

A Field Study on Static-99R Interrater Reliability by Canadian Police Personnel

By Heather Burke, MA, PhD Candidate

Department of Psychology

Date: 2024-10-18

Acknowledgments



The points of view expressed in this presentation do not necessarily represent the views of the Edmonton Police Service. This presentation draws on research supported by the Social Sciences and Humanities Research Council.



Social Sciences and Humanities
Research Council of Canada

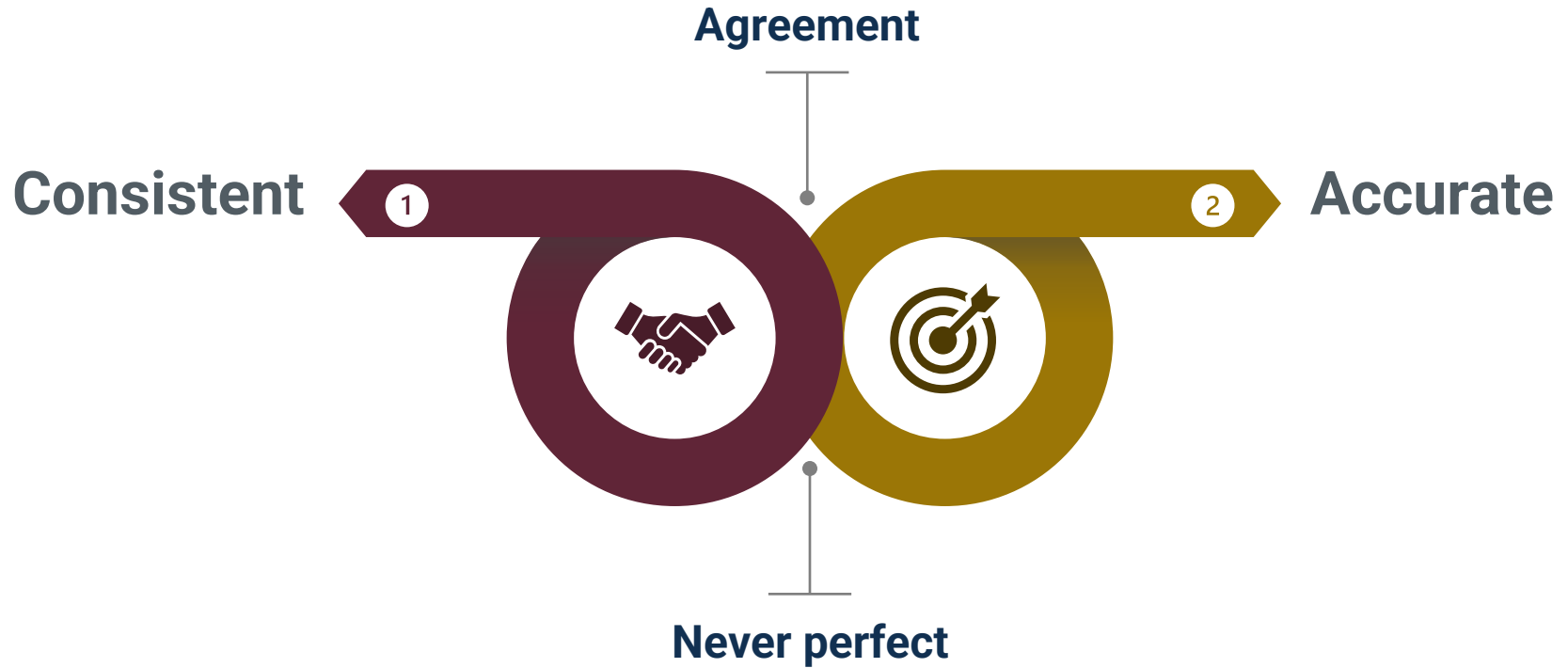
Conseil de recherches en
sciences humaines du Canada

