# HOW DO PROFESSIONALS ASSESS RISK? AN UPDATED SURVEY OF PRACTICES

**Sharon Kelley, Psy.D.**
**ATSA: Kansas City, MO**
**October 26, 2017**

# Background

- Risk assessment should be informed by research
- Clinicians and evaluators will need to periodically revise their assessment methodology in light of new research findings and best practice guidelines
- Adopting new methodologies can be difficult:
  - Learning new measures takes time and effort
  - Keeping up to date with research is time-consuming and potentially expensive
  - Instruments used in forensic settings must meet legal standards for admissibility (Daubert; Frye)
  - Employment context may limit this
- Surveys allow us to compare our methods with overall trends

# Background

- In 2013, surveyed predominately ATSA members on use of static actuarial measures, mechanical dynamic measures, and Structured Professional Judgment (SPJ) measures
- N = 158 participants
- Mostly from United States (*n* = 109)
- Included participants who completed sexual risk assessments for the court (*n* = 73) and SVP evaluators (*n* = 56)

Kelley, S.M., Barahal, R. M., Thornton, D., & Ambroziak, G. (2017). How do professionals assess sexual recidivism risk? An international survey of practices. The Forum Newsletter of the Association for the Treatment of Sexual Abusers, 29(1), 1-13.

# 2013 Survey results - limitations

- ATSA-list participants may represent a subgroup who keep up to date with research
  - What are other professionals doing?
- There have been notable advances since 2013 so the data may already be stale:
  - 2015 Static-99R norms paper
  - 2016 Static-99R coding manual
  - 2014 ATSA Practice Guidelines for the Assessment, Treatment, and Supervision of Individuals with Intellectual Disabilities and Problematic Sexual Behaviors
- Numerous questions we wished we had asked
  - How are they choosing a Static-99R reference group? What norms are they using?
  - Is selection limited by institutional requirements?
  - Is there a difference when people work in private practice versus within institutions?

# 2017 Survey

- Electronic survey sent out to members of
  - ATSA
  - SOCCPN (Sex Offender Civil Commitment Program Network)
  - AP-LS (American Psychology and Law Society / Division 41 of APA)
  - IATSO (International Association for the Treatment of Sexual Offenders)
- Some participants forwarded emails and other professional groups are included
- 34 questions about risk assessment practices
- Responses March 2017 – June 2017

# 2017 Survey

- Has risk assessment usage changed since 2013?
- Is risk assessment usage changing with empirical advances?
  - For example: Are evaluators using the most current norms?
- What influences evaluators' choice of instruments?

# 2017 SURVEY RESULTS

# Participants (N = 145)

PRIMARY ROLE

Evaluator = 103 (71.0%)
Treatment Provider = 32 (22.1%)
Researcher = 3 (2.1%)
Other = 7 (4.8%)

AGE OF CLIENTELE

Adults =137 (94.5%)
Adolescents = 42 (29.0%)
Children = 6 (4.1%)

NOTE: Participants could endorse working in more than one role and with more than one age group of clientele

