Agreement between court decisions and evaluator opinions in SVP trials

Rachel Kahn, Ph.D. & Sharon Kelley, Psy.D.

Annual Wi-ATSA Conference

June 9, 2017

Credit

- The data presented here is based in part on a study by Richard Elwood, Rachel Kahn, James Mundt, and Sharon Kelley (manuscript in process). Agreement between court decisions and evaluator opinions in SVP commitment hearings.

What do courts think of expert witnesses?

- Edens et al. (2012): Examined legal cases published in the past 20 years for evidence of disparaging remarks towards psychologists and psychiatrists (i.e., bias; charlatan).
  - N = 160 cases with 245 derogatory statements from attorneys and judges.
- Statements included:
  - being for sale (28%)
  - partisan/advocate (27%)
  - biased (21%)
  - pseudoscience (14%)
  - mysticism (6%)

What do courts think of expert witnesses?

- Criminal trials more frequently referenced being for sale (51% vs. 29%)
- Civil trials more frequently referenced partisan/advocate (69% vs. 27%)
- Judges made statements referencing bias 41% of the time, although this was usually triggered by attorneys and may have reflected actual bias.

(Edens et al., 2012)

What do jurors think of expert witnesses?

- Jurors in criminal child sex offense trials find witnesses to be credible when (1) they had relevant professional experience; (2) did not appear biased; and (3) communicated with clarity (Blackwell & Seymour, 2015).
- Perceived impartiality and clarity frequently cited as being most influential (Freckelton et al., 1999; Young et al., 1999)
- Perceived credibility and efficacy are cited as predictive of legal outcomes (Brodsky et al., 2010; Cramer et al., 2010)

What do jurors think of expert witnesses?

- Cramer, Parrott, Gardner, Stroud, Boccaccini, & Griffin (2014)
  - N = 324 mock jurors rated experts seen in a trial via video recording
  - Credibility – Witness Credibility Scale
    - Measures likeability, trustworthiness, confidence, & knowledge
  - Efficacy – Observed Witness Efficacy Scale
    - Measures poise and communication style
  - Personality – Five-Factor Mini-Markers Scale
    - Measures neuroticism, extraversion, openness, agreeableness, & conscientiousness

What do jurors think of expert witnesses?

- Cramer, Parrott, Gardner, Stroud, Boccaccini, & Griffin (2014)
  - Character but not efficacy was associated with jurors agreement with the expert on sentencing
  - Character was correlated with
    - Openness
    - Low neuroticism
    - Agreeableness
    - Conscientiousness
    - Likeability
    - Trustworthiness
    - Knowledge
    - Confidence

What do jurors think of SVP evaluators?

- Boccaccini et al. (2014)
  - Surveyed actual jurors (N = 161) from 14 SVP trials in Texas
- When trials had both a state and defense expert, most jurors believed
  - Testimony from the experts was important (98%)
  - Experts were either OK (42%) or Very Good (56%) at identifying people who are high risk
  - Actuarial measures help experts make accurate decisions (67%)*
  - The experts were honest (74%)
  - The reason the experts did not agree was because one expert did a better job than another (50%) and because it was a difficult and complicated case (34%)
  - Some thought at least one of the experts was biased (23%) and attributed the outcome to this bias (13%)

How do jurors make decisions in SVP trials?

- Mock juror studies –
  - More influenced by unstructured clinical expert testimony vs. actuarial expert testimony (Krauss et al., 2011; McCabe et al., 2010)
  - Mock jurors credit actuarial results when it confirms perception of risk and discredits actuarial results when it indicates low risk (Scurich & Krauss, 2013)
  - Most influenced by past offenses (Krauss et al., 2011)
  - Being referred for SVP may be sufficient grounds for commitment for most mock jurors (Scurich & Krauss, 2014)
  - Female jurors more strongly in favor of civil commitment (McCabe et al., 2010), especially when respondent is described as "a psychopath" (Guy & Edens, 2006)

"Don't spread it around, but on the really tough ones, I just go with 'eenie, meenie, minie, moe.'"

How do jurors make decisions in SVP trials?

