

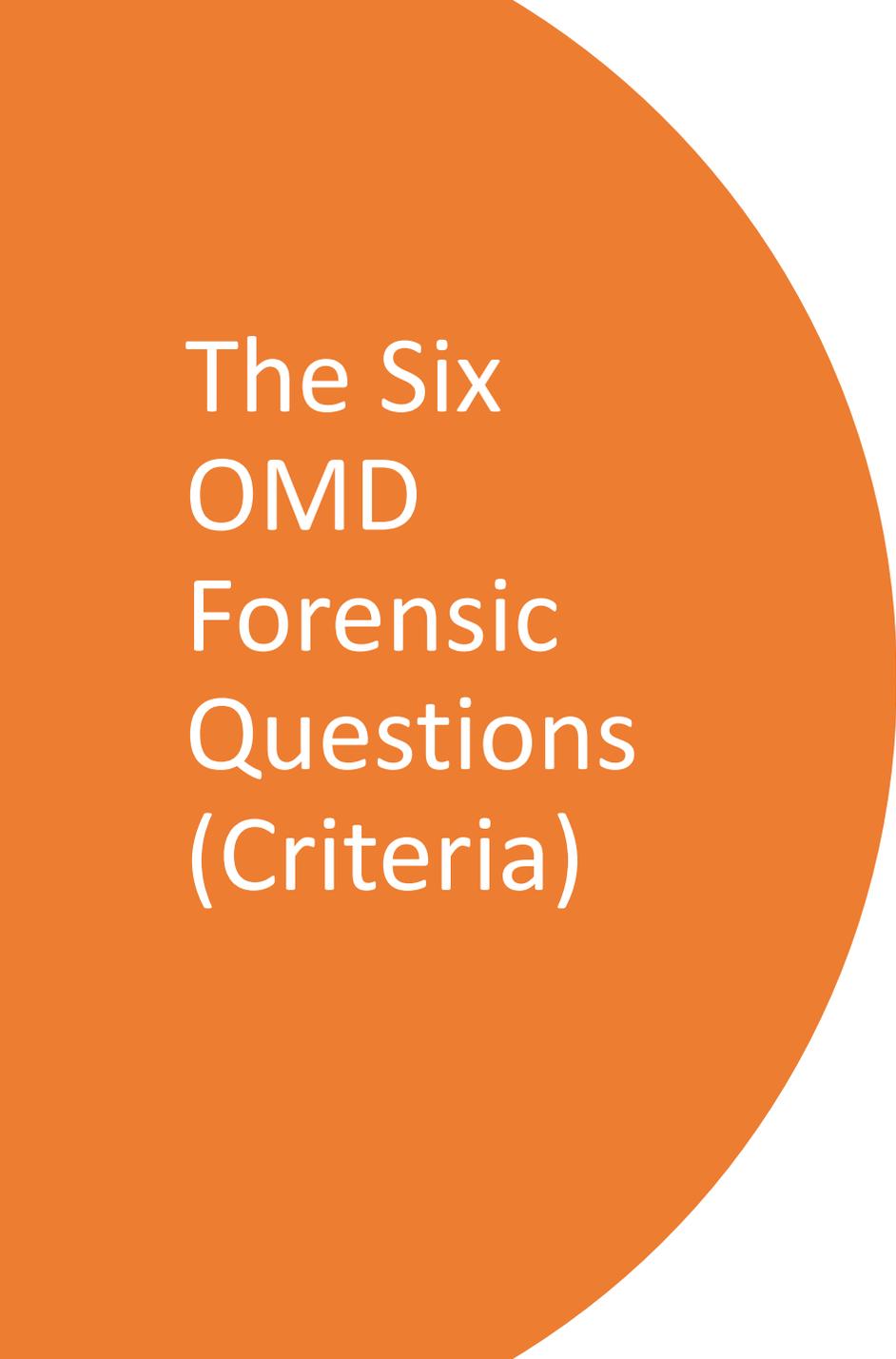
Interrater Reliability and Judicial Agreement:

