factor evidence may:
  - help identify targets for new interventions that aim to modify the course of a disease
  - enhance the design & analysis of intervention trials
- Understanding course of the disease

PERGAMON

European Journal of Cancer 39 (2003) 19–30

www.ejconline.com

Review

A systematic review of molecular and biological markers in tumours of the Ewing's sarcoma family

R.D. Riley[a,*], S.A. Burchill[b], K.R. Abrams[a], D. Heney[c], A.J. Sutton[a], D.R. Jones[a], P.C. Lambert[a], B. Young[a], A.J. Wailoo[d], I.J. Lewis[e]

[a]*Department of Epidemiology and Public Health, University of Leicester, 22–28 Princess Road West, Leicester, LE1 6TP, UK*
[b]*Cancer Research UK Clinical Centre, St James's University Hospital, Beckett Street, Leeds, LS9 7TF, UK*
[c]*Department of Medical Education, University of Leicester, University Road, Leicester, LE1 9HN, UK*
[d]*School of Health and Related Research, University of Sheffield, Regent Street, Sheffield, S1 4DA, UK*
[e]*Department of Paediatric Oncology, St James's University Hospital, Beckett Street, Leeds, LS9 7TF, UK*

Received 10 April 2002; received in revised form 19 August 2002; accepted 6 September 2002

Abstract

The aims of this study were to perform the first systematic review of molecular and biological tumour markers in tumours of the Ewing's sarcoma family (ESFT), and evaluate the current evidence for their clinical use. A well-defined, reproducible search strategy was used to identify the relevant literature from 1966 to February 2000. Papers were independently assessed for tumour markers used in the screening, diagnosis, prognosis or monitoring of patients with ESFT. Eighty-four papers studying the use of 70 different tumour markers in ESFT's were identified. Low-quality, inconsistent reporting limited meta-analysis to that of prognostic data for 28 markers. Patients with tumours lacking S-100 protein expression have a better overall survival (OS) (hazard ratio (HR) = 0.41, 95% confidence interval (CI) 0.19, 0.89) than those with expression; patients with high levels of serum LDH had a worse OS and disease-free survival (DFS) (OS: HR = 2.92, CI 2.16, 3.94, DFS: HR = 3.38, 95% CI 2.28, 4.99); patients with localised disease and tumours expressing type 1 *EWS-FLI1* fusion transcripts had an improved DFS compared with those with other fusion transcript types (HR = 0.17, 95% CI 0.079, 0.37). The knowledge base formed should facilitate more informative future research. Improved statistical reporting and large, multicentre prospective studies are advocated.

Serum Levels of Phosphorus, Parathyroid Hormone, and Calcium and Risks of Death and Cardiovascular Disease in Individuals With Chronic Kidney Disease

A Systematic Review and Meta-analysis

Suetonia C. Palmer, MB ChB, PhD
Andrew Hayen, PhD
Petra Macaskill, PhD
Fabio Pellegrini, MSc
Jonathan C. Craig, MB ChB, PhD
Grahame J. Elder, MB BS, PhD
Giovanni F. M. Strippoli, MD, PhD

Context Clinical practice guidelines on the management of mineral and bone disorders due to chronic kidney disease recommend specific treatment target levels for serum phosphorus, parathyroid hormone, and calcium.

Objective To ass
serum phosphorus
cular mortality, an
disease.

Data Sources Th
(1947 to Decembe
also were conducte
cal guidelines alon

Study Selection
ies (N=327 644 pa

Data Extraction
covariates togethe
ity, and nonfatal cardiovascular events at different levels of serum phosphorus, parathyroid hormone, and calcium were analyzed within studies. Data were summarized across studies (when possible) using random-effects meta-regression.

Data Synthesis The risk of death increased 18% for every 1-mg/dL increase in serum phosphorus (relative risk [RR], 1.18 [95% confidence interval {CI}, 1.12-1.25]). There was no significant association between all-cause mortality and serum level of parathyroid hormone (RR per 100-pg/mL increase, 1.01 [95% CI, 1.00-1.02]) or serum level of calcium (RR per 1-mg/dL increase, 1.08 [95% CI, 1.00-1.16]). Data for the association between serum level of phosphorus, parathyroid hormone, and calcium and cardiovascular death were
justment for con

Conclusions The evidentiary basis for a strong, consistent, and independent associa-

Objective: To assess the quality of evidence for the association between levels of serum phosphorus, parathyroid hormone, and calcium and risks of poor outcomes in individuals with chronic kidney disease.

Palmer et al., JAMA. 2011;305(11):1119-1127

CLINICAL PRACTICE GUIDELINES are a powerful influence on management strategies for common conditions. When evidence that treating risk factors reduces disease and improves health outcomes is provided by large and well-conducted randomized controlled trials, guidelines can appropriately summarize the relevant data and widely disseminate recommendations for best practice. However, when practice guidelines promote therapeutic strategies without sufficient evidence of effectiveness or harms, overtreatment and widespread inappropriate use of medications, services, or devices may occur. Accordingly, guidelines may rec-

Quality of Reporting of Cancer Prognostic Marker Studies: Association With Reported Prognostic Effect

Panayiotis A. Kyzas, Despina Denaxa-Kyza, John P. A. Ioannidis

Background Issues of reported study quality have not been addressed empirically with large-scale data in the cancer prognostic literature.

Methods Eight quality measures pertaining to study design and assay methods (i.e., blinding, prospective versus retrospective design, power calculations, outcomes' definitions, time of enrollment, reporting of variables, assay description, and assay reference) were evaluated in cancer prognostic marker studies included in meta-analyses identified in Medline and EMBASE. To be eligible, meta-analyses had to include at least six studies and to examine binary outcomes. We estimated the ratios of relative risks, which compared the overall prognostic effects (summary relative risks) between poor-quality and good-quality studies for each quality item. Between-study heterogeneity was tested with the *Q* statistic (statistically significant at $P<.10$). All statistical tests were two-sided.

Results We identified 20 meta-analyses that included 331 cancer prognostic marker studies published between 1987 and 2005. Only three (0.9%) of the 331 studies presented power calculations, 129 (39.0%) studies stated that analyses were blinded, and 73 (21.5%) stated that they were prospective. Time of enrollment was defined in 232 (70.0%), 234 (70.7%) gave lists of candidate variables, and 254 (76.7%) defined outcomes. The assay used was described in 317 (95.8%), but only 177 (53.5%) provided the assay reference. Estimates of prognostic effects from poor-quality studies varied considerably and could be larger or smaller than summary estimates derived from meta-analyses. Summary ratios of relative risks of poor- versus good-quality studies for the seven quality measures ranged from 0.95 to but 1.26, but none was statistically significantly. There was statistically significant heterogeneity ($P<.10$) between the ratios of relative risk estimates across meta-analyses for blinding, defining endpoints, and stating variables and assay references.

