Field Validity of Static-99R and STABLE-2007: A 15-year Follow-Up Study of 4,511 Men Sentenced for Sexual Offences in British Columbia

Joel Morrison, R. Karl Hanson, Audrey M. Villeneuve & Kelly M. Babchishin

Canadian Psychological Association Annual Convention • Montreal, Québec, June 6, 2026

Sexually Harmful Behaviours Research Lab

SAARNA
Safety in Numbers

Carleton University

1

Field Validity

- Consider factors that could influence the performance of risk tools

Staff Training

Data Availability

Funding

Systemic Incentives

Carleton University

2

Why this Sample is Unique

- BC Corrections → 4,173 men → Community Supervision → 15-year Follow-up

Real-World Field Scoring

Community Supervision Officers

The Largest STABLE-2007 Dataset

N = 4010 with at least one STABLE-2007 assessment

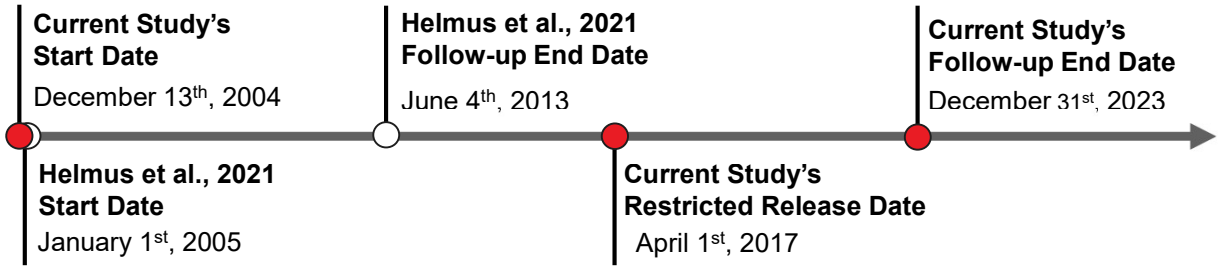
Carleton University

3

A Replication of Helmus et al. (2021)

& 10-Year Static-99R Calibration Using Recidivism Norms From Lee & Hanson (2021)

Same sample with longer follow-up



Carleton University

4

Some Other Publications with the Same Dataset

Villeneuve, A., Morrison, J., Babchishin, K. M., & Hanson, R. K. (2025). *Evaluating the real-world performance of the STABLE-2007: A 15-year field study of dynamic risk assessment for sexual recidivism* [Poster presentation]. CPA Convention, Montréal, QC, Canada.

- Ahmed, S., & Helmus, LM. (2026). Comparing Indian and White men charged or convicted of sexual offences on the Static-99R and STABLE-2007. *Psychology, Crime and Law*, *32*(1), 135-156. DOI:10.1080/1068316X.2023.2227313
- Babchishin, KM, Hanson, RK, & Lee, SC. (2024). Risk to reoffend changes over time: Improving correctional programming through progress monitoring. *Psychological Assessment, 36*(10), 595-605. DOI:10.1037/pas0001335
- Babchishin, KM, Dibayula, S, McCulloch, C, Hanson, RK, & Helmus, LM. (2023). ACUTE-2007 and STABLE-2007 predict recidivism for men adjudicated for child sexual exploitation material offending. *Law and Human Behavior, 47*(5), 606-618. DOI:10.1037/lhb0000540
- Helmus, LM, Hanson, RK, Murrie, DC, & Zabarauckas, CL. (2021). Field validity of Static-99R and STABLE-2007 with 4,433 men serving sentences for sexual offences in British Columbia: New findings and meta-analysis. *Psychological Assessment, 33*(7), 581-595. DOI:10.1037/pas0001010
- Lee, SC, Brankley, AE, & Hanson, RK. (2023). There is no such thing as zero risk for sexual offending. *Canadian Journal of Criminology and Criminal Justice, 65*(3), 1-31. DOI:10.3138/cjccj-2022-0025
- Lee, SC, Babchishin, KM, Mularczyk, KP, & Hanson, RK. (2024). Dynamic risk scales decay over time: Evidence for reassessment. *Assessment*, *31*(3), 698-714. DOI:10.1177/10731911231177227

Carleton University

5

Sexual Recidivism Risk Tools

Static-99R
10 items
Static factors
Score: -3 to 12

STABLE-2007
13 items
Dynamic factors
Score: 0–26

Standardized Risk Levels

I	II	III	IVa	IVb

Static-99R Scores	STABLE-2007 Scores																				
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+
-3	I	I	I	I	I	I	I	I	II	II	II	II	III	III	III	III	III	III	III	III	III
-2	I	I	I	I	I	I	I	I	II	II	II	II	III	III	III	III	III	III	III	III	III
-1	I	I	II	II	II	II	II	II	II	II	III	III	III	III	IVa	IVa	IVa	IVa	IVa	IVa	IVa
0	I	I	II	II	II	II	II	II	II	II	III	III	III	III	IVa	IVa	IVa	IVa	IVa	IVa	IVa
1	II	II	II	III	III	III	III	III	III	III	III	III	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb
2	II	II	II	III	III	III	III	III	III	III	III	III	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb
3	II	II	II	III	III	III	III	III	III	III	III	III	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb
4	III	III	III	III	III	IVa	IVa	IVa	IVa	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb	IVb	IVb	IVb
5	III	III	III	III	III	IVa	IVa	IVa	IVa	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb	IVb	IVb	IVb
6	III	III	III	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
7	III	III	III	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
8	III	III	III	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
9	III	III	III	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
10+	III	III	III	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb

Carleton University

6

Current Study

Discrimination

1) REASSESSED discrimination of Static-99R and Static/STABLE combination

Incremental Validity

2) REASSESSED the incremental validity of STABLE-2007 to the Static-99R total score

Calibration

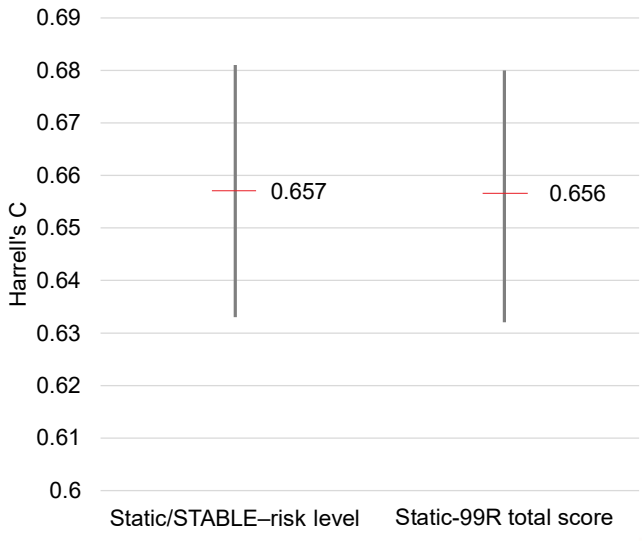
3) NEW 10-year calibration for Static-99R using the same norms for routine/complete samples from Lee & Hanson (2021)

4) UPDATED 5-year calibration for Static-99R and Static/STABLE combination using the same norms from Brankley et al. (2017)