**Factors Related to Decertification of
Mentally Disordered Offenders**

PC 2962

Offenders with Mental Health Disorders Statute and Genesis

- Enacted 1985, effective 07-01-1986
- ➤ Penal Code Sections 2960-81
- ➤ End of indeterminate sentencing (1977) Same crime = same time.
- Previously, BPH had wide discretion: shielded from judicial & legislative control.
- LPS-was the mechanism to transfer dangerously ill patients to a mental hospital. Screening was not routine.
- ➤ Public outcry regarding mentally ill offenders -SaldanaArthur Jackson was delusional and wanted to kill Saldana.
- Prior to and following release, Jackson was intent on Saldana.
- After release, Jackson attacked her (3-15-82 --stabbed 10 times).
- Saldana advocated for a change in the law.
- Jackson released before the OMHD (MDO) law took effect.
- Jackson deported to UK-and murders someone else.

A large orange circle is positioned on the left side of the slide, partially overlapping the text.

The Six OMD Forensic Questions (Criteria)

C1 Severe Mental Health Disorder (SMHD)

C2 Violent Crime

C3 SMHD as Cause or Aggravator

C4 Not in Remission

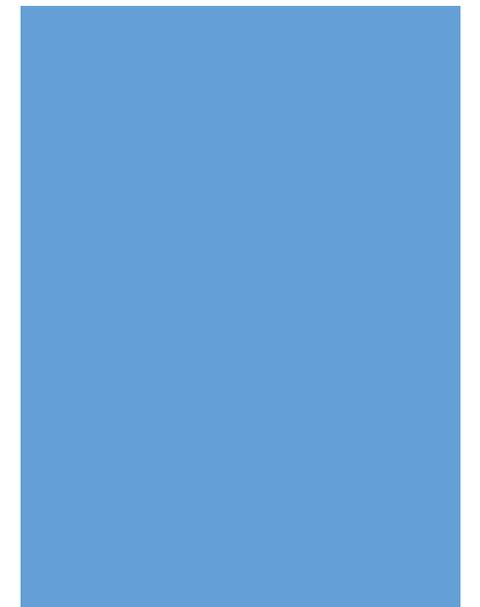
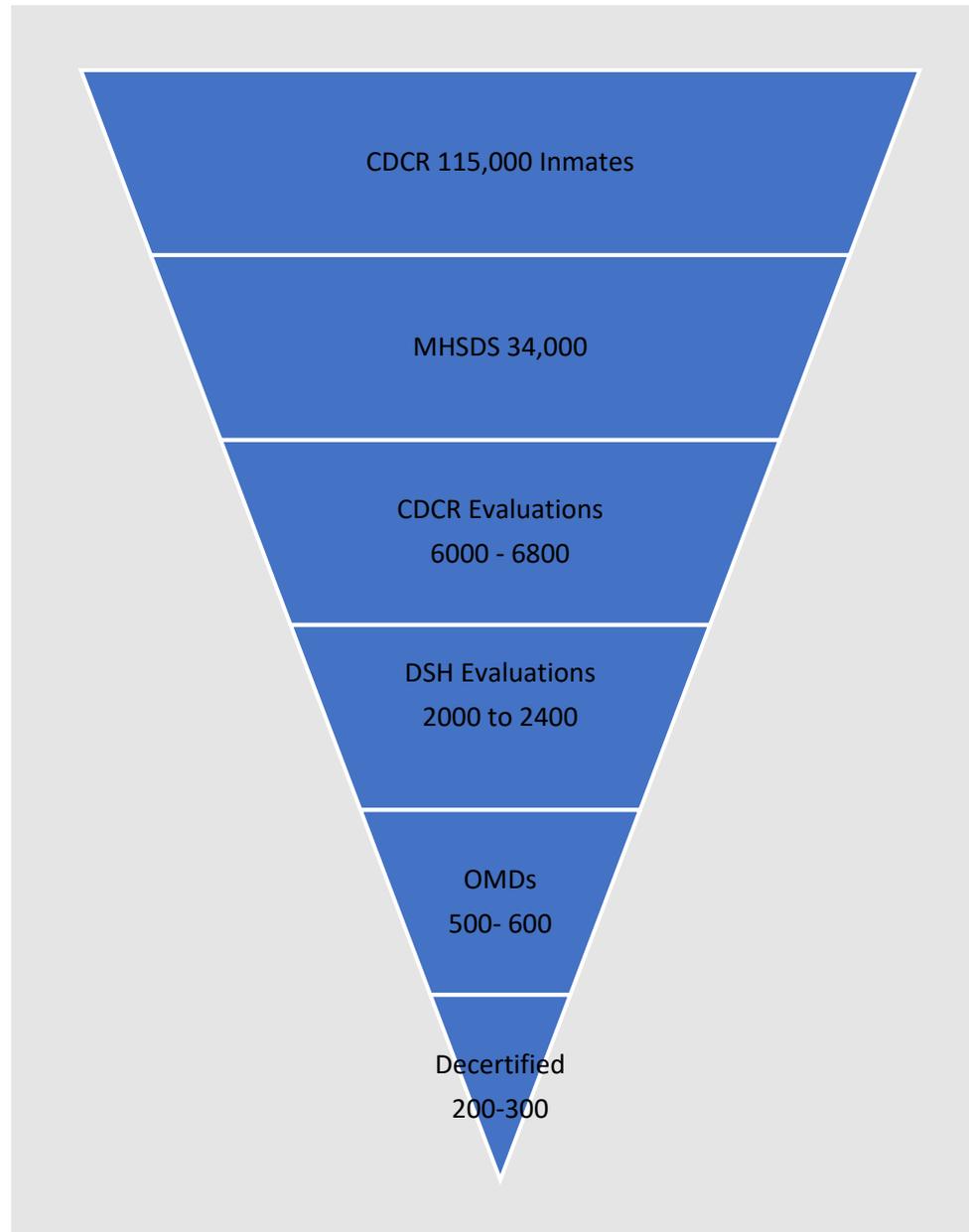
C5 90 days of Treatment of SMHD