# EXCLUSIONS

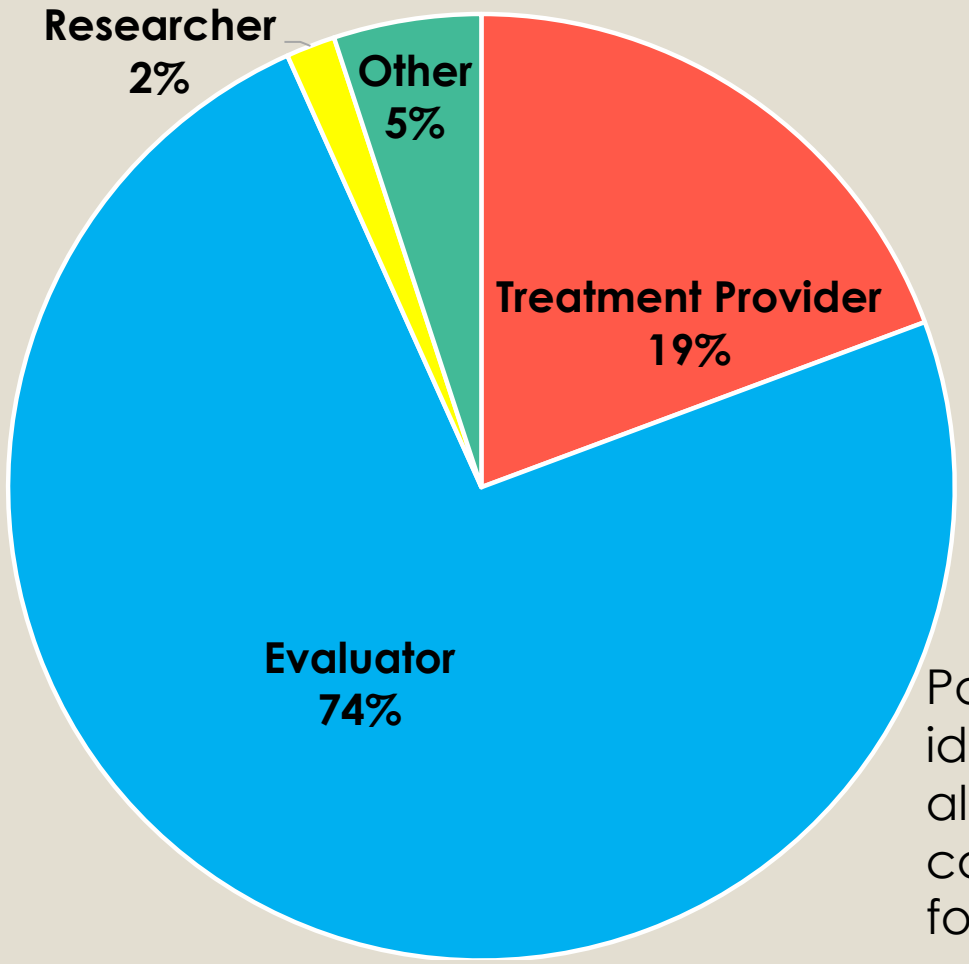
**No adult clients (*n* = 8)**

**OR**

**Not completing risk assessments for court (*n* = 18)**

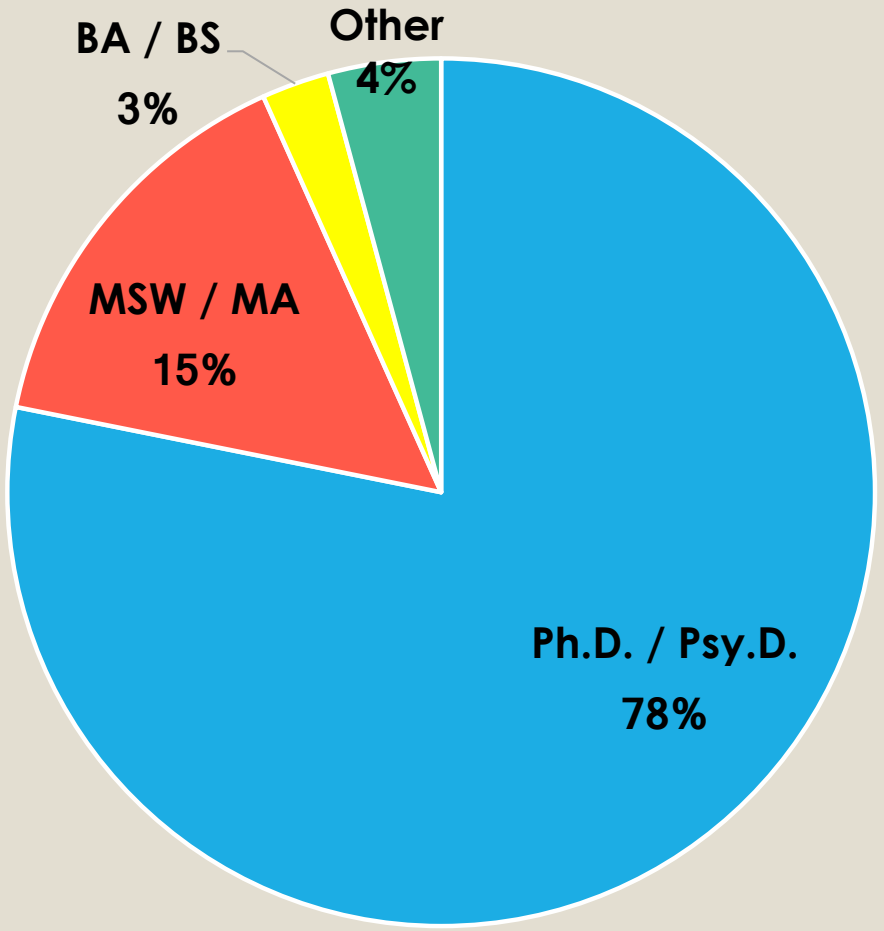
**N = 119**

# Primary role (N = 119)




Participants were asked to identify their *primary* role, although everyone included completes risk assessments for the court