Conclusions Among cancer prognostic marker studies, reporting quality of design and assay information often appears suboptimal, indicating that this literature may be largely unreliable. Given the potential clinical importance of prognostic marker information, improved design and reporting of these studies are warranted.

Prognostic (risk prediction) models

1. Developing a prognostic model
2. Validate the model in other subjects
   - not necessarily patients
3. Update existing model to local situation
4. Quantify model's impact on doctor's decision making and patient outcome (cost-effectiveness)

BMC Medical Informatics and Decision Making

Research article

Systematic review of prognostic models in traumatic brain injury

Pablo Perel*, Phil Edwards, Reinhard Wentz and Ian Roberts

Address: Nutrition and Public Health Intervention Research Unit, Epidemiology and Population Health Department, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Email: Pablo Perel* - pablo.perel@lshtm.ac.uk; Phil Edwards - phil.edwards@lshtm.ac.uk; Reinhard Wentz - reinhard.wentz@lshtm.ac.uk; Ian Roberts - ian.roberts@lshtm.ac.uk

* Corresponding author

Published: 14 November 2006

BMC Medical Informatics and Decision Making 2006, **6**:38 doi:10.1186/1472-6947-6-38

Received: 03 August 2006
Accepted: 14 November 2006

This article is available from: http://www.biomedcentral.com/1472-6947/6/38

Abstract

Background: Traumatic brain injury (TBI) is a leading cause of death and disability world-wide. The ability to accurately predict patient outcome after TBI has an important role in clinical practice

Collins *et al. BMC Medicine* 2011, **9**:103
http://www.biomedcentral.com/1741-7015/9/103

BMC Medicine

RESEARCH ARTICLE Open Access

Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting

Gary S Collins*, Susan Mallett, Omar Omar and Ly-Mee Yu

Abstract

Background: The World Health Organisation estimates that by 2030 there will be approximately 350 million people with type 2 diabetes. Associated with renal complications, heart disease, stroke and peripheral vascular disease, early identification of patients with undiagnosed type 2 diabetes or those at an increased risk of developing type 2 diabetes is an important challenge. We sought to systematically review and critically assess the conduct and reporting of methods used to develop risk prediction models for predicting the risk of having undiagnosed (prevalent) or future risk of developing (incident) type 2 diabetes in adults.

Methods: We conducted a systematic search of PubMed and EMBASE databases to identify studies published before May 2011 that describe the development of models combining two or more variables to predict the risk of prevalent or incident type 2 diabetes. We extracted key information that describes aspects of developing a prediction model including study design, sample size and number of events, outcome definition, risk predictor selection and coding, missing data, model-building strategies and aspects of performance.

Results: Thirty-nine studies comprising 43 risk prediction models were included. Seventeen studies (44%) reported the development of models to predict incident type 2 diabetes, whilst 15 studies (38%) described the derivation of models to predict prevalent type 2 diabetes. In nine studies (23%), the number of events per variable was less than ten, whilst in fourteen studies there was insufficient information reported for this measure to be calculated. The

Heart Online First, published on December 18, 2011 as 10.1136/heartjnl-2011-300734

Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: a systematic review

S van Dieren,[1] J W J Beulens,[1] A P Kengne,[1,2,3] L M Peelen,[1] G E H M Rutten,[1] M Woodward,[3] Y T van der Schouw,[1] K G M Moons[1]

ABSTRACT

Context A recent overview of all CVD models applicable to diabetes patients is not available.

Objective To review the primary prevention studies that focused on the development, validation and impact assessment of a cardiovascular risk model, scores or rules that can be applied to patients with type 2 diabetes.

Design Systematic review.

Data sources Medline was searched from 1966 to 1 April 2011.

Study selection A study was eligible when it described the development, validation or impact assessment of a model that was constructed to predict the occurrence of cardiovascular disease in people with type 2 diabetes, or when the model was designed for use in the general

nation (ability to discriminate between patients who will get the disease and those who will not) and calibration (ability to correctly quantify the absolute risk), but the outcomes have varied widely.[7–9]

A systematic review by Chamnan *et al*[6] provides an overview of CVD prediction models that have been developed in diabetes populations, and prediction models for the general population that have been validated in a diabetes population. However, new prediction models for the diabetes population have been developed since this review, and many more prediction models exist that can be applied to people with diabetes. Moreover, it is unknown whether applying a certain prediction model in clinical practice affects the treatment of patients with diabetes

ppendices are
ne only. To view
ase visit the
(http://heart.bmj.
arly/recent).

for Health
Primary Care,
dical Center
ht, The

Medical
ncil & University of
ape town, South

nstitute for Global
sity of Sydney,
alia

SR: discuss 3 main issues

- 1. Review question
- 2. Search strategy
- 3. Data extraction and Critical appraisal (Risk of Bias)

Importance to the Review Question

- Clearly framed question will guide the review
- Guide the reader:
  - Initial assessment of relevance
- Guide the reviewer on how to:
  - Collect studies
  - Check eligibility of studies
  - Conduct the analysis

Issues to consider to frame review question (1)

- Population
  - Disease or health condition of interest
  - Setting
  - Special population defined by specifc factor(s) or characteristic(s)
- Factors
  - Definition of subgroups (OP); prognostic factor ; risk prediction model of interest
- Outcome
  - Type of scale
  - Timing of measurement

Issues to consider to frame review question (2)

- Clearly specify review objective
  - Describe average (overall) prognosis
  - Strength of prognostic factor(s)
  - Accuracy of risk prediction models
- Define the scope of review
  - Broad systematic reviews
    - All prognostic factors/risk prediction models
    - Good overview of topic area (exploratory)
    - May be diffiicult to synthesize
  - Focused systematic reviews
    - One prognostic factor/domain or risk prediction model
    - May allow more thorough assessment/interpretation of evidence

Item	Comments and examples
1. Type of prognosis studies (overall prognosis, prong factor studies, prog model studies)	Focus on studies addressing overall prognosis; prognostic factors; model development, model validation or combination.
2. Target population to whom overall prognosis, prognostic factor(s), or prognostic model under review may apply	•Overall survival of women diagnosed with breast cancer; •prognostic factors for adult healthy women in the general population to predict the life time risk of developing breast cancer; •prognostic models for predicting the risk of postoperative 30-day mortality in patients that underwent cardiac surgery.
3. Outcome (endpoint)	Which endpoint/outcome; e.g. all cause mortality, cause-specific mortality, combined events.
4. Type of prognostic factors or models under study and timing of their measurement	Focus can be on specific type or timing of measurement of predictor(s). - focusing on the added value of CRP measurement to traditional predictors in prediction of future CVD events; - models to predict 30-day postoperative complications using only predictors measured preoperatively (and not intra- or post-operatively)
5. Intended use or purpose of prognostic factor or model	e.g. on factors or model predicting the risk of future events to direct individual's behaviour or life style changes ; or physicians' therapeutic decision making ; or more targeted (enriched) RCT designs.