C6 Danger of Physical Harm by Reason

The OMD Process

The Funnel

The OMD Process



Reliability

- Over the past two decades, the reliability of forensic mental health evaluations has drawn considerable interest, particularly for evaluations with real-life impacts on civil liberties and public safety.
- Research has revealed marked disparities in expert opinions on issues such as trial competency, legal sanity, and conditional release.
- Murrie and colleagues (2020) found marked variability in forensic mental health evaluator opinions in a study of 3,644 court-ordered competency to stand trial evaluations. In that study, the rate at which individual evaluators found a defendant incompetent to stand trial ranged from 9.1% to 76.88%.
- In studies of evaluator agreement in competency to stand trial evaluations, pairwise percentage agreements ranged from 57% to 100%, with reliability coefficients that ranged from fair ($\kappa = .28$) to perfect ($\kappa = 1.0$) agreement (Acklin et al., 2005; Gowensmith et al., 2013; Skeem et al., 1998).
- In a study of evaluator agreement in criminal responsibility cases, Fuger and colleagues (2014) found 46% triowise agreement among evaluators resulting in only fair agreement ($\kappa = .51$) Similarly, Nguyen and colleagues (2011) found that trios of forensic mental health evaluators reached consensus in only 40% of conditional release recommendations, resulting in poor agreement (ICC = .068).

Judicial-Evaluator Agreement

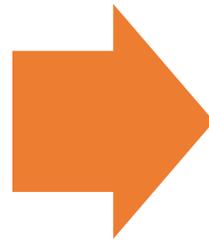
The rate that triers of fact agree with evaluators' opinions reflects the credence they lend to and the import they place upon the findings. When agreement is high, the utility and weight of the findings is high. Consequently, reliability becomes more important. Judges typically defer to the opinions of forensic examiners, with judicial-examiner agreement rates that often exceed 90% (Hart & Hare, 1992; Reich & Tookey, 1986).

Judicial agreement with examiners in cases of competency to stand trial evaluations was "substantial" (Acklin, Kennedy, Robinson, Dunkin, Dwire, & Lees, 2005).

Judicial decisions in competency to stand trial cases most often favored the "majority opinion," yet when the judges disagreed with the majority opinion, it was typically in a conservative manner. In other words, when judges disagreed with the majority of evaluators, it was to find defendants "incompetent to stand trial." Judges were also more likely to disagree with the majority opinion when there was disagreement amongst evaluators (Gowensmith, Murrie, & Boccaccini, 2013).

Judicial agreement with examiners in sanity cases was only “fair” (Fuger, Acklin, & Gowensmith, 2014). Judicial decisions in sanity cases were less likely to favor the majority opinion than was the case in evaluations of competency to stand trial.

