Pretrial Risk Assessment: Presentation to Whatcom County

Zachary Hamilton, Ph.D.

2017-09-25

Outline

- History, development and use of risk assessments
- Current risk assessment use in pretrial settings
- Assessment functionality
- Development of Spokane and King County Tools
- Implementing a tool: What to consider

HISTORY, DEVELOPMENT AND USE OF RISK ASSESSMENTS

History

- Psychology of Criminal Conduct (Andrews & Bonta, 1994)
  - Risk, Needs, and Responsivity (RNR)
    - *Risk* – level of service/supervision should match an offender's risk for recidivism
    - *Needs* – dynamic risk factors that, when changed, are associated with changes in the probability of recidivism (Andrews & Bonta, 2010)
    - *Responsivity* – matching intervention to an offender's learning style, abilities, & strengths initial formulation of generalized risk assessment tools.
  - 4 generations of assessment (Andrews and Wormith, 2005)
    - 1G – Clinical Judgment (lacked structure)
    - 2G – Static Items (criminal history)
    - 3G – Dynamic/Needs Items
    - 4G – Responsivity & Automation
    - Note – increased generations do not necessarily equate to increased performance

Can assessments improve human judgement?

- 1G tools are based on human discretion/Expertise
  - Generational moves were not in response to inherent bias
  - Still need for human judgement
- Tools attempt to guide humans and remove idiosyncratic nature of day-to-day decision
  - Humans are notoriously inconsistent –tools found to have higher predictive accuracy
  - Bring with them good and bad baggage that can impact decision making
    - Family or community impacted by substance abuse, violence, mental health issues
    - Political pressures
    - In ability to calculate impact of many predictive factors simultaneously
- Assessments have become and evidence-based Practice
  - Still, will not apply to every case
  - Low and high risk anomalies will exist
  - Human discretion still needed

What makes up and assessment

- Risk items
  - Increase probability of recidivism (any item type)
- Static Items
  - Client's criminal history, which are included in a risk assessment to predict future behavior
- Dynamic Items
  - Client's current circumstances, including needs regarding education, employment, housing, and others, to predict future behavior and inform case management.
- Domains
  - Common areas of interest, included in needs assessments, to predict criminal behavior
  - Used as predictors in a risk assessment
  - Can be used to predict a variety of outcomes
  - Often used to identify areas of programming needs

Common Domains

assessments Variations to consider

- Items predicted by the risk assessment vary greatly
  - Some tools as small as 13 items, others have over 100
- Vary by system stage
  - Recidivism - Violent, Property, Drug, Sex, 'Any'
  - Failure to appear (FTA)
  - Technical violations
  - Needs
- Duration of prediction can vary as well
  - Recidivism – commonly 2-3 years for correctional tools
  - FTA – can vary depending on the length of time in the community
- Gender Responsivity
  - Males and females come into crime through differing pathways
  - Van voorhis and colleagues (2010) identified risk assessment variations differed
  - Creating separate models for each gender improves context and prediction (Hamilton et al., 2016)

Selecting, maintaining, and updating justice assessment tools

- Difficult landscape to navigate
  - Many studies but no *Consumer Reports*

- Practitioners seek to use 'validated' tools
  - Tools are typically developed of a singular population/jurisdiction
  - Have been applied liberally 'off the shelf'
  - Tantamount to 'jamming a square peg in a round hole'

- Adopting or retaining a less-than-optimal tool can:
  - Increases implementation issues (Taxman & Belenko, 2011),
  - Reduce accuracy of prediction, and
  - Reduce staff buy-in
  - Result = stuffed in a drawer after collection (see Vigalone et al., 2015)

CURRENT RISK ASSESSMENT USE IN PRETRIAL SETTINGS

Two broad types of assessments

- General population
  - Proposed to be used in any jurisdiction
  - Psychological traditions – like a mental disorder, recidivism and other outcomes are similar regardless of setting
  - Item sand weights often developed with a singular population/study and assumed to transfer when implemented elsewhere

- Specific population
  - Tailored and customized based on a specific population's needs and context

Advantages and disadvantages of tools

- General population tools
  - Easier, quicker, and often cheaper to implement
  - outcomes not specified to jurisdiction base rates and prevalence
  - Few pretrial tools exist and most, recently developed
  - Overall, less accurate

- Specific population (customized) tools
  - More labor intensive to develop and implement
  - Utilize localized data to improve accuracy and stakeholder buy-in
  - Ability to add relevant items over time

Existing Tools : What's out there?

