THE MYTHOLOGY OF RISK ASSESSMENTS

JOAKIM STURUP, PH.D.

RÄTTSMEDICINALVERKET
NATIONAL BOARD OF FORENSIC MEDICINE

Institute for
FUTURES STUDIES

Karolinska Institutet

Prediction

”Half of my patients would go straight out in the community and commit a major violent crime. The problem is that I do not know what half….”

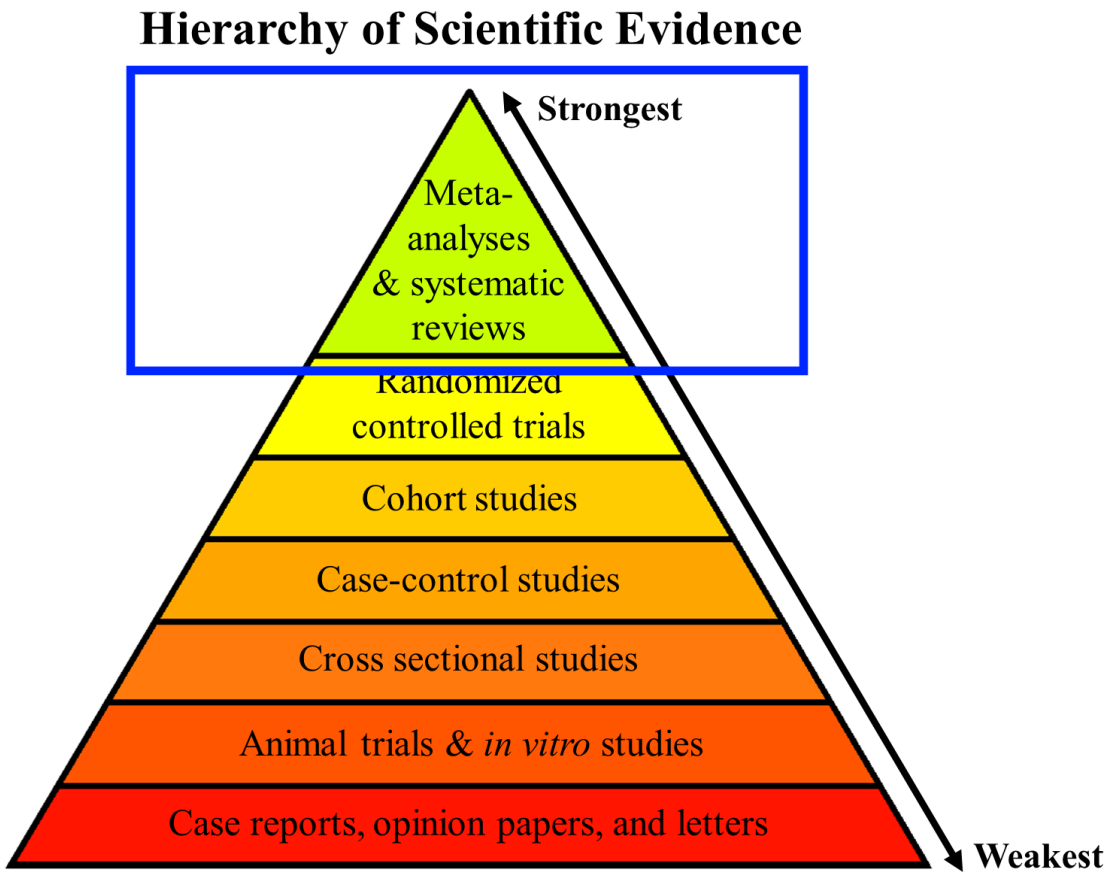
// Unknown clinician at Maudsley Special Hospital