Judges were far more likely to find a defendant sane than insane, even when the majority of evaluators believed the defendant was insane (Gowensmith, Murrie, & Boccaccini, 2013).



Judges were the least likely to agree with majority opinion in conditional release cases, as compared to competency or legal sanity cases, and were more likely to retain a hospitalized patient even when the majority of evaluators recommended conditional release (Gowensmith, Murrie, & Boccaccini, 2013)

Potentially Biasing and Influencing Factors

- Brain Cognitive Biases
 - Confirmation bias
 - Anchoring
- Individual Biases
 - Thresholds
 - Risk tolerance
 - Training
- Situational Factors
 - Time
 - Location
 - COVID-19 Lockdowns
 - Organizational philosophy
- Noise

Study One

Reliability

Method

- This research used archival data; there was no interaction between researchers and human subjects. Therefore, there were no actual participants and informed consent was not required.
- This sample consisted of matched-set reports written about 98 male prisoners who appealed their OMD commitment in court between February 2018 and April 2018 ($N = 98$)
- this study is focused on evaluations from three state organizations. Evaluations from Organization One were completed prior to those completed by Organization Two, and the evaluations from Organization Two were completed before those completed from Organization Three
- In preparation for this undertaking, we developed a coding manual, in which we specifically identified report quality factors for all MDO criteria except for Criterion Two. Two researchers then coded five cases together (15 reports), to learn the coding rules, as well as identify and refine any ambiguity in the coding manual. These two researchers then coded five cases independently, in order to establish acceptable percentage agreement
- The percentage agreement for these cases (15 reports, 357 codes) was 96%. Once strong percentage agreement was established, the same two researchers coded the remaining 88 cases independently. The overall percentage agreement for this study was respectable at 87%. The third researcher resolved coding disagreements by independently coding disputed variables.

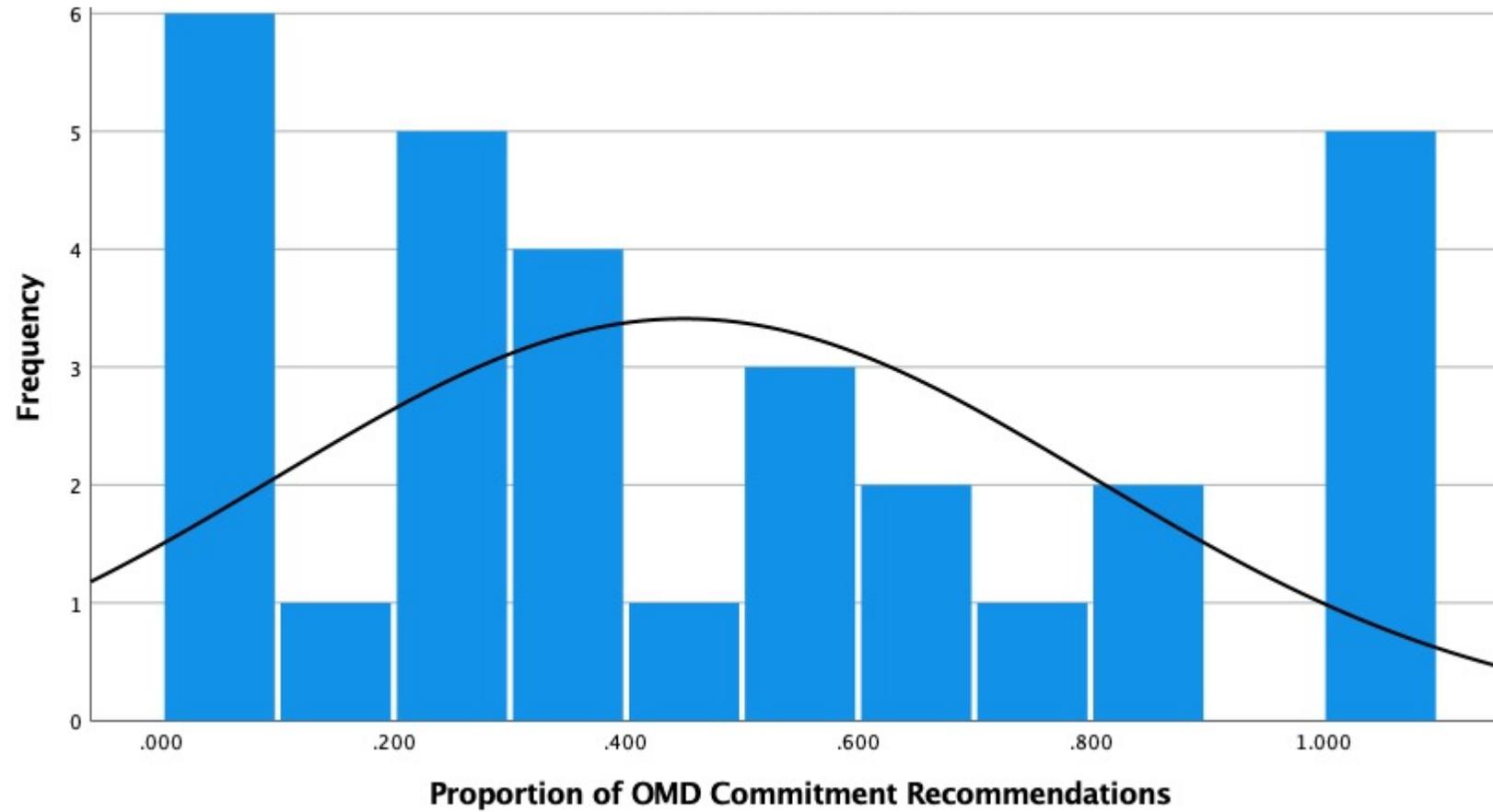
- We first investigated the rates at which OMD commitment recommendations came from each different organization