- Turner et al. (2015): surveyed actual jurors (N = 462) across 40 SVP trials in Texas
  - Most influenced by past sex crimes, sexual behavior during crimes, and respondent's character (e.g., given opportunities to change but he hasn't)
  - More influenced by clinical than actuarial based testimony
  - More influenced by diagnosis, lack of remorse, and failure to take advantage of treatment opportunities as compared to results from actuarial risk assessments or PCL-R
  - Importance of PCL-R increased with higher scores
  - Female jurors more influenced by PCL-R rating
- Note: in Texas SVP trials, respondent has no 5th Amendment right in civil trials and must testify in front of a jury; trials are only done in one county; all initial SVP trials resulted in commitment since law enacted in 1999

The current study

- Exploratory in nature
- How often do the courts agree with a "neutral" evaluator in Wisconsin SVP trials?
- Do jury trials and bench trials have a differing rate of agreement?
- Are commitments and discharges predicted by empirical data: Static-99R score, progress in treatment, etc.?

Some context

- In Wisconsin, individuals undergo two separate risk assessments to determine their need for commitment as SVP: a pre-probable cause evaluation and a post-probable cause evaluation.
- Both evaluators testify at the initial commitment hearings
- Following commitment, an annual risk assessment to determine their continued need for commitment is required (re-commitment evaluation)
  - When respondent petitions for release, the evaluator who completed the re-commitment report testifies

More context

- This study focuses on the following evaluations:
  - The initial post-probable cause evaluation (**.04**) used in the initial commitment trial
  - The annual report to the court (**.07**) used in the re-commitment trial
  - The evaluator is assigned to these cases at random as part of their state employment and is considered a "neutral" evaluator.
  - Either side can retain additional expert witnesses (they are not considered "neutral" for the purposes of this study).
- The threshold for determining an individual's need for commitment is "more likely than not" to commit a sexually violent act, which has been interpreted by case law to mean over 50%.

Current Study

Study Description

- N = 205 trial outcomes from 2012 – 2016
  - n = 73 were excluded for trial ending in stipulation or dismissal
    - 35 stipulated to commitment
    - 25 stipulated to discharge
    - 9 dismissed by state
    - 3 stipulated to SR
    - 1 mistrial

FINAL SAMPLE: *n* = 132 trials

Patient Descriptives

- Patients were all adult male sex offenders residing at SRSTC or on Supervised Release (SR).
  - Ages 24 to 80 ($M = 51.04$, $SD = 10.71$)
  - Ethnicity: 61.1% White/Caucasian; 31.1% Black; 6.1% Native American; 1.6% Other
  - Static99R Score: 1 – 9 ($M = 5.40$, $SD = 1.66$)
  - PCL-R Score: 9 – 37.5 ($M = 25.23$, $SD = 5.00$)

Evaluators

- N = 13
- 61.5% male / 38.5% female
- Number of trial cases
  - Range: 4 – 22 cases
  - *M (SD)* = 6.49 (4.11); Median = 5 cases
- Evaluator Conservativeness *(percentage based on commitment recommendation rate)*
  - Range: 29% - 100%
  - *M* = 60%

Description of Main Analyses

- McNemar tests comparing paired proportions in evaluator recommendation and court decision.
- *Kappa* calculation of agreement rate between evaluators and courts and adjusts for chance rate of agreement
  - Ranges from -1 to 1 (< 0 equivalent to less than chance agreement, 0 being no agreement, and 1 equivalent to perfect agreement)
- Logistic regression: outcome is prediction of court decision

Interpretation of Kappa

Table 2

Interpretation of Kappa

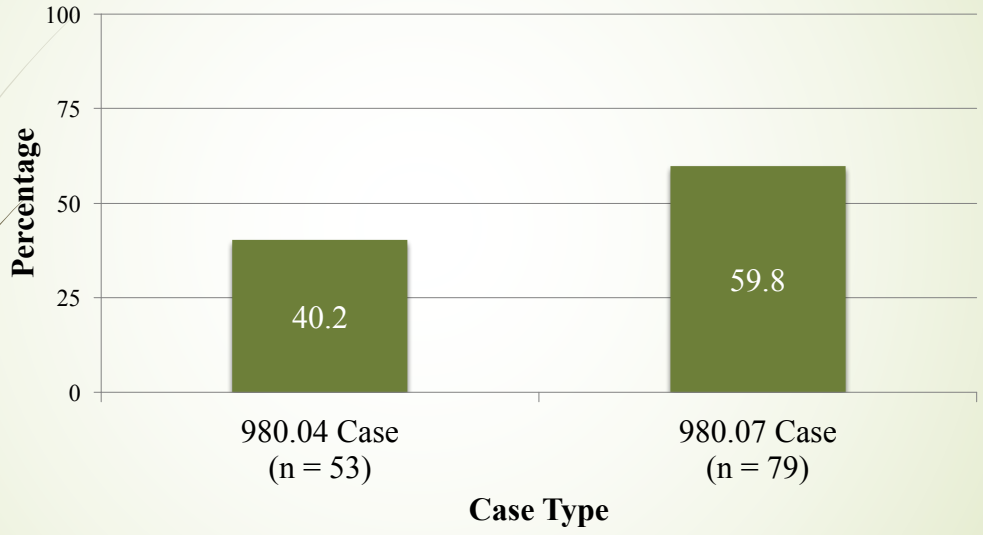
Poor Slight Fair Moderate Substantial Almost perfect