Carleton University

7

Discrimination

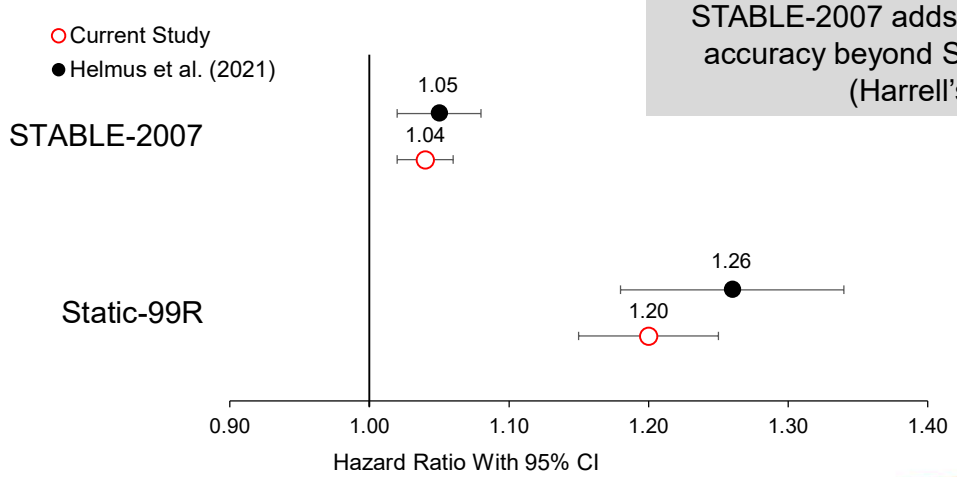


Moderate predictive accuracy with no significant difference between predictors

Carleton University

8

Incremental Validity

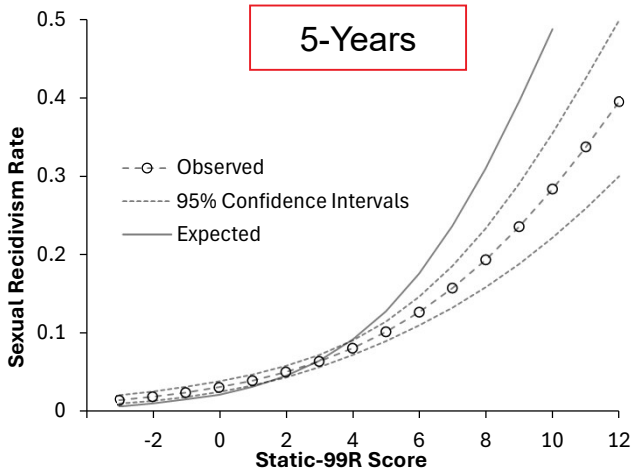


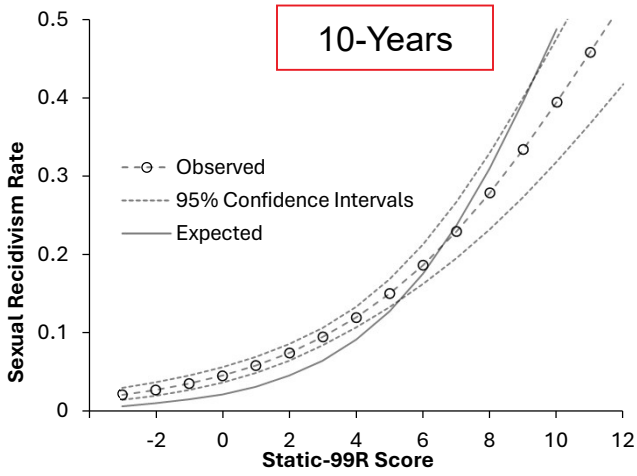
STABLE-2007 adds incremental predictive accuracy beyond Static-99R total scores (Harrell's *C* = .671)

Carleton University

9

Calibration of Static-99R





Static-99R is well calibrated to the norms; however, the norms underpredict for the lowest risk groups and overpredict for the highest risk groups

Carleton University

10

Calibration of Static-99R and STABLE-2007 at 5-Years

	Expected	Observed		
Risk Level	Recidivists (%)	Recidivists (%)	E/O	[95% CI]
I	2.8	1.9	1.50	[0.62, 3.61]
II	5.3	1.9	**2.73**	**[1.67, 4.46]**
III	7.5	6.0	**1.26**	**[1.03, 1.55]**
IVa	13.6	8.4	**1.62**	**[1.28, 2.05]**
IVb	26.8	14.8	**1.81**	**[1.46, 2.25]**
Overall	10.8	6.6	**1.65**	**[1.46, 1.86]**

Note. E/O values in bold are statistically significant (p <.05).

The norms significantly overpredicted by 65% for overall sexual recidivism

Carleton University

11

A Comparison with Helmus et al. (2021) and Lee & Hanson (2021)

Feature	Helmus et al. (2021)	Lee & Hanson (2021)
Discrimination	**Consistent:** Moderate discrimination for both Static-99R and Static/STABLE risk levels.	**Different:** Discrimination parameter for Static-99R was significantly lower in the BC sample.
	Different: • Lower, but more precise than in Helmus study. • The risk tools performed better for violent and general criminal recidivism than for sexual recidivism. • The association between the risk tools and the recidivism outcomes were not constant over time.	
Incremental Validity	**Consistent:** STABLE-2007 adds significant incremental predictive accuracy beyond Static-99R for sexual, violent, and any criminal recidivism.	
Calibration	**Consistent:** • Static-99R underpredicted for the lowest risk groups and overpredicted for the highest risk groups. • Static/STABLE risk levels significantly overpredicted sexual recidivists.	**Consistent:** The base rates for sexual recidivism at 5 and 10 years were very similar to the expected norms.
	Different: The current study found Static-99R was well calibrated to the norms, whereas Helmus et al. found it to overpredict sexual recidivism.	

12

Implications & Future Directions

- Continued use in BC
- Risk tools continue to discriminate
- Incremental utility
- Long-term recidivism trends
- Need for local norms

Carleton University

13

Thank You

Carleton University

14

Measurement Invariance
Over Time in Dynamic Risk Assessment

Evidence from STABLE-2007 and ACUTE-2007

Seung C. Lee, R. Karl Hanson, & Kelly M. Babchishin

Canadian Psychological Association Annual Convention • Montreal, Québec, June, 2026

SAARNA
Safety in Numbers

Carleton UNIVERSITY
Canada's Capital University

15

BACKGROUND

Background & Research Questions

Dynamic risk assessment tools are designed for repeated administration — used multiple times across supervision to track changes in an individual's risk level. STABLE-2007 (stable dynamic factors) and ACUTE-2007 (acute dynamic factors) are among the most widely adopted in correctional settings.

Instruments & Samples

STABLE-2007
n = 669 • 13 items • 2 time points

ACUTE-2007
n = 2,289 • 7 items • 3 time points

Research Questions

RQ1 What factor structure best represents STABLE-2007 and ACUTE-2007?

RQ2 Are these factor structures consistent across multiple time points (measurement invariance)?

RQ3 Do observed score changes reflect true changes in risk, or shifts in measurement?