2. Searching and selecting

- Electronic search
  - Include at least MEDLINE & EMBASE, plus any topic-specific databases
  - Available search strategies (plus content area terms)
    - PubMed Broad search filter for PF SRs
    - (Updated) Ingui strategy for RPMs
  - Caution using exposure-related terms
- Supplement the search!
  - Bibliography screening, citation-tracking, hand-searching relevant journals, content experts
- Search strategies should be made available to readers
- Report complete citation/study flow chart

Further research needed (better indexing & more sensitive strategies)

OPEN ACCESS Freely available online

PLoS one

Search Filters for Finding Prognostic and Diagnostic Prediction Studies in Medline to Enhance Systematic Reviews

Geert-Jan Geersing[1]*, Walter Bouwmeester[1], Peter Zuithoff[1], Rene Spijker[2,4], Mariska Leeflang[3,4], Karel Moons[1]

1 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands, **2** Medical Library Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, **3** Department of Clinical Epidemiology and Bio-Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, **4** Dutch Cochrane Center, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

Abstract

Background: The interest in prognostic reviews is increasing, but to properly review existing evidence an accurate search filer for finding prediction research is needed. The aim of this paper was to validate and update two previously introduced search filters for finding prediction research in Medline: the Ingui filter and the Haynes Broad filter.

Methodology/Principal Findings: Based on a hand search of 6 general journals in 2008 we constructed two sets of papers. Set 1 consisted of prediction research papers (n = 71), and set 2 consisted of the remaining papers (n = 1133). Both search filters were validated in two ways, using diagnostic accuracy measures as performance measures. First, we compared studies in set 1 (reference) with studies retrieved by the search strategies as applied in Medline. Second, we compared studies from 4 published systematic reviews (reference) with studies retrieved by the search filter as applied in Medline. Next – using word frequency methods – we constructed an additional search string for finding prediction research. Both search filters were good in identifying clinical prediction models: sensitivity ranged from 0.94 to 1.0 using our hand search as reference, and 0.78 to 0.89 using the systematic reviews as reference. This latter performance measure even increased to around 0.95 (range 0.90 to 0.97) when either search filter was combined with the additional string that we developed. Retrieval rate of explorative prediction research was poor, both using our hand search or our systematic review as reference, and even combined with our additional search string: sensitivity ranged from 0.44 to 0.85.

Conclusions/Significance: Explorative prediction research is difficult to find in Medline, using any of the currently available search filters. Yet, application of either the Ingui filter or the Haynes broad filter results in a very low number missed clinical

Study Selection

- Complete (reproducible) definitions of:
  - Populations
  - Prognostic factor(s), models
  - Outcomes of interest
- Studies excluded clearly described:
  - Reasons for exclusions (summary)
  - Studies where eligibility may be unclear should be listed
- Design-related selection criteria applied and presented separately

3. critical appraisal

Recommendations for Intervention Reviews

- Risk of bias assessment
- Domain-based evaluation
  - Assessments are made separately for different bias domains
  - Judgment
- Use of scales “explicitly discouraged”
  - Weighting of items difficult to justify
  - Often based on reporting rather than conduct
  - Less transparent

Critical Appraisal: Prognosis Reviews

- Similar approach (consider potential biases)
- Different study types require different considerations:
  - Studies of overall prognosis
  - Prognostic factor studies
  - Risk prediction modeling studies
- Some similarities; start with PF studies

Ad. RoB Prognostic Factor Studies

- Similar to RCT RoB: Domain-based
- Reflect basic epidemiological principles
- Potential biases related to:
  1. Study participation
  2. Study attrition
  3. Prognostic factor measurement
  4. Outcome measurement
  5. Covariate measurement and account
  6. Analysis and reporting

Annals of Internal Medicine. 2006;144:427-437

Assessing Bias: The QUIPS* Instrument

Journal and year of publication				
Study identifier				
First author				
Reviewer				
Biases	**Issues to consider for judging overall rating of "Risk of bias"**	**Study Methods & Comments**	**Rating of reporting**	**Rating of "Risk of bias"**
Assess the risk of each potential bias	These issues will guide your thinking and judgment about the overall risk of bias within each of the 6 domains. Some 'issues' may not be relevant to the specific study or the review research question. These issues are taken together to inform the overall judgment of potential bias for each of the 6 domains.	Provide comments or text exerpts, as necessary, to facilitate the consensus process that will follow.	Click on each of the light yellow cells and choose from the drop down menu to rate the adequacy of reporting as yes, partial, no or unsure.	Click on the yellow cells; choose from the drop-down menu to rate potential risk of bias for each of the 6 domains as High, Moderate, or Low considering all relevant issues
1. Study Participation	**Goal: To judge the risk of selection bias (likelihood that relationship between *joint damage* and *long-term disability* is different for participants and eligible non-**			
Source of target population	The source population or population of interest is adequately described for key characteristics.			
Method used to identify population	The sampling frame and recruitment are adequately described, including methods to identify the sample sufficient to limit potential bias (number and type used, e.g., referral patterns in health care)			
Recruitment period	Period of recruitment is adequately described			
Place of recruitment	Place of recruitment (setting and geographic location) are adequately described			
Inclusion and exclusion criteria	Inclusion and exclusion criteria are adequately described (e.g., including explicit diagnostic criteria or "zero time" description).			
Adequate study participation	There is adequate participation in the study by eligible individuals			
Baseline characteristics	The baseline study sample (i.e., individuals entering the study) is adequately described for key characteristics.			
Summary Study participation	**The study sample represents the population of interest (adults with early RA) on key characteristics, sufficient to limit potential bias of the observed relationship between joint damage and long-term disability.**			
2. Study Attrition	**Goal: To judge the risk of attrition bias (likelihood that relationship between *joint damage* and *long-term disability* are different for completing and non-completing participants).**			
Proportion of baseline sample available for analysis	Response rate (i.e., proportion of study sample completing the study and providing outcome data) population) is adequate.			
Attempts to collect information on participants who dropped out	Attempts to collect information on participants who dropped out of the study are described.			
Reasons and potential impact of subjects lost to follow-up	Reasons for loss to follow-up are provided.			
Outcome and prognostic factor information on those lost to follow-up	Participants lost to follow-up are adequately described for key characteristics.			
	There are no important differences between key characteristics and outcomes in participants who completed the study and those who did not.			
Study Attrition Summary	**Loss to follow-up (from baseline sample to study population analyzed) is not associated with key characteristics (i.e., the study data adequately represent the sample) sufficient to limit potential bias to the observed relationship between joint damage and long-term disability.**			