Canada 

Introduction

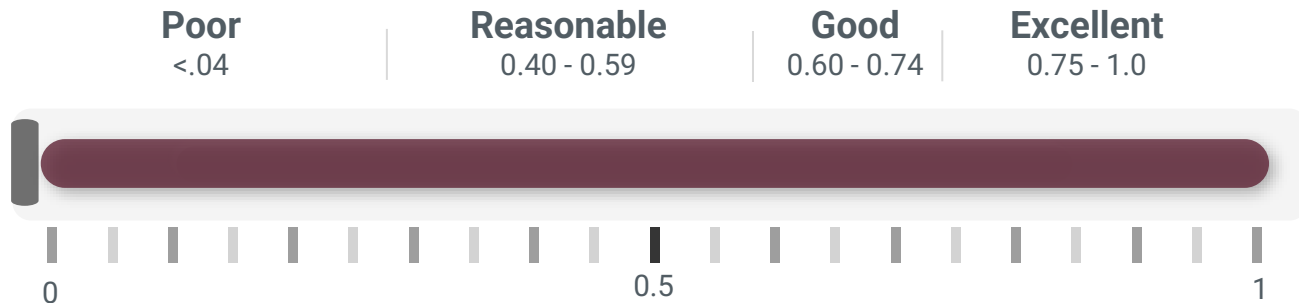


Interrater Reliability



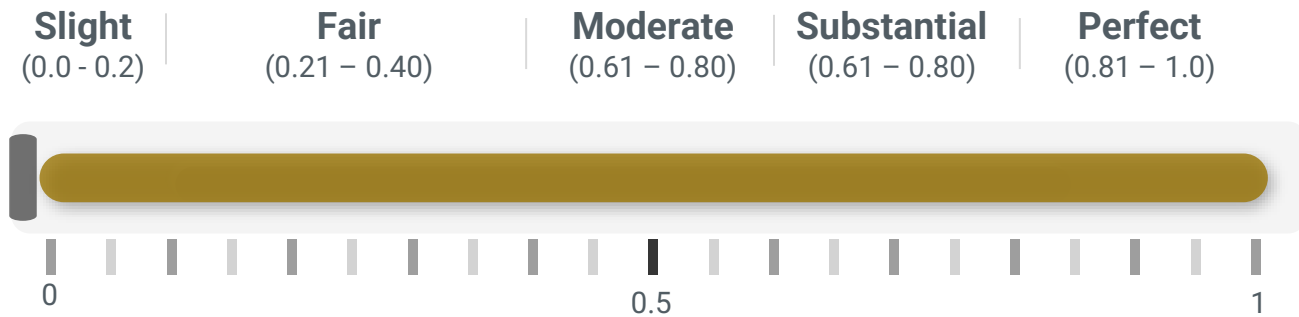
Interrater Reliability Statistics

Total Score



**Intraclass
Correlation
Coefficient
(ICC)**

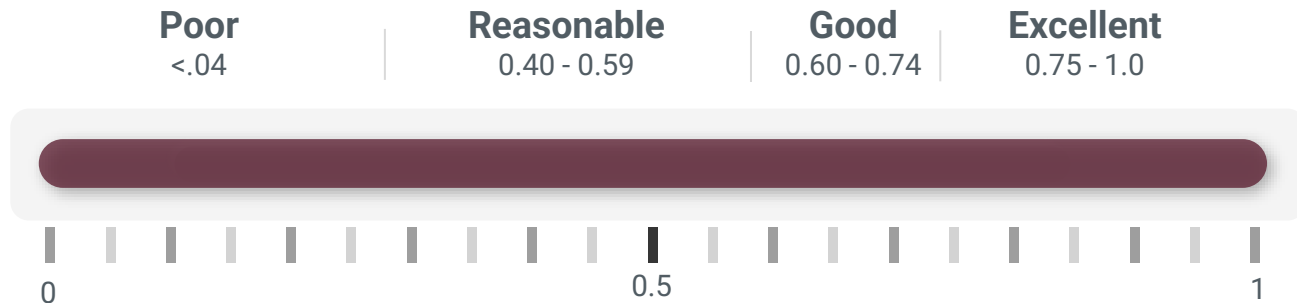
Item Level



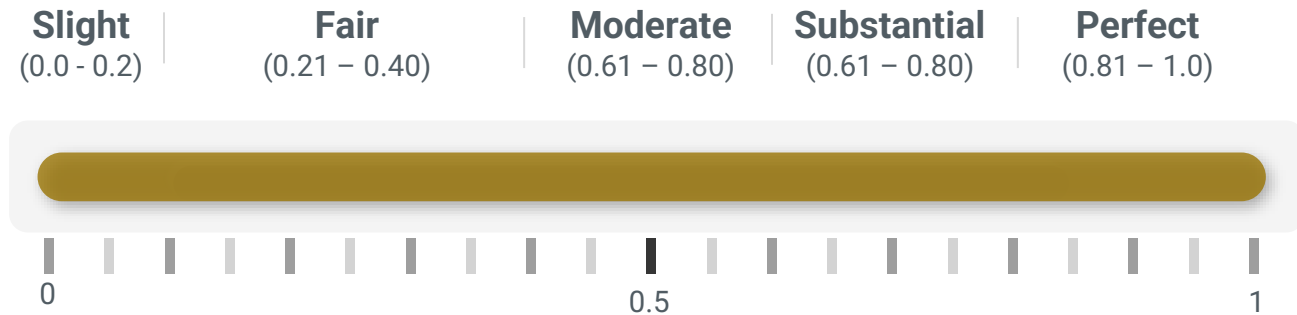
**Cohen's
Kappa**

Interrater Reliability Statistics

Total Score



Item Level



Percent Agreement

SEM (Precision)