16

BACKGROUND

Previous Research on ACUTE-2007 Factors/Subscales

Study	Sample	Analysis	Results
Hanson et al. (2007)	Dynamic Supervision Project *N* = 744 cases *N* ≈ 7,000 ratings	PC FA, Pearson *r* Eigenvalue > 1 Scree plot (overruled) Varimax Rotation	Total (7 items) Sex/Violence Subscale (4 items)
Babchishin & Hanson (2020)	Dynamic Supervision Project *N* = 632 cases Three assessments each	EFA, Polychoric Eigenvalue > 1, PA, MAP Measurement Invariance Model fit indices	One factor (all 7 items)

17

BACKGROUND

Previous Research on STABLE-2007 Factors/Subscales

Brouillette-Alarie & Hanson (2015)

Brouillette-Alarie, S., & Hanson, R. K. (2015). Comparaison de deux mesures d'évaluation du risque de récidive des délinquants sexuels [Comparison of two recidivism risk measures for sexual offenders]. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 47*(4), 292–304. https://doi.org/10.1037/cbs0000019

Etzler et al. (2020)

Etzler, S., Eher, R., & Rettenberger, M. (2020). Dynamic risk assessment of sexual offenders: Validity and dimensional structure of the Stable-2007. *Assessment, 27*(4), 822–839. https://doi.org/10.1177/1073191118754705

Brien-Robidoux (2022)

Brien-Robidoux, E. (2022). Dynamic risk assessment of sexual offenders in the real world: Study of predictive validity and dimensionality of the Static-99R and Stable-2007 on a French-speaking Canadian sample. *Doctoral Dissertation*, Université de Montréal.

Ditchfield et al. (2025)

Ditchfield, R. E., Azizian, A., D'Orazio, D., & Chankin, L. (2025). Assessing the dimensional structure and measurement invariance of STABLE-2007 scores in a California sample of adult men on parole for sex crimes. *Unpublished manuscript*.

18

BACKGROUND

Analytic Choices Across STABLE-2007 Studies

Comparison of methodological decisions — prior studies vs. current study

Analytic Choice	Brouillette-Alarie & Hanson (2015)	Etzler et al. (2020)	Brien-Robidoux (2022)	Ditchfield et al. (2025)	Current Study
Framework	EFA	EFA	EFA	ESEM	ESEM
Correlations	polychoric	polychoric	polychoric	polychoric	polychoric
Estimation	WLSMV	WLSMV	WLSMV	WLSMV	WLSMV
Factors	scree, PA, MAP	E > 1, scree, PA, MAP, model fit	E > 1, PA, model fit	PA, MAP, model fit, parsimony	E > 1, scree, PA, parsimony
Rotation	Geomin	Oblique Equamax Ortho. Equamax	Geomin	Oblimin	Geomin
Loadings	> .40	> .40	> .40	> .40	> .40
Fit Indices	RMSEA < .06 CFI > .95	RMSEA < .06, .08 CFI > .95 Tucker-Lewis > .95	RMSEA < .06 CFI > .95 Tucker-Lewis > .95	R-RMSEA < .05, .08 R-CFI > .95 SRMR < .08	RMSEA < .08 CFI > .90, .95 SRMR < .08 Tucker-Lewis > .90, .95
Variables	+STATIC sexual crim. +STATIC general crim. - Rejection/Lonely	- Significant Social Influences			

19

BACKGROUND

STABLE-2007 Factor Assignments Across Studies

How each item was classified by factor across prior studies

STABLE Item	Brouillette-Alarie & Hanson	Etzler et al. (2020)	Brien-Robidoux (2022)	Ditchfield et al. (2025)
No. of Factors	2	3	2	2
Sig Social Influences	Antisociality	—	Antisociality	General Criminality
Relationship Stability	No factor	Antisociality	No factor	No factor
Emo ID with Kids	Sexual Criminality	Sexual Deviance	Sexual Deviance	Sexual Criminality
Hostility to Women	Antisociality	Antisociality	Antisociality	General Criminality
Loneliness/Rejection	—	Antisociality	Antisociality	General Criminality
Lack of Concern	Antisociality	Antisociality	Antisociality	General Criminality
Impulsive Acts	Antisociality	Antisociality	Antisociality	General Criminality
Poor Cognitive Problem-Solving	Antisociality	Antisociality	Antisociality	General Criminality
Neg Emotionality	Antisociality	Antisociality	Antisociality	General Criminality
Sex Preoccupation	Sexual Criminality	Hypersexuality	Sexual Deviance	Sexual Criminality
Sexualized Coping	Sexual Criminality	Hypersexuality	Sexual Deviance	Sexual Criminality
Deviant Sex Interests	Sexual Criminality	Sexual Deviance	Sexual Deviance	Sexual Criminality
Cooperation w/ Superv.	Antisociality	Antisociality	Antisociality	General Criminality

20

FACTOR ANALYSIS

Current Study: Model Selection Rationale

Why these factor solutions? Criteria: parsimony, interpretability, theoretical coherence, admissibility

ACUTE-2007 → 1-Factor Selected	STABLE-2007 → 2-Factor Selected
3-factor Excellent fit (RMSEA ≈ .009) but only ~2 items per factor — psychometrically unstable for 7 items	**4–5 factor** Excellent fit (RMSEA ≈ .03–.04) but fewer than 3 items per factor — statistical over-extraction
2-factor Admissible; however, 2-factor lacks theoretical basis	**3-factor** Borderline defensible; theoretically inconsistent factor structure across time points
1-factor ✓ Most parsimonious and theoretically coherent — all 7 items reflect unified acute risk escalation (loadings .53–.72)	**2-factor ✓** Best balance F1: General Criminality (8 items) F2: Sexual Criminality (4 items); Stable across T1 & T2