Theoretical aspects

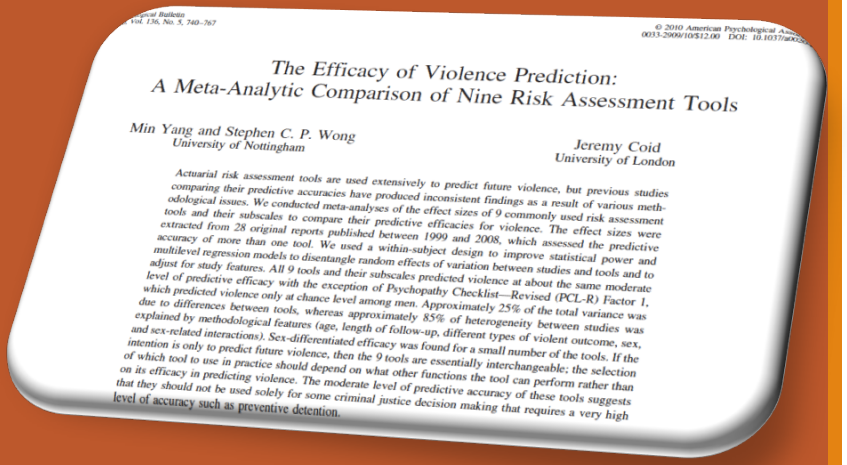
According to Paul Falzer (2013) there are three aspects of interest:

- The practical aspect (are they better than chance?)
- The differential aspect (is one method superior to others?)
- The epistemological aspect (how accurate is accurate enough?)

What can recent meta-analytical work tell us?



Are risk assessments better than chance?