# Degree (N = 119)

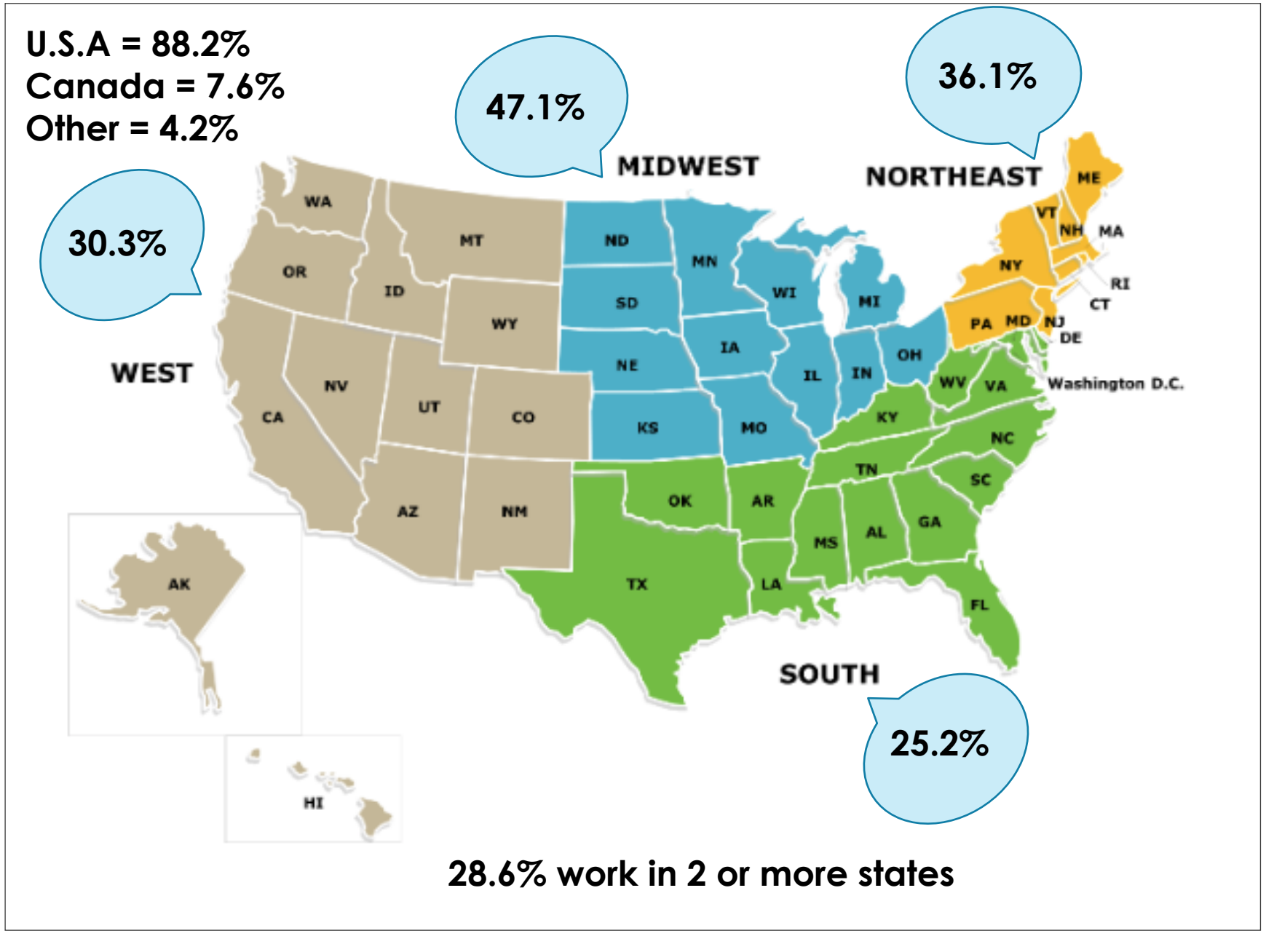


# Years of experience (N = 119)

Range = 0.5 – 40
Mean = 13.2 (SD = 9.3)
Median = 12

63.9% had ≥ 10 years of experience

# Where do participants work?



# Professional membership

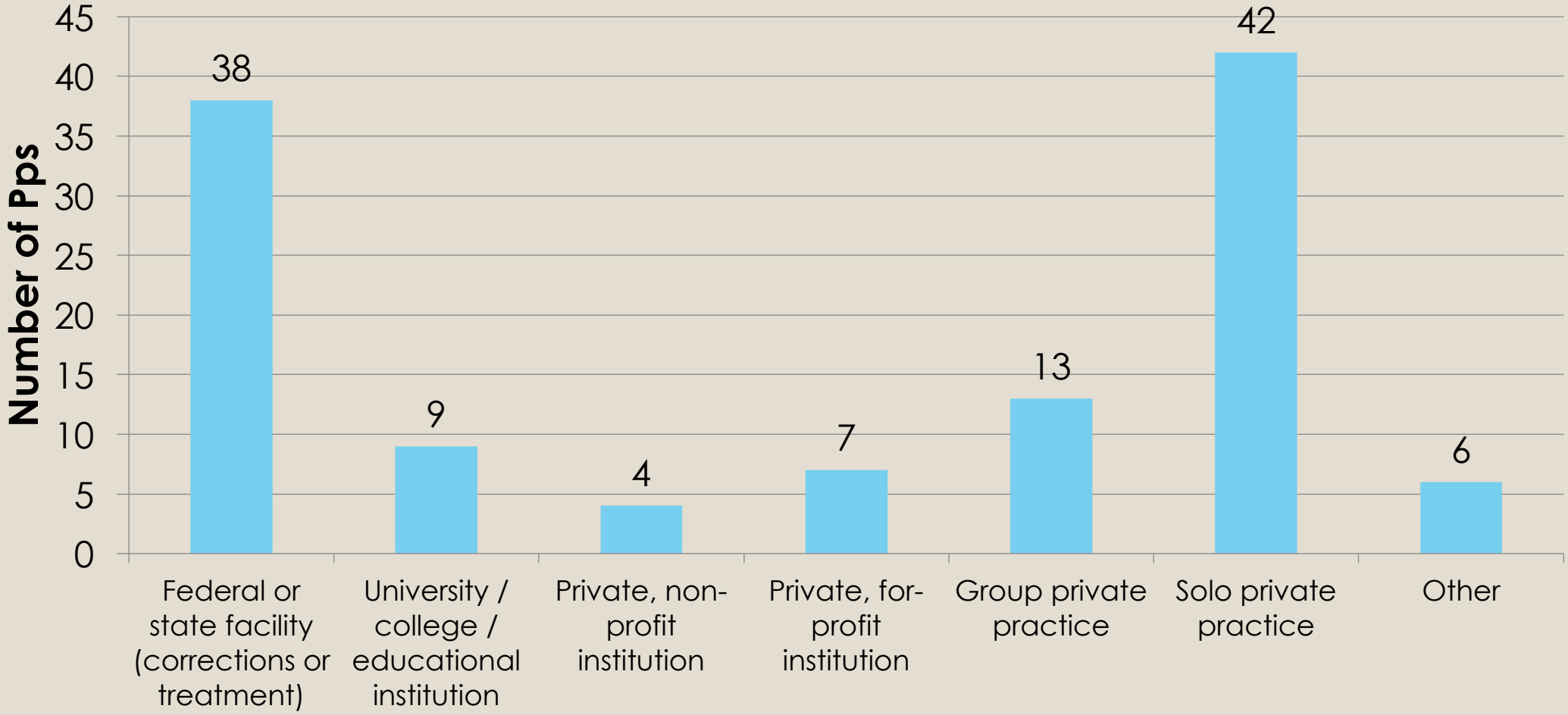
- ATSA = 78 (65.5%)
- IATSO = 4 (3.4%)
- AP-LS = 77 (64.7%)
- SOCCPN = 13 (10.9%)

- ATSA only = 32 (26.9%)
- AP-LS only = 35 (29.4%)
- SOCCPN only = 1 (0.8%)
- Mixed membership = 47 (39.5%)
- None = 4 (3.4%)

# Client population

- Court System / Charged = 80 (67.2%)
- Incarcerated = 37 (31.1%)
- Parole / Probation = 48 (40.3%)
- Any SVP / SDP = 59 (49.6%)
  - Committed = 52 (43.7%)
  - Post-Probable Cause = 39 (32.8%)
- Outpatient = 26 (21.8%)

# Employment


Number of Pps
45
40
35
30
25
20
15
10
5
0
38
9
4
7
13
42
6
Federal or state facility (corrections or treatment)
University / college / educational institution
Private, non-profit institution
Private, for-profit institution
Group private practice
Solo private practice
Other

# Research & training: Keeping up to date

- Regularly read research articles = 104 (87.4%)
- National training and conferences = 87 (73.1%)
- Local training and conferences outside worksite = 85 (71.4%)
- Webinars = 73 (61.3%)
- Team meetings at worksite = 45 (37.8%)
- Presented at professional conferences = 47 (39.5%)
- In-house training by worksite = 42 (35.3%)
- Completed research / published articles = 34 (28.6%)
- Peer reviewer for journal = 27 (22.7%)
- Journal editorial board = 13 (10.9%)
- 73.1% of Pp rely on 4 or more of the methods to keep up to date
  - **Range = 1 – 9**
  - **Median = 5**

# ATSA conference attendance

- Recently = 45 (37.8%)
  - 2016 = 21 (17.6%)
  - 2015 = 16 (13.4%)
  - 2014 = 8 (6.7%)
- Less Recently = 24 (20.2%)
  - Within the last 5 years = 14 (11.8%)
  - With the last 10 years = 8 (6.7%)
  - > 10 years ago = 2 (1.7%)
- Never = 50 (42.0%)

# Assessment methods

| Methodology | Frequency | % |
|---|---|---|
| Independently choose & change from case to case | 61 | 51.3 |
| Independently choose & does not change from case to case | 30 | 25.2 |
| Chosen, but approved in advance & different methodologies for different cases | 5 | 4.2 |
| Chosen, but approved in advance & does not change from case to case | 2 | 1.7 |
| Fixed methodology by the institution or contract, but negotiable depending on the case | 17 | 14.3 |
| Fixed methodology by the institution or contract & non-negotiable | 4 | 3.4 |
| Total | 119 | 100.0 |

# Static risk assessment (N = 119)

# Static risk instrument use

Note: Participants could choose more than one

| Instrument | Use in Past Year | | Routine Use | |
|---|---|---|---|---|
| | *Frequency* | *%* | *Frequency* | *%* |
| Static-99 | 9 | 7.6 | 7 | 5.9 |
| Static-99R | 96* | 80.7 | 98* | 82.4 |
| Static-2002 | 4 | 3.4 | 1 | 0.8 |
| Static-2002R | 36 | 30.3 | 23 | 19.3 |
| VRS-SO Static | 15 | 12.6 | 4 | 3.4 |
| MnSOST-R | 6 | 5.0 | 4 | 3.4 |
| MnSOST-III | 2 | 1.7 | 2 | 1.7 |
| MATS-1 | 2 | 1.7 | 1 | 0.8 |
| RRASOR | 9 | 7.6 | 7 | 5.9 |
| Risk Matrix 2000 | 9 | 7.6 | 5 | 4.2 |
| SVR-20 | 32 | 26.9 | 17 | 14.3 |
| CPORT | 10 | 8.4 | 4 | 3.4 |