Risk Assessment in Policing

01

**Traditional
Practice**

02

**Gap in
Policing**

03

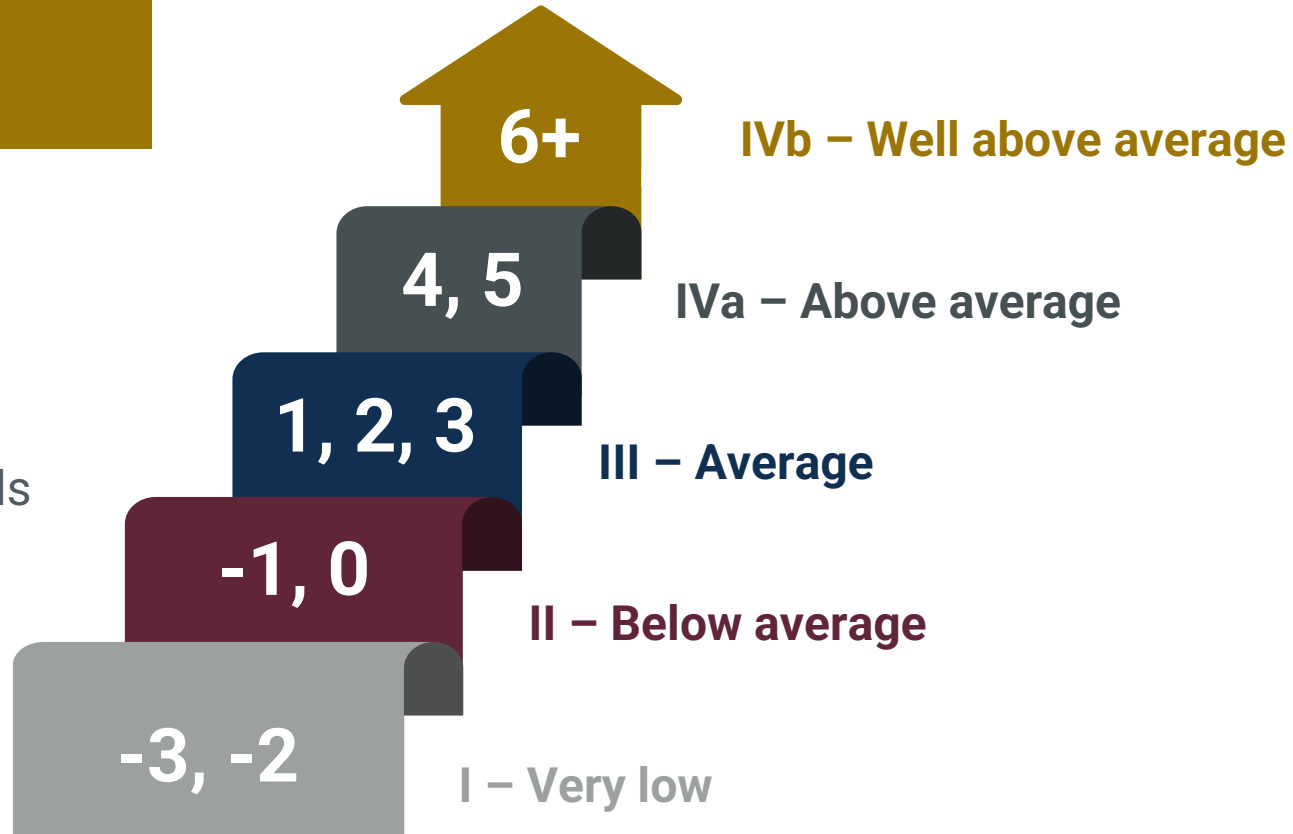
**Tools in
Policing**

04

**Potential
Solution**

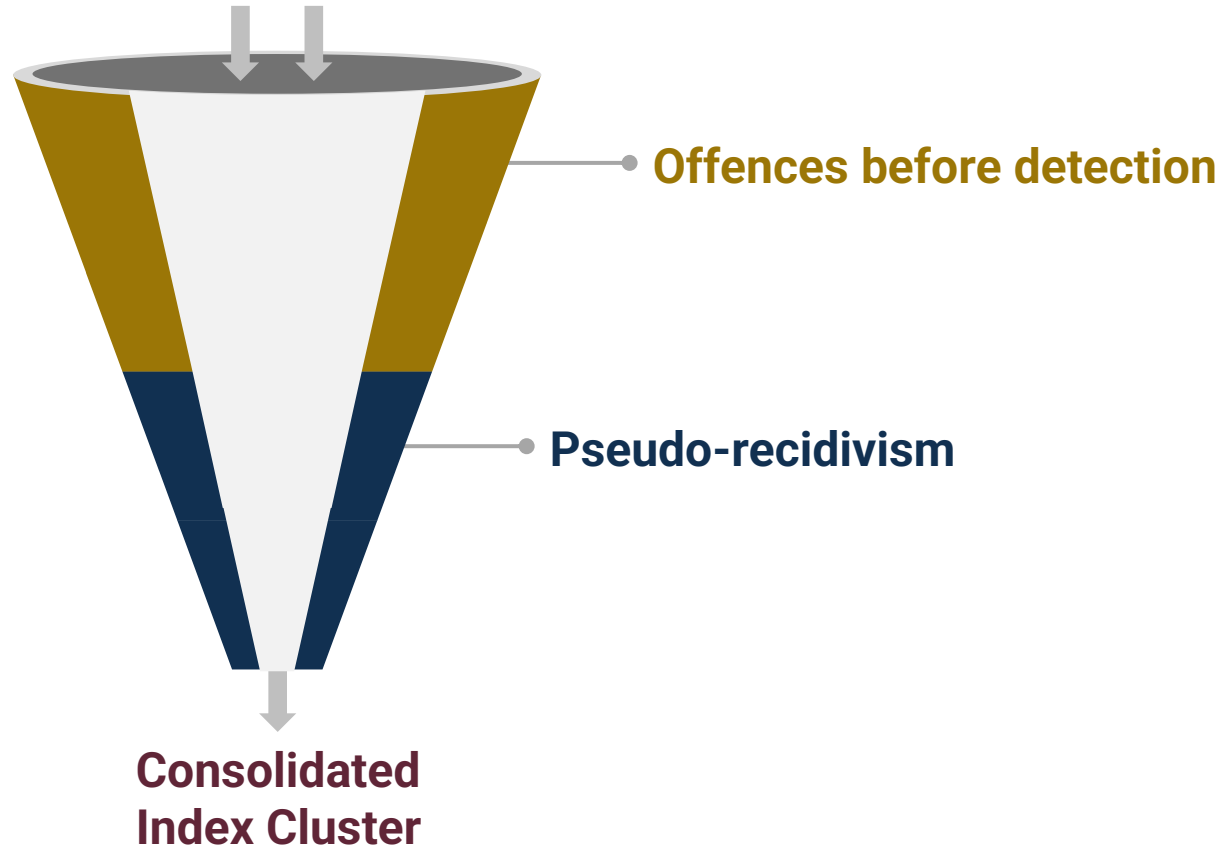
Static-99R

