

# Developing and Validating An Evidence Based Framework for Grading and Assessment of Predictive Tools for Clinical Decision Support

Mohamed Khalifa, MD, MSc, MRCSEd, CPHIMS  
AIHI - Australian Institute of Health Innovation

Health Data Analytics 2018 Conference  
Health Informatics Society of Australia

Melbourne, 22-23 October 2018

# Background

---

- Clinical decision support (CDS) systems improve healthcare cost-effectiveness through enhancing evidence-based practice.
- Among CDS are the clinical predictive tools, which quantify contributions of relevant patient characteristics to derive likelihood of diseases or predict clinical outcomes.
- Clinicians are challenged when choosing among a growing number of tools, most of which have never been implemented or assessed for comparative performance or impact.

# Aim and Objectives

---

- The aim is to develop and validate a framework to Grade and Assess Predictive tools (Abbreviated as GRASP).
- The objective is to provide clinicians with a standardised, evidence-based grading system to support their search for and selection of best clinical predictive tools for their tasks.
- The GRASP framework is based on the critical appraisal of published evidence about the performance, potential effect, usability, and impact of predictive tools.

# Study 1 – Developing the GRASP

---

- To develop the framework; a focused review of the literature was conducted to extract criteria to evaluate predictive tools.
- An initial framework was designed and applied to assess and grade five predictive tools: LACE Index, Centor Score, Well's Criteria, the Modified Early Warning Score, and Ottawa knee rule.
- After peer review, the GRASP framework was revised and the grading of the five tools was updated.

# Results

---

- The GRASP framework grades predictive tools based on published evidence across three dimensions:
  - 1) Phase of evaluation (before, during and after implementation)
  - 2) Level of evidence (assigning a numerical score)
  - 3) Direction of evidence (positive, negative or mixed)
- The final grade of a tool is based on the highest phase of evaluation, supported by the highest level of positive evidence, or mixed evidence that supports a positive conclusion.

# GRASP Framework - Grading and Assessment of Predictive Tools for Clinical Decision Support

## Assigned Grades

## Direction of Evidence

## Level of Evidence

- Positive
- Negative
- ◐ Mixed supporting positive conclusion
- ◑ Mixed supporting negative conclusion