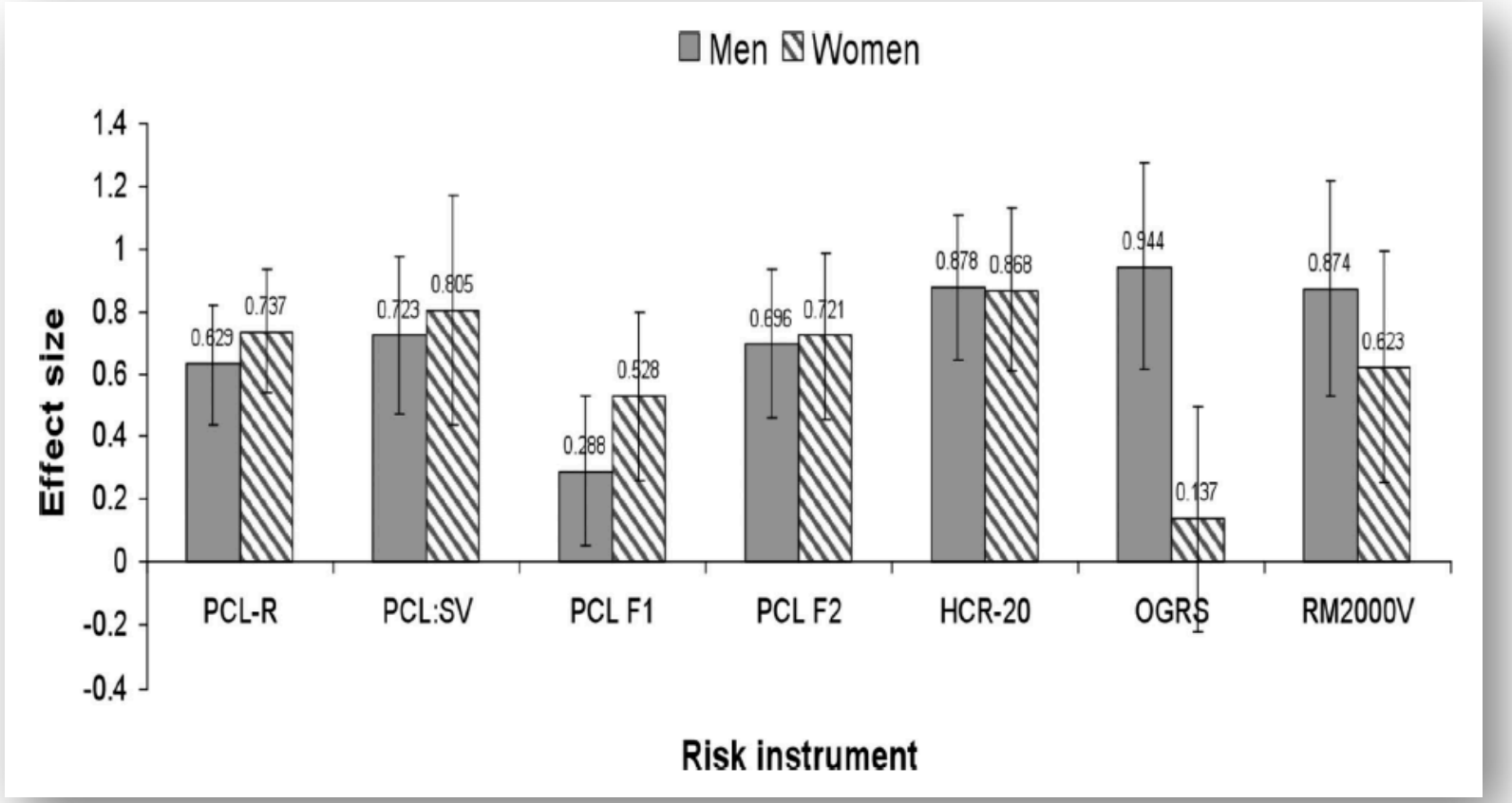


The Efficacy of Violence Prediction:
A Meta-Analytic Comparison of Nine Risk Assessment Tools

Min Yang and Stephen C. P. Wong
University of Nottingham

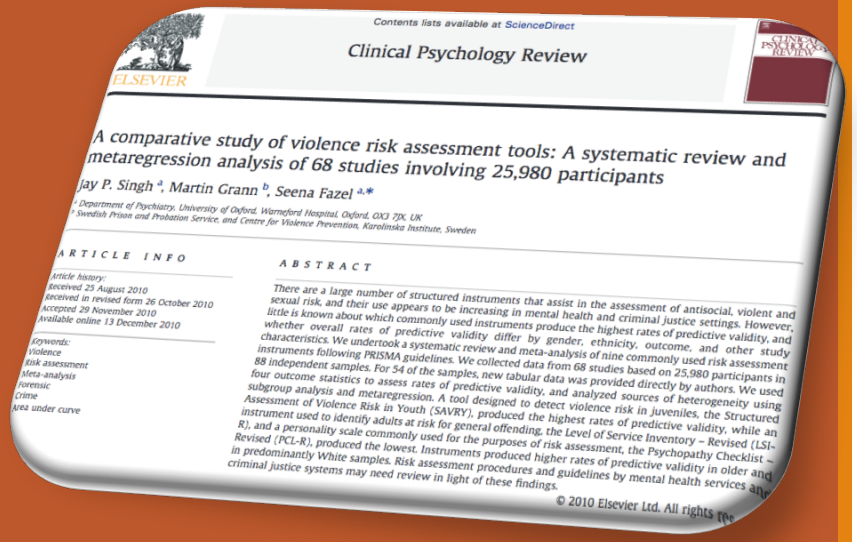
Jeremy Coid
University of London

Actuarial risk assessment tools are used extensively to predict future violence, but previous studies comparing their predictive accuracies have produced inconsistent findings as a result of various methodological issues. We conducted meta-analyses of the effect sizes of 9 commonly used risk assessment tools and their subscales to compare their predictive efficacies for violence. The effect sizes were extracted from 28 original reports published between 1999 and 2008, which assessed the predictive accuracy of more than one tool. We used a within-subject design to improve statistical power and multilevel regression models to disentangle random effects of variation between studies and tools and to adjust for study features. All 9 tools and their subscales predicted violence at about the same moderate level of predictive efficacy with the exception of Psychopathy Checklist—Revised (PCL-R) Factor 1, which predicted violence only at chance level among men. Approximately 25% of the total variance was due to differences between tools, whereas approximately 85% of heterogeneity between studies was explained by methodological features (age, length of follow-up, different types of violent outcome, sex, and sex-related interactions). Sex-differentiated efficacy was found for a small number of the tools. If the intention is only to predict future violence, then the 9 tools are essentially interchangeable; the selection of which tool to use in practice should depend on what other functions the tool can perform rather than on its efficacy in predicting violence. The moderate level of predictive accuracy of these tools suggests that they should not be used solely for some criminal justice decision making that requires a very high level of accuracy such as preventive detention.