- File information
- No advanced degree
- Total score: -3 to 12
- Five nominal risk levels



Static-99R Scoring

**Index
Clusters**



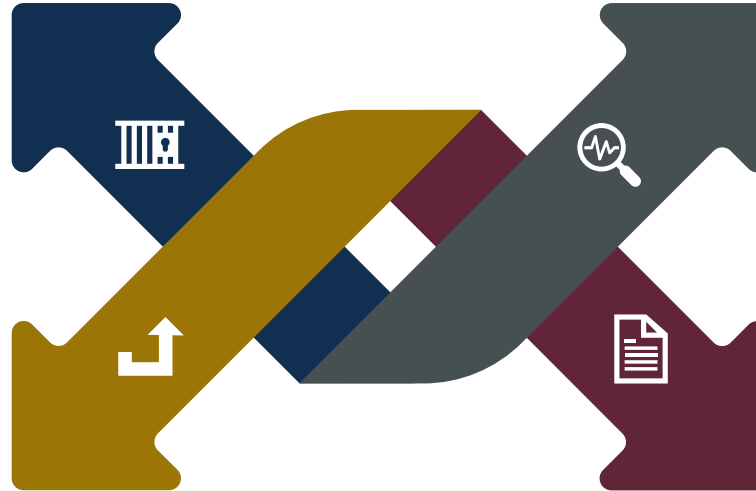
Static-99/99R Interrater Reliability Field Studies

