

What is the Evidence for Evidence-Based Offender Programs?

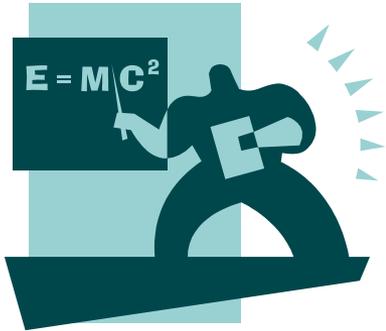
Sheldon Zhang, SDSU

David Farabee, UCLA

Rehabilitation Back in Vogue

- The correctional continuum:
 1. Punishment and incapacitation
 2. Treatment and rehabilitation
- The current pendulum is swinging toward rehabilitation, brought about mostly by court orders and financial pressure (i.e., AB109 and fiscal problems)



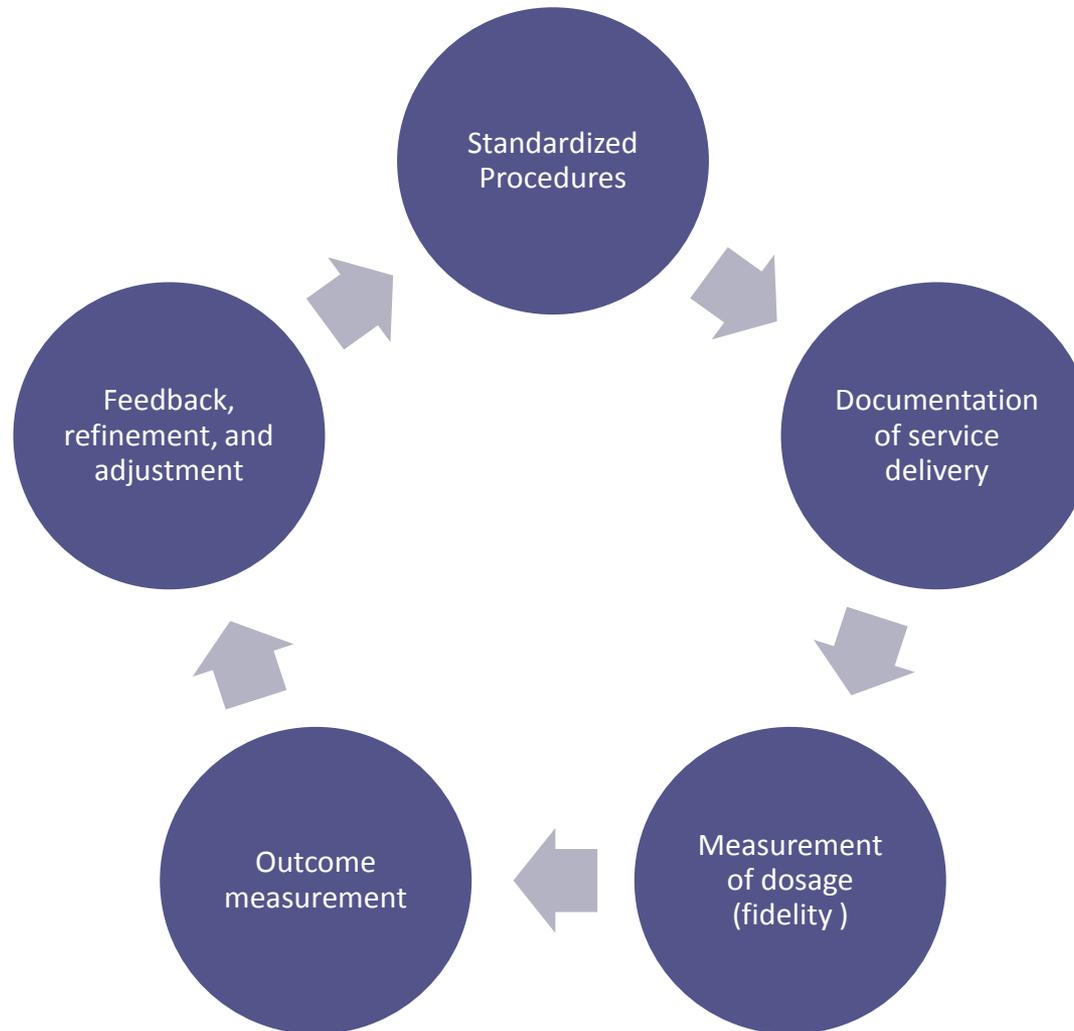


The Ideal Model

- An intervention needs to be based on
 - A sound rationale (i.e., make sense), and
 - Empirical evidence (i.e., proven effective)