21

FACTOR ANALYSIS

FA: Final Factor Loadings — Selected Models

ACUTE-2007 | 1-Factor (*n* = 2,289)

Item	F1 Loading
Hostility	.724
Social Supports	.711
Rejection of Supervision	.681
Emotional Collapse	.655
Victim Access	.588
Sexual Preoccupation	.542
Substance Abuse	.533

All 7 items: single unified acute risk escalation construct

STABLE-2007 | 2-Factor (*n* = 669)

Item	F1	F2
Cooperation w/ Supervision	.809	—
Problem Solving	.711	—
Social Influences	.674	—
Lack of Concern	.664	—
Impulsivity	.666	—
Negative Emotionality	.624	—
Hostility Toward Women	.654	—
Loneliness	.433	—
Relationship Stability	(0.346)	—
Sexual Preoccupation	—	.795
Sex as Coping	—	.796
Deviant Sexual Interests	—	.622
Emotional Identification	—	.503

F1: General Criminality | F2: Sexual Criminality | — = loading < .40

22

FACTOR ANALYSIS

Internal Consistency of Factors

Ordinal alpha (α); α ≥ .70 adequate • α ≥ .80 good • α ≥ .90 excellent

ACUTE-2007 | 1-Factor (*n* = 2,289)

Time Point	Items	Ordinal α	Interpretation
Time 1	7	.82	Good
Time 2	7	.82	Good
Time 3	7	.83	Good

STABLE-2007 | 2-Factor (*n* = 669)

Factor / Time	Items	Ordinal α	Interpretation
F1 — T1	8	.88	Good
F1 — T2	8	.87	Good
F2 — T1	4	.79	Adequate
F2 — T2	4	.78	Adequate

Ordinal alpha values indicated good internal consistency for ACUTE-2007 (α = .82–.83) and STABLE-2007 Factor 1: General Criminality (α = .87-.88). Factor 2: Sexual Criminality showed adequate consistency (α = .78–.79). These values support the reliability of the selected factor solutions.

23

From Factor Structure to Temporal Stability

Factor Analysis

What is the factor structure?
ACUTE-2007: 1 factor
STABLE-2007: 2 factors
Loadings ≥ .40 and theoretically meaningful — confirmed

→

leads to

Measurement Invariance

Does that structure hold over time?
Are the same items measuring the same constructs at each assessment?
Is score change real — or a measurement artifact?

FA establishes the factor structure at a single time point. MI tests whether that structure is equivalent across time — a prerequisite for valid longitudinal comparisons of dynamic risk scores.

24

MEASUREMENT INVARIANCE

Why Does Measurement Invariance Matter?

STABLE-2007 and ACUTE-2007 are designed for repeated administration to monitor changes in risk over time. Valid score comparisons require that the tools measure the same underlying constructs at each time point — this is measurement invariance.

Configural Invariance	Metric Invariance	Scalar Invariance
Do the same items load on the same factors across time points?	Are factor loadings (item–factor relationships) equal across time?	Are item intercepts equivalent, enabling comparison of latent means?

Without invariance: score changes may reflect shifts in how the tool measures — not genuine changes in risk — undermining the validity of repeated assessments and correctional decisions.

25

MEASUREMENT INVARIANCE

MI: Analytic Decisions

How invariance was tested in this study

Framework
Factor structures from FA entered into CFA-based longitudinal invariance models

Estimation
WLSMV — appropriate for ordinal polytomous items; used consistently across FA and MI stages

Invariance Sequence
Configural → Metric → Scalar (hierarchical; each model progressively more constrained)

Comparison Criteria
ΔRMSEA, ΔCFI, ΔSRMR across nested models — chi-square difference tests avoided (sensitivity with large *N*)

Invariance Decision Thresholds

Cheung & Rensvold (2002); Chen (2007)

Criterion	Threshold for Non-Invariance
Δ RMSEA	> +.015
Δ CFI	> −.010
Δ SRMR (metric)	> +.030
Δ SRMR (scalar)	> +.015

Target: full scalar invariance | Partial invariance = specific items non-equivalent

26

MEASUREMENT INVARIANCE

ACUTE-2007: Measurement Invariance Results

n = 2,289 • Three time points • 7 items • 1-Factor Model (Selected)

1-Factor Model (Selected)

Model	χ^2 (df)	RMSEA	CFI	SRMR	ΔRMSEA	ΔCFI	ΔSRMR
Configural	647.04 (42)	.079	.960	.074	—	—	—
Metric	655.71 (54)	.070	.961	.075	**−.009**	**+.001**	**+.001**
Scalar	662.55 (80)	.056	.962	.075	**−.014**	**+.001**	**.000**

Configural: Same 1-factor structure at all 3 time points. **Metric:** Item loadings equivalent across time (ΔRMSEA = −.009).
Scalar: Item intercepts also equivalent (ΔRMSEA = −.014) — latent mean comparisons valid.