Non-Clinicians

Clinicians

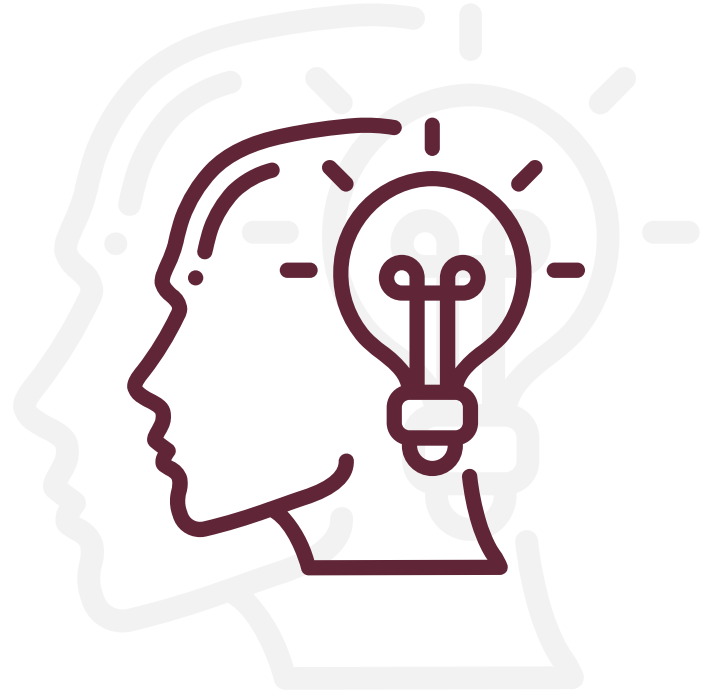
Experience

**Researcher:
Policing**



Research Objectives

- 01 Interrater reliability**
- 02 Score discrepancies on risk categorization**
- 03 Training comparison**
- 04 Post-training analysis**



Evaluators and Static-99R Training

Formal Training

- Led by Static-99R co-creator R. Karl Hanson
- **Attendees:** 11 police personnel

1

- 1 graduate level
- 1 undergraduate level

Researchers

2

- 8 **formally** trained officers
- 2 **informally** trained officers
- 1 formally trained administrator

Researcher/ Police

Case Sample

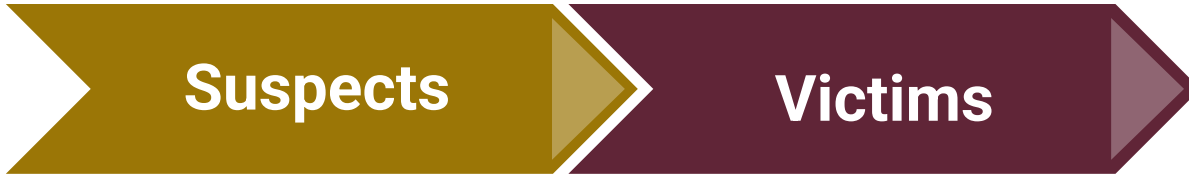
105 cases



M age (earliest): 23.19 years

Age range: 3 to 53 years

Most female ($n = 125$; 95.4%)



M age (at report): 36.67 years

Age range: 17 to 89 years

Most male ($n = 103$; 98.1%)



131 victims

Sources of Information

Municipal Police Records

• Niche Records Management System

Federal Police Records

• Canadian Police Information Centre
(CPIC)