Program Implementation



What Constitutes Meaningful Outcome Measures for Our Parolee Population?

- Broadly speaking—“improved social functioning”:
 - Get a job
 - Have a stable residence
 - Form a good social support network (e.g., friends, family, spouse, etc.)
 - Engage in pro-social activities (e.g., sports, civic activities)
 - Pursue education or vocational training
 - Stay out of trouble



Disclosure: Our Bias

- Behavioral outcomes not changes in attitude
- It is important to work on changing/adjusting one's attitude, perception, or viewpoint—hence cognitive therapies, psychological counseling, group interactions, etc.
- Only behaviors bear consequences.



Bottom Line

- Don't come back to the criminal justice system
 - If you don't come back, you are not my problem.
- Varied ways of counting recidivism
 - Parole violation
 - Parole revocation
 - New arrest
 - New prison term
- All recidivism measures are valid!
 - Depend on what stakeholders care the most
 - CA is not worse than most other big states, except we measure it differently.



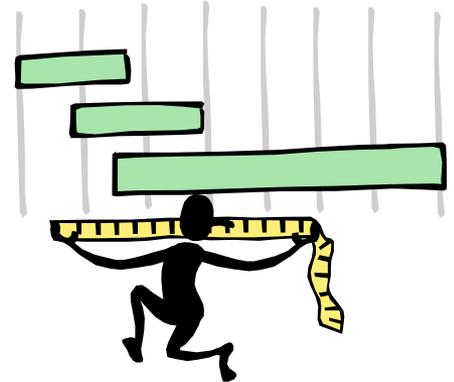


Community-Based Programs are Specialized—Hence Varied Outcomes

- Substance Abuse Treatment:
 - Licensed Residential Treatment
 - Sober Living Environments
 - Outpatient Counseling
- Cognitive Behavioral Treatment:
 - Criminal Thinking
 - Anger Management
 - Family Relationships
- Structured Housing Opportunities
- Education (Literacy, Computer Literacy, GED preparation and testing)
- Employment-related services (job readiness training, transitional jobs, and job development)

Each Program Has Its Own Measures

- Substance abuse looks at “relapse” in use
- Housing looks at residential stability
- Employment looks at job retention
- Education looks at grades/GED diplomas
- All seem to contribute to social functioning, none are individually held accountable for the failure (because humans are complicated).
- Under current treatment model, program completion becomes the most straightforward outcome measure. Whether a particular program is helpful to the overall social functioning of the parolee is another matter.



Politics in Defining Program Success

- Providers want to receive funding (in addition to genuinely wanting to produce positive change), thus pressured to project a positive image.
- Varied outcome measures have been used by providers, often after the fact.
- We frankly know little about what configurations of services are most helpful to the offender population—even the most influential correctional researcher can't find a model program!



A Holistic Approach?



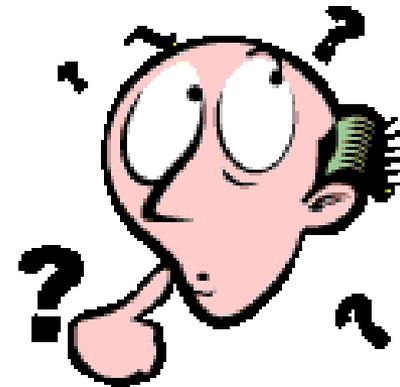
- The Cleveland Clinic Model

- Group practices are better and cheaper
- Collaborative medicine is more effective
- Big data can be harnessed to improve the quality of care and lower costs
- Apply scientifically proven strategies, not just “new” or expensive techniques
- Cooperative practices lead to innovation
- Focus on wellness or healthcare, not “sickcare”
- Emphasis on patient care and patient experience

