

The Level of Service Assessments: A Question Confusion, Selectivity, and Misrepresentation of Evidence in Baird (2009)

Don Andrews

ICCA, Sept 15, 2009, Orlando Florida

Baird, Arling, Lerner et al in the 1980s/90s

- Baird et al were giants in the early development of applications of the psychology of criminal conduct (The Andrews & Bonta [2006, 2010] list of major contributors)
- The Wisconsin group
 - Risk. (8-12 items with site-specific cut-off scores and weights).
 - Need. A few dynamic items with criminogenic factors undifferentiated.
 - Responsivity: Client Management Classification Subgroups based on variable intervention strategies and intermediate targets

Now we have Baird (2009): “the point ... is to use actual data to clarify issues for ... those ...who select and implement assessment systems”

- To make the Wisconsin/NCCD products look good to those who purchase assessment instruments and services
- The LS evidence was not reviewed in Baird (2009). The LS/CMI manuals are not even referenced
- The meta-analytic evidence re the relative validity of Wisconsin risk and LS risk/need was not reviewed by Baird
- So, what “actual data” were used in Baird (2009)?
Very little, as you will see.

Misrepresentation of Evidence and Confusion in Baird (2009): Baird (2009) has damaged the reputation of LS assessment instruments

- 1) A highly selective review with deliberate misrepresentation of the evidence re
 - the Wisconsin/NCCCD instruments
 - the Level of Service (LS) approach
- 2) Confusion in Baird (2009)

Confusion in Baird (2009)

- 2) Confusion in Baird (2009) re
 - 3rd and 4th generation instruments
 - static and dynamic risk factors
 - criminogenic needs
 - strengths / protective factors
 - risk reduction
 - risk and need scores should not be combined (a Baird rule)

Confusion is Dramatic in Baird (09)

- Baird (2009, p. 4) agrees with Gottfredson and Moriarty (06) ``entirely`` --- “risk and needs should not be combined as a composite measure”
- In fact, **G & M (06) did not say that static and dynamic risk factors should not be combined**
- In fact, **G & M (2006, p. 195) say “criminogenic needs should be assessed”**
- G & M agree with and recommend the Bonta (02) guidelines for risk assessment, including ‘derived from relevant theory,’ ‘sample multiple domains,’ ‘employ multi-method assessment,’

The Confusion is Dramatic

- LS assessment instruments follow the Bonta guidelines that are accepted by Gottfredson and Moriarity (2006)
- The LS/CMIs and LS/RNR include risk and need and carefully differentiate among
 - major, moderate, and mild criminogenic factors
 - criminogenic and non-criminogenic factors
 - specific responsivity factors
 - normative concerns (e.g., ethical, legal, humanitarian concerns)
 - case management: service planning and delivery through case closure and follow-up.

Confusion: Two Different Types of Prediction and Studies

- 1) Validity in the prediction of recidivism
e.g., The mean correlation of LS risk/need and recidivism is .36
 - 2) Validity of the principles of effective service
 - e.g., Prediction: Use of the Wisconsin system is associated with reduced recidivism
 - e.g., Prediction: The delivery of RNR-based services is associated with reduced recidivism relative to the effects of non-adherence
- Control Group 68% RNR Services 32%
- The effect of RNR-based services is a 36 percentage point reduction in the recidivism rate (68 minus 32 = 36)

Failure to Provide Comprehensive and Balanced Reviews of Evidence in Baird (2009)

- Failure to attend to evidence in regard to
 - 1) The psychology of criminal conduct:
Theoretically relevant knowledge of a) predictive validity of risk/need factors and b) effective intervention
 - 2) The validity and utility of the Risk-Need-Responsivity (RNR) model of correctional assessment and treatment
 - 3) The validity of Level of Service Risk/Need
 - 4) The limited utility of the Wisconsin/NCCCD

The Level of Service Approach

- Comprehensive survey of the risk, need and specific responsivity factors that possess the strongest
 - research support
 - theoretical support
 - concordance with opinions of correctional professionals
 - grounds for causal inferences
 - predictive validity
 - dynamic predictive validity
 - effect of deliberate induced change
 - assessment of relevant factors increases adherence with the principles of RNR
 - adherence with RNR increases crime prevention effects
- Wide applicability: general & violent offending, misconducts; gender, race, age, class...; community, ½ way houses, prisons