Kappa 0.0 .20 .40 .60 .80 1.0

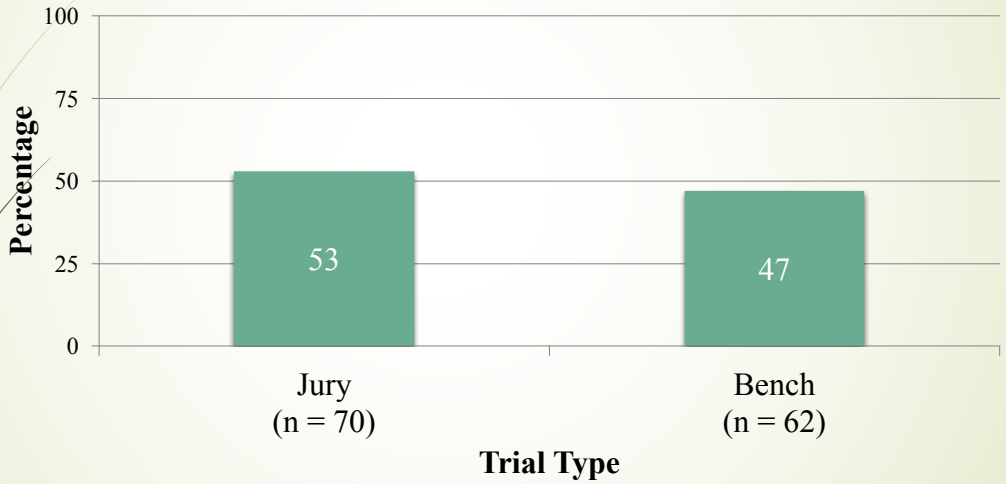
Kappa Agreement
< 0 Less than chance agreement
0.01–0.20 Slight agreement
0.21– 0.40 Fair agreement
0.41–0.60 Moderate agreement
0.61–0.80 Substantial agreement
0.81–0.99 Almost perfect agreement

