



Measuring Up:

Performance monitoring and statistical analysis in youth justice reform

Over the last 20 years, youth justice systems increasingly used data to inform decision-making and figure out what interventions work with the ultimate goal of reducing the number of youth held in secure facilities. Agencies adopted performance monitoring systems with descriptive data to regularly, quickly, and easily collect, track, and analyze data. At the same time, data is also used for statistical analyses to determine evidence-based programming or the validation of detention screening tools to make informed release decisions. This factsheet describes different kinds of uses of data in the effort to measure and advance progress in youth justice reform.

Using Data for Performance Monitoring and Measurement

What are performance measurement and monitoring?

- Performance measures are established to determine which data elements to collect and analyze. Performance measures might include both outcomes and outputs, or the number of times an activity happens. For example, the number of admissions to detention might be a performance measure.
- Performance monitoring is the continual and repeated analysis of performance measures. For example, regular review of an analysis of the number of referrals is performance monitoring.
- Together, performance measures and monitoring are meant to document how or if decisions or services are happening and measure outcomes that might be related to those decisions or services.¹
- A system of performance monitoring, which includes established performance measures, can be described as an “ongoing process and

¹ Bazemore, G. (2006). Performance measures: Measuring what really matters in juvenile justice. Alexandria, VA: American Prosecutors Research Institute.

outcome evaluation”² that help a system keep track of progress toward a goal.

How is performance monitoring and measurement used by youth justice systems?

- Regularly collecting and monitoring system data can help an agency or system ensure that it is operating in line with established policy, practice, and process. A system seeking to reduce racial and ethnic disparities in detention would regularly analyze admissions to data broken down by racial and ethnic groups.
- Breaking down the data into small, measurable bites at specific decision-points allows stakeholders to identify problems or determine where a different decision could be made to reach the overall goal of the system or agency.
- Reviewing data regularly improves the quality of data for more robust monitoring or for future evaluation for evidence-based practice or intervention or validation of a detention screening tool.

What questions can be asked about performance measuring and monitoring?

- **Goal:** What is the goal to be achieved? Do the performance measures, including both outputs and outcomes, relate to the outcome? Can decisions related to those outputs and outcomes be reviewed and changed?
- **Measurements:** Are both counts and percentages included? What is the denominator when calculating percentages? How many cases are included? Are averages and medians included?
- **Descriptive statistics:** Are cases broken down by youth characteristics across all measures?
- **Replicability:** How often can this analysis be completed? Can it be repeated easily? Can it be reviewed frequently?
- **Current relevance:** What is the time period from which these cases are drawn? How old are the included cases?

² Mears, D.P. and Butts, J.A. (2008). “Using Performance Monitoring to Improve the Accountability, Operations, and Effectiveness of Juvenile Justice.” *Criminal Justice Policy Review* 19(3): 264-284. <https://diginole.lib.fsu.edu/islandora/object/fsu%3A640602/datastream/PDF/view>

Using Data for Evaluation or Statistical Analysis

What is an evaluation or statistical analysis?

- Typically, an evaluation or statistical analysis requires a research design and statistical methods to determine whether certain activities would create or contribute to a particular outcome. An example might be whether or not mentoring reduces arrests for youth involved in the youth justice system.
- With statistical tools, an evaluation takes into account an array of factors, including characteristics of the youth involved, system features, or attributes of the program.
- An evaluation may require a research design that considers the outcomes of youth involved in a particular intervention compared to youth who are not using that intervention.

What are the benefits of an evaluation or statistical analysis?

- Evaluations can help systems determine what works or does not work to improve outcomes for youth, which in addition to helping young people stay out of trouble, can also help systems direct resources efficiently. For example, evaluations of outcomes related to secure confinement, including detention have shown that youth who are detained are more likely to commit another offense, have lower educational attainment, and earn less over a lifetime.³
- Evidence-based programming, such as Multisystemic Therapy (MST), which was determined to be effective at reducing reoffending for youth who participate, has been an important program for helping youth justice systems reduce the number of youth held in secure facilities.⁴

What questions can be asked to assess a statistical analysis?

- **Research question:** What was the research question? What is the purpose of the analysis?

