

A QUESTION OF EVIDENCE: RISK ASSESSMENTS MODELS USED IN THE JUSTICE SYSTEM

August 3, 2009

Christopher Baird



426 South Yellowstone Drive, Suite 250, Madison, WI 53719
Phone (608) 831-8882 / Fax (608) 831-6446
www.nccd-crc.org

The National Council on Crime and Delinquency is a nonprofit social research organization.

Major Problems With Current Systems

1. Composite measures of risk and needs
2. Inadequate measures of validity
3. Completely inadequate tests of reliability
 - Wrong measure
 - Lack of quality studies
4. Definition/identification of “criminogenic needs” and “protective factors”



History of Risk Assessment in Corrections

| | |
|--------------|--|
| 1928 | Illinois Parole Board, L. Burgess |
| 1960's | California Base Expectancy Tables Federal Parole Guidelines I-Level Quay |
| 1975–80 | Wisconsin Case Classification |
| 1980's | NIC Model CMC/SJS Gottfredson Study |
| 1990–Present | “Third- and Fourth-generation” Systems Third generation: <ul style="list-style-type: none">• Static vs. dynamic risk factors• Criminogenic needs Fourth generation: <ul style="list-style-type: none">• Protective factors |

Risk and Needs Principles

Risk: Target resources to highest risk cases

Needs: Services should address criminogenic needs
(dynamic risk factors)

- Peer relationships
- Family factors
- Substance abuse
- Education
- Employment
- Anti-social attitudes toward authority

Questions:

How should the efficacy of the risk assessment model be evaluated?

Do current systems provide adequate measures of risk, particularly for “high stakes” decisions such as sentencing or release from prison?

LSI-R

1

15

30

54

Low

Moderate

High

- Correlations measure association both within and across risk levels (AUC)

- The real issue is the degree of separation attained between risk levels
 - » Proportionality
 - » Separation/Discrimination

Outcome Comparisons by Risk Level: Pennsylvania Parolees

Outcome Comparisons by Risk Level: Pennsylvania Parolees

| Risk Level | Full LSI-R | | Eight Factors From LSI-R | |
|------------|------------|--------------------|--------------------------|--------------------|
| | N | Rate of Recidivism | N | Rate of Recidivism |
| Low | 86 (9%) | 43% | 146 (15%) | 34% |
| Moderate | 398 (40%) | 51% | 614 (65%) | 53% |
| High | 522 (52%) | 58% | 186 (20%) | 69% |

Source: Austin, Coleman, Peyton, & Johnson, 2003.

Risk Scale Comparisons: Outcomes by Risk Level

| Risk Instrument | Risk Level | Percent of Sample | Recidivism Rate | Spread |
|-----------------------------------|------------|-------------------|-----------------|--------|
| LSI-R ¹ (54 items) | Low | 9% | 43% | } 15% |
| | Moderate | 40% | 51% | |
| | High | 52% | 58% | |
| LSI-R ² (54 items) | Low | 58% | 20% | } 23% |
| | Moderate | 31% | 41% | |
| | High | 11% | 43% | |
| Nevada ³ (11 items) | Low | 18% | 8% | } 43% |
| | Moderate | 47% | 22% | |
| | High | 35% | 51% | |

¹ Austin, Coleman, Payton, and Johnson; 2003

² Hollen and Palmer, 2006

³ Wagner, Quigley, Ehrlich, and Baird; 1998

Vose, Cullen, and Smith (Federal Probation, December 2008):

Twenty-nine of 47 studies reported samples of less than 500.

Twenty-two outcomes based on samples of less than 200.

All results based on correlations.

Correlations as low as .137 are cited as evidence of predictive validity.

THE BAR FOR VALIDITY HAS BEEN SET VERY LOW

A Question of Evidence: COMPAS (Skeem and Eno Louden, 2008)

- There is no evidence that the original COMPAS basic scales, higher order scales, or risk scales predict recidivism.
- The available data provide no evidence that the original COMPAS risk scales predict reoffending of any sort.
- The authors do not include any external validated measures to examine the construct validity of COMPAS.
- The original scales were initially “validated” using past criminal behavior.