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

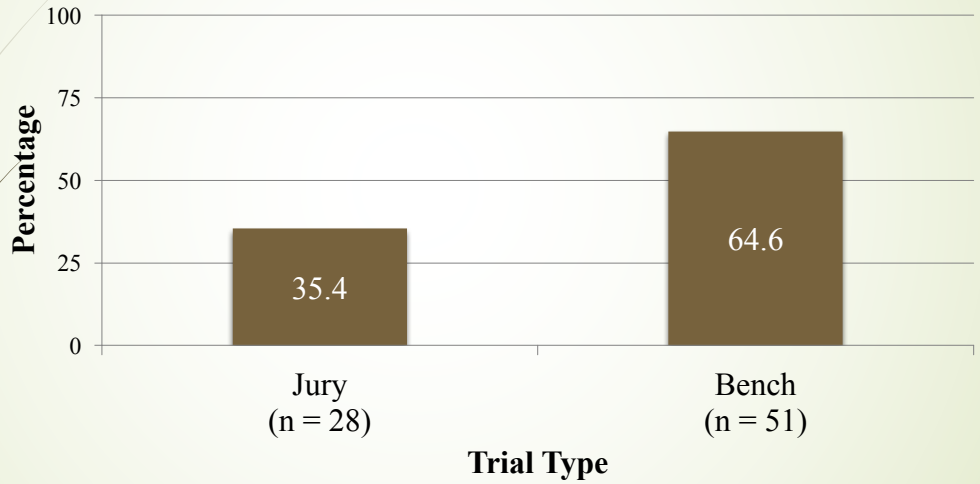
Case Type



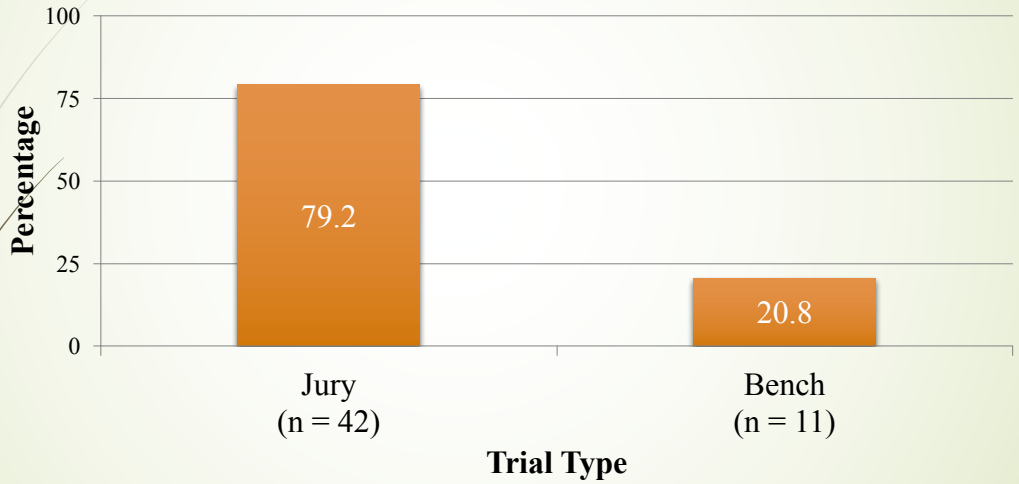
Jury or Bench Trial



980.07 Cases (*n* = 79) by Trial Type



980.04 Cases (n = 53) by Trial Type



Outcome Summary

Evaluator Recommendation:

47 (36%) Recommended for Dismissal/Discharge

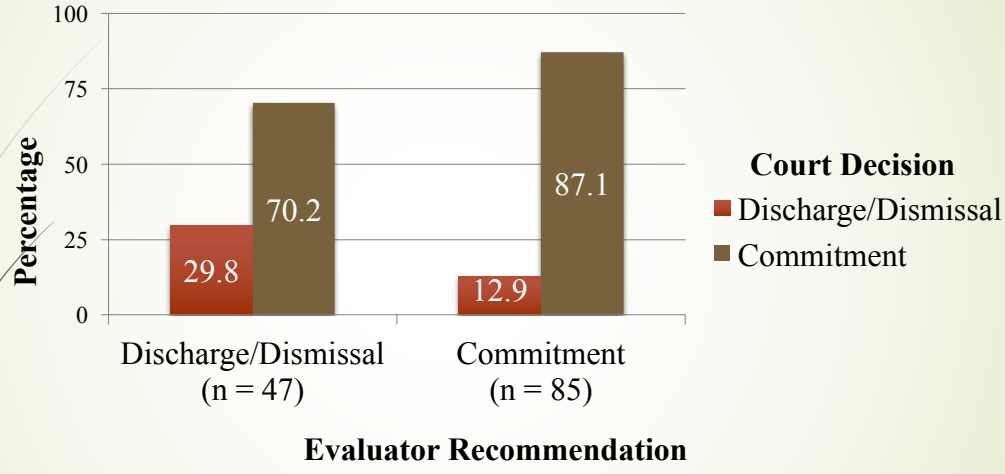
- 12 (26%) 04 trials
- 35 (74%) 07 trials
- 29 (62%) Jury trials
- 18 (38%) Bench trials

Court Outcome:

25 (19%) Dismissed/Discharged

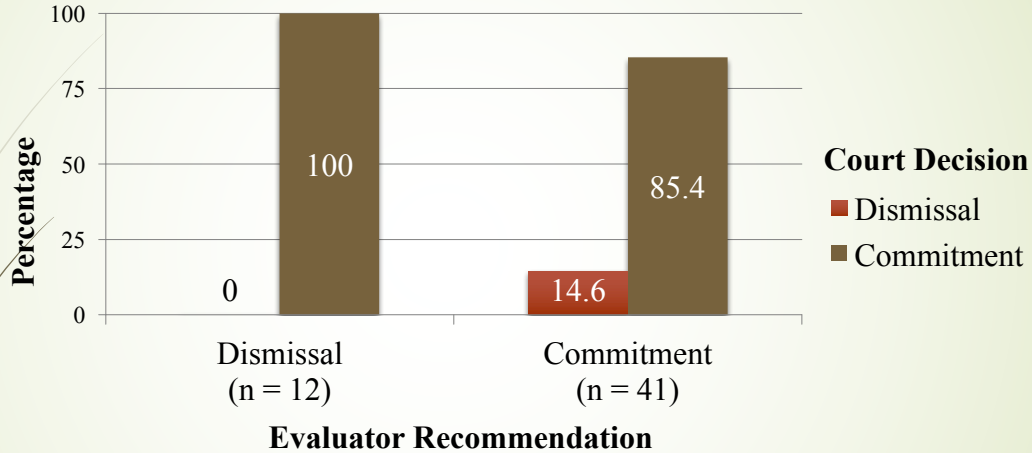
- 6 (24%) 04 trials
- 19 (76%) 07 trials
- 14 (56%) Jury trials
- 11 (44%) Bench trials

Agreement between Court and Evaluators (*n* = 132)



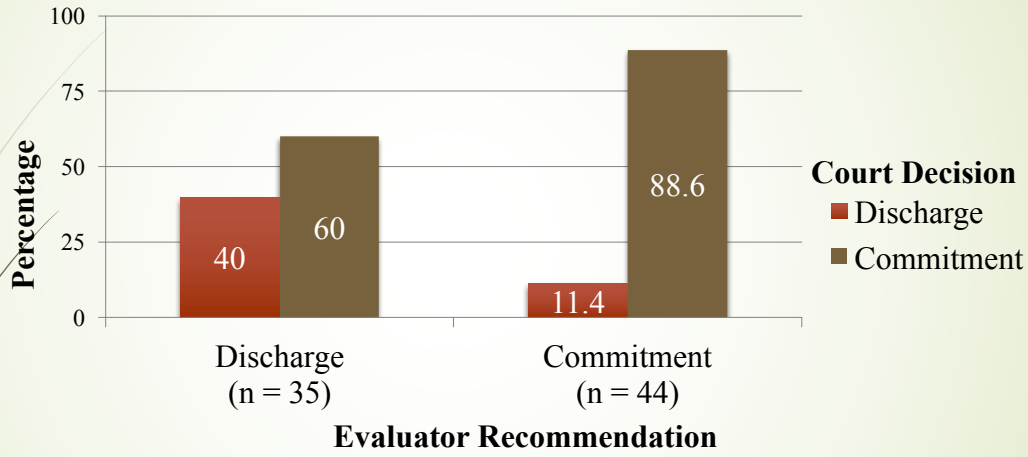
McNemar test, (p = .001); OR: 3.00 (95% CI: 1.52 – 5.94)
Kappa = .188 (Slight agreement) ✂ *.239 (Fair agreement)*

Breakdown by Case Type: 980.04 (*n* = 53)



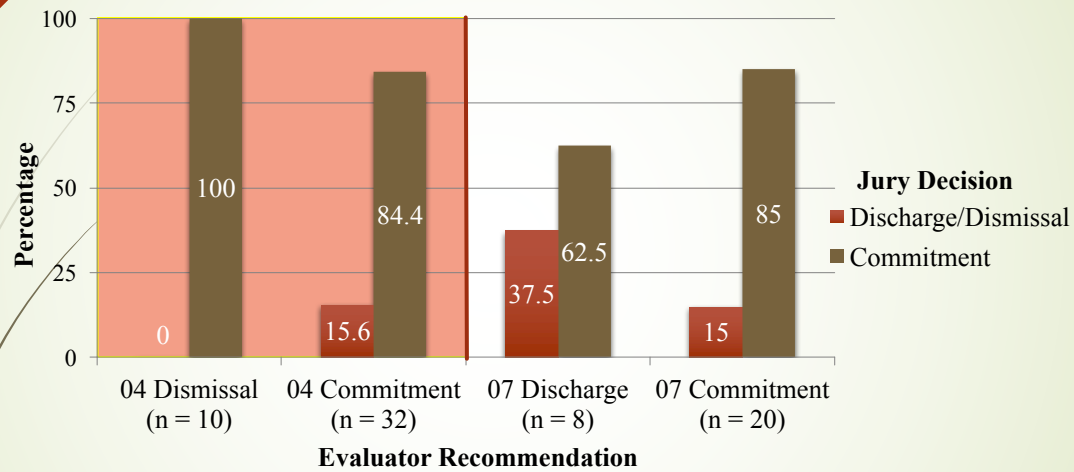
McNemar test, (p = .24); OR: 2.00 (95% CI: 0.75 – 5.32)
Kappa = -.178 (Less than chance) ✂ *-.158 (Less than chance)*