## Phase of Evaluation

### Phase A: Post Implementation Impact

### Phase B: During Implementation

### Phase C: Pre Implementation Performance

Clinical Effectiveness, Patient Safety or Healthcare Efficiency

Experimental studies

Observational studies

Based on subjective studies

Reported usability testing

Estimated potential effect on healthcare

Tested for external validity multiple times

Tested for external validity only once

The predictive tool has been tested for internal validity

A1

A2

A3

B1

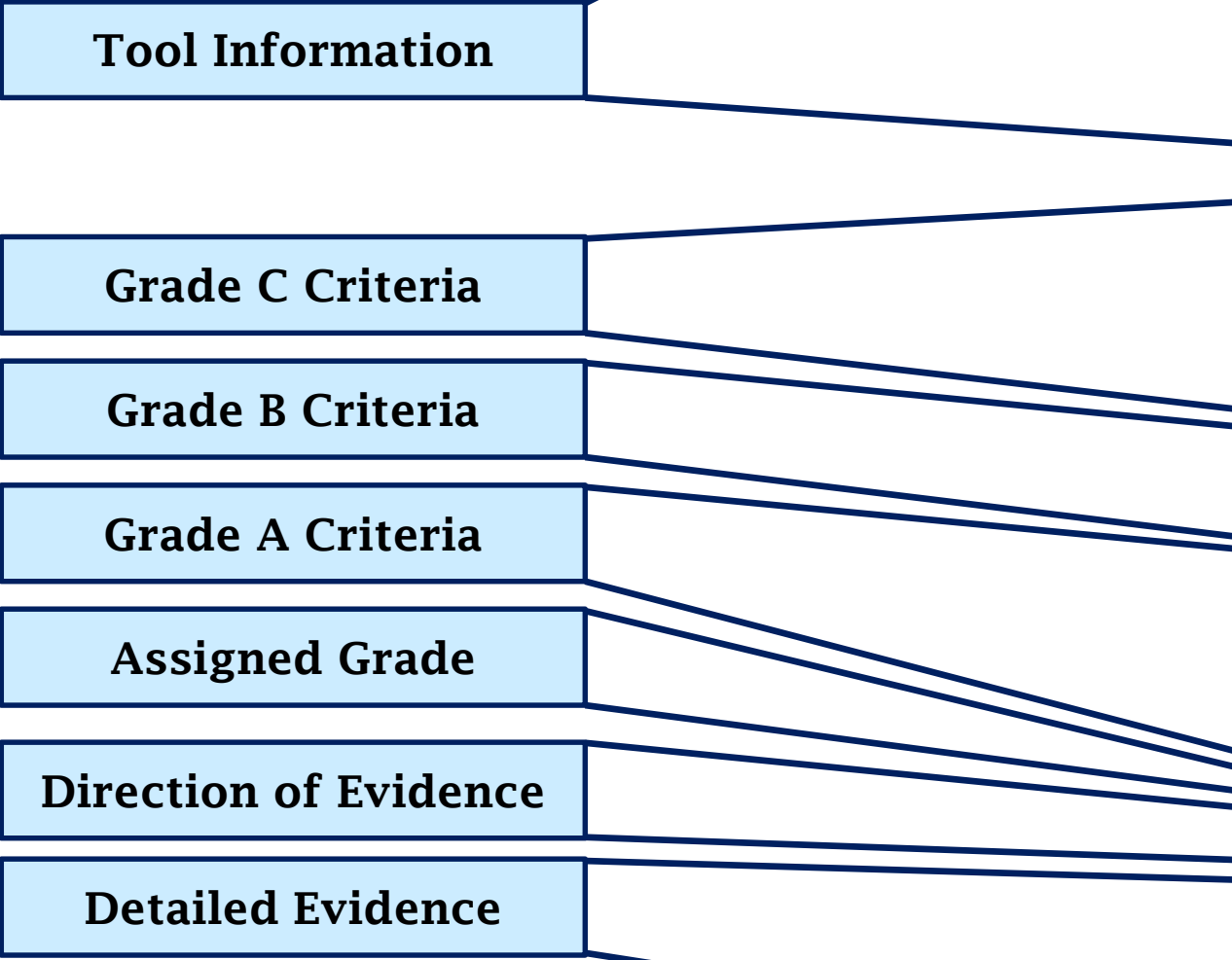
B2

C1

C2

C3

# GRASP Framework Report



Name	Name of predictive tool (report tool's creators and year in the absence of a given name)									
Authors/Year	Name of developer, country and year of publication									
Intended use	Specific aim/intended use of the predictive tool									
Intended user	Type of practitioner intended to use the tool									
Category	Diagnostic/Therapeutic/Prognostic/Preventive									
Clinical area	Clinical specialty									
Target Population	Target patient population and health care settings in which the tool is applied									
Target Outcome	Event to be predicted (including prediction lead time if needed)									
Action	Recommended action based on tool's output									
Input source	<ul style="list-style-type: none"><li>Clinical (including Diagnostic, Genetic, Vital signs, Pathology)</li><li>Non-Clinical (including Healthcare Utilisation)</li></ul>									
Input type	<ul style="list-style-type: none"><li>Objective (Measured input; from electronic systems or clinical examination)</li><li>Subjective (Patient reported; history, checklist ...etc.)</li></ul>									
Local context	Is the tool developed using location-specific data? (e.g. life expectancy tables)									
Methodology	Type of algorithm (e.g. parametric/non-parametric)									
Endorsement	Organisations endorsing the tool and/or guidelines recommending its utilisation									
Automation Flag	Automation status (manual/automated)									
Phase of Evaluation	Level of Evidence	Grade	Evaluation Studies							
Phase C:  Before implementation  Is it possible?	Insufficient internal validation	C0	Tested for internally validity but was either insufficiently internally validated or validation was insufficiently reported.							
	Internal validation	C3	Tested for internally validity (reported calibration & discrimination; sensitivity, specificity, positive and negative predictive values & other performance measures).							
	External validation	C2	Tested for external validity, using one external dataset.							
	Extensive external validation	C1	Tested extensively for external validity, using more than one external dataset.							
Phase B:  During implementation  Is it practicable?	Potential effect	B2	Reported estimated potential effect on clinical effectiveness, patient safety or healthcare efficiency.							
	Usability	B1	Reported usability testing (effectiveness, efficiency, satisfaction, learnability, memorability, and minimizing errors).							
Phase A:  After implementation:  Is it desirable?	Evaluation of post implementation impact on Clinical Effectiveness, Patient Safety or Healthcare Efficiency	A3	Based on subjective studies; e.g. the opinion of a respected authority, clinical experience, a descriptive study, or a report of an expert committee or panel.							
		A2	Based on observational studies; e.g. a well-designed cohort or case-control study.							
		A1	Based on experimental studies; a systematic review of randomised/nonrandomised controlled trials or at least one properly designed, widely applied randomised/nonrandomised controlled trial.							
Final Grade	Grade ABC,123		A1	A2	A3	B1	B2	C1	C2	C3
Direction of Evidence	● Positive Evidence		● Mixed Evidence Supporting Positive Conclusion							
	○ Negative Evidence		● Mixed Evidence Supporting Negative Conclusion							
Justification	Explains how the final grade is assigned based on evidence; which conclusions were taken into consideration, as positive evidence, and which were considered negative.									
References	Details of studies that support the justification: phase of evaluation, level of evidence, direction of evidence, study type, study settings, methodology, results, findings and conclusions (highlighted according to the colour code).								These two sections are included in the full GRASP report on each tool.	
Label/Colour Code	<ul style="list-style-type: none"><li>Positive Findings</li><li>Negative Findings</li></ul>		<ul style="list-style-type: none"><li>Important Findings</li><li>Less Relevant Findings</li></ul>							