Inter-rater reliability tests have been totally inadequate

- Eighteen offenders received LSI-R from two different officers over a 16-month period (Girard, Wormith; 2004).
- One-hundred sixty-seven correctional practitioners scored a single hypothetical case during LSI-R training (Lowenkamp et al., 2004).
- THE FULL YASI/PACT HAS NEVER UNDERGONE RELIABILITY TESTING (R. Baranowsky 2009)

OR have addressed the wrong issue: INTERNAL CONSISTENCY.

Reliability (continued)

- Internal consistency is a measure of the validity of scales, inventories, and tests that measure a construct.
- Important in the field of psychology.
Why?

Depression: There is no method for definitively measuring the existence of depression.

The more covariance in these measures, the more certain a clinician can be of his/her diagnosis.

Many depressed people have sleeping problems.

Many lose interest in job, family, or friends.

Many become inactive or uninvolved.

Many have thoughts of self harm, etc.

But...

Recidivism is not a construct but an observable, measurable outcome.

The less covariance there is among predictors, the stronger the overall prediction.

- Prior arrests and prior convictions may both be good predictors, but they are both measures, to a degree, of the same thing.
- Peer relations, substance abuse, and employment are not as related to prior convictions and therefore “contribute” more to the actuarial formula.



Need Principle: Services Should Address Criminogenic Needs (Dynamic Risk Factors)

- Peer relationships
- Family factors
- Substance abuse
- Education
- Employment
- Anti-social attitudes toward authority

Questions:

Why are these needs “criminogenic” and others are not?

What **evidence** is available to suggest they are criminogenic?

If a need is associated with future criminality for the general offender population, is this relevant to case planning?

The term criminogenic infers causality...

Measures of association do not represent causality.

Less than half of nine studies on the relationship of LSI-R domains and outcomes found correlation between family/marital, leisure and recreation, substance use, and emotional/personal domains and recidivism (Andersen, 2008).



A Question of Evidence: COMPAS Criminogenic Needs (Parole)

| Parole | Previous Arrests | | |
|-------------------------|------------------|-------------|-----------|
| | Michigan | California* | Georgia |
| Scale | N = 920 | N = 785 | N = 3,809 |
| Criminal Associates | 0.04 | 0.15 | 0.04 |
| Substance Abuse | 0.17 | 0.27 | 0.21 |
| Financial Problems | 0.08 | 0.11 | 0.10 |
| Vocational Education | 0.03 | 0.04 | -0.03 |
| Family Criminality | 0.03 | 0.05 | 0.03 |
| Social Environment | -0.10 | 0.08 | -0.06 |
| Leisure/Boredom | 0.06 | 0.08 | 0.01 |
| Residential Instability | 0.01 | 0.02 | 0.02 |
| Social Isolation | 0.06 | 0.10 | 0.04 |
| Criminal Attitudes | 0.09 | 0.06 | -0.06 |
| Criminal Personality | 0.07 | 0.12 | 0.00 |

Note: With N = 920, a correlation of .055 is significant; with N = 785, a correlation of .07 is significant; and with N = 3,809, a correlation of .032 is significant at $p < .05$ (2-tailed).

*For California Spearman correlations are presented, for Michigan and Georgia Pearson correlations are provided.

Taken from Northpointe – Evolving Practice Through Scientific Innovation, 2007.

© 2009 by NCCD, All Rights Reserved

- **Theory is not evidence; potential is not evidence. Literature on generation three and four models is rife with statements like the following:**

“Adding an evaluation of protective factors that may reduce the risk to reoffend has created a more advanced and potentially more accurate predictive model.” (PACT; Florida DJJ)

- **Combining risk and need factors in a single scale diminishes the accuracy of risk assessment and thus, violates the risk principle.**

- **The standard for measuring the efficacy of risk indices should be the level of discrimination attained between risk levels.**
- **The inter-rater reliability of most third- and fourth-generation systems has not been adequately established. Claims of reliability are grossly overstated.**

- **Internal consistency is not an appropriate measure of risk assessment reliability.**
- **There is nothing inherent in most generation three and four systems to accurately identify criminogenic needs or protective factors. The only systems with any evidence that they identify criminogenic needs are CMC and SJS.**