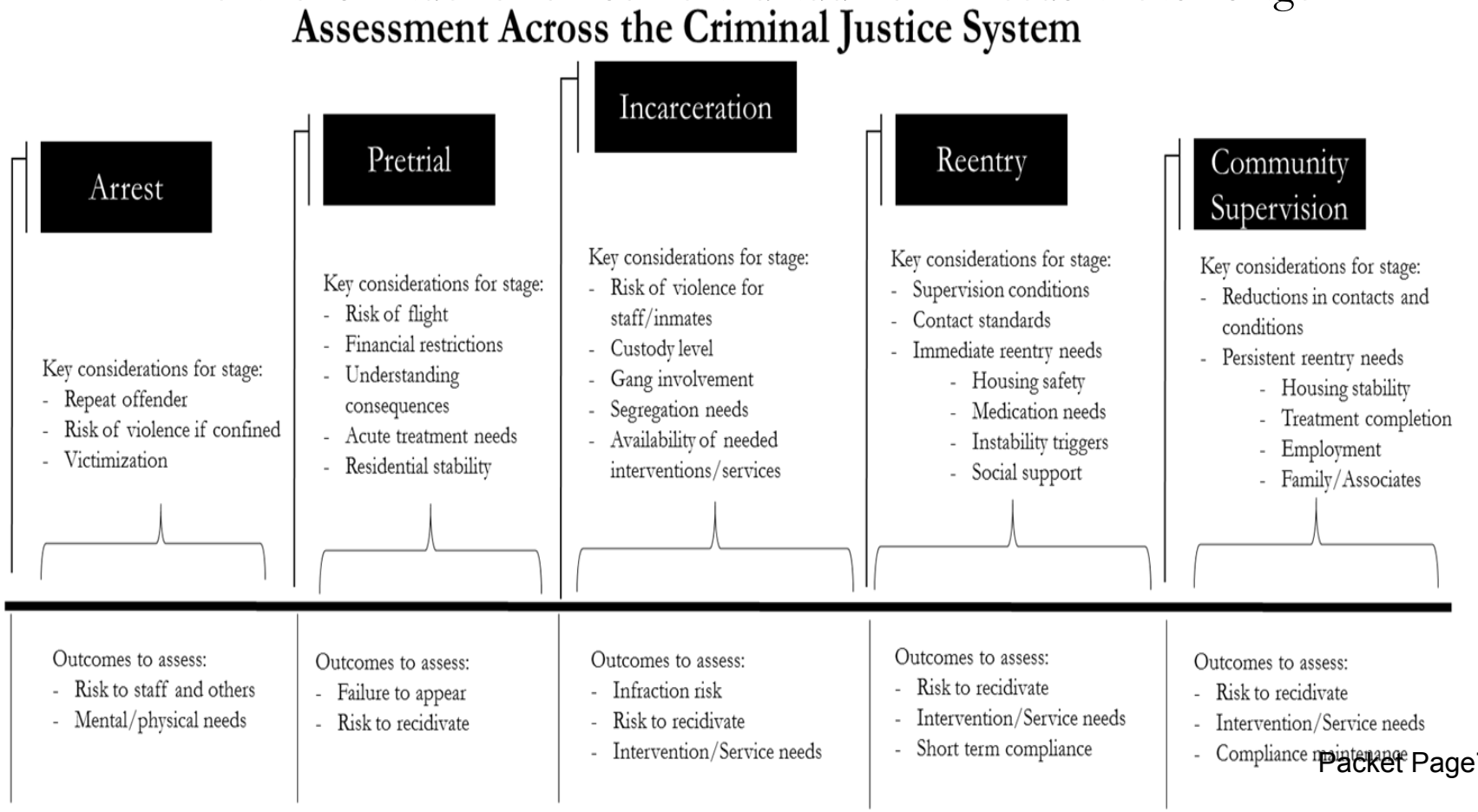
- General Population Tools
  - Level of Service/Case Management Inventory (LS/CMI)
  - Ohio Risk Assessment System (ORAS)
  - Federal Pretrial Risk Assessment Instrument (PTRA)
  - Public Safety Assessment (PSA) tool (LJAF, 2013)