QUIPS= **QU**ality **I**n **P**rognosis **S**tudies;

Adapted from: *Hayden JA. Methodological issues in systematic reviews of prognosis and prognostic factors: Low back pain. Doctoral Thesis, University of Toronto. 2007.*

Assessing Bias: The QUIPS Instrument

Page 2

3. Prognostic Factor Measurement	**Goal: To judge the risk of measurement bias related to how joint damage was measured (differential measurement of joint damage related to the level of long-term disability).**			
Definition of the PF	A clear definition or description of 'joint damage' is provided (e.g., including dose, level, duration of exposure, and clear specification of the method of measurement).			
Valid and Reliable Measurement of PF	Method of joint damage measurement is adequately valid and reliable to limit misclassification bias (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and limited reliance on recall)			
	Continuous variables are reported or appropriate cut-points (i.e., not data-dependent) are used.			
Method and Setting of PF Measurement	The method and setting of measurement of joint damage is the same for all study participants.			
Proportion of data on PF available for analysis	Adequate proportion of the study sample has complete data for joint damage variable.			
Method used for missing data	Appropriate methods of imputation are used for missing 'joint damage' data.			
PF Measurement Summary	***Joint damage* is adequately measured in study participants to sufficiently limit potential bias.**			
4. Outcome Measurement	**Goal: To judge the risk of bias related to the measurement of long-term disability (differential measurement of disability related to the baseline level of joint damage).**			
Definition of the Outcome	A clear definition of disability is provided, including duration of follow-up and level and extent of the outcome construct.			
Valid and Reliable Measurement of Outcome	The method of disability measurement used is adequately valid and reliable to limit misclassification bias (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and confirmation of outcome with valid and reliable test)			
Method and Setting of Outcome Measurement	The method and setting of disability measurement is the same for all study participants.			
Outcome Measurement Summary	***Long-term disability* is adequately measured in study participants to sufficiently limit potential bias.**			
5. Study Confounding	**Goal: To judge the risk of bias due to confounding (i.e. the effect of joint damage is distorted by another factor that is related to joint damage and long-term disability).**			
Important Confounders Measured	All important confounders, including treatments (key variables in conceptual model), are measured.			
Definition of the confounding factor	Clear definitions of the important confounders measured are provided (e.g., including dose, level, and duration of exposures).			
Valid and Reliable Measurement of Confounders	Measurement of all important confounders is adequately valid and reliable (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and limited reliance on recall).			
Method and Setting of Confounding Measurement	The method and setting of confounding measurement are the same for all study participants.			
Method used for missing data	Appropriate methods are used if imputation is used for missing confounder data.			
Appropriate Accounting for Confounding	Important potential confounders are accounted for in the study design (e.g., matching for key variables, stratification, or initial assembly of comparable groups).			
	Important potential confounders are accounted for in the analysis (i.e., appropriate adjustment).			
Study Confounding Summary	**Important potential confounders are appropriately accounted for, limiting potential bias with respect to the relationship between *joint damage* and *long-term disability*.**			
6. Statistical Analysis and Presentation	**Goal: To judge the risk of bias related to the statistical analysis and presentation of results.**			
Presentation of analytical strategy	There is sufficient presentation of data to assess the adequacy of the analysis.			
Model development strategy	The strategy for model building (i.e., inclusion of variables in the statistical model) is appropriate and is based on a conceptual framework or model.			
	The selected statistical model is adequate for the design of the study.			
Reporting of results	There is no selective reporting of results.			
Statistical Analysis and Presentation Summary	**The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid or spurious results.**			

Ad. RoB Overall Prognosis Studies

- Sample selection
- Recruitment method,
- Completeness of follow-up
- Timing of diagnosis
- Blinding
- Analysis for covariates was not assessed as we were not investigating predictors of outcomes.

(adapted from Hayden et al. 2006)