1

**Use of multiple static instruments in the same evaluation**

- **33 (27.7%)**

2

**Use of Static-99R and Static-2002R in the same evaluation**

- **23 (19.3%)**

3

**Use of an OLD and NEW in the same evaluation**

- **9 (7.6%)**

4

**Static instruments in same evaluation**

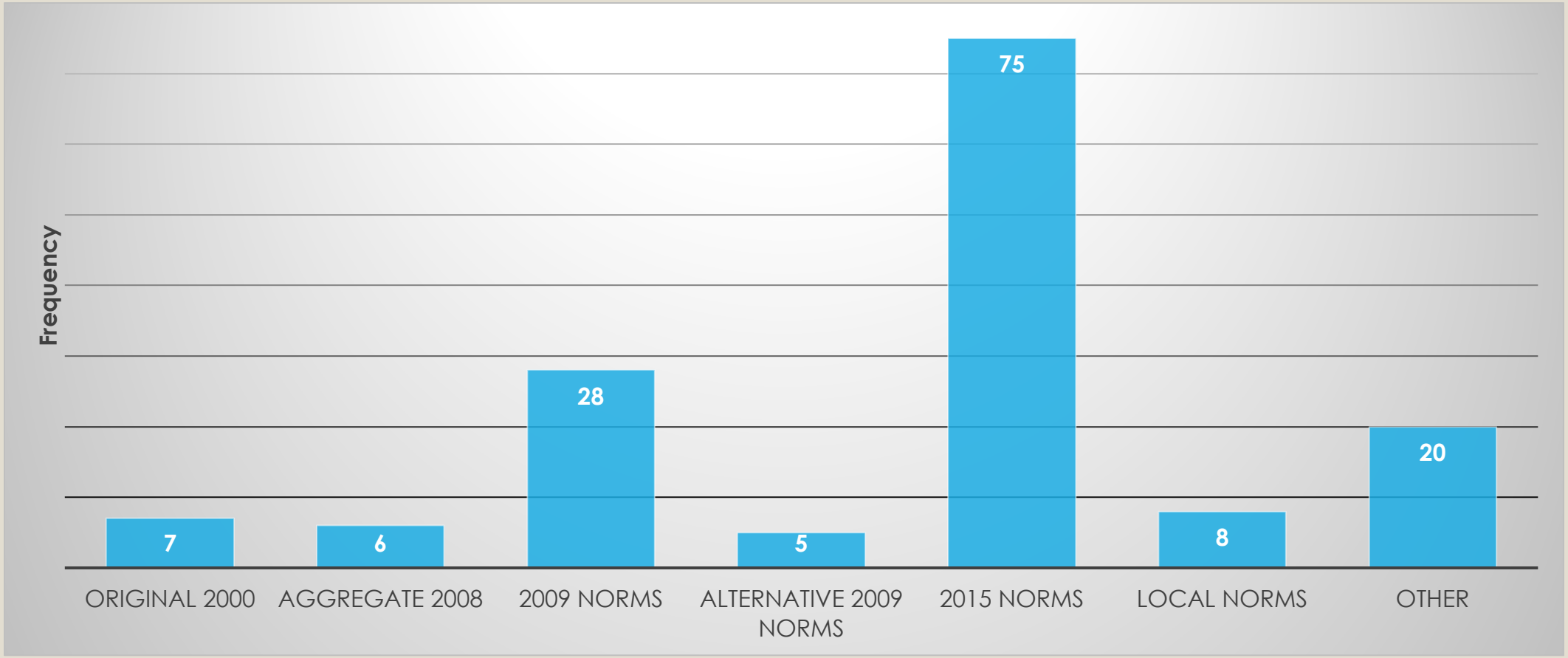
- **0 = 17 (14.3%)**
- **1 = 69 (58.0%)**
- **2 = 27 (22.7%)**
- **3 or 4 = 6 (5.1%)**

# Use of multiple static risk instruments

# Static-99/R coding manual

- **2003 publication = 16 (13.4%)**
- **2016 publication = 85 (71.4 %)**
- **N/A = 18 (15.1%)**

# Use of norms for Static-99/R



# Use of norms for Static-99/R

**Use of multiple norms in same evaluation**

| Norms | Frequency | % |
|---|---|---|
| 0 | 17 | 14.3% |
| 1 | 80 | 67.2% |
| 2 | 19 | 16.0% |
| 3 | 2 | 1.7% |
| 4 | 0 | - |
| 5 | 1 | 0.8% |
| Total | 119 | 100% |

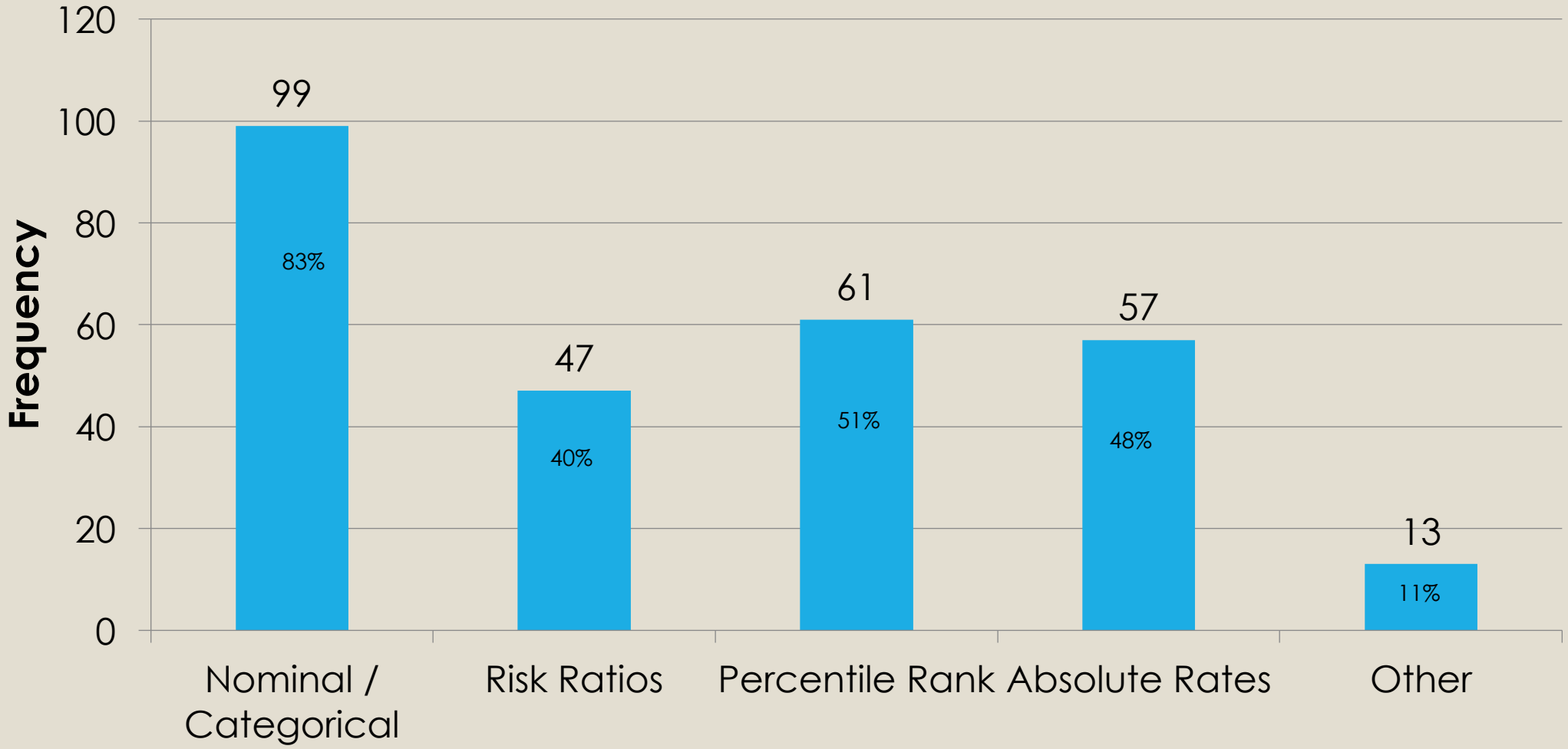
**Use of local norms**

| Local norm | Frequency |
|---|---|
| New York | 2 |
| Florida (SVP) | 1 |
| California | 1 |
| Washington | 1 |
| Unspecified | 3 |
| Total | 8 |

# Static-99R: reference groups

| Selection of Reference Group | Frequency | % |
|---|---|---|
| "Matching" based on historical selection factors | 14 | 11.8 |
| "Matching" based on a current case formulation / clinical judgment of external risk factors | 21 | 17.6 |
| **"Matching" combined** | **35** | **29.4%** |
| Use a mechanical measure of psychological risk | 25 | 21.0 |
| Only use the Routine/Complete group | 30 | 25.2 |
| Other | 7 | 5.9 |
| N/A | 22 | 18.5 |
| Total | 119 | 100.0 |