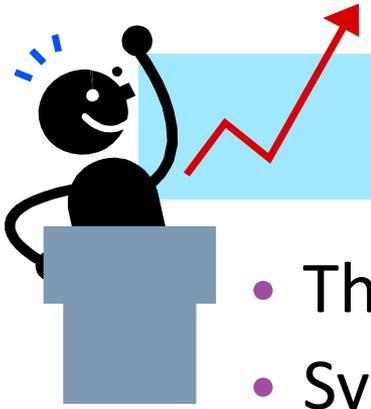


Evidence Based Practices by Definition

- What constitutes evidence?
- Where do we find these evidence-based programs?



Science vs. Pseudo-Science



- The Challenge of Finding Effective Treatment
- Systematic Measurement vs. Anecdotal Stories
- Science vs. Belief (gut feeling, conviction)
- RCT vs. Comparison



Why Randomized Controlled Trial (RCT)?

- RCT is the only responsible and scientific way to develop and assess intervention programs.
 - Background noise
 - Selection bias



Outperforming One's Social Environment (i.e., Background Noise)



Background
Noise

Random
Impact

Logistical Barriers to RCT

- Political barriers
- Ethical considerations
- Unfamiliarity with the procedures
- But the situation is changing!



Perks & Problems of RCTs

- Perks

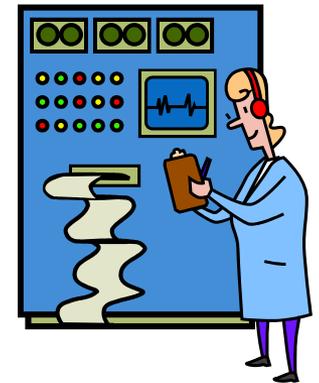
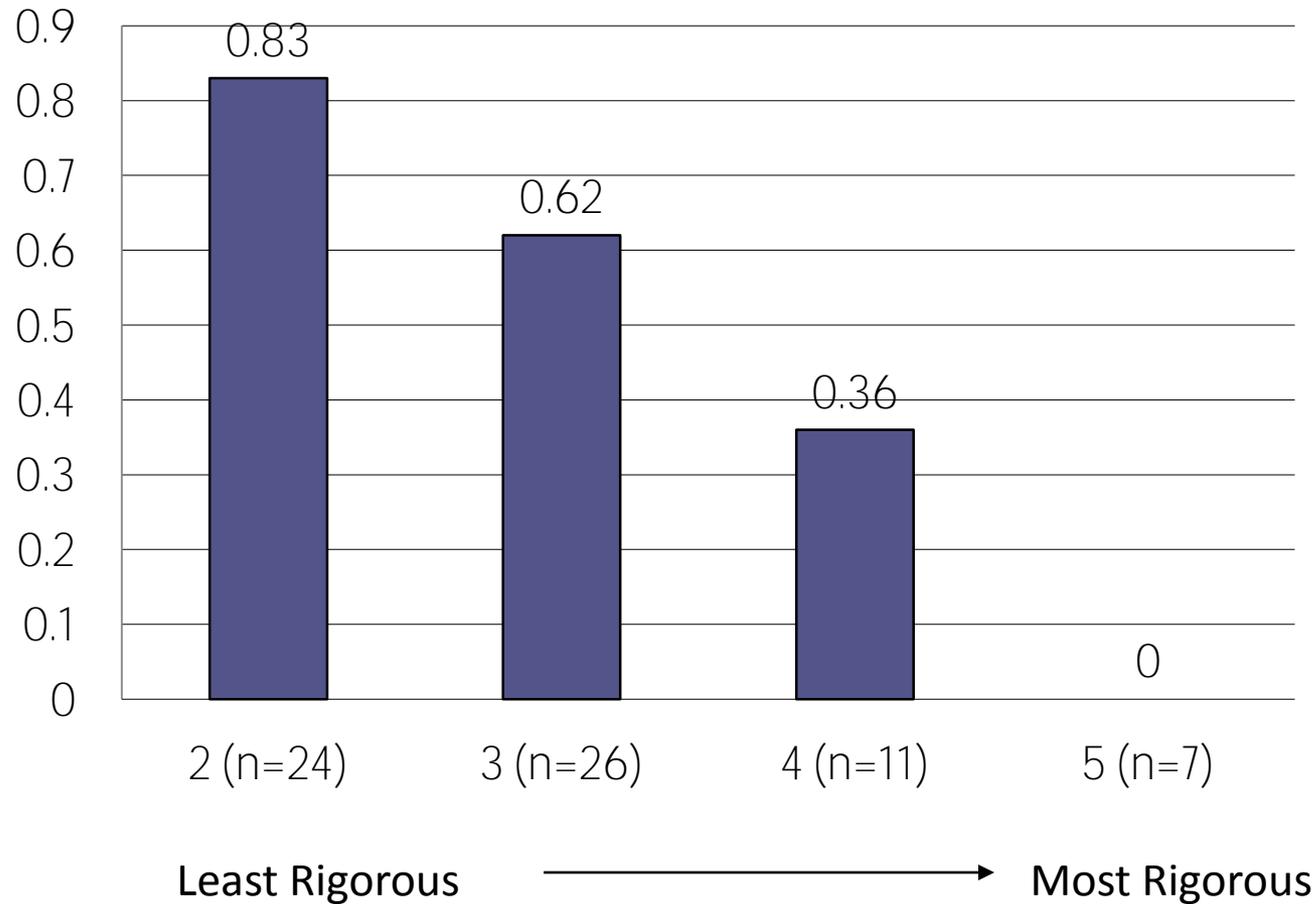
- RCT is “gold standard”
- Random assignment is non-discriminatory
- Culture is changing—higher standards of evidence
- Genuine interest in how well programs works