The expanded LS/CMI, YLS/CMI and LS/RNR Instruments are not for every agent or agency

- Not of added value if you are not interested in effective crime prevention through correctional service delivery
- Not of added value if you are satisfied with minor crime prevention effects
- Not of added value if you reject the risk-need-responsivity (RNR) model of correctional assessment and treatment

The "Central Eight" Risk Factors	The "Big Four" Risk Factors	History of antisocial behaviour
		Antisocial personality pattern
		Antisocial cognition
		Antisocial associates
	The "Central Eight" Risk Factors	Family and/or marital
		School and/or work
		Leisure and/or recreation
		Substance abuse

Grand Mean of the Mean Validity Estimates from Eight Meta-Analyses (including Bonta, Hanson, Lipsey, Gendreau)
Andrews & Bonta (2006, p.66)

- Grand Mean r for the Big Four
.26 (CI = .22 / .30) $k = 24$
 - Grand Mean for the Moderate Four
.17 (CI = .13 / .20) $k = 23$
 - Grand Mean for the Minor Risk/Need Factors
.03 (CI = -.02 / .08) $k = 16$
- LS/CMI Risk/Need (composite of central 8)**
.35 to .45 (mean estimates)

A Summary of the Meta-Analytic Evidence of the Effects on Recidivism of Adherence with RNR (K = 374)

- Rely on Incarceration - 10% point diff
- Rely on Prison or Community - 03%
- Introduce Human Service + 12%
- Human Service plus RNR + 26%
- Human Service + RNR + Staffing and Integrity +36% point diff

Meta-analytic conclusions of Mark Lipsey and Andrews et al. are highly similar

One of the First Randomized Tests of Effects of Risk Level by Level of Service (Andrews & Kiessling, 1980)

- | LSI | Level of Service | | |
|-----------------------------------------------------------------------------------------------|------------------|-----------|---------------------|
| Risk Level | Minimal | Augmented | |
| Higher (12+) | 67% | 33% | 34 point diff |
| Lower (11 -) | 07% | 14% | -07 point diff |
| Almost exactly the same effect as the meta-analytic mean value of a 36 % age point difference | 68% | 32% | 36 point difference |
- A single study, but one representative of the meta-analytic findings.

The Effect of Adherence with Wisconsin/NCCCD CMC: Revocation Rates

- 1) Milwaukee (High Risk cases)
Intensive CMC compared with Regular Supervision
10 % point diff Significant (30% vs 20%)
Intensive Supervision with CMS vs. IS without CMS
04% Non-significant (24 vs 20)
- 2) Texas (High Risk cases)
08% Significant (23 vs 15)
- 3) 4) At least two other positive examples of reduced revocation rates

The Effect of Adherence with Wisconsin/NCCCD CMC (Harris et al, 2004)

A South Central US County

- **Revocation Rates** **09 % Significant**
- **Other technical violation -10 % Significant**
Increased technical violations
- **Arrested** **00 percentage**
point difference
- **SELECTIVITY: Harris study is not cited by Baird**

No Positive Effect on Offender Behavior
Any Effect is on Officer's Revocation Behavior

- There is little evidence of a positive effect on offender behavior

Technical violations actually increase

No Difference in Arrests

- There is some evidence of a positive effect on officer behavior

Officers less likely to initiate revocation procedures following CMC training

Misrepresentation and Selectivity in Baird (2009)

- **MISREPRESENTATION:** “Evaluations of the CMC have shown it can have extraordinary impact on recidivism (Leninger, 1998; Eisenberg & Markly, 1987; McManus, Stagg, & McDuffie, 1988).

In fact, weak effect on revocation rates (at best 10% points)

relative to mean effect on recidivism of RNR (26% to 36%)

In fact, increase in technical violations

In fact, nil effect on arrests

In fact, non-significant effect on prison misconducts

- **SELECTIVITY:** Harris, Gingerich and Whittaker (2004) is not described (nor is the nil effect of CMC on prison misconducts)

A Recent Narrative Summary of LS: Vose, Cullen and Smith (2008)

- Summary of 47 studies of the validity of LS risk/need
- 98% (46/47) found that recidivism increases with LS risk/need scores
- With r as the measure of predictive validity statistically significant effect
83% of the time (29/35)