Table 2 Criteria used for Risk of Bias assessment

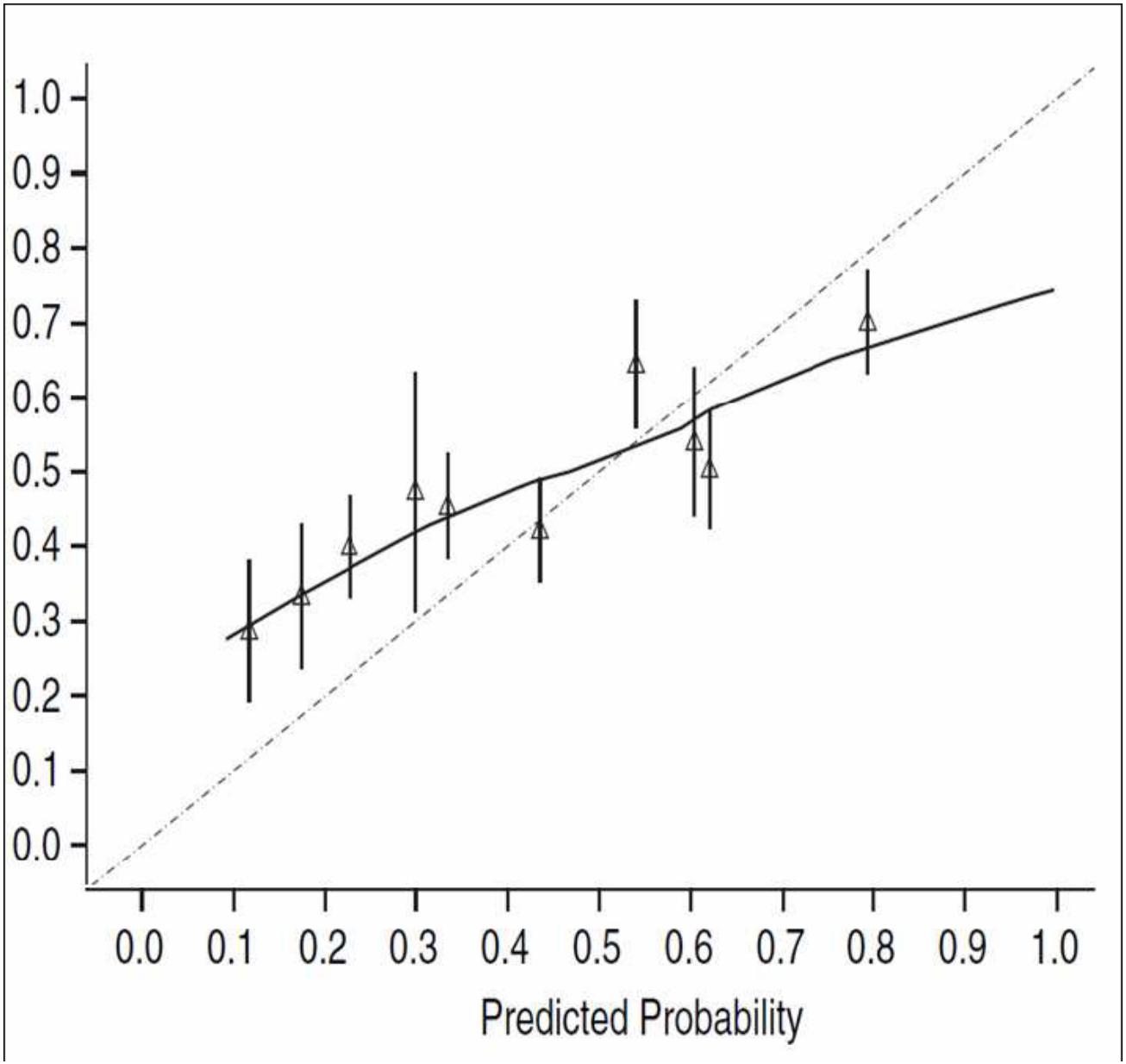
Criteria	High risk of Bias	Low risk of Bias	Unclear
Sample	Clinical	Population	Unclear
Recruitment	Retrospective	Prospective	Unclear
Follow Up	<80%	≥ 80%	Unclear
Timing of Diagnosis	At the conclusion of the study	At baseline or before recruitment to study	Unclear
Blinding	Not blinded	Blinding adequate	Unclear

Risk of Bias -at a glance

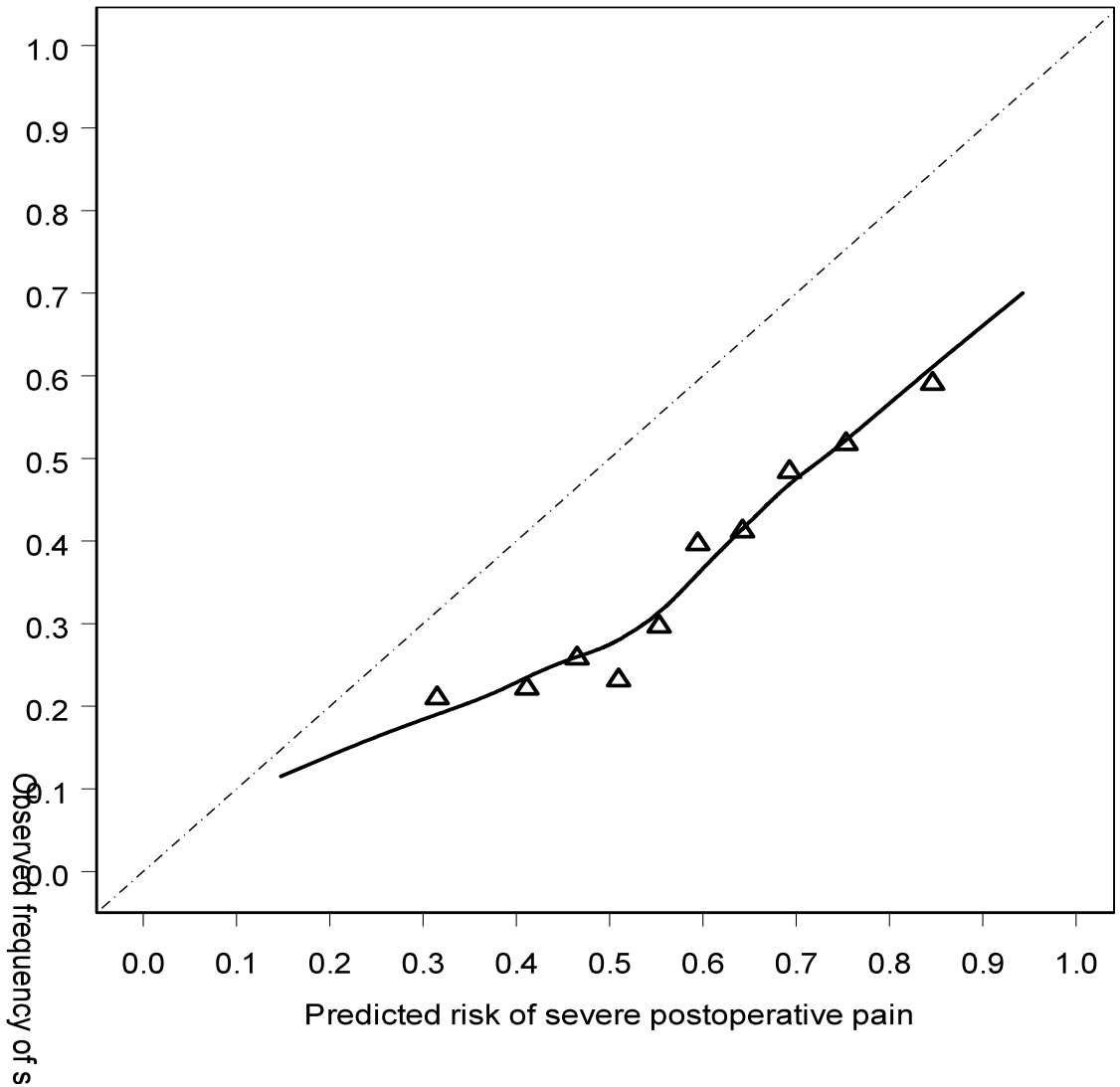
Author	year	Sample	Recruitment	Follow Up	Timing of Diagnosis	Blinding
Sutera	2006					
Billstedt	2005					
Knorring	1993					
Jonsdottir	2006					
Kleinman	2008					
Itzchak	2009					
Stone	2003					
Lord 2006						
Gonzalez	1993					
Paul	2008					
Freeman	2004					
Eaves	2004					
Venter	1992					
Chawarska	2009					
Turner	2007					
Mc Govern	2005					
Cederlund	2008					
Takeda	2005					
Eaves	1996					
Piven	1996					
Mesibov	1989					