- Problems

- A priori assumption of effectiveness
- Control subjects mean lost revenue
- Resistance to counting dropouts/no-shows as treatment Ss
- Increased rigor → smaller effects



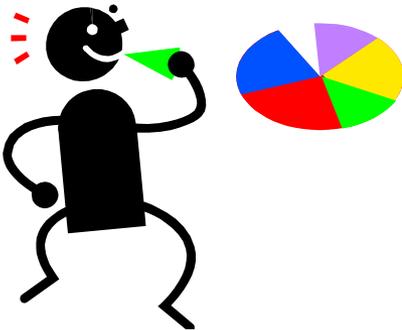
Study Quality & Effectiveness of Offender Programs (N=68)



Weisburd et al. (2001)

Intent to Treat vs. Post Hoc Analysis

- Whole group as a totality (IT)
- Or partition study subjects in different ways so something positive can be found (post hoc)



Selected Experiments

(from Farrington & Welsh, 2005)

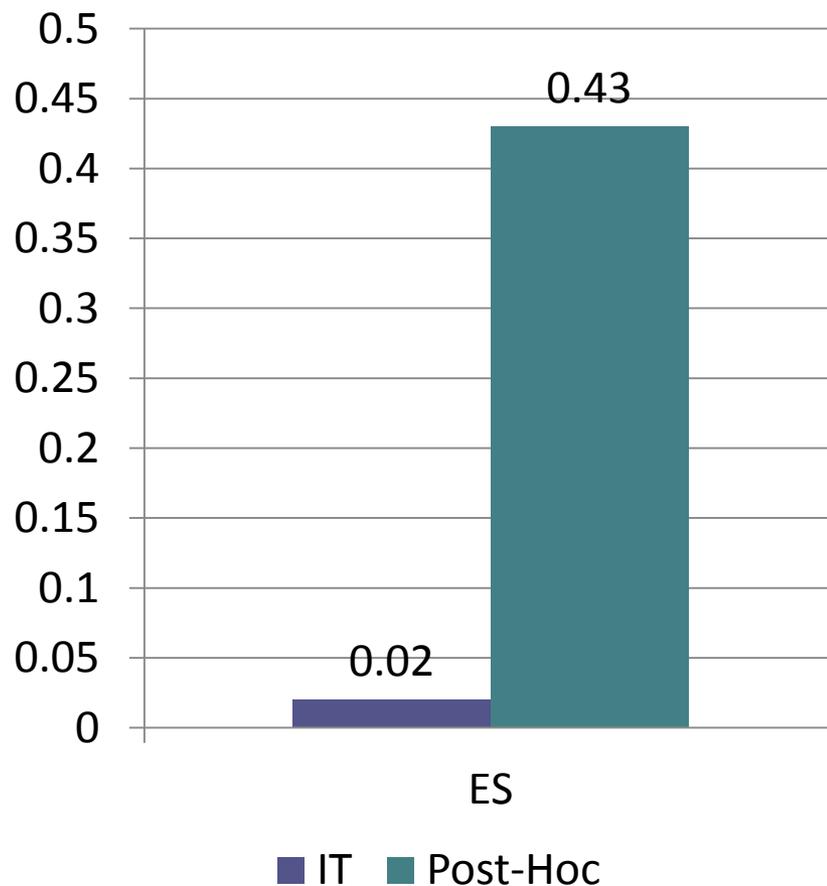
Authors	Intervention	Outcomes
Robinson et al. (1995)	CBT	Convictions
Armstrong (2003)	MRT	Convictions
Wexler et al. (1999)	TC	Re-incarceration
Marques et al. (1994)	CBT	Arrests for sex offenses
Lewis (1983)	Scared Straight	Arrests
Greenwood & Turner (1993)	CBT	Recidivism

Measure of Effect Size

- Cell sizes were not always available, so we used relative difference in percentages:
 - E.g., if 50% of the Control Ss and 40% of Experimental Ss reoffended, the relative decrease would be 20% $[(50-40)/50]$.
 - Positive values indicate lower recidivism among Ss in the treatment group.



Effect Size Comparisons



- Mean post hoc ES was 20x larger than the IT ES.
- Two of the IT results showed worse results for treatment group.
- None of the post hoc differences were negative.
- IT range: $-.20$ to $.22$; PH range: $.12$ to $.83$.

Post Hoc and Positive Findings

- On average , the relative differences between treatment and control Ss was 20x larger for post-hoc analyses than IT analyses.
- Perry et al. (2010) rated 83 correctional studies on eight descriptive criteria and found that CJ researchers performed worst reporting—
 - Methods of subgroup/adjusted analyses
 - Descriptions of whether IT analysis was followed.