Breakdown by Case Type: 980.07 (*n* = 79)



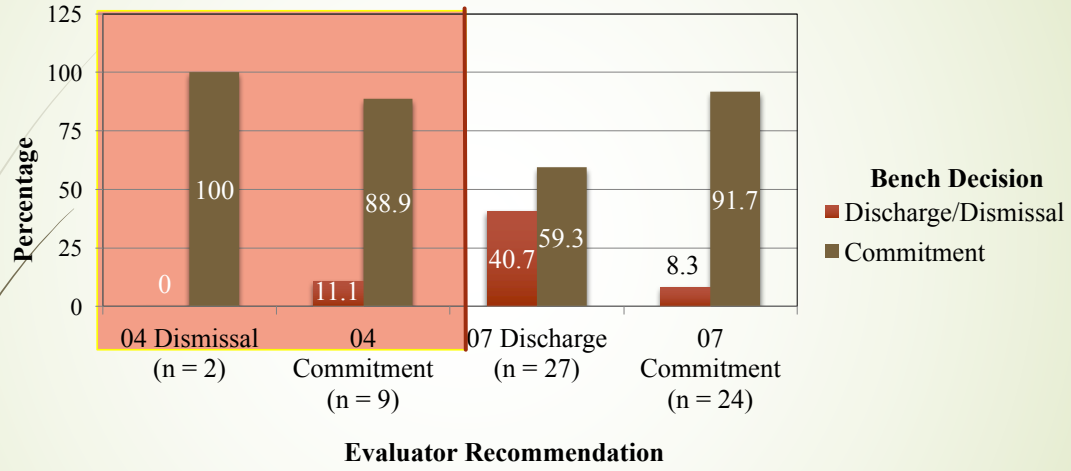
McNemar test (p < .01), OR: 4.2 (95% CI: 1.58 – 11.14)
Kappa = .300 (Fair agreement) *.334 (Fair agreement)*

Breakdown by Trial Type: Jury (*n* = 70)



04: McNemar test (p = .30), OR: 2.00 (95% CI = 0.68 – 5.85); Kappa = - .189 (Less than chance) *-.152 (Less than chance)*
07: McNemar test (p = .72), OR: 1.67 (95% CI = 0.39 – 6.97); Kappa = .243 (Fair agreement) *.213 (Fair agreement)*

Breakdown by Trial Type: Bench (n = 62)



04: Kappa = - .138 (Less than chance) ✂ .000 (Less than chance)
07: McNemar test (p = .001), OR: 8.00 (95% CI = 1.84 – 34.79); Kappa = .314 (Fair agreement) ✂ .411 (Moderate agreement)

Summary Points

- 47 (36%) patients were recommended for dismissal/discharge
- 25 (14%) patients were dismissed/discharged
- Across all trials there is "slight agreement" between courts and state evaluators
  - Agreement is rather poor for 04 trials regardless of jury or bench trials
  - Substantially higher (fair to moderate) in 07 trials for both jury and bench trials

Does evaluator recommendation predict the court decision? (n = 132)

	B (SE)	Lower	Odds Ratio	Upper
Jury / Bench	-.06 (.53)	.34	.94	1.28
Type of Case (04/07)	-.91 (.59)	.13	.41	1.28
Static99R	-.02 (.18)	.69	.98	1.39
PCL-R Total	-.07 (.06)	.83	.93	1.04
Evaluator sex	.87 (.57)	.99	2.38	7.19
Evaluator Recommendation	1.10* (.56)	.99	2.99	9.04

For 07 trials alone? (n = 79)