”Overall, our results showed that all of the nine tools predicted violence at above-chance levels, with medium effect sizes…thus the instruments included in this study demonstrated medium effects for predicting violence risk.” Yang et al (2010) p. 754 & 757

Is one method superior to others?


Contents lists available at ScienceDirect

Clinical Psychology Review

ELSEVIER

A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants

Jay P. Singh [a], Martin Grann [b], Seena Fazel [a,*]

[a] Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, OX3 7JX, UK
[b] Swedish Prison and Probation Service, and Centre for Violence Prevention, Karolinska Institute, Sweden

ARTICLE INFO

Article history:
Received 25 August 2010
Received in revised form 26 October 2010
Accepted 29 November 2010
Available online 13 December 2010

Keywords:
Violence
Risk assessment
Meta-analysis
Forensic
Crime
Area under curve

ABSTRACT

There are a large number of structured instruments that assist in the assessment of antisocial, violent and sexual risk, and their use appears to be increasing in mental health and criminal justice settings. However, little is known about which commonly used instruments produce the highest rates of predictive validity, and whether overall rates of predictive validity differ by gender, ethnicity, outcome, and other study characteristics. We undertook a systematic review and meta-analysis of nine commonly used risk assessment instruments following PRISMA guidelines. We collected data from 68 studies based on 25,980 participants in 88 independent samples. For 54 of the samples, new tabular data was provided directly by authors. We used four outcome statistics to assess rates of predictive validity, and analyzed sources of heterogeneity using subgroup analysis and metaregression. A tool designed to detect violence risk in juveniles, the Structured Assessment of Violence Risk in Youth (SAVRY), produced the highest rates of predictive validity, while an instrument used to identify adults at risk for general offending, the Level of Service Inventory – Revised (LSI-R), and a personality scale commonly used for the purposes of risk assessment, the Psychopathy Checklist – Revised (PCL-R), produced the lowest. Instruments produced higher rates of predictive validity in older and in predominantly White samples. Risk assessment procedures and guidelines by mental health services and criminal justice systems may need review in light of these findings.

© 2010 Elsevier Ltd. All rights re

Table 3
Median area under the curve produced by nine risk assessment tools ranked in order of strength.

Tool	*n*	*k*	Median AUC	*IQR*
SVR-20	380	3	0.78	0.71–0.83
SORAG	1599	6	0.75	0.69–0.79
VRAG	2445	10	0.74	0.74–0.81
SAVRY	915	8	0.71	0.69–0.73
HCR-20	1320	8	0.70	0.64–0.76
SARA	102	1	0.70	–
Static-99	8246	12	0.70	0.62–0.72
LSI-R	856	3	0.67	0.55–0.73
PCL-R	2645	10	0.66	0.54–0.68

Note. *n* = sample size; *k* = number of samples; *AUC* = area under the curve; *IQR* = interquartile range.

”Our study found no evidence that, compared with SCJ tools, actuarial instruments produced better levels of predictive validity.” Singh et al. 2011

” …and no tool predicted violence significantly better than any other. In sum, all did well, but none came first.” Yang et al p. 757

Author bias?