Tool	Grade	Impact After Implementation			During Implementation		Performance Before Implementation		
		Experimental Studies	Observational Studies	Subjective Studies	Usability	Potential Effect	External Validation Multiple Times	External Validation Only Once	Internal Validation
		A1	A2	A3	B1	B2	C1	C2	C3
LACE Index	C1						◐		●
Centor Score	B1	◑			●		●		●
Wells' Criteria	A2		●		●		●		●
Modified Early Warning Score	A2		◐				◐		●
Ottawa Knee Rule	A1	●					●		●
Evidence Direction	● Positive Evidence			◐ Mixed Evidence Supporting Positive Conclusion					
	○ Negative Evidence			◑ Mixed Evidence Supporting Negative Conclusion					



# Study 2 – Expert User Validation

---

- The study examines the content validity as well as the inter-rater reliability of GRASP framework.
- For content validity, expert users complete an online survey on the criteria used to grade predictive tools.
- For the inter-rater reliability, expert researchers grade eight predictive tools using GRASP and published studies on the tools.
- Levels of agreements and consensus will be evaluated as well as feedback suggestions on adding, changing or removing criteria.

# Study 3 – End User Validation

---

- The study examines the impact of using GRASP on improving the decisions made by end user clinicians regarding predictive tools.
- A group of emergency department clinicians are requested to answer critical questions about clinical predictive tools, with and without using the GRASP framework reports.
- The levels of efficiency, consistency, and accuracy are measured comparing the results of the two scenarios.

# Study 4 – Applying the Framework

---

- Using GRASP, all available 14 paediatrics head injury predictive tools used at the emergency department are graded based on the critical appraisal of their published evidence.
- The study will discuss: the correlation between tools assigned grades and their country, year of development, number of citations, studies, patient sample size, number of authors, and support by dedicated and well-funded research networks, programs, and professional groups or their appearance on clinical guidelines or endorsement by professional organisations.

	Tool Information				Study Indices			Study Quality Indicators				Tool Grade	Impact After Implementation			During Implementation		Performance Before Implementation		
	Country	Year	Citations	Studies	Citation Index	Publication Index	Literature Index	Sample Size	Journal Impact	Number of Authors	Dedicated Support		Experimental Studies	Observational Studies	Subjective Studies	Usability	Potential Effect	External Validation Multiple Times	External Validation Only Once	Internal Validation
													A1	A2	A3	B1	B2	C1	C2	C3
PECARN	USA	2009	885	24	88.5	2.40	21.24	42,412	53.3	32	Yes	A2		🟡			🟢		🟢	
CHALICE	UK	2006	309	15	23.8	1.15	4.64	22,772	3.3	6	Yes	B2					🟡		🟢	
CATCH	USA	2006	319	12	24.5	0.92	3.83	3,866	6.8	14	Yes	C1						🟢	🟢	
NEXUS II	USA	2005	124	6	8.9	0.43	0.74	1,666	5.7	8	Yes	C1						🟡	🟢	
Palchak	USA	2003	248	3	15.5	0.19	0.74	2,043	5.4	10	No	C2							🟢	
Greenes	USA	1999	237	2	11.9	0.10	0.47	422	5.7	2	No	C3							🟢	
Haydel	USA	2003	118	1	7.4	0.06	0.12	175	5.4	5	No	C3							🟢	
Atabaki	USA	2008	111	1	10.1	0.09	0.11	1,000	10.8	8	No	C3							🟢	
Da Dalt	Italy	2006	85	1	6.5	0.08	0.09	3,806	1.8	8	No	C3							🟢	
Klemetti	Finland	2009	18	1	1.8	0.10	0.02	485	1.1	4	No	C3							🟢	
Buchanich	USA	2007	4	1	0.3	0.08	0.00	97	1.0	1	No	C3							🟢	
Quayle	USA	1997	291	1	13.2	0.05	0.29	322	5.7	7	No	C0							🟡	
Dietrich	USA	1993	220	1	8.5	0.04	0.22	324	5.4	5	No	C0							🟡	
Güzel	Turkey	2009	17	1	1.7	0.10	0.02	916	1.0	6	No	C0							🟡	

# Study 4 – Conclusions

---

- All the tools used the same development methodologies and almost the same clinical variables.
- However, they showed variable predictive performances, which were not correlated with their assigned grades.
- The quality of tools' development studies, the experience and credibility of their authors, and the support by dedicated and well-funded research programs were more significantly influential, than the predictive performance, on the acceptance and successful implementation of tools



# Thank You