# Static risk communication



# Dynamic risk assessment (N = 119)

# Dynamic risk instrument use

Note: Participants could choose more than one

| Instrument | Use in Past Year | | Routine Use | |
|---|---|---|---|---|
| | *Frequency* | *%* | *Frequency* | *%* |
| STABLE-2007 | 60 | 50.4 | 50 | 42.0 |
| SVR-20 | 27 | 22.7 | 20 | 16.8 |
| VRS-SO | 19 | 16.0 | 15 | 12.6 |
| RSVP | 19 | 16.0 | 15 | 12.6 |
| SOTIPS | 12 | 10.1 | 9 | 7.6 |
| SRA-FV | 11 | 9.2 | 10 | 8.4 |
| ARMIDILO-S | 7 | 5.9 | 5 | 4.2 |
| MIDSA | 4 | 3.4 | 1 | 0.8 |
| SARN | 2 | 1.7 | 2 | 1.7 |
| None | 25 | 21.0 | 26 | 21.8 |
| Other | 13 | 10.9 | 15 | 12.6 |

# Dynamic risk assessment: Routine use

- Mechanical Dynamic Risk Assessment
  - 72 (60.5%)
- SPJ Dynamic Risk Assessment
  - 35 (29.4%)
- ANY Dynamic Risk Assessment
  - 85 (71.4%)

# Use of multiple dynamic risk instruments within same evaluation

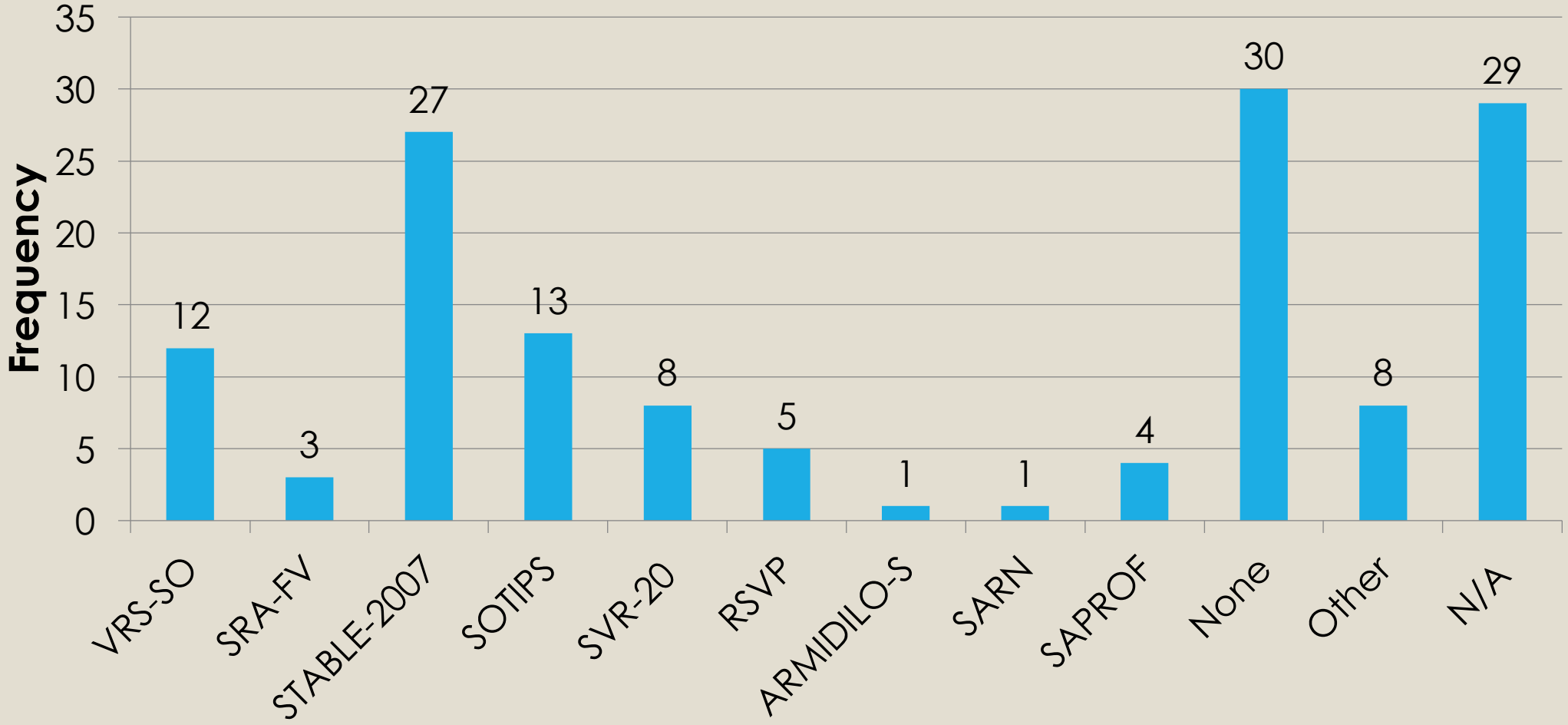
- Use of multiple DRF instruments in same evaluation
  - 18 (15.1%)
- Use of both MECHANICAL and SPJ in same evaluation
  - 9 (7.6%)

# Dynamic risk assessment: Now & then

- N = 96
- Excludes:
  - N = 18 (not doing risk assessments in 2013)
  - N = 5 (could not recall)