	B (SE)	Lower	Odds Ratio	Upper
Jury / Bench	-.11 (.64)	.26	.90	3.14
Static99R	-.00 (.22)	.65	1.00	1.53
PCL-R Total	-.08 (.07)	.80	.92	1.07
Evaluator sex	.39 (.67)	.40	1.48	5.44
Evaluator Recommendation	1.84* (.71)	1.57	6.29	25.17

For Jury trials alone? (n = 70)

	B (SE)	Lower	Odds Ratio	Upper
Case Type (04/07)	-.86 (.70)	.11	.42	1.67
Static99R	-.04 (.29)	.54	.96	1.70
PCL-R Total	-.07 (.08)	.79	.93	1.09
Evaluator sex	1.09 (.84)	.57	2.97	15.55
Evaluator Recommendation	.26 (.89)	.23	1.30	7.40

For Bench trials alone? (n = 62)

	B (SE)	Lower	Odds Ratio	Upper
Case Type (04/07)	-.99 (1.23)	.03	.37	4.11
Static99R	-.03 (.25)	.60	.97	1.58
PCL-R Total	-.08 (.09)	.78	.92	1.09
Evaluator sex	.55 (.85)	.33	1.73	9.07
Evaluator Recommendation	1.90* (.85)	1.27	6.71	35.53

Summary Points

- Evaluator recommendation is a significant predictor of ultimate court decision
  - After controlling for Static99R, PCL-R, case type (04/07), and evaluator gender
- **Particularly true for 07 bench trials**
- No variables predicted jury trial outcome
- Risk assessment instruments were not predictive of trial outcome.

Is there anything unique about those who were dismissed/discharged (n = 25)?

Age, Static99R, PCL-R, Time at SRSTC or in Treatment?

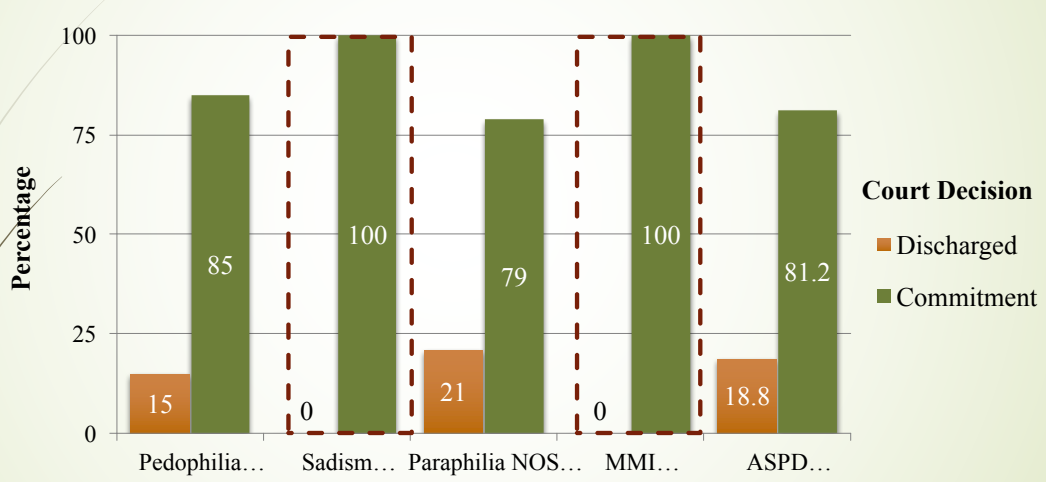
	Discharged/Dismissed (n = 25)		Committed (n = 107)			
	M	*SD*	*M*	*SD*	*T*-test	*Cohen's d*
Age	50.00	10.52	51.28	10.79	.54	.12
Static99R	5.12	1.45	5.47	1.70	.94	.22
PCL-R Score	25.51	5.23	25.17	4.96	-.31	.07
Days since Admission	3651.12	1960.20	2891.55	2144.04	-1.62	.39
Days in Treatment	1817.36	2031.77	1655.41	1803.32	-.40	.08

Differences in Race?