	Findings by Organization				
Organization	OMD	SMHD	SMHD + Crime	Remission	SMHD + Danger
One	.95	.99	.94	.00	.96
Two	.52	.81	.85	.07	.67
Three	.38	.87	.71	.39	.53
Effect Size (Cohen, 1998)	Medium	Small	Medium	-----	-----

Agreement Among Organizations

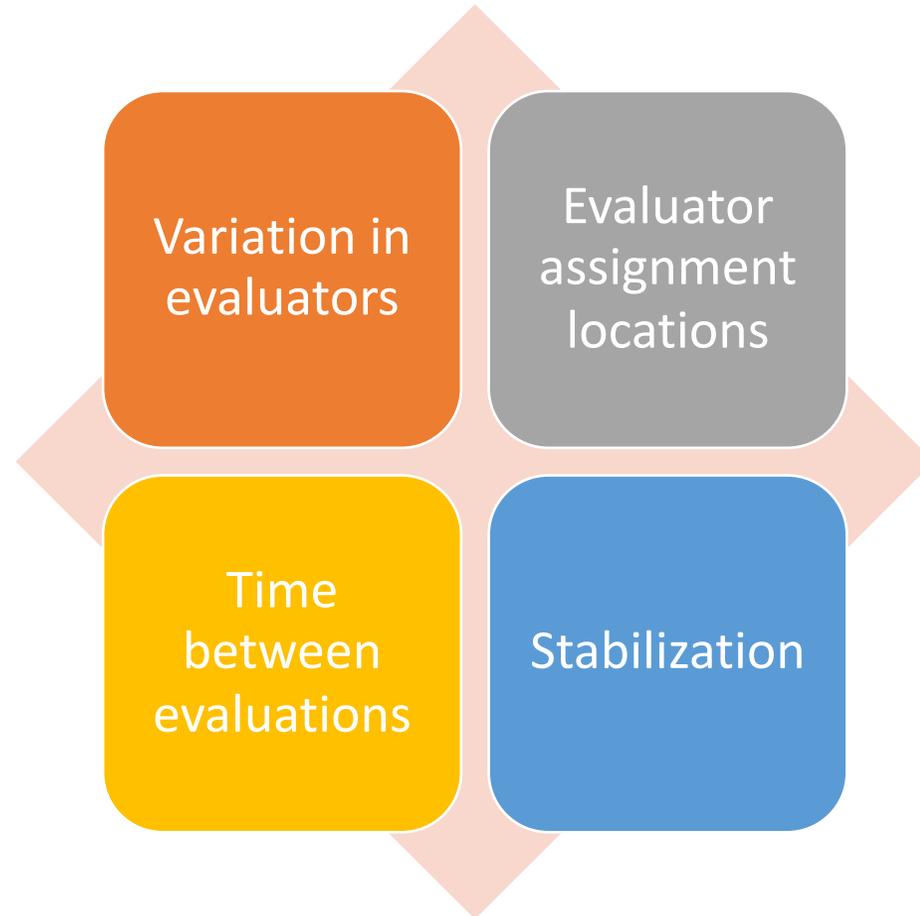
	Organizations One and Two	Organizations One and Three	Organizations Two and Three
OMD-Final Outcome	45.9	33.7	51.0
Severe Mental Health Disorder	79.6	84.7	80.6
Severe Mental Health Disorder + Crime	82.1	61.4	62.0
Remission	92.3	49.4	58.2
Severe Mental Health Disorder + Violence Risk	61.5	51.8	63.3