Authorship Bias in Violence Risk Assessment? A Systematic Review and Meta-Analysis

Jay P. Singh[1,2], Martin Grann[3], Seena Fazel[4]*

Abstract

Various financial and non-financial conflicts of interests have been shown to influence the reporting of research findings, particularly in clinical medicine. In this study, we examine whether this extends to prognostic instruments designed to assess violence risk. Such instruments have increasingly become a routine part of clinical practice in mental health and criminal justice settings. The present meta-analysis investigated whether an authorship effect exists in the violence risk assessment literature by comparing predictive accuracy outcomes in studies where the individuals who designed these instruments were study authors with independent investigations. A systematic search from 1966 to 2011 was conducted using PsycINFO, EMBASE, MEDLINE, and US National Criminal Justice Reference Service Abstracts to identify predictive validity studies for the nine most commonly used risk assessment tools. Tabular data from 83 studies comprising 104 samples was collected, information on two-thirds of which was received directly from study authors for the review. Random effects subgroup analysis and metaregression were used to explore evidence of an authorship effect. Overall, studies authored by tool designers reported predictive substantial and statistically significant authorship effect. We found a validity findings around two times higher those of investigations reported by independent authors (DOR = 6.22 [95% CI = 4.68–8.26] in designers' studies vs. DOR = 3.08 [95% CI = 2.45–3.88] in independent studies). As there was evidence of an authorship effect, we also examined disclosure rates. None of the 25 studies where tool designers or translators were also study authors published a conflict of interest statement to that effect, despite a number of journals requiring that potential conflicts be disclosed. The field of risk assessment would benefit from routine disclosure and registration of research studies. The extent to which similar conflict of interests exists in those developing risk assessment guidelines and providing expert testimony needs clarification.

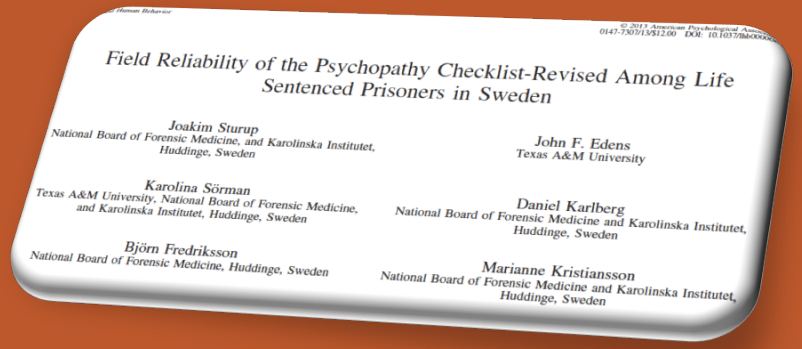
Table 3. Subgroup and metaregression analyses of diagnostic odds ratios (DORs) produced by nine commonly used risk assessment tools when a tool designer was a study author versus independent investigations.

Analysis	Subcategory	Authorship status	*DOR* (95% CI)	Metaregression
Overall	Translators not	Tool designer as study author	6.22 (4.68–8.26)	$\beta = 0.83$, $SE = 0.36$, $p = 0.02$
	included as "designers"	Tool designer not study author	3.08 (2.45–3.88)	
	Translators included	Tool designer as study author	4.45 (3.06–6.47)	$\beta = 0.39$, $SE = 0.26$, $p = 0.13$
	as "designers"	Tool designer not study author	3.04 (2.36–3.91)	
Type of	Actuarial	Tool designer as study author	5.38 (3.82–7.58)	$\beta = 0.78$, $SE = 0.48$, $p = 0.11$
tool[a]		Tool designer not study author	2.56 (1.98–3.30)	
	SCJ	Tool designer as study author	8.60 (5.15–14.35)	$\beta = 0.59$, $SE = 0.51$, $p = 0.26$
		Tool designer not study author	5.07 (3.27–7.84)	
Publication	Journal	Tool designer as study author	6.13 (4.59–8.20)	$\beta = 0.79$, $SE = 0.38$, $p = 0.04$
source[a]		Tool designer not study author	3.09 (2.39–3.98)	
	Gray literature	Tool designer as study author	8.73 (2.06–36.94)	$\beta = -1.03$, $SE = 1.05$, $p = 0.34$
		Tool designer not study author	3.07 (1.93–4.90)	