- Specific Population Tools
  - Colorado Pretrial Assessment Tool (CPAT)
  - San Francisco Pretrial Diversion Project Pretrial Risk Assessment (SFPDP PTA)

ASSESSMENT FUNCTIONALITY

A Systems Approach to Assessment

- Items and outcomes are correlated
- General recidivism risk tool predict reasonably well for: violence, infractions, FTA, compliance
- General pool of items
- Alterations made for outcome and additional needs at each stage



HOW TO CREATE A CUSTOMIZED TOOL

- Any tool needs
  - Criminal history indictors collected from agency records
  - Outcomes – either general or specific
  - Interview items – less frequent but often provide defendant context, monitoring considerations, and programming needs

- Customize assessment for each jurisdiction
  - Develop a general pool of items
    - Option - If no interview items at pretrial, can develop a matched sample from a similar jurisdiction
      - Select subsample from King that 'look like' defendants in Thurston
  - Build model for jurisdiction
    - Select and weight items for the jurisdiction's population variations
    - Remove or adjust item definitions based on local statues and policy

Mechanics of a tool

- Developing a tool from scratch indicates a dedication to customization
  - It will serve as a foundational element for release, supervision, and programming decisions
  - Built on data and prior research
- Many steps for construction and validation
- Requires subject matter Experts (SMEs)

Item Pools

- Begins with an item pool or predictors (larger the pool = greater prediction)
  - Static, Dynamic, Risk, Need, and Protective
- Set of outcomes needing predicted
  - Most 'general' recidivism prediction tools – arrest, convictions, return to incarceration
  - Others provide outcome specificity – FTA, violent, property, drug, DV
- Selection of items that predict a given outcome
  - Give weight/values to those that improve prediction
  - Remove those that do not (noise)