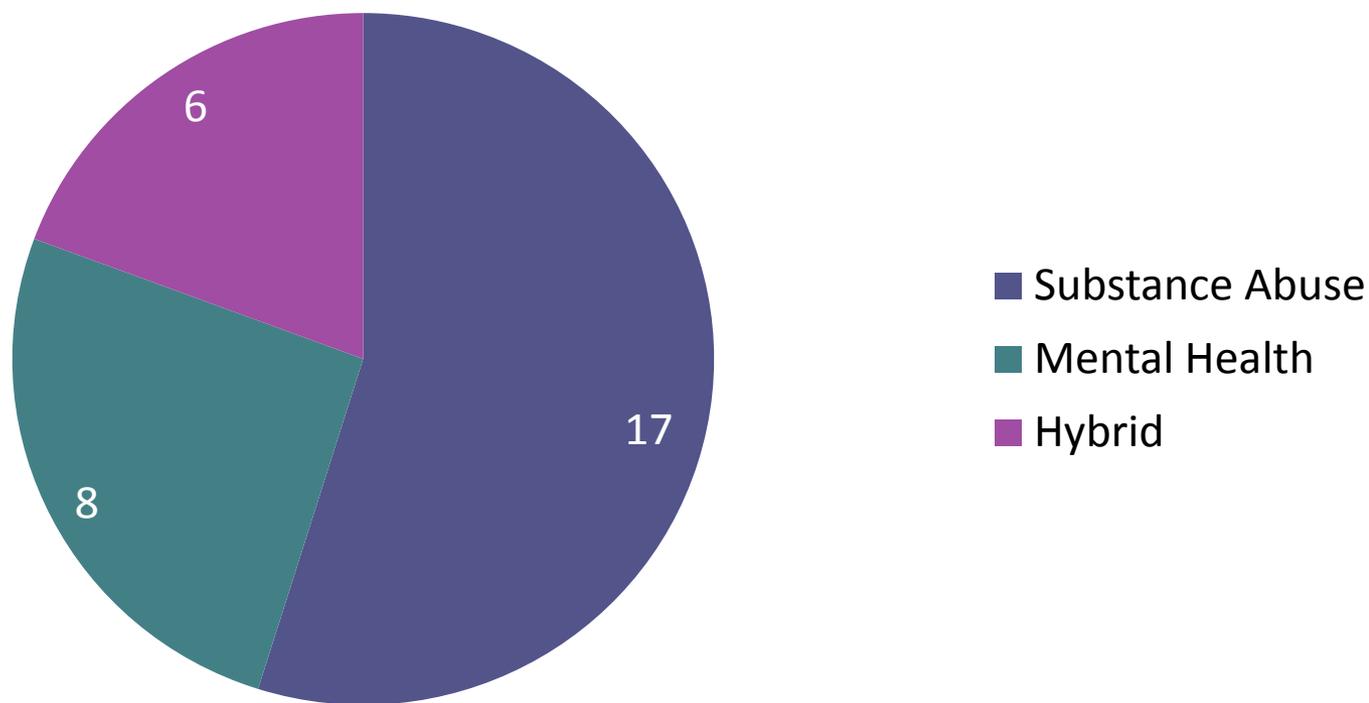


Where Do We Find Evidence-Based Programs?

- National Registry of Evidence-Based Programs and Practices (NREPP)
- Managed and funded by SAMHSA
- Began in 1997
- Purpose: “[T]o assist the public in identifying approaches to preventing and treating mental and/or substance abuse disorders that have been *scientifically* tested and that can be readily disseminated to the field.” (NREPP, 2009)



Types of Interventions Included in This Study (N=31)