Note. β = unstandardized regression coefficient; SE = standard error; SCJ = structured clinical judgment; DOR = diagnostic odds ratio; CI = confidence interval; Gray literature = doctoral dissertations, Master's theses, government reports, and conference presentations.
[a]Authorship operationally defined as being an author of the English-language version of the instrument under investigation.
doi:10.1371/journal.pone.0072484.t003

"Having explored this issue in the growing violence risk assessment literature, we have found evidence of both an authorship effect as well as lack of disclosure by tool designers and translators."
Singh et al. (2013)

An example from the real world



Field Reliability of the Psychopathy Checklist-Revised Among Life Sentenced Prisoners in Sweden

Joakim Sturup
National Board of Forensic Medicine, and Karolinska Institutet, Huddinge, Sweden

John F. Edens
Texas A&M University

Karolina Sörman
Texas A&M University, National Board of Forensic Medicine, and Karolinska Institutet, Huddinge, Sweden

Daniel Karlberg
National Board of Forensic Medicine and Karolinska Institutet, Huddinge, Sweden

Björn Fredriksson
National Board of Forensic Medicine, Huddinge, Sweden

Marianne Kristiansson
National Board of Forensic Medicine and Karolinska Institutet, Huddinge, Sweden

Table 1
PCL-R Sco
Assessment

PCL-R			(95% CI)
Total			**[.44, .85]**
Factor 1#			**[.32, .81]**
Facet 1			[.28, .80]
Facet 2			[.29, .79]
Factor 2+			**[.54, .88]**
Facet 3			[.21, .76]
Facet 4			[.80, .95]