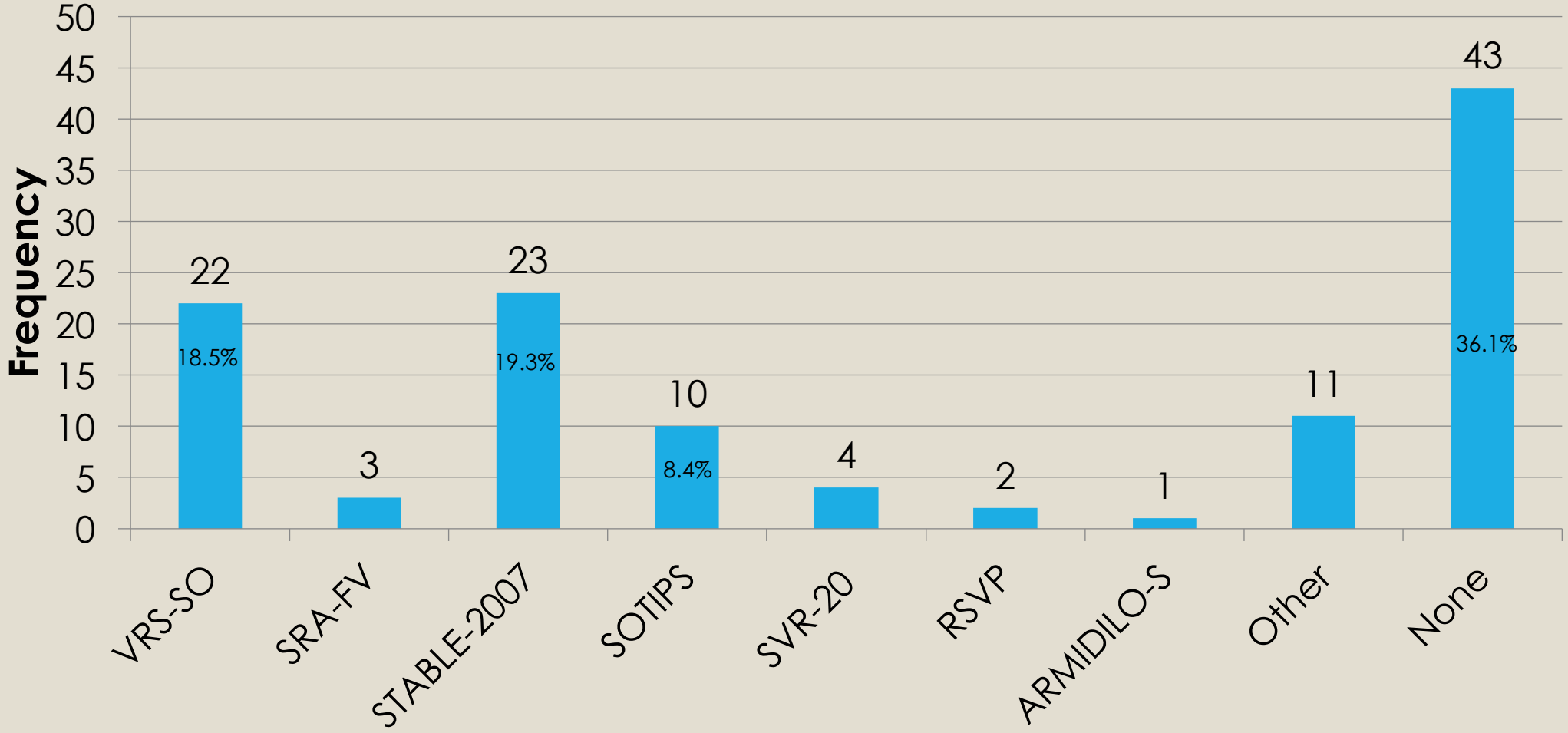
| Instrument | Used in 2013 | | 2017 Routine Use | | % CHANGE |
|---|---|---|---|---|---|
| | *Frequency* | *%* | *Frequency* | *%* | |
| VRS-SO | 6 | 6.3 | 12 | 12.5 | **6.2** |
| SRA-FV | 10 | 10.4 | 9 | 9.4 | **-1.0** |
| STABLE-2007 | 44 | 45.8 | 40 | 41.7 | **-4.1** |
| SOTIPS | 6 | 6.3 | 7 | 7.3 | **1.0** |
| SVR-20 | 20 | 20.8 | 15 | 15.6 | **-5.2** |
| RSVP | 9 | 9.4 | 10 | 10.4 | **1.0** |
| MIDSA | 0 | 0.0 | 1 | 1.0 | **1.0** |
| ARMIDILO-S | 3 | 3.1 | 3 | 3.1 | **0.0** |
| SARN | 1 | 1.0 | 1 | 1.0 | **0.0** |
| None | 22 | 22.9 | 23 | 24.0 | **1.1** |
| Other | 8 | 8.3 | 12 | 12.5 | **4.2** |

# Measuring treatment gains (N = 119)




Note: Participants could choose more than one

# Which measure do you think has the best research support? (N = 119)



# Dynamic risk assessment: reasons not used

| Why not using DRF | Frequency | % |
|---|---|---|
| Not enough research to support use | 25 | 21.0 |
| Available norms not large enough | 11 | 9.2 |
| Available norms not representative of relevant population | 7 | 5.9 |
| Too time consuming | 1 | 0.8 |
| Lack of training | 6 | 5.0 |
| Other | 7 | 5.9 |
| Not applicable | 83 | 69.7 |

# Protective factors

# Protective factors (N = 119)

| Instrument | Frequency | % |
|---|---|---|
| SAPROF | 12 | 10.1 |
| START | 2 | 1.7 |
| DUNDRUM | 0 | - |
| IORNS | 2 | 1.7 |
| Qualitative Description | 70 | 58.8 |
| No Protective Factors Assessment | 22 | 18.5 |
| Other Protective Factor Assessment | 11 | 9.2 |

Note: Participants could choose more than one

# Protective factors: Now & then

| Instrument | Used in 2013 | | 2017 Routine Use | | % CHANGE |
|---|---|---|---|---|---|
| | *Frequency* | *%* | *Frequency* | *%* | |
| SAPROF | 4 | 4.9% | 6 | 7.4% | **+2.5** |
| USE OF ANY PF SCALE *(includes measures for youth)* | 17 | 21.0% | 18 | 22.2% | **+1.2** |
| No protective factors assessment | 48 | 59.3% | 13 | 16.1% | **-43.2** |

- N = 81
- Excludes:
  - n = 19 (not doing risk assessments in 2013)
  - n = 19 (could not recall)

# Differences in methods

- Professional memberships?
- Freedom to select methods?
- Type of employment?
- Involvement in research and training activities?

# There were no statistically significant differences for the following:

1. Amount of freedom (low v. high) in choice of methodology and use of
   - Old static instruments ($\chi 2\ (1) = 1.169,\ p = .280$)
   - New static instruments ($\chi 2\ (1) = 1.498,\ p = .221$)
   - Any dynamic risk instruments ($\chi 2\ (1) = 1.133,\ p = .287$)
2. Amount of research & training* activities and use of
   - Old static instruments ($\chi 2\ (2) = 4.528,\ p = .104$)
   - New static instruments ($\chi 2\ (2) = 0.176,\ p = .916$)
   - Any dynamic risk instruments ($\chi 2\ (2) = 4.470,\ p = .107$)

*categorized as limited, moderate, and extensive

# Does professional membership make a difference?

- ATSA members may have more specialized knowledge of sexual risk assessment than AP-LS only members
- Therefore, ATSA members might make more use of new static instruments and dynamic risk instruments

# Effect of membership on use of new static instruments

| Membership | *Using NEW Instrument* | |
|---|---|---|
| | *Frequency* | *%* |
| ATSA only | 28 | 87.5 |
| AP-LS only | 25 | 71.4 |
| Mixed Membership | 43 | 91.5 |
| Total | 96 | 84.2 |