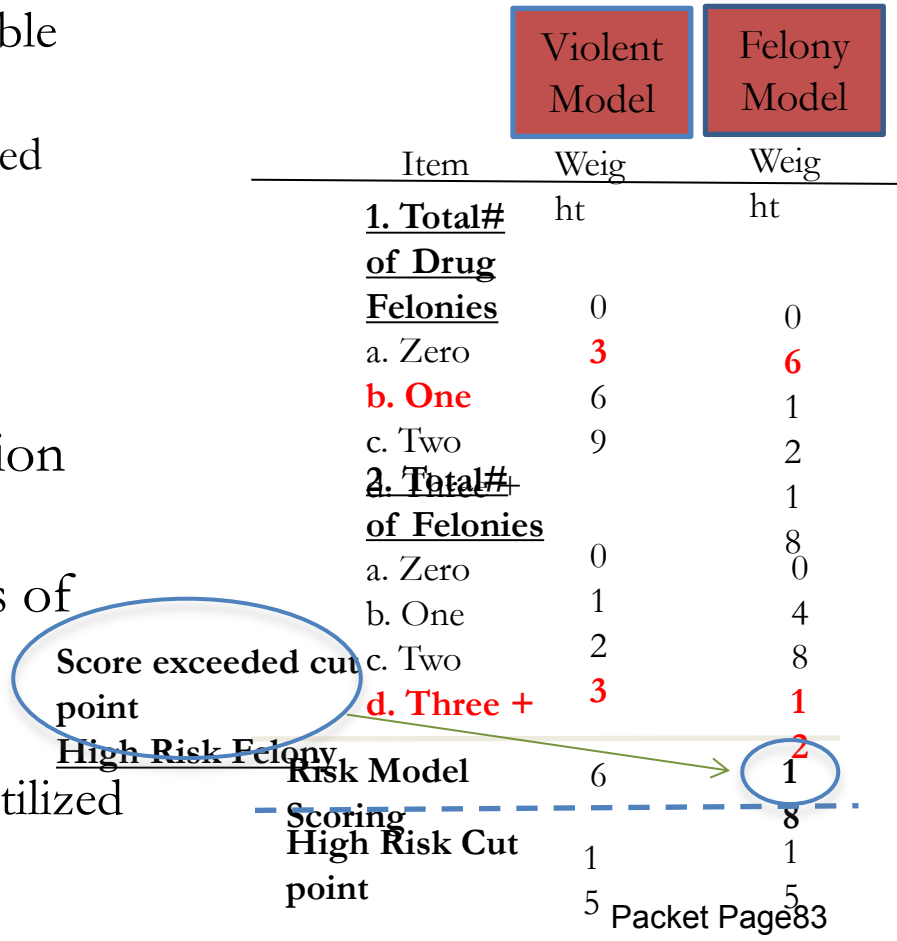
Risk Assessment Item Scoring

Risk assessments intended goals

- Predict recidivistic events
- Use ANY and ALL items that are feasible and ethically viable
  - i.e. static, dynamic, criminal history, or need

Items selected & weighted

- Vary based on outcome of interest
- Model distinctions improve prediction strength
- Cut points use to identify categories of risk
  - i.e. Low, Moderate, High
  - Risk categories are primary element utilized by users
    - Recommendations based on category
  - Can *customize cut points* to suit jurisdiction