What the indices show: RMSEA/CFI confirm each constrained model fits well in isolation. Negative ΔRMSEA means fit improves as constraints are added — a strong indicator of invariance. ΔCFI and ΔSRMR are negligible.

Full scalar invariance confirmed — ACUTE-2007 1-factor model across all three time points.
ΔRMSEA ≤ .014 • ΔCFI ≤ .001 • ΔSRMR ≤ .001 — all well within acceptable thresholds.

27

MEASUREMENT INVARIANCE

STABLE-2007: Measurement Invariance Results

n = 669 • Two time points • 13 items • 2-Factor Model (Selected)

2-Factor Model — F1: General Criminality | F2: Sexual Criminality

Model	χ^2 (df)	RMSEA	CFI	SRMR	ΔRMSEA	ΔCFI	ΔSRMR
Configural	458.82 (106)	.072	.966	.081	—	—	—
Metric	462.55 (116)	.068	.967	.081	**−.004**	**+.001**	**.000**
Scalar	463.16 (126)	.064	.968	.081	**−.004**	**+.001**	**.000**

Factor 1 – General Criminality (8 items)
cooperation w/ supervision, problem solving, social influences, lack of concern, impulsivity, negative emotionality, hostility toward women, loneliness

Loadings .43–.81 • Consistent across T1 & T2

Factor 2 – Sexual Criminality (4 items)
sexual preoccupation, sex as coping, deviant sexual interests, emotional identification

Loadings .50–.80 • Consistent across T1 & T2

Full scalar invariance confirmed — STABLE-2007 2-factor model across two time points.
ΔRMSEA = −.004 • ΔCFI ≤ .001 • ΔSRMR ≤ .001 — fit improved with added constraints; factor structure is highly stable.

28

SUMMARY

Summary & Implications

Factor Analysis

ACUTE-2007: 1-factor structure — unified acute risk escalation construct (loadings .53–.72)
STABLE-2007: 2-factor structure — General Criminality (8 items; loadings .43–.81) + Sexual Criminality (4 items; loadings .50–.80)
Both solutions psychometrically defensible and theoretically meaningful

Measurement Invariance

Full scalar invariance: ACUTE 1-factor + STABLE 2-factor across all time points
Δ fit indices: all well within thresholds (ΔRMSEA ≤ .014, ΔCFI ≤ .001)
Factor structures replicated consistently — not time-point-specific artifacts

Implications for Evidence-Based Practice

Score changes over time reflect genuine changes in dynamic risk
STABLE-2007 subscales indicate relative placement on general criminality and sexual criminality

FORTHCOMING: Percentile Tables for subscales

Thank You

Questions

Seung C. Lee
Research Advisor
Public Safety Canada
seungclee18@gmail.com

Measurement Invariance Over Time in Dynamic Risk Assessment Tools
CPA 2026

Risk to Reoffend Changes Over Time: A study of the ACUTE-2007 and STABLE-2007

Scan for slides

QR code here

Kelly M. Babchishin, Seung C. Lee, & R. Karl Hanson

Saturday June 6th, 2026 | 9AM–9:55AM

CPA, Montreal, Québec

SEXUALLY HARMFUL BEHAVIOURS Research Lab

SAARNA *Safety in Numbers*

Carleton University

31

Progress Monitoring in the Criminal Justice System

WHY PROGRESS MONITORING?

- Is the intervention working?
- Is your client getting better or worse?
- Foundation of evidence-based intervention *(CPA, 2018)*

INVOLVES REPEATED ASSESSMENT THROUGHOUT

FOCUS IN CRIMINAL JUSTICE

- Outcome of interest (recidivism) is rarely observable.
- Instead, focus is on intermediate targets.
- Dynamic tools: assumed to be "intermediate targets"
  - Changeable and related to the outcome
  - But... are they?

Progress monitoring helps us make timely, data-informed decisions to improve outcomes.

Carleton University

32

Limitations of Previous Research

- Few studies, some with **small samples**
- Few **sex offending** studies
- Few studies with **more than two time points** and **multiple tools**

Carleton University

33

Purpose of the Current Talk

Are the ACUTE-2007 and STABLE-2007 intermediate targets?

- i.e., Are they changeable?

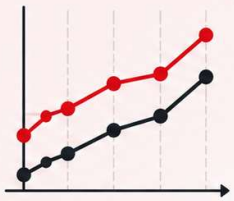
Why should we care? Understanding who changes and who does not can improve supervision and intervention planning.

Carleton University

34

Study Objectives

Our study aims to understand change in **ACUTE-2007** and **STABLE-2007** scores over time and the factors that influence it.