Most Recent Meta-Analyses of the Validity of LS Risk/Need: Mean r

• Female Offenders				k
• Andrews (09)	LS/CMI	.46		09
• Schwalbe (08)	YLS/CMI	.40		03
• Andrews (09)	LSI, CMIs	.36		41
• Olver et al (09)	YLS/CMI	.36		09
• Smith et al (09)	LSI-R	.35		27
				N = 14,737

Most Recent Meta-Analyses of the Validity of LS Risk/Need

- **Male Offenders** k
- Andrews et al (06) LS/CMI .41 19
- Andrews (09) LS/CMI .38 09
- **Andrews et al (06) LSI-R .36 04**
(grand mean from 4 meta-analyses by Gendreau et al and by Hare et al)
- Olver et al (09) YLS/CMI .33 09
- Schwalbe (08) YLS/CMI .32 04

Mean r of .35 / .36:
Binomial Effect Size Display (BESD)
of Recidivism Rates by Level of Risk

- Higher Risk Group (highest scoring 1/2 of offender sample)

$$50 + 36/2 = 50 + 18 = 68\% \text{ recidivism rate}$$

- Lower Risk Group (lowest scoring 1/2 of offender sample)

$$50 - 18 = 32\% \text{ recidivism rate}$$

- BESD: the r value directly reflects the magnitude of the difference in the recidivism rates of lower and higher risk cases $r = .36$ 36% percentage point difference between 68% and 32%

Most Recent Meta-Analyses of the Validity of LS Risk/Need and Violence

- **Male Offenders Predominately**

- Campbell et al (09) LSI-R k

- **Violent Recidivism .28 19**

- **Violent Misconduct .24 06**

LS risk/need as valid or more valid than

PCL-R, HCR-20, including 2nd generation instruments keyed to violence (VRAG)

LS risk/need scores have wide-applicability

The Most Recent Summary of Meta-Analyses: Hanson (2009, Table 1)

	Mean r	
	General	Violence
LSI	.40	.27

Validity of LS Equals or Exceeds that of Other Risk and Related Assessments

- Mean validity (Gendreau, Goggin & Little, 1996)

	Mean r	k	N
	.35	28	04,579
	.29	15	09,850
	.28	09	01,400
	.27	14	14,092
	.16	16	03,420

What else is not reported in Baird (2009)?

- Mean validity (Gendreau, Goggin, Little, 1996)

Mean r k

- LSI-R .35 28
- Wisconsin .27 14

LSI-R validity exceeds the validity of the Wisconsin

and it does so 76% of the time

(Common Language Effect Size)

What is Reported in Baird (2009): Odd Rhetoric that Misrepresents the Evidence

- “There is substantial evidence available to suggest that relatively brief risk indices outperform longer, more complex models.”

What is cited as evidence?

a single study by Wagoner

appeals to regression solutions

Finagling: Sample-specific

selection of the minimum number items

weighting of items

setting of cut-off scores

Baird (09) Table 1: Austin et al (03) Item-Analysis on Pennsylvania Parolees

	Full LSI-R	Eight Item LS
• High	58%	69%
• Moderate	51%	53%
• Low	43%	34%

Serious reliability problems in Penn. Improvement with training of officers Pre-training data employed in Baird (2009)

Statistical Manipulation and Finagling

- 1) Select a study with relatively weak validity (that is, Austin et al)
- 2) Employ regression and analogous approaches
 - Select the minimum number of items required to maximize predictive accuracy in a particular sample
 - In the name of efficiency
 - Finagle with a) weighting of items and
 - b) cut-off scores for low vs higher risk groups
- Regression results must be cross-validated / replicated / wide-applicability must be demonstrated

Baird (09) Table 2: Results of 2 Separate Studies conducted in 2 different regions

	Onifade et al YLS/CMI	Nevada Wisc.	Nevada WiscRevision
• High	39%	45%	51%
• Moderate	26%	24%	22%
• Low	11%	9%	8%

Statistical Manipulation and Finagling

- Once again, 1) Select an LS study that yielded below average validity estimates
- 2) This time employ regression and analogous approaches to make the Wisconsin approach appear better than it actually performed

Select the minimum number of items required to maximize predictive accuracy in a particular sample

Finagle with a) weighting of items

b) setting of cut-off scores for low vs high risk

Odd rhetoric continues re Table 1 and 2 in Baird (2009)

- “how can change on a non-risk factor reduce risk”
“can the inclusion of non-risk factors actually reduce predictive accuracy”

Odd rhetorical questions intended to undermine LS assessment instruments

Not a balanced and complete review of the relevant evidence!