Ad. RoB Prognostic Model

- Largely same as prognostic factor studies
  - Some specifics of course
- Bias in prognostic model development exhibited in:
  - Wrong relative risks (predictor weights)
  - Wrong intercept
  - Overfitted models (too large ROC area, too optimistic calibration plot or outcome classification)
- Unfortunately: often don't know from development study → until model validation



- Slope plot < 1.0
  - Low prob too low
  - High prob too high



Bias &/or applicability problem

Domain	Applic-ability Bias	Items
Source of data (study design)	A/B	• Source of data or study design (e.g. cohort, case-control, existing registry, randomized trial participants) • Key study dates (e.g. start and end of accrual, and of follow-up if applicable)
Participants	A	• Participant eligibility criteria (e.g. suspected of having breast cancer; having rheumatoid arthritis; undergone a particular type of surgery) • Setting (e.g. primary care, secondary care, general population) and geographical location • Single or multicenter (if so, number of centres) • Received treatments (if applicable)
Outcome (event or target condition) to be predicted	A/B	• Outcome definition and method of assessment • Outcome assessed without knowledge of (blinded for) the candidate predictors or index test results • Candidate predictors or index test results part of outcome • Patient relevant outcome (rather than process outcomes such as length of hospital stay or duration of surgery) • How was outcome analysed (e.g. continuous, categorized (at which cut-off points), binary, count, time to event)

Domain	Applic-ability Bias	Items
Candidate predictors (or index test results) under study	A/B	• Number and which candidate predictors (e.g. from patient history, physical examination, additional tests including genetic testing), and measurement methods (if applicable) • How were they selected (e.g. explicit systematic review, cited clinical judgement/relevance) • Time of predictor measurement (e.g. at presentation, at diagnosis, at treatment initiation) • Predictors assessed blinded for outcome and each other (if relevant) • How were they analysed (e.g. continuous, linear, non-linear transformations, categorized (at which cut-off points) • Predictor interactions tested or specific subgroups analysed
Sample Size	B	• Number of participants and outcome events (in relation to the number of candidate predictors)
Missing data	B	• Number participants with any missing values or complete data, and which variables (predictors and outcomes, i.e. including loss-to-follow-up) • Handling missing data (complete-case analysis, indicator method, single or multiple imputation

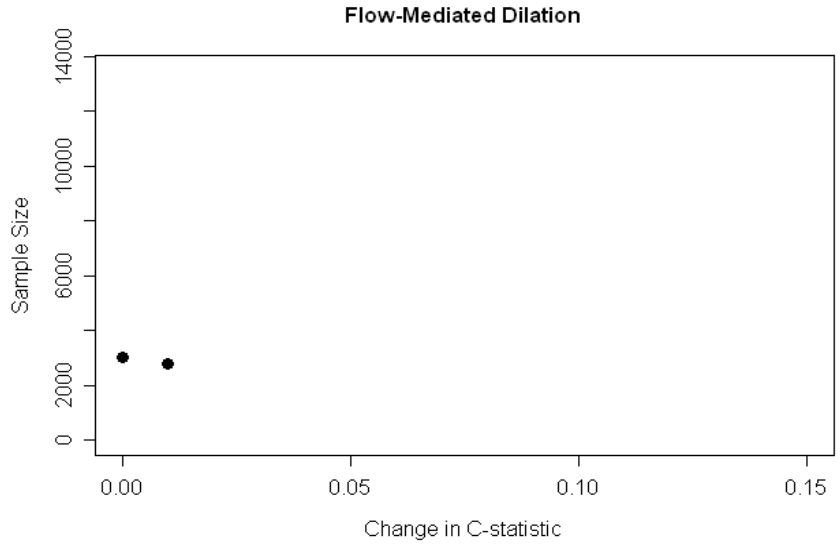
Domain	Applic-ability Bias	Items
Model development	B	• Modelling method (e.g. logistic or survival model) and assumptions checked • Selection of predictors for inclusion in multivariable analysis (e.g. all candidate predictors, pre-selection based on univariable analysis with specific criterion, factor or principal component analysis) • Selection of predictors during multivariable modelling (e.g. backward elimination, forward selection, forced in model, added value of a particular predictor) • Criteria for predictor selection in multivariable modelling, if applicable (e.g. p-value, Akaike or Bayesian Information Criterion, explained variance, c-statistic or reclassification measures) • Shrinkage of regression coefficients (no shrinkage, uniform shrinkage, penalized estimation, Lasso) • Different models developed and/or compared (e.g. basic and extended models) • Risk groups defined, if so, how
Model Performance	A/B	• Calibration (e.g. Hosmer-Lemeshow test, calibration plot) and discrimination measures (C-statistic, D-statistic, log-rank) • (Re-)classification measures (e.g. sensitivity, specificity, predictive values, net reclassification improvement, integrated discrimination improvement) and for which cut-off points • Overall performance measures (e.g. R-squared, Brier score) • Clinical usefulness measures (e.g. decision curve analysis)
Model testing (validation)	A/B	• Model performance tested/quantified beyond development set (none, using resampling methods or other participant data, or combination) • Type of resampling method or internal validation (e.g. none, bootstrap, cross-validation, random split-sample) • Type of other participant data or external validation (e.g. none, temporal, geographical, setting, other or same investigators) • Model updated or recalibrated after (poor) external validation (e.g. intercept/ baseline hazards recalibrated, predictor effects adjusted, new predictors added)