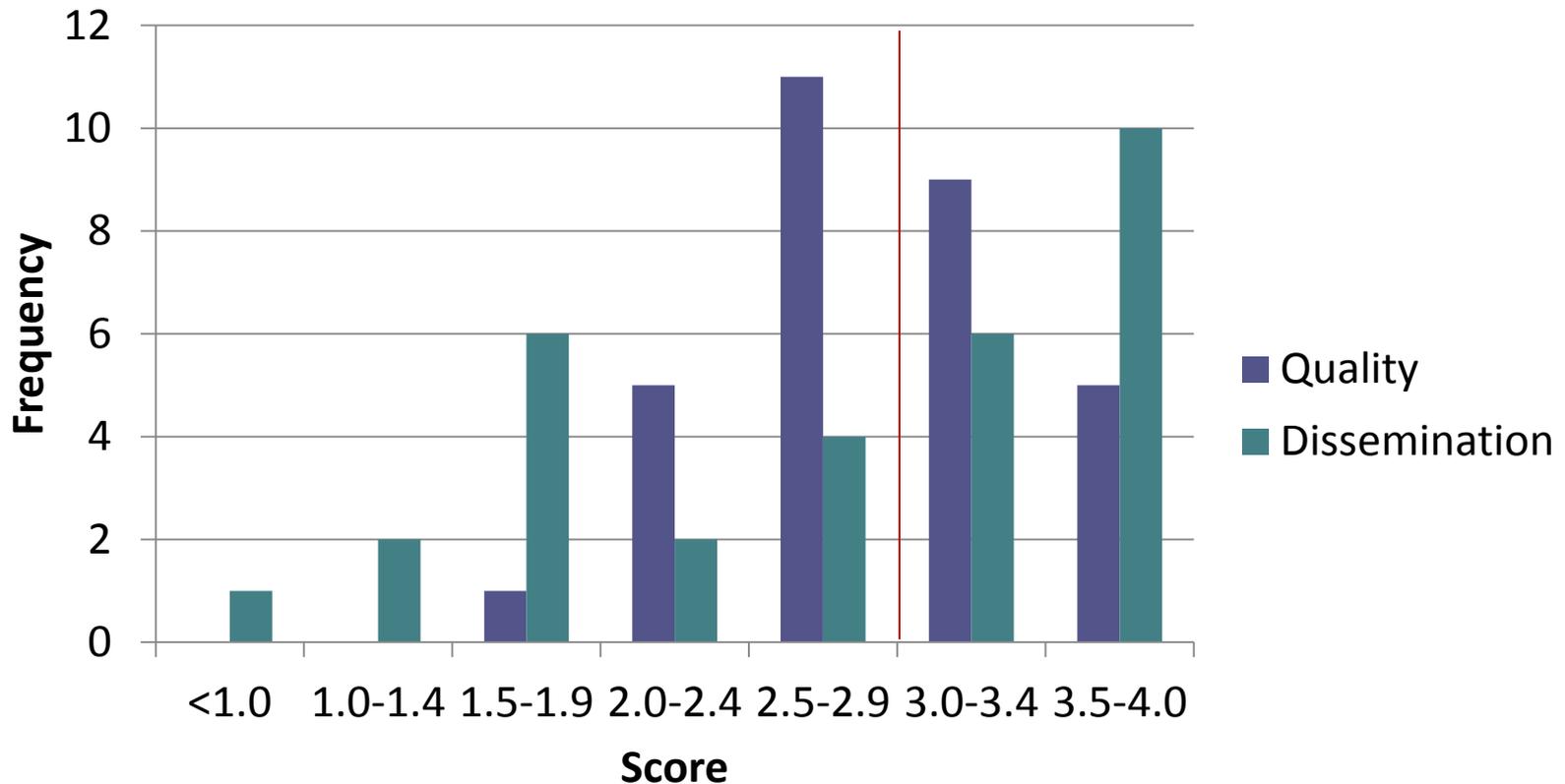


NREPP Review Criteria

- Quality of research is reported results using the following six criteria:
 - Reliability of measures
 - Validity of measures
 - Intervention fidelity
 - Missing data and attrition
 - Potential confounding variables
 - Appropriateness of analysis
- Reviewers use a scale of 0.0 to 4.0, with 4.0 being the most favorable.



Distribution of NREPP Quality and Dissemination Readiness Scores (N=31)



Ratings by Program Type (N=31)

Program Type	Quality	Dissemination
Substance Abuse (n=17)	Mean=2.8 (SD=0.52) Range=1.7-3.8	Mean=2.6 (SD=0.87) Range=0.8-4.0
Mental Health (n=8)	Mean=3.0 (SD=0.37) Range=2.2-3.5	Mean=2.7 (SD=0.92) Range=1.3-4.0
Hybrid (n=6)	Mean=3.00 (SD=0.49) Range=2.1-3.5	Mean=3.17 (SD=0.82) Range=1.5-3.9

One third of the EBPs had not been replicated.