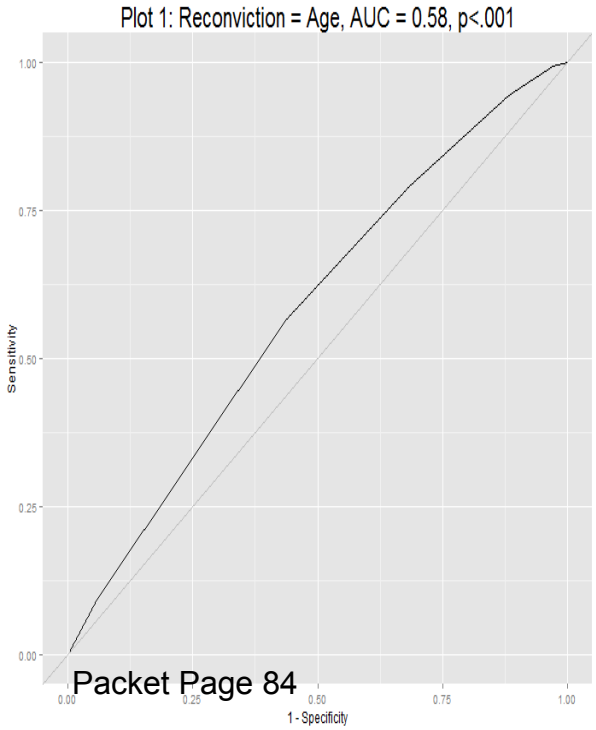


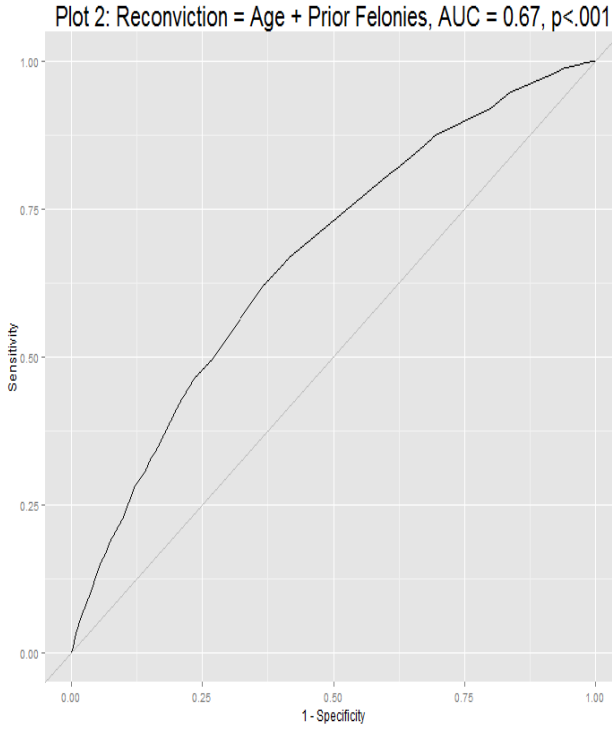
Validation and Validation Strength

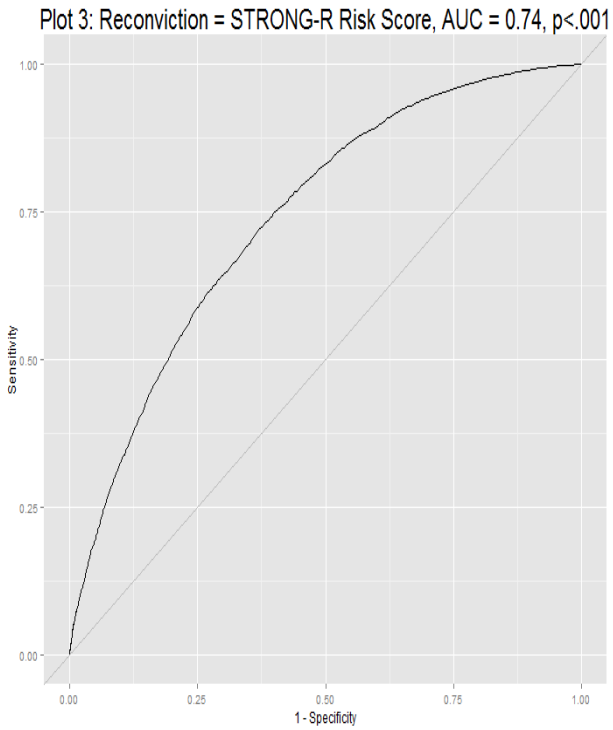
- Predictive validity is another term for instrument accuracy
- Validation reflects that the Area Under the Curve (AUC) is not .5
  - 0.5 – like flipping a coin or "betting the base rate" to predict recidivism
  - 1.0 – perfect prediction (Minority Report)
- Creating a "validated" is very easy criteria to achieve
  - Strength of prediction is more important

Table 2. Area Under the Curve Strength

AUC	Strength
Below 0.55	Negligible
0.56 to 0.63	Small
0.64 to 0.70	Moderate
Above 0.71	Strong







DEVELOPMENT OF SPOKANE AND KING COUNTY TOOLS

Spokane PTA

- Spokane's blueprint for change created a need for an assessment
  - Jail overcrowding
  - City and citizens did not want to build a new jail
  - Analysis reviled that municipal, district, and county were inconsistently placing defendants in jail
    - What is high risk for district court, Superior would give OR
    - Need for consistent assessment of risk across all jurisdictions
- Development of a customized tool and add pretrial staff

King PTA

- Multiple agencies requesting tools for variety of reasons
  - Prosecutors concerned with violent/DV recidivism
  - Defense wanting better assessment of treatment needs
  - Judges wanting a better use of alternative to detention programming

- Requested a universal Assessment for all court system
  - Dr. Barnoski began process in 2011-2012 as

Development of Spokane and King PTA

- Gathered large samples of pretrial defendants
  - Spokane N~14K
  - King N~10K

- Acquired AOC criminal history data
  - Priors - convictions, FTA, FTC, community supervision
  - Current charges
  - FTA, recidivism, days-to-outcome event

- Local pretrial data collection form both sites

Assessment Models created

- ASRA used in Spokane previously
  - not satisfied with results and reports
  - Make sure improvements over ASRA
- Created separate models for males and females
- Models predicting FTA and Recidivism
  - Low, moderate, and high risk with recommendations for OR, monitoring and 'set bail/detain'
  - Separate models for high risk of violent and non-violent risk (felony, property, drug)
  - Additional DV model for King
- Cut points

