Appendix 1 Study protocol Systematic review

Research question: What is the reliability and validity of record review to quantify adverse events (AEs)?

Used criteria for analyzing and reporting
COSMIN[1] → The COSMIN-checklist is used for analyzing the methodological quality of the included studies and for reporting the results.
PRISMA[2] → The PRISMA guidelines are used when conducting the systematic review and for publishing the results of the systematic review.

Data Sources:
1**.** Original research in chronological order:
1) PubMed, 2) EMBASE, 3) CINAHL, 4) PsycINFO and 5) the Cochrane Library.

After searching databases, we will use the following methods to complete our search strategy:
- Snowballing of reference list of included studies (result of the systematic search)
- Hand-searching

Selection criteria:

Selectie criteria	**Inclusion criteria**	**Exclusion criteria**
Study design	No good methodology filter available[13] → no selection of study designs	
Settings	Healthcare	High-risk industries with developed methods/techniques of accident investigation
Instruments	Record review	Instruments that analyse patient safety in healthcare or diagnose, treat or prevent diseases of patients
Outcomes	Publications that reports measurements related to patient safety outcomes, e.g. (preventable) adverse events	
Psychometrics properties	Publications that report on: 1) Reliability and/or; 2) Validity.	
Time frame	All studies published until February 2015	
Languages	All	

[1] 1. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist manual. Amsterdam: VU University Medical Centre 2009.
[2] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine 2009;151(4):264-69.

A. DATA EXTRACTION FORM

Reviewers	
Name Reviewer	
Date	

Study	
ID Study	
Authors	
Title	
Source and year	

Objective and methods	
Aim of the study	
Aim of the method/instrument	
Publication data (period of data collection)	
Name of the instrument	
Description of the instrument/method	
Country of origin	
Target population/participants	
Setting	○ primary care ○ secondary care ○ tertiary care
Study design	
Sample size	
Inclusion and exclusion criteria	

Features of the instrument	○ Identify
	○ Quantify
	○ Reliability study
	○ Level of measurement:
	  ○ patient level
	  ○ professional level
	  ○ organizational/institutional level
	○ Format /description instrument and used scale:

Psychometric properties / Outcomes	
Validity	
Reliability	
Raters	
Feasibility	
Qualitative outcomes	E.g. limitations and recommendations for the use of the instrument
Outcomes (definition of AE)	
Limitations of the study	
Authors' key conclusions	
Other comments/remarks	

B. METHODOLOGICAL QUALITY ASSESSMENT FORMS

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)	
ID Study:	
Authors:	
Title:	
Source and Year:	
Design requirements	**excellent-good-fair-poor**
1. Was the percentage of missing items given?	
2. Was there a description of how missing items were handled?	
3. Was the sample size included in the analysis adequate?	
4. Were at least two measurements available?	
5. Were the administrations independent?	
6. Was the time interval stated?	
7. Were patients stable in the interim period on the construct to be measured?	
8.Was the time interval appropriate?	
9. Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	
10. Were there any important flaws in the design or methods of the study?	
Statistical methods	
11. for continuous scores: Was an intra class correlation coefficient (ICC) calculated?	
12. for dichotomous/nominal/ordinal scores: Was kappa calculated?	
13. for ordinal scores: Was a weighted kappa calculated?	
14. for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	

Box D. Content validity (including face validity)	
ID Study:	
Authors:	
Title:	
Source and Year:	
Design requirements	**excellent-good-fair-poor**
1. Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?	
Design requirements	**excellent-good-fair-poor**
2. Was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)	
3. Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)?	
4. Was there an assessment of whether all items together comprehensively reflect the construct to be measured?	
5. Were there any important flaws in the design or methods of the study?	

Box H. Criterion validity	
ID Study:	
Authors:	
Title:	
Source and Year:	
Design requirements	**excellent-good-fair-poor**
1. Was the percentage of missing items given?	
2. Was there a description of how missing items were handled?	
3. Was the sample size included in the analysis adequate?	
4. Can the criterion used or employed be considered as a reasonable 'gold standard'?	
5. Were there any important flaws in the design or methods of the study?	
Statistical methods	
6. for continuous scores: Were correlations, or the area under the receiver operating curve calculated?	
7. for dichotomous scores: Were sensitivity and specificity determined?	