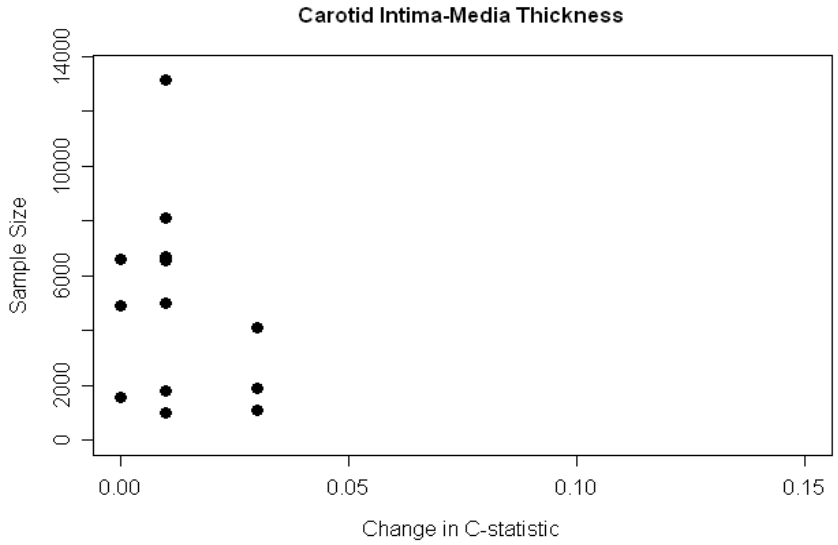
Domain	Applic-ability Bias	Items
Results	A/B	• Distribution of predictors (including missing data) for development and validation data sets (if applicable) • Final and other (e.g. basic or extended) multivariable models presented (e.g. regression coefficients, including intercept or baseline hazard, model performance measures, all with standard errors or confidence intervals) • Any alternative presentation of the final prediction model (e.g. sum score, nomogram, score chart, predictions for specific risk subgroups)
Discussion	A/B	• Interpretation of presented models (e.g. confirmatory, i.e. model useful for practice, or exploratory, i.e. more research needed) • Where were conclusions based on (e.g. predictor effects, P-values, performance measures, results of model validation) • Comparison with other studies, discussion of generalizability, limitations

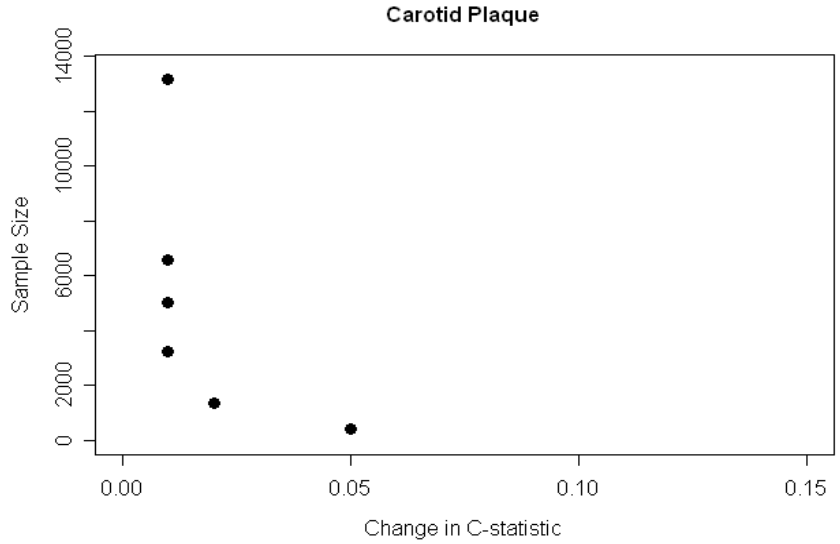
Ciritical Appraisal

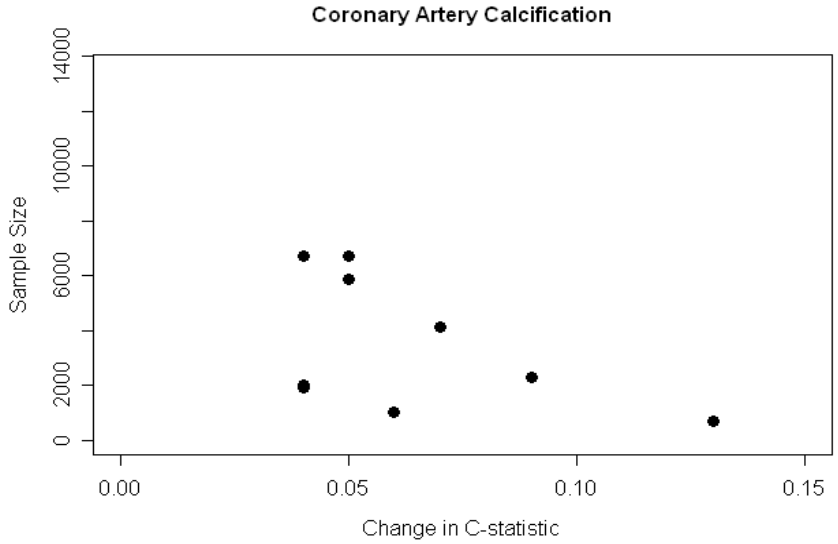
Study Feature	Study participants					Attrition			Prognostic variable			Outcome			Analysis
Subfeature	Relevant domain	Participant selection	In- and exclusion criteria	Baseline characteristics	Participation rate	Response rate	Data collection drop-outs	Reasons loss to follow-up	Definition marker	Measurement method	Missing values	Definition outcome	Outcome ascertainment	Missing values	Presentation
Flow-mediated dilation															
Yeboah[19]	x	x	x	x	x	x	-	-	x	x	-	x	x	-	x
Yeboah[11]	x	x	x	x	x	x	-	-	x	x	-	x	x	-	x
Carotid intima-media thickness															
Anderson[20]	x	x	x	x	-	x	-	-	x	x	-	x	x	-	x
Cao[21]	x	x	x	x	x	-	-	x	x	x	x	x	x	x	x
Chambless[22]	x	x	x	x	x	x	-	-	x	x	x	x	x	x	x
Chambless[23]	x	x	x	x	x	x	-	-	x	x	-	x	x	x	x
Elias-Smale[29]	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Folsom[25]	x	x	x	x	x	x	-	-	x	x	x	x	x	x	x
Lorenz[28]	x	x	x	x	x	-	-	-	x	x	-	x	x	-	x
Mathiesen[26]	x	x	x	x	x	x	-	-	x	x	-	x	x	-	x
Nambi[9]	x	x	x	x	x	x	-	-	x	x	x	x	x	x	x
Polak[30]	x	x	x	x	x	x	-	-	x	x	x	x	x	x	x
Price[27]	x	x	x	x	x	x	x	-	x	x	-	x	x	-	x
del Sol[24]	x	x	x	x	x	x	-	-	x	x	x	x	x	-	x
Carotid plaque															
Cao[21]	x	x	x	x	x	-	-	x	x	x	x	x	x	-	x
Mathiesen[26]	x	x	x	x	x	x	-	-	x	x	-	x	x	-	x
Nambi[9]	x	x	x	x	x	x	-	-	x	x	x	x	x	x	x
Prati[31]	x	x	x	x	x	x	-	-	x	x	-	x	x	-	x

‘Publication bias’









Small Group Discussion

- Break up into groups related to a question of interest
  - Overall prognosis
  - Prognostic Factor
  - Risk prediction model
- Introduce from the group an example
- Discuss:
  - Framing review question
  - (search strategy)
  - Critical appraisal (RoB)

Conducting Prognosis Reviews

- Methods should be guided by review question
- Cautious interpretation and careful attention to methods and reporting in future reviews
- Need for further methodological work in the area of prognosis systematic reviews to investigate potential biases

If you would like to contribute to this work…

Contact: Greta Ridley(Coordinator PMG)

ridleyresearch@aapt.net.au

Proposal for a new Cochrane Systematic Review of Prognosis

Please provide brief answers to each point so the editors can assess the proposal.

