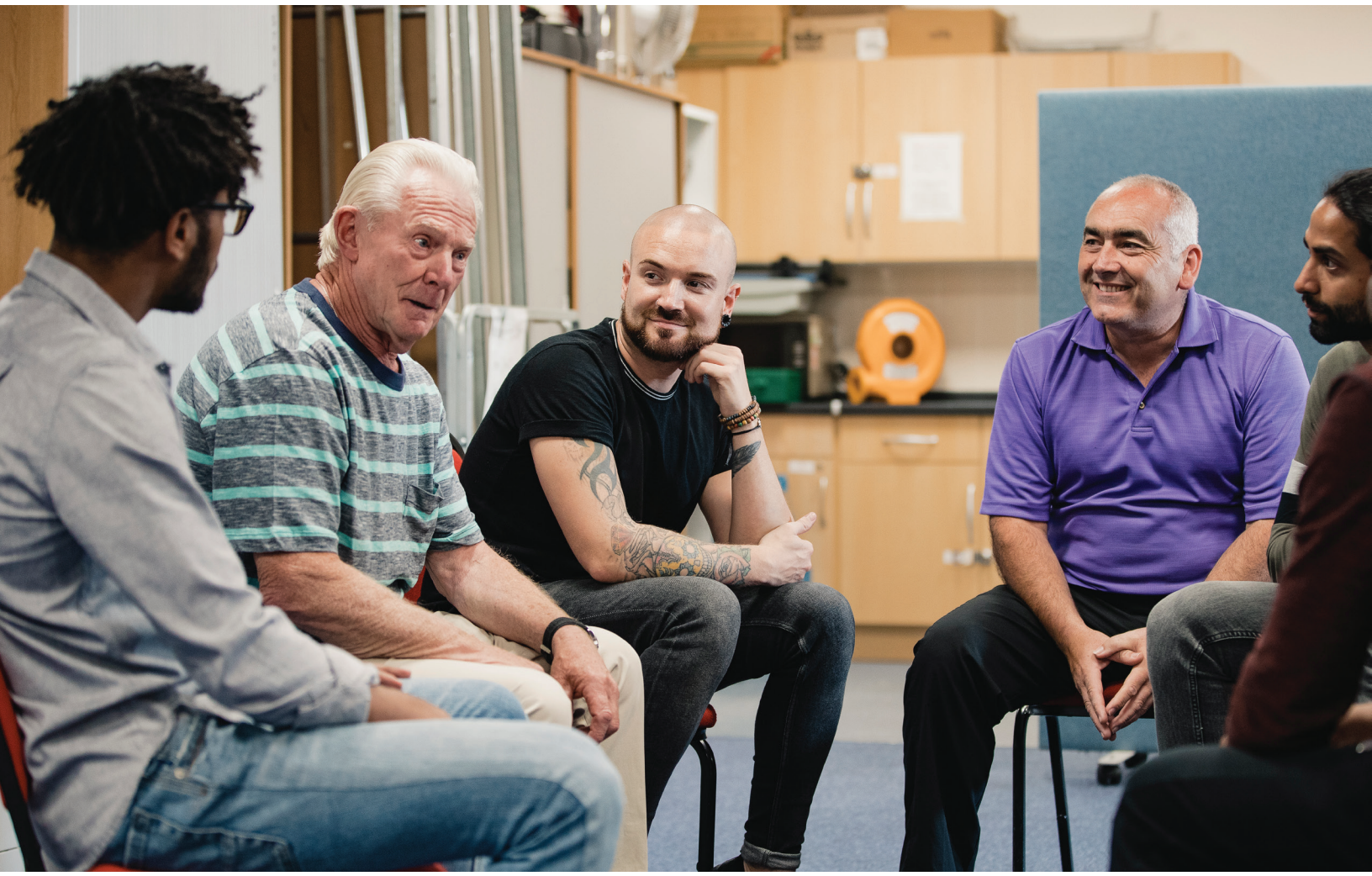
Jail Diversion Program Evaluation Resource Guide