Results

- Models indicated predictive accuracy ranging from moderate-to-strong
  - Compared to other created PTA tools – San Francisco, Utah, Kentucky, Ohio,
  - FTA prediction greater than most tools created
  - All other outcomes exceed other pretrial tools
  - Exceeded ASRA on ever model, up to 36%, average of 17% increase in accuracy

Table 2. Area Under the Curve Strength

AUC	Strength
Below 0.55	Negligible
0.56 to 0.63	Small
0.64 to 0.70	Moderate
Above 0.71	Strong

	King		Spokane		ASRA (in Spokane)	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
FTA	**.71**	.70	**.72**	.70	.64	.62
Recidivism	.70	**.71**	.65	.65	.59	.56
Felony	**.76**	**.77**	.70	.68	.63	.59
Violent	**.73**	**.73**	**.71**	**.74**	.65	.66
Property	**.78**	**.76**	.69	.67	.68	.66
Drug	**.77**	**.79**	.70	.68	.66	.65
DV	**.77**	**.79**	--	--	--	--

Report Functionality

Assessment Information

Subject:	LAGOY, JONATHAN	**Assessment:**	SAFER		
Subject ID#:	333444	**Completed By:**	Aaron Stromberger	**Completed:**	5/24/2017
FBI ID#:		**Updated By:**	Aaron Stromberger	**Updated:**	5/24/2017
DOB:	1/27/1984				

Assessment Results and Data

Risk of Failure to Appear	Probabilty Range	Gen Pop
High	12.00%	8.00%

Risk to Recidivate	Probabilty Range	Gen Pop
High	37.38%	22.00%

Risk to Recidivate by Categories

Risk of Recididivate	Risk Categories	Probability Range	Gen Pop
High	Felony	35.37%	15.00%
High	Drug	23.74%	7.00%

Failure to Appear Risk Rating

Recidivism Risk Rating

Sex Offender Substance Abuse Mental Health Domestic Violence Drug Court Recommendation	- Substance Issues Indicated Mental Health Issues Indicated Domestic Violence Risk Indicated - Detain and Set Bail

Implementing a tool: what to consider

resources needed

- May options – from moped to Cadillac
  - Off the shelf tool will be quick to implement, paper and pencil (if needed), will devote little to quality assurance
  - Often have annual licensing fee and/or charge per assessment
- Customized tools have more up-front costs
  - North of $200k
  - Includes research-developed models, software, integration and training
  - Validate and update in 1-2 years
  - Annual Maintenance and quality assurance
- is a customized tool worth it?

Decrease jail, increase monitoring

- What are the advantages and disadvantages?
  - Increased liability
  - Decreased costs and decreased overcrowding
  - Research indicates greater positive outcomes from release than detainment

- Release and monitoring occurs now, nothing new
  - Courts other than superior monitored
  - Several thousand a year on OR

- System approach to assessment
  - Trend toward a uniform system
  - would have advantages for monitoring in Whatcom

Liability with community monitoring

- Court rule 3.2
  - Release OR if not at risk for violent crime -> assessment provide a % likelihood to guide decision
  - Likely to FTA – least restrictive conditions imposed
    - May consider criminal record, nature of charge, danger to the witness/victim, use of weapons, mental health, substance abuse issues, community members opinions/vetting
    - Restrict travel, periodic return to custody, EHM, other relevant release conditions

Liability with community monitoring, cont.

- VERA's summary of research
  - defendants who secure release in 1 day, defendants who spend time in jail before pretrial release are more likely to commit new offenses
  - Detaining low & moderate risk defendants, even just 2-3 days, is correlated with higher rates of new criminal activity
  - no proven relationship between a particular charge & risk of flight or new offenses
  - posting bond, there is generally no capacity for supervision to minimize risk
  - There is no duty when a crime is not foreseeable

Who should be involved in the process?