01 Examine longitudinal trajectories of **ACUTE-2007** and **STABLE-2007** scores.

02 Estimate variability in individual change trajectories using **Hierarchical Linear Modelling (HLM)**.

03 Identify **predictors of change** (e.g., Static-99R, baseline risk, supervision factors).

04 Examine meaningful change using **Reliable Change Index (RCI)** and risk category transitions.

Carleton University

35

Participants

4,063
men with a history of sexual offences

At least one ACUTE-2007 or STABLE-2007 assessment
Assessed by community supervision officers

Average age: 41 years
(SD = 14)

Unselected sample from BC Corrections
Community supervision between 2005–2013
(not treatment)

Measures

ACUTE-2007

- Posited to change **faster**

Example item
Victim access

VS.

STABLE-2007

- Posited to change **slower**

Example item
Lack of concern for others

Carleton University

36

At Least Three Assessments

STABLE-2007

Total assessments
8,192

Total persons
1,947

RESTRICTED TO
≥ 3 Assessments
ACUTE and/or STABLE

ACUTE-2007

Total assessments
57,647

Total persons
3,584

This restriction reduces our sample to ensure **reliable measurement of change over time.**

Carleton University

37

Analyses

Hierarchical Linear Modeling (HLM)

- A way of modeling change
- HLM models a growth curve for each individual
- Allows comparisons of individual trajectories/ growth curves according to group membership

Quantifying "Meaningful Change"

1 Index 1:
Changing risk categories

2 Index 2: $RCI = \frac{X_2 - X_1}{S_{diff}}$

- ≥ 1.96: reliable **deterioration**
- ≤ -1.96: reliable **improvement**
- -1.96 to 1.96: **no change**

Key Definitions

X_1	Initial score
X_2	Follow-up score
S_{diff}	Expected measurement error

In my study:
$S_{diff} = \sqrt{2} \times \sigma_{residual}$
(from HLM model)

Carleton University

38

Unconditional Models
(Intercept-Only Models)

Assess the degree of between-person variability in scores to justify the use of multilevel modeling.

What was estimated?
The overall mean (M), standard deviation (SD), and intraclass correlation coefficient (ICC).

Outcomes

ACUTE-2007 Total Scores	ICC = .704
STABLE-2007 Total Scores	ICC = .829
STABLE-2007 Factor 1 (General Criminality)	ICC = .818
STABLE-2007 Factor 2 (Sexual Criminality)	ICC = .806

ICC reflects the **proportion of variance** in **scores due to between-person differences.**

Note: M and SD are produced in intercept-only models but are not interpreted. Our focus is on the ICC to assess between-person variability.

Carleton University

39

Unconditional Growth Models

These models estimate the **average change in scores over time** and the **variability in change across individuals.**

Time Metric

Time in months

Model Specification

- Random intercept and random slope models estimated using restricted maximum likelihood (REML).
- For the STABLE-2007 total score and Factor 2 (Sexual Criminality) models, convergence was achieved using the *bobyqa* optimizer.

Carleton University

40

Findings

Where do individuals start? → How does score change over time? → What explains the variance?

Measure	Intercept (SE)	Time Slope (SE)	Intercept–Slope r	Marginal R^2 (Fixed Effects Only)	Conditional R^2 (Fixed + Random Effects)
ACUTE Total	2.186 (0.040)***	-0.024 (0.002)***	-.47	.015	.862
STABLE Total	8.024 (0.113)***	-0.018 (0.003)***	-.30	.006	.884
F1 General Criminality	4.619 (0.082)***	-0.010 (0.002)***	-.28	.003	.871
F2 Sexual Criminality	1.680 (0.038)***	-0.008 (0.001)***	-.29	.005	.896

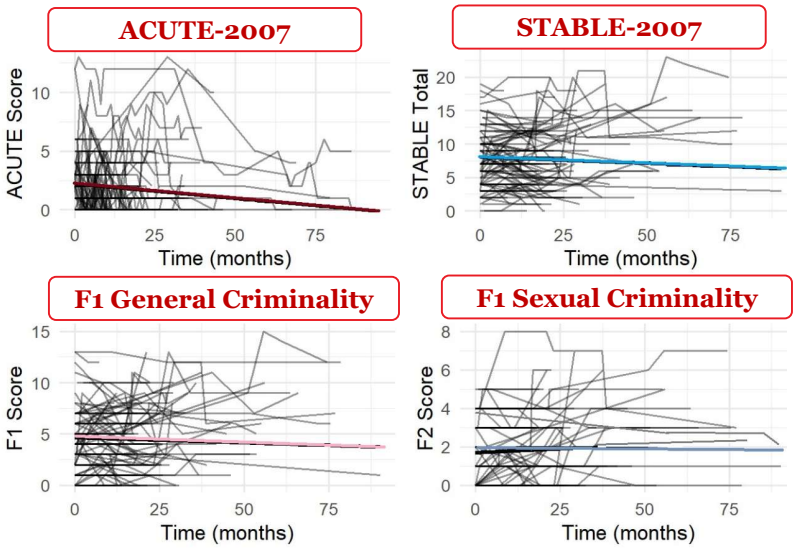
Note: Time was measured in months. Marginal R^2 reflects variance explained by fixed effects only; conditional R^2 reflects variance explained by both fixed and random effects.