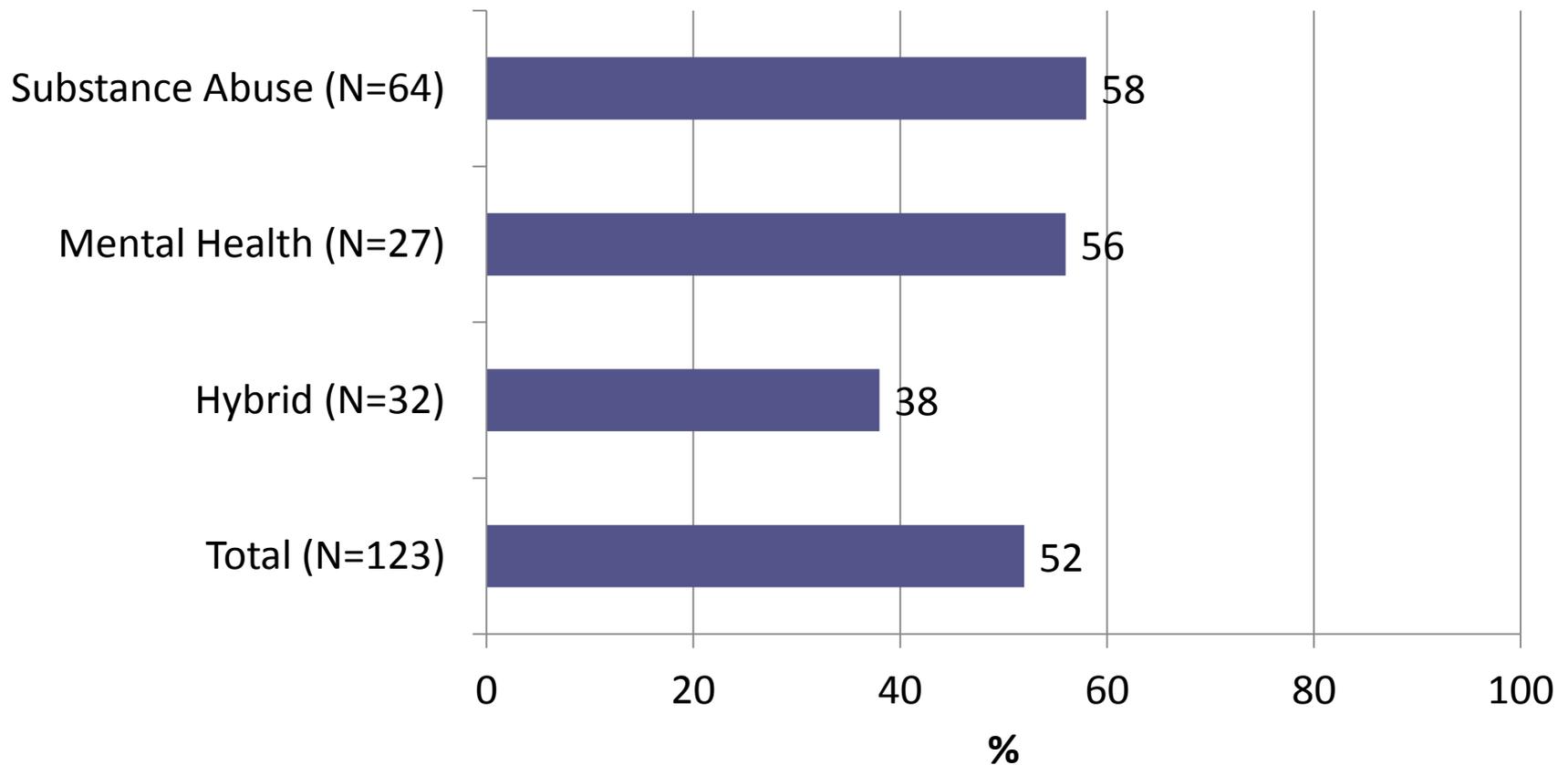
Case Example (3 of 4 Rating)

“After controlling for time at risk in the community and other covariates, youth who received treatment in the ___ program showed a significant reduction in the prevalence of recidivism compared with youth who were admitted to the program briefly for assessment or stabilization services and then returned to the referring secured correctional institution.”

Case Example (3 of 4 Rating)

“Among program participants, the group receiving a low dosage of the intervention was compared with the group receiving a medium or high dosage. At posttest, the medium- to high-dosage group had a significant reduction in AOD use on the quantity-frequency index ($p = .001$) and number-of-drugs index ($p = .035$) compared with the low-dosage group.”

Developers as Evaluators



Thoughts on “Evidence-Based” Programs

- Tremendous variation in the quality/dissemination readiness of the NREPP studies we reviewed.
- Lack of evaluator independence undermines confidence in findings.
- One third of the EBPs had yet to be replicated.
- Many of the NREPP consumers lack time or expertise to weigh the details of the evaluation methods.



Where Do We Go from Here?



- Vendor to provide evidence for their services
- Careful review of the evidence before investing serious money into programs
- Rigorous evaluation to ascertain efficacy
- Establishment of minimum number of studies that must be submitted to constitute “evidence”
- A mandatory disclosure of researchers’ involvement with the program being evaluated
- Encourage innovation and explore alternative ways to improve parolee success