Categories of Variables

• Offence, suspect, and victim
characteristics

Procedure

07 – Researcher B

06 – Researcher A

05 – PDF Retrieval



01 - Ethics

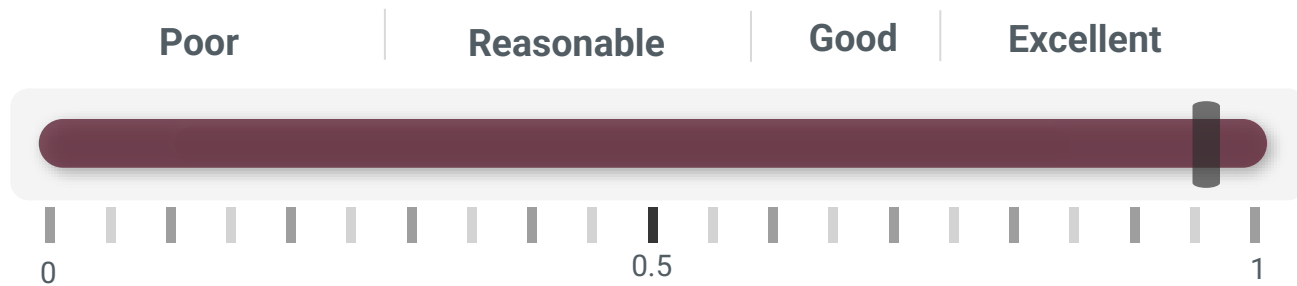
02 – Training

03 – Police Coding

04 – Case Logging

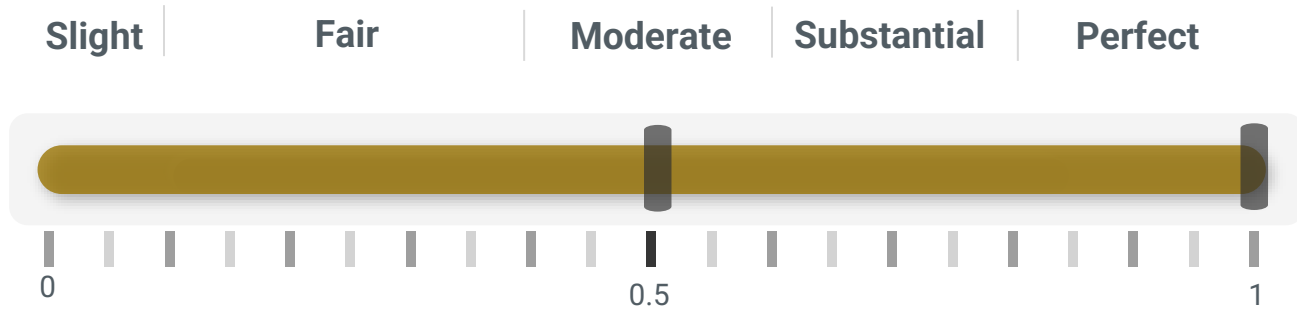
Agreement Between **Researchers**

Total Score (ICC)



- **80%** exact agreement
- **SEM** = .493

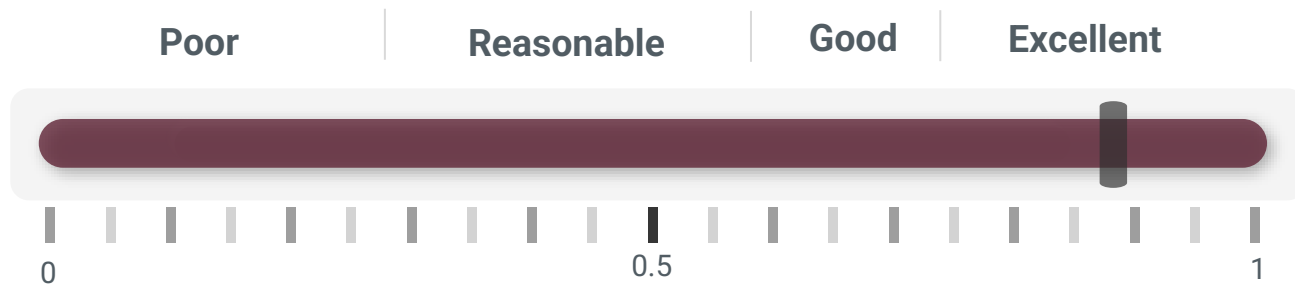
Item Level (κ)



- **90% – 100%** agreement

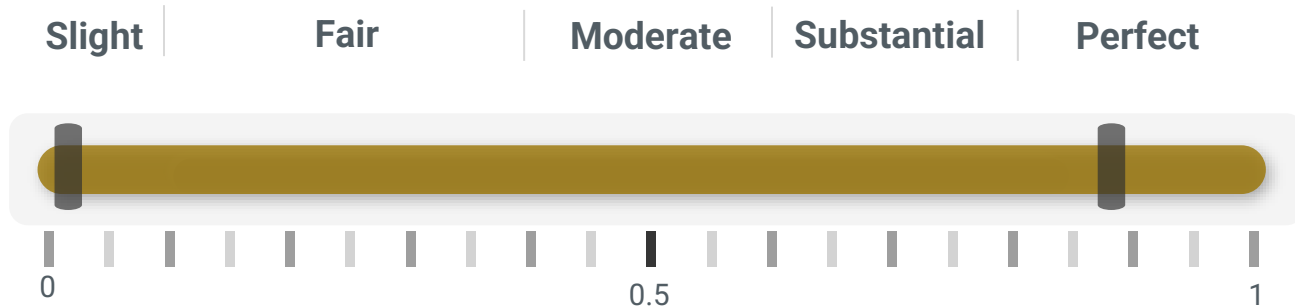
Agreement Between a **Researcher and Police**

Total Score (ICC)



- **47%** exact agreement ↓
- **SEM** = .752 ↑
- **24%** risk difference

Item Level (κ)



- **72%** item discrepancy
- **60% – 96%** agreement ↓

Total Score and Item Level Agreement Between **Researcher and Police**

	Range	Range	Range
	.019 – .896	60 – 96	Slight – Almost Perfect
Static-99R Item	Kappa (ICC _{2,1})	% Agreement	Interpretation
1. Age at index report	.896***	93	Almost perfect
2. Ever lived with a lover	.019	60	Slight
3. Index non-sexual violence (charges or arrests)	.349***	82	Fair
4. Prior non-sexual violence (any convictions)	.768***	91	Substantial
5. Prior sex offences	.665***	92	Substantial
6. 4+ sentencing dates (excluding index)	.843***	94	Almost perfect
7. Any convictions for non-contact sex offences	.314**	96	Fair
8. Any unrelated victims	.266**	76	Fair
9. Any stranger victims	.533***	90	Moderate
10. Any male victims	.294**	92	Fair
Total score	(.893)***	47	Excellent

Note. $n = 105$; * $p < .05$; ** $p < .01$; *** $p < .001$. ICC = intraclass correlation coefficient.