Range of Individual Evaluator Recommendations: Organizations Two and Three



			Court Disposition	
Evaluator Agreement	N of Cases	% of Cases	OMD	Not OMD
EVALUATORS AGREE:	56	57.1%		
Both Agree Not OMD	33	33.6%	0	100%
Both Agree OMD	23	23.4%	78.3%	21.7%
EVALUATORS DISAGREE:	42	42.8%		
Agency One Not OMD/Agency Two OMD	14	14.2%	57.1%	42.9%
Agency Two Not OMD/Agency One OMD	28	28.6%	0%	100%

Reasons for Outcomes



Ways to mitigate



JOINT AGENCY TRAINING



STANDARDIZED
ASSESSMENT PROTOCOL

Study Two

Judge Agreement

Agency	Proportion Agreement	Fleiss' kappa	Agreement	Standard Error	Lower 95% CI	Upper 95% CI	p-value
One/Two	.57	.13	slight	.06	.02	.25	.02
One/Judge	.58	.12	slight	.06	.01	.24	.03
Two/Judge	.89	.74	substantial	.06	.63	.86	.00
One/Two/Judge	.68	.33	fair	.06	.21	.44	.00

Outcomes

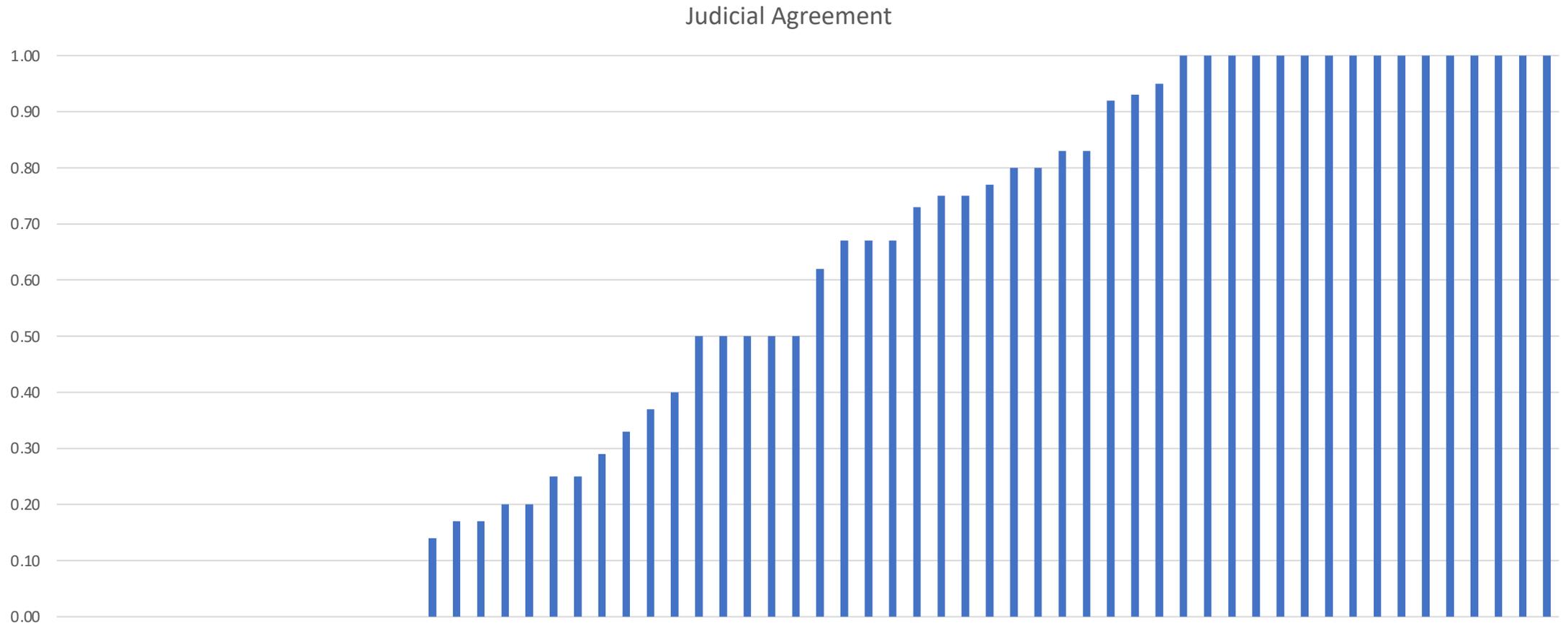
- It became evident that Judges were more likely to agree when evaluators from both state agencies recommended against commitment as an OMD. The Court agreed 100% of the time when evaluators from both state agencies determined that the criteria for OMD commitment were not met. Whereas, the Judge only agreed with evaluators 78.3% of the time when they both recommended commitment as an OMD.
- When there was disagreement amongst the evaluators from the state agencies, the Judge tended to agree with the evaluators from Agency Two, rather than Agency One. When Agency Two recommended commitment, the Judge agreed 57.1% of the time. When Agency Two recommended against commitment, the Judge agreed 100% of the time.