Despite the Odd Rhetoric LS Change Scores are Valid (although not presented by Baird, 2009)

Intake Risk	Retest Risk	
	Lower	Higher
LSI R (A & R 84)		
Higher	00	57
Lower	04	29
LSI R (Raynor 07)		
Higher	54	76
Lower	29	59

The validity of one time assessments of LS risk was already demonstrated

Contrary to the regression and analogous predictive solutions

- Sets of items may be formed that are theoretically-relevant and of equal or greater validity than the regression solution
- The regression solutions may select items without any causal significance

items with misleading significance Examples:

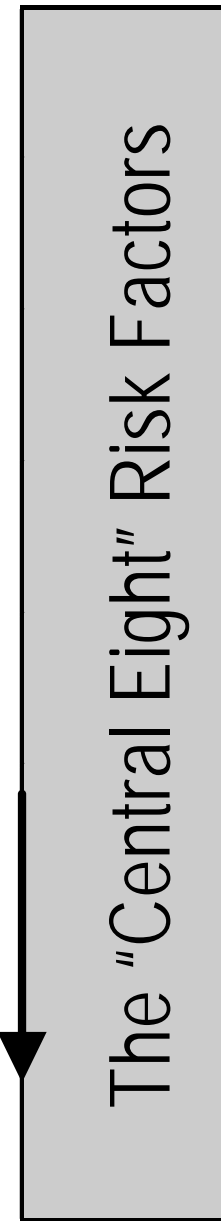
being schizophrenic is an indicator of low risk

(so: turn non-schizophrenics into schizophrenics?)

having victimized a female is an indicator of low risk

(so: encourage victimization of females?)

- * **As reviewed by Goffredson and Moriarty, employ an assessment system that is theoretically relevant , samples multiple domains, includes criminogenic needs, and**



The "Central Eight" Risk Factors	The "Big Four" Risk Factors	History of antisocial behaviour
		Antisocial personality pattern
		Antisocial cognition
		Antisocial associates
		Family and/or marital
		School and/or work
		Leisure and/or recreation
		Substance abuse

- Almost all of the predictive value of the Central 8 is carried by the Big 4
- Changes in the Moderate 4 may assist in the production and/or maintenance of changes in the big 4

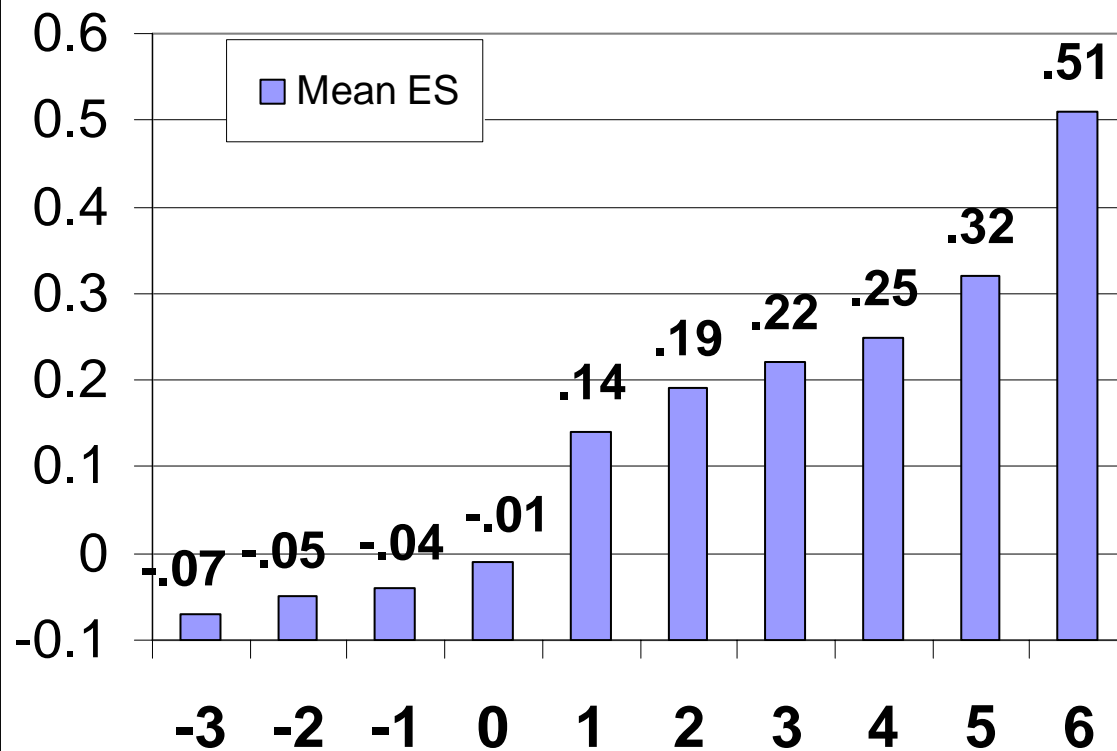
- For the vast majority of humanity
 - the central eight risk/need factors are related to criminal futures
 - **targeting of the central eight is followed by reduced offending**
 - if there is also adherence with the risk and general responsivity principles**