Total Score and Item Level Agreement Between **Researcher and Police**

	Range	Range	Range
	.019 – .896	60 – 96	Slight – Almost Perfect
Static-99R Item	Kappa (ICC _{2,1})	% Agreement	Interpretation
1. Age at index report	.896***	93	Almost perfect
2. Ever lived with a lover	.019	60	Slight
3. Index non-sexual violence (charges or arrests)	.349***	82	Fair
4. Prior non-sexual violence (any convictions)	.768***	91	Substantial
5. Prior sex offences	.665***	92	Substantial
6. 4+ sentencing dates (excluding index)	.843***	94	Almost perfect
7. Any convictions for non-contact sex offences	.314**	96	Fair
8. Any unrelated victims	.266**	76	Fair
9. Any stranger victims	.533***	90	Moderate
10. Any male victims	.294**	92	Fair
Total score	(.893)***	47	Excellent

Note. $n = 105$; * $p < .05$; ** $p < .01$; *** $p < .001$. ICC = intraclass correlation coefficient.

Comparing Training Methods

Formally Trained

84 cases

Excellent ($ICC_{2,1} = .913$)

44%

75% ($n = 63$)

44%: 21 of 47 cases

SEM = .679

VS

Informally Trained

18 cases

Excellent ($ICC_{2,1} = .884$)

67%

56% ($n = 10$)