³ Holman, B. and Ziedenberg, J. (2006), Dangers of Detention: The Impact of Incarcerating Youth in Detention and Other Secure Facilities. Washington, DC: Justice Policy Institute. www.justicepolicy.org/images/upload/06-11_rep_dangersofdetention_ji.pdf ; Petteruti, A., Schindler, M. and Ziedenberg, J. (2014), Sticker Shock: Calculating the Full Price Tag for Youth Incarceration. Washington, DC: Justice Policy Institute.

http://www.justicepolicy.org/uploads/justicepolicy/documents/sticker_shock_final_v2.pdf

⁴ Blueprints for Health Youth Development, "Multisystemic Therapy (MST)", April 7, 2021. www.blueprintsprograms.org/programs/32999999/multisystemic-therapy-mst/

- **Sample size and make-up:** How many cases were included? Were they representative of the population? Were there enough cases spread across interventions and outcomes?
- **Research design:** Is there a control group? Was it a mathematical comparison or an experiment?
- **Replicability:** How often can this analysis be completed? Can it be repeated easily?
- **Descriptive statistics:** How are all the characteristics of the participants spread across the interventions and the outcomes?
- **Statistical methods:** What statistical tests were used? Are they appropriate to the analysis?
- **Current relevance:** When did these interventions take place? How old are the included cases?
- **Research information:** Who conducted the research?

What is a validation of a screening tool?

- While a validation of a screening tool is not the same as an evaluation for effectiveness, validation requires following a process to confirm that the tool is accurately predicting certain outcomes. Often, validation is checking to see if a person released per screening tool recommendation commits a new offense or fails to appear in court.
- Validation requires a large sample of cases to complete the validation. For example, the Juvenile Detention Alternatives Initiative recommends a sample size of at least 300 cases, in addition to an array of factors about the youth and their cases, to conduct a validation of a screening instrument.
- Similar to an evaluation, external parties are typically involved in the final analysis and recommendations stemming from a validation of a detention screening tool.⁵

⁵ Steinhart, D. (2006), Juvenile Detention Risk Assessment: A Practice Guide to Juvenile Detention Reform. Baltimore, MD: Annie E. Casey Foundation. www.aecf.org/m/resourceimg/aecf-juviledetentionriskassessment1-2006.pdf

Comparing Performance Measurement and Statistical Analysis

How is performance measurement different than evaluation or other statistical analysis?⁶

- **Performance measurement is ongoing and repeated at short intervals while a statistical analysis is discrete and can take some time to complete.** Performance measurement is done as close to real time as possible to give stakeholders immediate feedback on the program so that changes can be made to meet the desired goals. An evaluation is usually done after some interval has passed to determine how outcomes might relate to the intervention.
- **Performance measurement is responsive and adaptive while statistical analysis answers a predetermined set of research questions.** Performance measurement and monitoring provides information about program implementation. Performance measurement usually raises more questions and can change as the program changes. Statistical analysis or evaluation answers questions that are determined before the analysis begins, with very little change during the analysis itself.
- **Performance measurement typically uses output and outcome data that can be collected quickly, while a statistical analysis usually involves an array of data collection and research methods that can take time.** Because performance measurement happens routinely and somewhat flexibly, the data that it uses must also be collected quickly and easily enough to support the rapid and responsive nature of the process. Assessment tools, use of services, admission data that can be found in a spreadsheet are the kinds of data that can be used in performance measurement. A statistical analysis may require data that requires hand tabulation or comes from a variety of sources. A full evaluation may also include qualitative methods such as interviews.
- **Performance measurement does not require many cases or instances, but a statistical analysis usually requires a certain number of cases for statistical significance to be included.** More cases can make for a more informative performance monitoring process, but even a few cases or data points make it possible to see

⁶Tatian, P.A. (2016), Performance Measurement to Evaluation. Washington, DC: The Urban Institute. www.urban.org/sites/default/files/publication/78571/2000555-performance-measurement-to-evaluation-march-2016-update_1.pdf

how a program might be functioning in relation to the program goals. A robust statistical analysis requires many cases to conduct tests for significance or draw conclusions and that could mean that the statistical analysis includes cases from a long period of time.

- **Performance measurement and monitoring does not require training in advanced research methods or design, in contrast a statistical analysis or evaluation requires specialized training or experience.** Performance measurement can be carried out by dedicated professionals who understand the program, the purpose and meaning of the data, and who can analyze the data. Statistical analysis or evaluation generally requires an understanding of research design and statistical methods in order to ensure a robust analysis.
- **Performance measurement is usually done by internal staff, while an evaluation or statistical analysis to show effectiveness or causation is typically conducted by external parties.** Internal staff are better suited to quickly and nimbly collect, analyze, report, and interpret performance data. They know the program and can be responsive if changes are needed. An outside evaluator is considered to be more objective with no connection to the outcome of the evaluation or statistical analysis.