Item 1, 2, 4 … 6, 7, 8, and
16. + Item … , 13, 14, and
15. ¥ Item

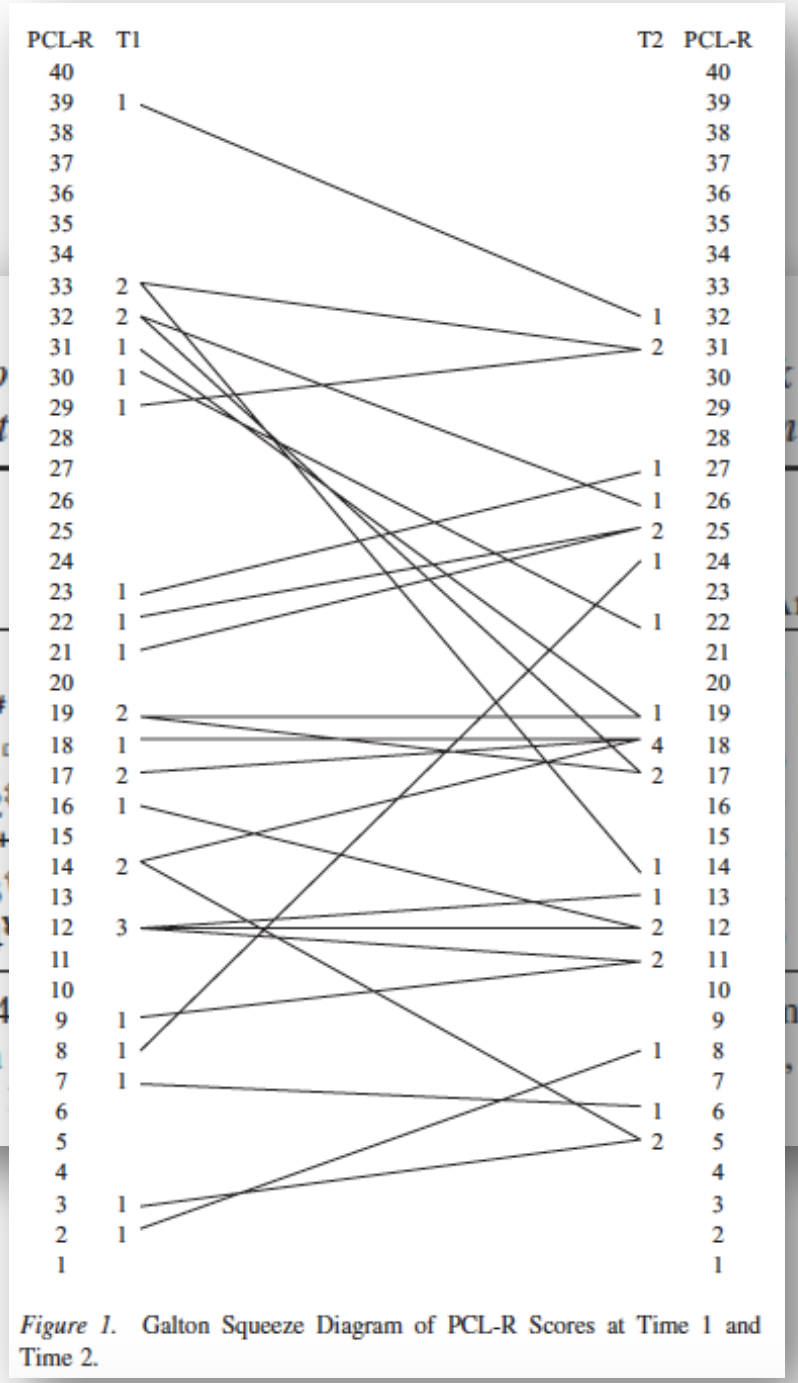
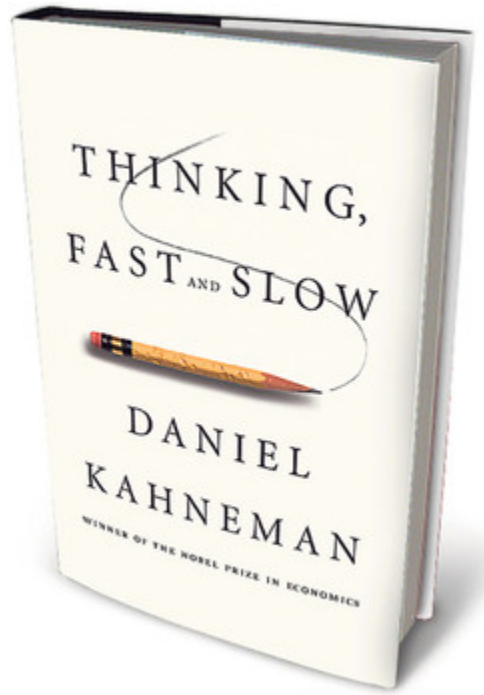


Figure 1. Galton Squeeze Diagram of PCL-R Scores at Time 1 and Time 2.

Other biases…

There are a number of biases that influence our assessments:

- Anchoring bias
- Confirmation bias
- Overconfidence
- Recency bias
- Etc.



So, where are we?

All in all median AUC is .75

=

25% better than chance or

25% less than perfect prediction?

Theoretical aspects revisited…

Falzers (2013) three aspects of interest:

- Are risk assessments better than chance? Yes
- Is one risk assessment method superior to others? No
- How accurate is accurate enough? Not only a scientific question!

What is practice/science/politics?

Within the field of risk assessments it is extra important to part the areas of science and moral:

- Science can tell us what is true!
- Moral (or politics) can tell us what is right!

A possible solution?

- With the current status of risk assessments should clinicians avoid situations where the individual assessed is at risk to be incarcerated due to what he or she is likely to do
- I argue that it is more acceptable and justifiable to conduct risk assessments when: I) they can lead to something positive (e.g. time determined life sentence, previously laid conditional release, etc), II) standardized, valid and reliable (as reliable as possible) methods are used, and III) that the assessor is well aware of the recent scientific literature…

Contact:

JOAKIM.STURUP@RMV.SE