67%: 4 of 6 cases

SEM = .785

Total Score Agreement

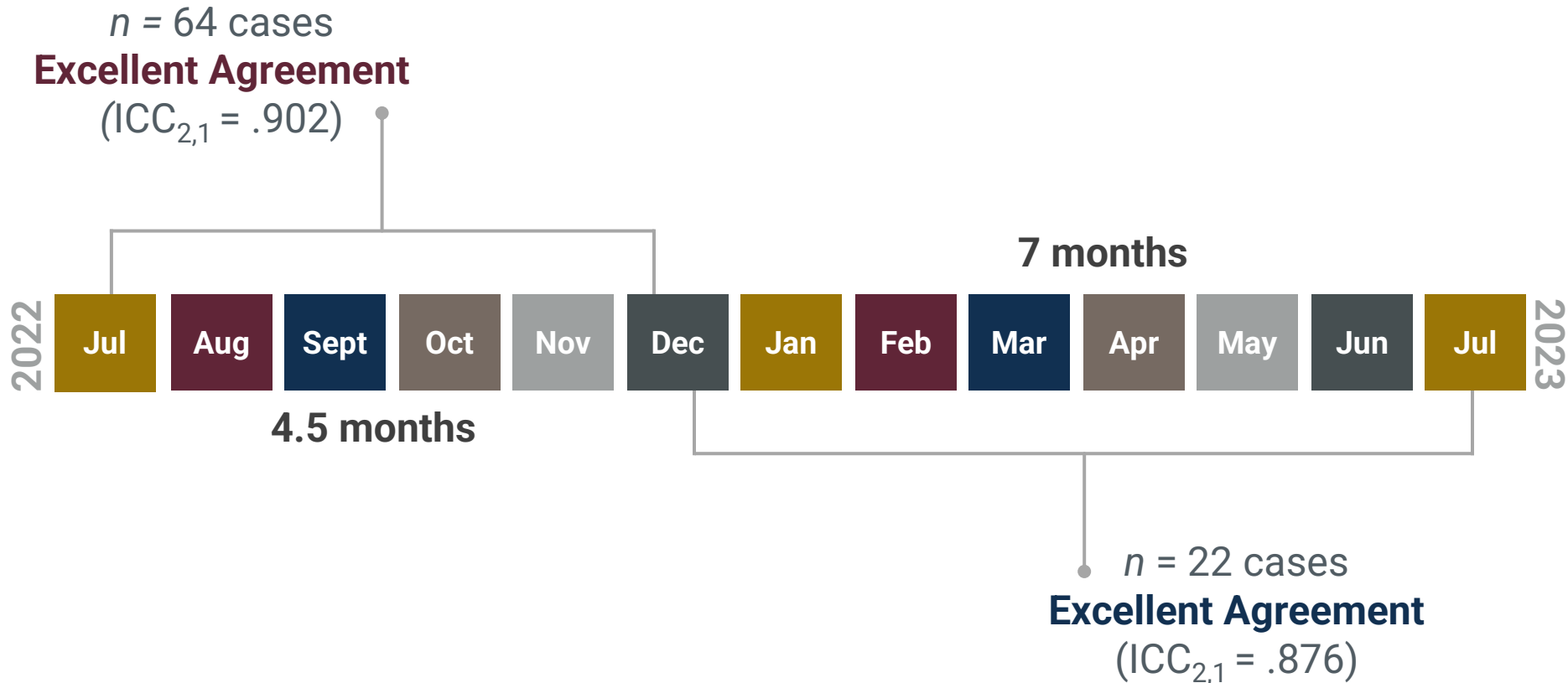
Identical Total Score

Item Discrepancy

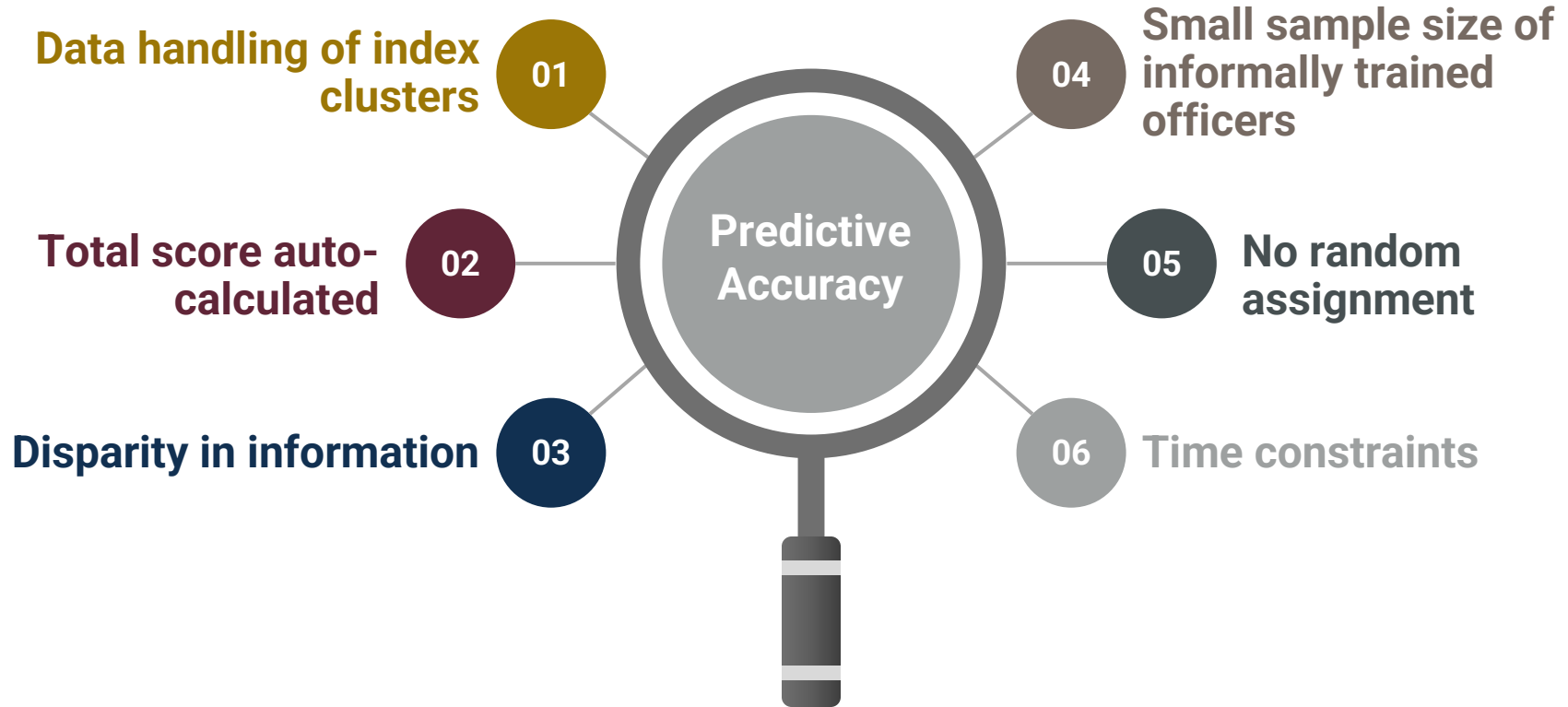
Risk Discrepancy

Precision

Total Score Reliability Among Police After **Formal Training**



Limitations and Future Directions



Conclusions

- Crucial **first step**
- Findings **generally support** the reliable use of the Static-99R
- Possible **cost-effective** training methods
- **Foundation for** future research on **predictive accuracy**

THANK YOU!

Heather Burke, Carleton University
heatherburke3@cmail.carleton.ca