Performance monitoring and evaluation: A comparison

	Performance Monitoring	Evaluation
Frequency	Ongoing	Discrete
Duration	Updated repeatedly over short time frames	Usually takes a long time to complete
Training Required	Basic Statistical Principles and Data Visualization	Research Design and Statistical Methods
Sample Sizes	Any size; no representative sample needed	Usually large/several hundred with a representative sample
Who conducts	Agency staff or stakeholders	External evaluators

Summary

Performance measurement/monitoring and determinations of statistical significance go hand-in-hand. Performance measurement is the first step to

ensuring a robust system of operations or interventions and data collection that could be ready for a statistical analysis or evaluation. In other words, performance monitoring supports consistency and quality data that is crucial to determining what works and what doesn't.

Performance measurement is an agile process that involves many stakeholders in an agency and organization. Ideally, performance measurement empowers program staff and stakeholders as active participants in not only collecting and analyzing data, but to also encourage innovative thinking to solve challenges that arise through the performance monitoring process, build momentum for change, and to regularly show how their hard work is paying off.

5 Tips for Preventing Statistical Analysis or Evaluation from Slowing Reform

Statistical analysis and evaluation can be a powerful tool for showing what's working and what isn't, in addition to potentially contributing valuable research to the field. However, statistical analysis can be intimidating, slow momentum, and undermine the confidence of program staff. Here are five tips for preventing statistical analysis or evaluation from impeding reform.

Statistical significance (or not) doesn't prove anything: Statistical analyses that include findings of significance or no significance do not necessarily mean that there is no effect between two variables. In an analysis of 791 articles across five journals, researchers found that slightly more than half mistakenly assumed that non-significance means no effect.⁷ For example, a statistical comparison of the race of a young person with detention admissions might show no significance, but that doesn't mean there is no difference between the treatment of white youth and youth of color.

Beware of making important decisions based on a single statistical analysis: More than 800 statisticians, clinical and medical researchers, biologists and psychologists signed onto an article in *Nature* magazine called for caution in using tests of significance to make important decisions.⁸ Statistical significance should not be applied in an all-or-nothing manner, but rather in a more nuanced approach that includes other evidence and methods.

Stop. Collaborate. Listen: Working closely with the evaluator or analyst to make sure recommendations or findings are feasible to implement and won't cause other harm, such as undermining credibility with the community or other stakeholders or demoralizing program staff that will still be responsible for implementing the program or reform.⁹

Evaluations aren't the only solution: Budgetary constraints, small numbers of participants, or unique needs of the community make it difficult to do evaluate a program to show that it works. Instead, using a robust performance monitoring plan to compare practices and indicators against evidence-based principles could be a compromise. For example, researchers at Vanderbilt University developed the Standardized Program Evaluation Protocol (SPEP) to help agencies develop effective, innovative programs, that work for their communities.¹⁰

Partner-up to get past the "know-do" gap: Knowing what works and doing it consistently and sustainably are often two different things. The expanding field of implementation science is an approach that gathers groups of internal and external partners, including researchers, evaluators, and data partners, to set goals, deliver training, offer coaching and support, gather data, and regularly monitor adherence to evidence-based practice.¹¹

⁷ Amrhein, V., Trafimow, D. and Greenland, S., "Inferential Statistics as Descriptive Statistics: There Is No Replication Crisis if We Don't Expect Replication," *The American Statistician*, (73), 2019. <https://www.tandfonline.com/doi/full/10.1080/00031305.2018.1543137>

⁸ Amrhein, V., Greenland, S. and McShane, B., "Scientists rise up against statistical significance," *Nature*, March 20, 2019. <https://www.nature.com/articles/d41586-019-00857-9#ref-CR3>

⁹ Callis, A. "5 Rules When Conducting Program Evaluation: First do no harm," February 20, 2019. <https://www.linkedin.com/pulse/5-rules-when-conducting-program-evaluation-first-do-harm-callis-mph#:~:text=While%20it%20is%20critically%20important,they%20can%20harm%20the%20program.>

¹⁰ Peabody Research Institute, "Standardized Program Evaluation Protocol: SPEP Information", June 15, 2021. <https://my.vanderbilt.edu/spep/spep-information/>

¹¹ Lily Gleicher, "Implementation Science in Criminal Justice: how Implementation of Evidence-based Programs and Practices Affects Outcomes," Illinois Criminal Justice Information Authority, June 15, 2021. <https://icjia.illinois.gov/researchhub/articles/implementation-science-in-criminal-justice-how-implementation-of-evidence-based-programs-and-practices-affects-outcomes#strategic-planning>