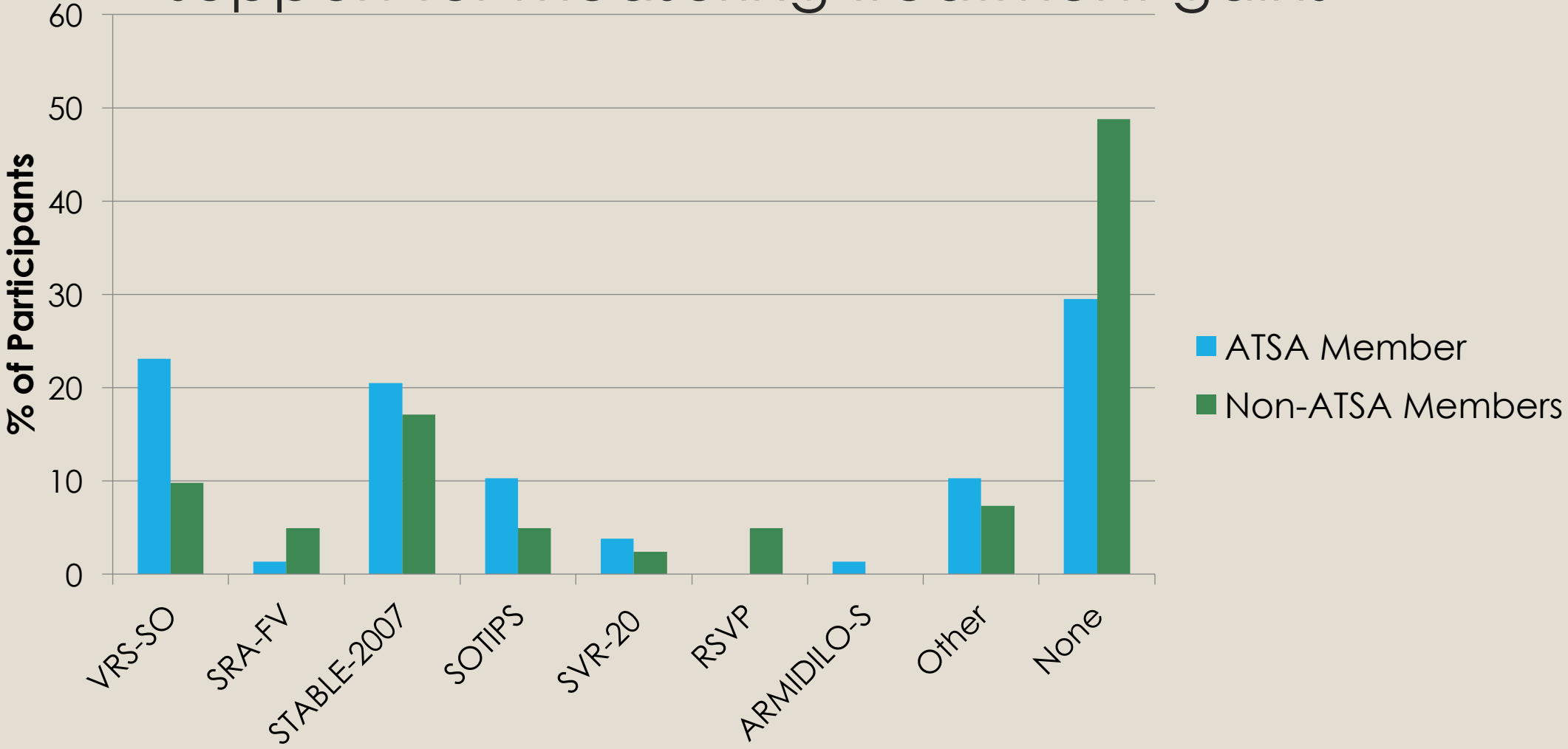
$\chi 2\ (2) = 6.434,\ p = .040$

# Effect of membership on use of mechanical dynamic instruments

| Membership | Routine Use of Mechanical DRF Instrument | |
|---|---|---|
| | *Frequency* | *%* |
| ATSA | 53 | 67.9 |
| Non-ATSA | 19 | 46.3 |
| Total | 72 | 60.5 |

$\chi^2(1) = 5.250,\ p = .022$

# Instruments judged to have the best research support for measuring treatment gains



# Does employment setting predict methodology?

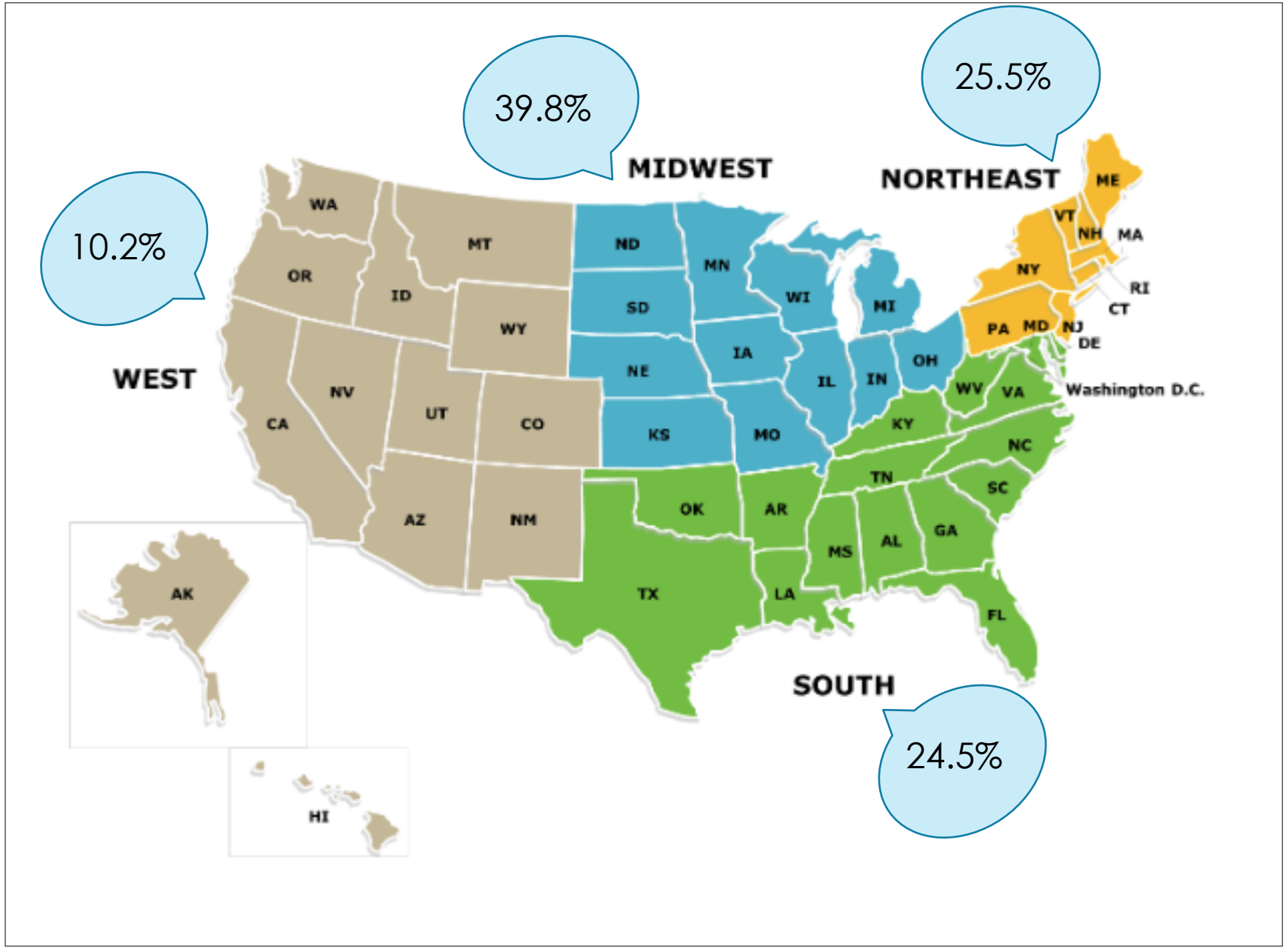
# Effect of employment on reference group choice

| Employment | Matching | | Use of Instrument | | Routine Only | |
|---|---|---|---|---|---|---|
| | *Frequency* | *%* | *Frequency* | *%* | *Frequency* | *%* |
| Private Practice | 13 | 31.0 | 9 | 21.4 | 20 | 47.6 |
| Other | 22 | 45.8 | 16 | 33.3 | 10 | 20.8 |
| Total | 35 | 38.9 | 25 | 27.8 | 30 | 33.3 |