Individual Evaluator Factors

In addition to differences in opinions about whether or not a particular case met criteria for commitment or not, there were also substantial differences in evaluators' opinions as to whether or not a severe mental health disorder was present in any given case. In fact, evaluators from the three agencies could not agree on the presence of a severe mental health disorder in approximately 38% (27 of 72 cases) of the cases that were decertified, as opposed to only 4% (1 of 26 cases) of cases that were denied or withdrawn.

Disagreement among evaluators was associated with a higher likelihood of decertification; $\chi^2 = 10.6, p < .01$.

Rate of judicial agreement with individual evaluators



Report Factors

	“False Positive”	Threshold	Reasoning
56% In remission at time of Court hearing		X	
10% Neither evaluator found a SMHD	X		
4% Presentence Treatment or Unclear Timing			X
3% Nexus unclear to Judge SMHD/Violent Crime			X
3% Nexus unclear to Judge SMHD/Violence Risk			X
1% Not in remission but only mild symptoms		X	

Reasons for Outcomes

Given the pattern we observed in the qualitative examination of these reports, which suggested threshold and reasoning issues, we chose to examine this further.

We calculated the average number of threshold variances and reasoning issues for each evaluator ($M = .51$, $SD = .61$). We then calculated the proportion of OMD commitment recommendations that resulted in decertification for each evaluator.

There was a positive association for individual evaluators between the average number of threshold variances and reasoning issues in their reports and the proportion of their OMD commitment recommendations that resulted in decertification, ($r(28) = .50$, $p < .01$).

In other words, cases with more threshold variances and reasoning issues were more likely to be decertified than cases that involved fewer threshold and reasoning issues.

Reasons for Outcomes

- We also observed that some individual evaluators experienced a lower rate of judicial agreement than other evaluators who recommended OMD commitments ($n = 24$). The rate of judicial agreement with individual evaluator OMD recommendations ranged from 17% to 100%. The average rate of judicial agreement with individual evaluators was .73% ($M = .73$, $SD = 0.25$). Of the 24 evaluators who recommended OMD commitment, five (21%) had perfect judicial agreement. The remaining 79% (19 of 24) had one or more cases they recommended for OMD commitment and the judge disagreed. A higher proportion of OMD commitment recommendations was significantly associated with a lower rate of judicial agreement, ($r(96) = .38$, $p < .001$). That is, judges were less likely to agree with evaluators who were most apt to recommended OMD commitment

Mitigation Strategies



Region assignment



Exposure to outcomes,
including court outcomes



Judicial education to evaluators

Implications



Training implications



Civil liberties
implications



Justice implications