- When developing a tool a team is needed
  - The researcher will be constructing the tool but the jurisdiction needs to be involved
  - SME team that will inform researcher of local concerns, policies, justice system processes
  - Described as a 'cross-sectional' or 'diagonal slice' of experts
    - Line staff, users, middle management, judges, prosecutors, defense and a coordinator
    - Needed to help select pool of items, outcomes, and definitions (i.e. charges vs. convictions)
    - Vet preliminary models created, suggest modifications
    - Direct the placement of cut points and recommendations for risk categories
- Continued effort
  - Team may change members over time but needs to have routine meetings
  - After creation, need to monitor bugs and fixes

Anticipated impact of tool

- Pretrial tools still relatively new
  - Few studies isolating their impact alone
  - Assessments generally, Evidence-based practice
    - "risk-needs assessment tools are valued as essential component for directing offender risk classification, increasing procedure justice, and helping resource allocation" (Taxman & Dezember, 2017)
- Jail
  - Setting cut points to account for proportion reduction will have a direct impact
    - Vera on Whatcom – 6% traffic offenses and 4% misdemeanors, - > 10% reduction seems feasible
    - Paired with court notification system, increased monitoring, and diversion options could greatly reduce use of jail pretrial
  - DC's implementation, increased monitoring and

Anticipated impact of tool, cont.

- Recidivism
  - Mesa County, CO found 88% of low risk remain crime free at pretrial
    - 93% make all their court appearances
    - For high risk 75% crime free and 87% make all court appearances
  - Vera
    - Moderate 40% reduction in FTA, high risk 33% in FTA
  - Correctional tools demonstrate reductions
    - See - Andrews, & Dowden, 2006;2005, Harris, Gingerich, & Whitaker, 2004, Luong & Wormith, 2011

Data needed to build a tool

- Customized tools require data to start, off-the-shelf tools do not
- Minimum requirements
  - Criminal history information
    - Used for priors and outcomes
    - Obtained from the AOC
  - Sample
    - Approximately 2k or more of whatcom defendants
    - All need released to the community while on pretrial
      - Or, monitoring
      - Bail/bonded out
- Increased prediction
  - Set of dynamic and locally relevant interview items
    - Married, local address, length of employment,

What are the options for customization?

1. Static only
   - Criminal history only
   - Tool similar to the Arnold Foundation tool used in Yakima and ASRA
   - Will be modestly effective (AUC in the .5s to .6s)

2. Pilot interview (to pair with static)
   - Collect data to add to the interview (similar to King)
   - Will take 6-12 months of collection, depending on Whatcom case load

3. Defendant paper files translated to data
   - Defense client interview one-sheet
     - Contact method, residency, relatives, married, children, education level, length of employment, bail post-able
   - Manual transfer to analyzable data