Proposed Title

Contact Author name

Motivation for the review (for example, is this going to be part of a PhD; is it part of a larger project; is it particularly topical at the present time?)

Description of proposal (please provide brief answers to each point, the aim is to allow the methodological editors, who may not be familiar with the clinical background of this topic, to assess the proposal).

(a) Objective(s)

What is the research question?

(b) Background (please answer questions below)

i) What is the clinical problem?

ii) Describe the clinical pathway (if relevant): A description of the existing clinical pathway of patients. Outline how patients might present, the point in the existing pathway that participants would be considered for testing with the prognostic factors/model,

iii) How might information about the prognostic factors/model be used to improve e.g. treatments and patient outcomes?

iv) Is there any other information required to understand the clinical problem?

(c) Rationale for review

Title

Authors*
Contact person
Dates

Background

State the rationale for the review and explain why it is important to undertake this review.

Review question

Present the review question in PECO format:

- Population defined by presence of a condition/disease or specific characteristics, setting, time period
- Exposure (if any) or simply time of follow-up
- Comparator (if any)
- Outcome defined by stage of condition/disease, timing of measure, type of scale

Examples:

For school aged children (population) diagnosed with autism (exposure1) and followed-up for one year or more (exposure 2), what percent still have autism (outcome)

Methods

The Methods section in a protocol should be written in the future tense.

Criteria for considering studies for this review

A priori decide on inclusion and exclusion parameters based on the information provided below. Describe selection process including number of reviewers involved and how disagreement will be resolved by consensus.

Types of participants

Clear and reproducible definition of population at risk of developing the outcomes of interest in terms of:

- presence of condition/ disease, procedure or circumstance including any

Reviews therapeutic Interventions (RCTs)	**Reviews DTA studies**	**Reviews Prognosis studies**
Background	**Background**	**Background**
Description of the condition	Target condition being diagnosed	Description of the condition
How the intervention might work	Index test Alternate test(s)	**Description of subgroups (overall prognosis) / predictors (PF)/ prediction models (PM)**
Why it is important to do this review	Rationale	Why it is important to do this review
Objectives	**Objectives**	**Objectives**
	Primary objective	Primary objective
	Investigation of sources of heterogeneity	Investigation of sources of heterogeneity (Clinical, Design, Statistical)
Methods	**Methods**	**Methods**
Criteria for considering studies for this review	Criteria for considering studies for this review	Criteria for considering studies for this review
Types of studies	*Types of studies*	*Types of studies*
Types of participants	*Types of participants*	*Types of participants*
Types of interventions	*Index tests*	***Types of subgroups (OP) predictors (PF) /prediction models (PM)***
Comparator (Control)	*Comparator tests*	***Predictor groups (values) / comparator models***
Types of outcome measures	*Target conditions*	*Type of outcome measures*
Primary outcomes	*Reference standards*	*Primary* outcome
Secondary outcomes	Secondary outcomes	*Secondary outcomes*
Search methods for identification of studies	Search methods for identification of studies	Search methods for identification of studies
Electronic searches	*Electronic searches*	*Electronic searches*
Searching other resources	*Searching other resources*	*Searching other resources ! (likely more publication bias)*
Data collection and analysis	Data collection and analysis	Data collection and analysis
Selection of studies	*Selection of studies*	*Selection of studies*
Data extraction and management	*Data extraction and management*	*Data extraction and management*
Assessment of risk of bias in included studies (RoB)	*Assessment of methodological quality (Quadas-2)*	***Assessment of methodological quality***
Measures of treatment effect	*Statistical analysis and data synthesis*	*Statistical analysis and data synthesis*
Unit of analysis issues	*Unit of analysis issues*	***Unit of analysis issues***

Reviews therapeutic Interventions (RCTs)	**Reviews DTA studies**	**Reviews Prognosis studies**
Dealing with missing data	Dealing with missing data	Dealing with missing data
Assessment of heterogeneity	*Investigation of heterogeneity*	*Investigation/description of statistical heterogeneity*
Assessment of reporting biases	Assessment of reporting biases	Discussing reporting bias (prediction models)
Data synthesis	Data synthesis	*Data synthesis*
Subgroup analysis and investigation of heterogeneity	*Subgroup analysis and investigation of heterogeneity*	*Description Subgroup effects and heterogeneity*
Sensitivity analysis	*Sensitivity analyses*	Not yet
Results	**Results**	**Results**
Results of the search	Results of the search	Results of the search
• Included studies	• *Included studies*	• *Included studies*
• Excluded studies	• *Excluded studies*	• *Excluded studies*
Methodological quality of included studies	Methodological quality of included studies	Methodological quality of included studies
Findings	Findings	Findings
Discussion	**Discussion**	**Discussion**
Summary of main results	Summary of main results	Summary of main results
Strengths and weaknesses of the review	Strengths and weaknesses of the review	Strengths and weaknesses of the review
Applicability of findings to clinical practice and policy	Applicability of findings to clinical practice and policy	Applicability of findings to clinical practice and policy
Authors' conclusions	**Authors' conclusions**	**Authors' conclusions**
Implications for practice	Implications for practice	Implications for practice
Implications for research	Implications for research	Implications for research
Acknowledgements	**Acknowledgements**	**Acknowledgements**
Contribution of authors	**Contribution of authors**	**Contribution of authors**
Declarations of interest	**Declarations of interest**	**Declarations of interest**
Differences between protocol and review	**Differences between protocol and review**	**Differences between protocol and review**
Published notes	**Published notes**	**Published notes**
Characteristics of studies	**Characteristics of studies**	**Characteristics of studies**
Tables	**Tables**	***Tables***
Summary of finding tables	Summary of findings (underway)	Too early
References to studies	**References to studies**	**References to studies**
Other references	**Other references**	**Other references**