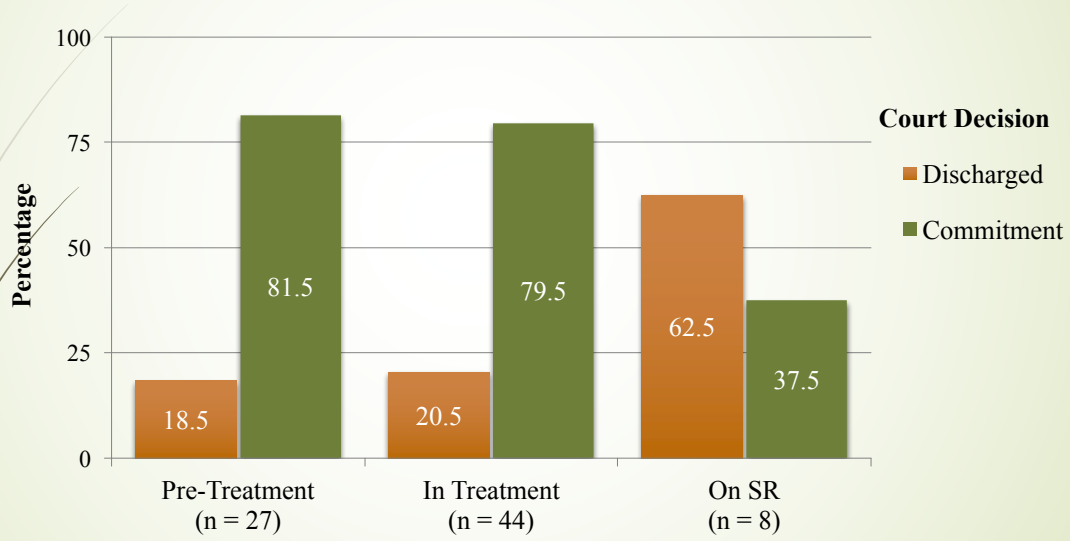
	Discharged/Dismissed		
	No	*Yes*	*Total*
Race			
Caucasian	65 (80.2%)	16 (19.8%)	81
Other	42 (82.4%)	9 (17.6%)	51
Total	107 (81.1%)	25 (18.9%)	132

$X^2(1) = .09$, $p = .76$, Cramer's V = .03

Predisposing Diagnoses



Treatment Phase (07 only)

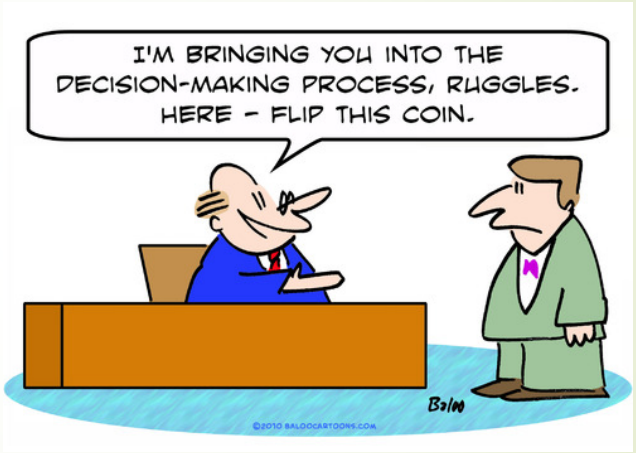


Summary

- Clearly commitment is a more likely outcome
- Agreement is variable
- Ranges from Less than chance to Fair
- State evaluator recommendation is a significant predictor for 07 bench trials
- Static99R is not a predictor of court outcome
  - Consistent with past research *(Boccaccini et al., 2013; Krauss et al., 2011; McCabe et al., 2010; Turner et al., 2015)*

Summary

- No meaningful differences between those discharged/dismissed and those who remained committed/committed
  - Potential diagnostic categories of interest (Sadism & MMI)
  - Supervised Release

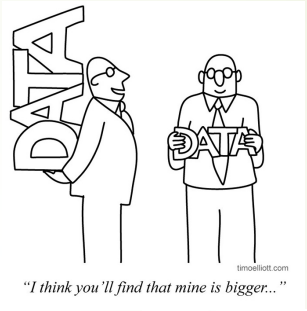


Main Limitations

- Cases sampled from 2012 to 2016 in Wisconsin
- Inability to capture all relevant variables
  - Other expert witnesses
  - Attorney experience
  - Conservativeness of court
  - Perceived quality of testimony
  - Severity of crimes
- Low base rate of discharge/dismissal
- Exclusion of agreement (i.e., stipulated agreements)

Future Directions

- Collection of additional years of trial outcomes
- Evaluation of perceived expert witness credibility and efficacy
- Jury / Judge decision making questionnaires



"I think you'll find that mine is bigger..."