Conclusion

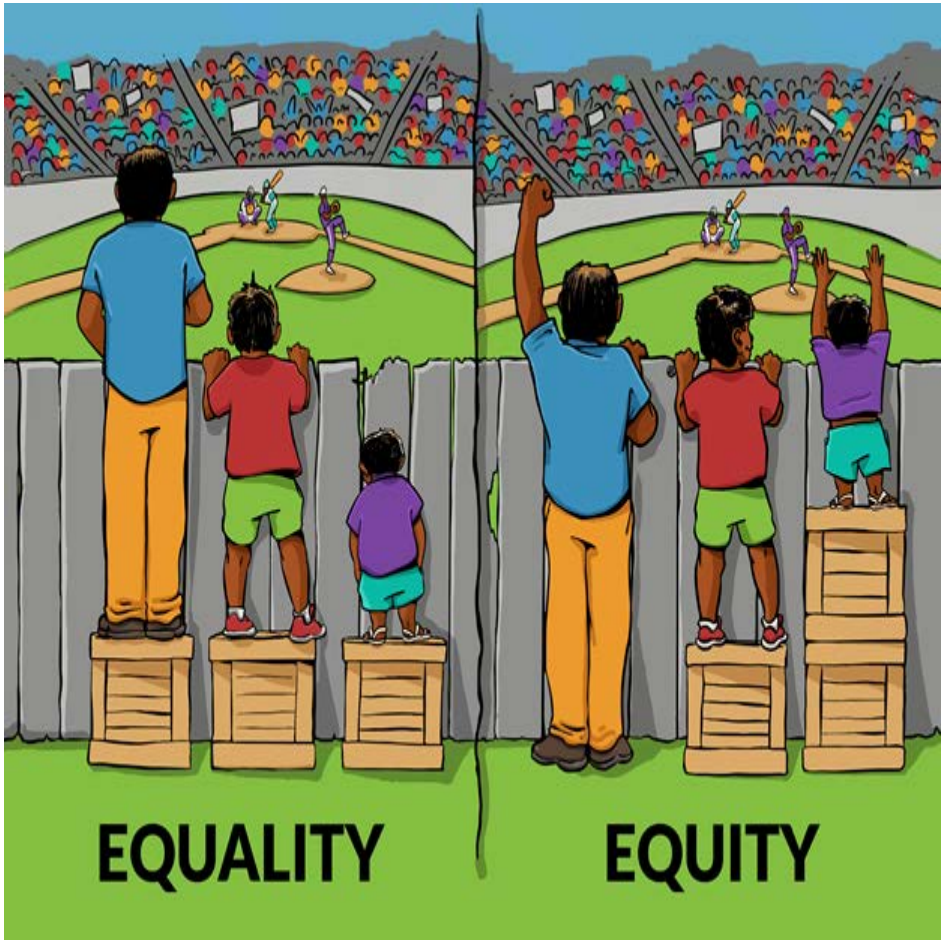
- There is a lot that goes into risk assessment
  - Terms, definitions, scoring, cut points, validation
  - Difficult landscape to navigate for experienced tem of SMEs
- Contemporary tools
  - Try to advance prediction with, gender specificity, outcome specification, improved recommendations and case management
  - Options to buy off the shelf or customize – advantages and disadvantages to each
    - Have to have buy-in or likely fail shortly after implementation
  - Multiple options for creating a customized tools (4 presented)
- While relatively recent, compared to correction

End

- Contact information
  - Zachary Hamilton, Ph.D.
  - Washington State University
  - Associate Professor - Department of criminal justice and criminology
  - Director – Washington State Institute for Criminal Justice
  - Email – Zachary.Hamilton@wsu.edu

DISCUSSION OF POTENTIAL ISSUES OF BIAS AND DMC

- Concern of 'perpetuating' racial/ethnic bias through the use of risk assessment
  - Controversial and difficult subject to obtain consensus
    - Risk assessment researchers developing tools that are absent race and attempting to ameliorate idiosyncratic release decisions (make as fair through statists)
    - On the other hand, justice system has inherent issues



Propublica articles

- First article specified 'potential' use of assessment for sentencing
  - Assessments are not to be used as a decision maker at sentencing, judge discretion
  - Risk is often computed using prior criminal events, which are correlated with SES, which is correlated with minority status
    - An assessment tool is designed to remove/identify idiosyncratic decisions
    - Cannot fix our country's enforcement and sentencing practices that revolve around DMC

- Second article, made substantial retractions
  - Statisticians 're-analyzing' Propublica findings

An evolving discussion

- Risk assessments have been shown to reduce recidivism and are an evidence-based practice
- However, there is potential for unintended use
  - Designed to be used for release decisions – pretrial, diversion, probation, and parole
    - Utility in identifying the appropriate 'level' of supervision and programming needs to improve public safety and reduce future criminal justice involvement
  - Should be avoided in use for sentencing decisions

- Tools can be improved with research

Conclusion and Future Directions

- Guideline not a requirement
  - Release decisions, not a sentencing tool
- Feedback loop of data
  - Customization allows you to validate locally
  - Expand assurances that the tool is consistently refined to assist the court room work group
- Need to have stakeholder ownership to develop system improvements
  - Customizable elements and implementation assistance is key for success
  - Training and QA are vey important

- Examine methods of mitigating DMC