Mean Effect Size by Specific Need Targeted: Criminogenic Needs

(Overall mean effect = .08, K = 374)

- Antisocial Attitudes .21 (78)
- Self-Control Deficits .22 (59)
- Antisocial Associates .21 (51)
- Non-Criminal Alternative
Behavior in High Risk Sits .22 (18)
- Family Process .29 (30)
- School/Work .15 (88)
- Substance Abuse .11 (36) *ns*
- Leisure Recreation not tested

Mean ES by Criminogenic - Non-Criminogenic Need



Mean ES by Relative Targeting of
Criminogenic and NonCriminogenic Needs

- The LSI-R and the LS/CMIs are recommended for agencies interested in crime prevention through the delivery of programs that are in adherence with the principles of the RNR model
- LS training and assessments are recommended
 - a) if reduced recidivism is a desired outcome
 - b) if there is an interest in increasing adherence with RNR principles
 - c) if crime prevention effects greater than 10 percentage points on revocation rates are valued
- The evidence is that RNR adherence does much better than the Wisconsin/NCCD approach.
- Enhanced adherence with RNR is the goal of LS assessments

In Conclusion

If you seek guidelines in the selection and use of assessment instruments with offenders

Avoid Baird (2009)

Turn to Bonta (2002)

Just as Stephen Gottfredson and Laura Moriarty (2006) did

It is a “question of evidence”

re the validity of LS risk/need

the power of adherence with the RNR model

Seriously, A Question of Evidence

- Mean validity (Gendreau, Goggin & Little, 1996)

	Mean r	k	N
	.35	28	04,579
	.29	15	09,850
	.28	09	01,400
	.27	14	14,092
	.16	16	03,420

Seriously, A Question of Evidence

- Mean validity (Gendreau, Goggin, Little, 1996)

Mean r k

- LSI-R .35 28
- Wisconsin .27 14

LSI-R validity exceeds the validity of the Wisconsin

and it does so 76% of the time

(Common Language Effect Size)

Seriously, A Question of Evidence

- | LSI
Risk Level | Level of Service | | |
|-------------------|------------------|-----------|----------------|
| | Minimal | Augmented | |
| Higher (12+) | 67% | 33% | 34 point diff |
| Lower (11 -) | 07% | 14% | -07 point diff |
- Almost exactly the same as the mean value of a 36 % age point difference revealed by the meta-analyses
A study with findings representative of the RNR meta-analyses

Seriously, join an international community of LS users and researchers with an Ongoing Commitment to Enhanced Reliability, Validity and Utility

- * Young offenders Hoge Luong and Wormith
- Adult offenders Bonta Bourgon and Armstrong
- Gender-informed, racially-informed, subgroup and pathway work
- Use of LS assessments and adherence with the principles of RNR
- Rewarded by enhanced crime prevention effects
- Unlike the Wisconsin/NCCD system, which has
 - No effect on re-arrests of offenders
 - The effect of increased technical violations by offenders
 - In some studies, modest decreases in revocation behavior on the part of correctional officers (no greater than 10 % points)
- The LS and RNR approach --- in theory, research and practice --- is very attractive