$\chi 2\ (2) = 7.240,\ p = .027$

# CLIENT POPULATION: SVP

*n* = 59




# Where do participants work?

# Static risk instruments in SVP: Routine use

| Instrument | Frequency | % |
|---|---|---|
| Static-99 | 2 | 3.4 |
| Static-99R | 52 | 88.1 |
| Static-2002R | 16 | 27.1 |
| VRS-SO Static | 3 | 5.1 |
| MnSOST-R | 0 | - |
| MnSOST-III | 1 | 1.7 |
| MATS-1 | 1 | 1.7 |
| RRASOR | 2 | 3.4 |
| Risk Matrix 2000 | 2 | 3.4 |
| SVR-20 | 10 | 16.9 |
| CPORT | 3 | 5.1 |

# Dynamic risk instrument in SVP: Routine use

| Instrument | Frequency | % |
|---|---|---|
| STABLE-2007 | 20 | 33.9 |
| VRS-SO | 11 | 18.6 |
| SVR-20 | 8 | 13.6 |
| SRA-FV | 7 | 11.9 |
| RSVP | 5 | 8.5 |
| SOTIPS | 2 | 3.4 |
| ARMIDILO-S | 1 | 1.7 |
| MIDSA | 0 | - |
| SARN | 0 | - |
| None | 17 | 28.8 |
| Other | 5 | 8.5 |

# Protective factors in SVP: Routine use

| Instrument | Frequency | % |
|---|---|---|
| SAPROF | 6 | 10.2 |
| START | 0 | - |
| DUNDRUM | 0 | - |
| IORNS | 1 | 1.7 |
| Qualitative Description | 40 | 67.8 |
| No Protective Factors Assessment | 7 | 11.9 |
| Other Protective Factor Assessment | 6 | 10.2 |

# Use of any DRF instrument in SVP evaluations

| Works with SVP Clients | Using DRF Instrument | |
|---|---|---|
| | *Frequency* | *%* |
| No | 48 | 80.0 |
| Yes | 37 | 62.7 |
| Total | 85 | 71.4 |

$\chi 2\ (1) = 4.357,\ p = .037$

# Use of absolute recidivism rates in SVP evaluations

| Works with SVP Clients | Reports Absolute Recidivism Rates | |
|---|---|---|
| | *Frequency* | *%* |
| No | 18 | 30.0 |
| Yes | 39 | 66.1 |
| Total | 57 | 47.9 |

$\chi 2\ (1) = 15.536,\ p < .001$

# Limitations

- Mostly respondents from USA
- Unclear what legal question they must answer
- Would be helpful to know whether respondents complete "neutral" evaluations or predominately work for defense/prosecution

# Co-collaborators

- Gina Ambroziak, B.S.
- Robert M. Barahal, Ph.D.
- David Thornton, Ph.D.

Contact:
Sharon M. Kelley, Psy.D.
Sand Ridge Evaluation Unit
SharonM.Kelley@dhs.wi.gov

# BRINGING IT ALL TOGETHER

Final thoughts and recommendations

# Final thoughts & recommendations

- This symposium explored decision-making by evaluators and the court

- Juries may have more difficulty using empirical data (i.e., Static-99R scores; estimated recidivism rates) when making decisions (Turner et al., 2015)
- Anecdotal evidence suggests juries in WI worry about being responsible for allowing a potentially dangerous individual back into their community and use a threshold lower than 50%.

# Final thoughts & recommendations

- Quality of expert testimony may help: character, credibility, impartiality, etc
- Our use of language in testimony may affect jury decision-making

- Use of labels activates the juror's stereotypes (Scurich & Krauss, 2014)
  - Ch. 980.04 Sexually Violent Person Evaluation Report
  - Sex Offender Treatment Program
- Use of person-first language may help to neutralize this effect (Willis, submitted)
  - Sexual Risk Assessment Pursuant to Ch. 980.04
  - Sex Offense-Specific Treatment Program

# Final thoughts & recommendations

- There appears to be an evaluator effect
- This affects how case information is weighed and what risk instruments are utilized

- While we are not advocating for a "right" way, we suggest reducing the effect of factors that degrade risk prediction
- Consider potential sources of bias:
  1. How might your current/past allegiances affect your decision-making? (e.g., employment)
  2. How do you review relevant materials?
     - First impressions matter; unbalanced assessments → confirmation bias
  3. How do you assign weight to factors outside a static risk instrument?
  4. Are you an informed consumer of the research and risk tools?

# Final thoughts & recommendations

- Being aware of potential bias is not enough
- Treat forensic evaluations like scientific inquiries: Test hypotheses
- Structured measures help to anchor ratings and increase reliability
- Write reports that are transparent
- Participate in trainings and read….

Zapf & Dror, 2017

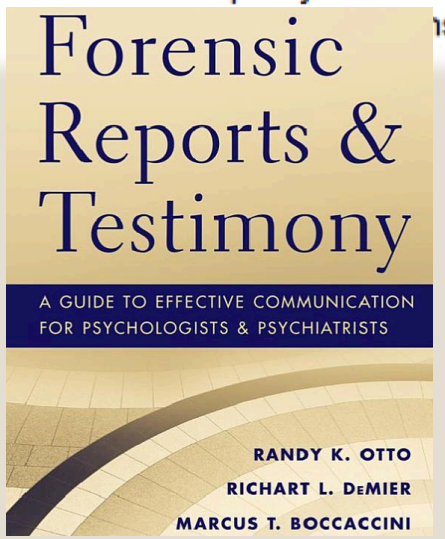
Open Access Journal of Forensic Psychology

http://www.forensicpsychologyunbound.ws/ – 2010.2: 102-115

## Guidance for Improving Forensic Reports: A Review of Common Errors

Thomas Grisso, University of Massachusetts medical School, Worcester, MA
Thomas.Grisso@umassmed.edu

**Abstract:** This study employed a n
tiqued by a panel of advanced forer
for the American Board of Forensic
types of faults that reviewers encou
scriptive statements to guide forens
frequent report-writing problems ir
describe the quality of forensic re
sic rep




INTERNATIONAL JOURNAL OF FORENSIC MENTAL HEALTH
https://doi.org/10.1080/14999013.2017.1317302

Routledge
Taylor & Francis Group

Check for updates

## Understanding and Mitigating Bias in Forensic Evaluation: Lessons from Forensic Science

Patricia A. Zapf[a] and Itiel E. Dror[b]

[a]Department of Psychology, John Jay College of Criminal Justice, The City University of New Yo
London, London, United Kingdom

**ABSTRACT**
Criticism has emerged in the last decade surrounding cognitive bias in forensic ex
National Research Council (NRC, 2009) issued a report that delineated weaknesse
forensic science domains. The purpose of this article is to examine and consi
influences that can bias observations and inferences in forensic evaluation and to
know from forensic science to propose possible solutions to these problems. We
Bacon's doctrine of idols—which underpins modern scientific method—to expar
five-level taxonomy of the various stages at which bias can originate within fore
create a seven-level taxonomy. We describe the ways in which biases can arise and
forensic evaluation at these seven levels, highlighting potential solutions and va
mitigating the impact of these biases, and conclude with a proposal for using scient
improve forensic evaluation.