*** $p < .001$.

Carleton University

41

Individuals Trajectories



Both acute and stable showed modest but **statistically significant decreases** over time, while also demonstrating **substantial individual variability** in both baseline scores and rates of change.

Carleton University

42

Summary of HLM Change Analyses

An average, ***slight*** improvement over the course of community supervision

Change was **linear**

But there was **significant variability** across individuals

...Are there any variables that can explain these different patterns of change?

Carleton University

43

Predictors of Patterns of Change

Variable	ACUTE-2007 Statistically significant?	STABLE-2007 Statistically significant?	Direction
Age	Yes (after controlling for risk tools: No)	Yes (after controlling for risk tools: No)	Older folks changed more than younger folks
ACUTE-2007/ STABLE-2007	Yes	Yes	Higher risk scores were associated with lower rates of change
Static-99R	Yes	Yes	

Yes = Statistically significant No = Not statistically significant

Carleton University

44

Summary of HLM Change Analyses

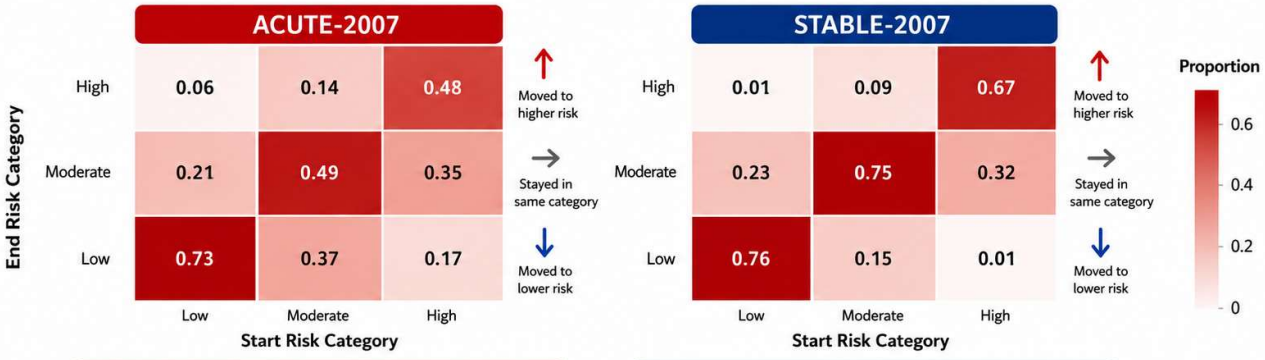
- An average, ***slight*** improvement over the course of community supervision
- Change was **linear**
- But there was **significant variability** across individuals
- Change differ depending on **age** and **risk scores**
- But is this level of change truly **meaningful**?

Carleton University

45

Meaningful Change in Risk Categories

Most individuals **remained stable**, but **meaningful transitions** across risk categories occurred.



Darker cells indicate a larger proportion of individuals.

Carleton University

46

Reliable Change Index (RCI) Findings

Meaningful individual-level change was observed beyond *measurement error*, particularly for ACUTE-2007 scores.

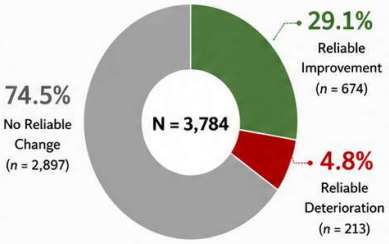
What is the RCI?

The Reliable Change Index (RCI) indicates whether change exceeds expected measurement error.

- **RCI < –1.96** Reliable improvement
- **–1.96 to 1.96** No reliable change
- **RCI > 1.96** Reliable deterioration

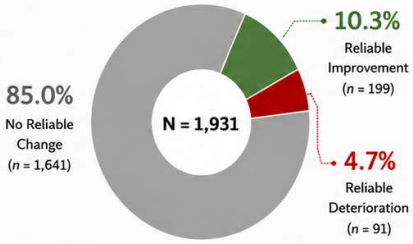
ACUTE-2007

Greater meaningful change observed



STABLE-2007

Greater stability overall



ACUTE-2007 demonstrated greater dynamic responsiveness, whereas STABLE-2007 scores showed comparatively greater stability over time.

Carleton University

47

Conclusions

What we learned about risk trajectories over time

1. Most Individuals Remained Stable
2. Meaningful Movement Was Observed
3. ACUTE-2007 Showed Greater Dynamic Responsiveness
4. STABLE-2007 Showed Greater Stability
5. Change Depends on Some Individual Characteristics

These findings support the importance of routine monitoring and repeated assessment across all risk levels.

Carleton University

48

Thank you.

Questions?

QR CODE

Scan for slides

Kelly.Babchishin@carleton.ca

https://carleton.ca/sexuallyharmfulbehaviourslab/

SEXUALLY HARMFUL BEHAVIOURS Research Lab

Carleton University

49