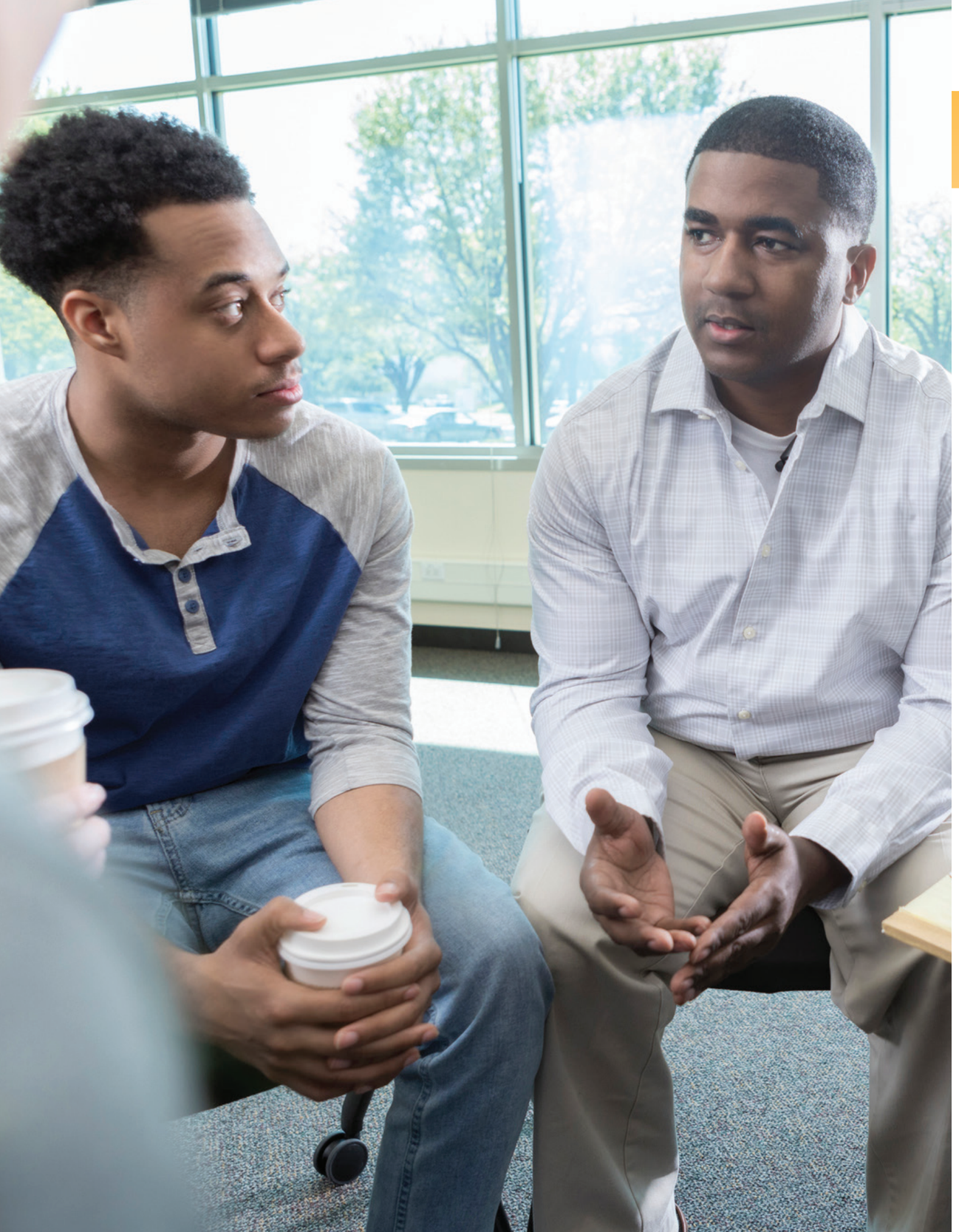
This Resource Guide is informed by the ASPE Research Project, State Interventions for Diverting Individuals with Serious Mental Illness and Co-occurring Disorders from the Justice System conducted by Abt Associates. This qualitative research study focused on the current field of jail diversion programs on the zero intercept of the Sequential Intercept Model. Findings from the study revealed a need for additional support for programs to pursue evaluation.

How to Use This Resource Guide

This Resource Guide is designed to provide jail diversion program staff and leadership with accessible tools to assist in the design and implementation of program evaluation in a way that best suits your needs and budget. Each section of the Resource Guide contains examples based on evaluations of the Law Enforcement Assisted Diversion (LEAD) program, as well as links to valuable resources for each topic. The Resource Guide contains the following sections:

1. Why Evaluate Your Program?
2. Planning Your Evaluation: Are You Ready?
3. Resources Needed for Evaluation
4. Identifying Your Goals for the Evaluation
5. Evaluation Designs
6. Sampling
7. Data Elements
8. Data Collection
9. Data Analysis
10. Conclusion

You can skip to sections of interest within the document or you can work through the entire Resource Guide section by section. The appendices include a glossary of useful evaluation-related terms along with additional evaluation resources, such as the Centers for Disease Control and Prevention's Framework for Program Evaluation which provides detailed steps and standards for evaluation.
This guide does not provide enough detail to serve as the sole source for completing an evaluation. However, reviewing it will help you and your leadership/staff understand the basic aspects and components of evaluation to prepare you for discussions with funders and evaluation partners about your program's evaluation readiness and needs.



1

Why Evaluate Your Jail Diversion Program?

Collecting data for evaluation can seem overwhelming or feel like unnecessary busy-work. But data collection and analysis are essential to know:

- If your program operations are running effectively and efficiently;
- Where resources are best allocated;
- Whether or not your program is reaching the goals and outcomes you expected; and
- Whether or not adjustments need to be made to the program approaches and/or implementation.

Evaluation data and findings can also be used to gain support for ongoing or expanded program funding, and can justify replication of your jail diversion program in other locales. For example, the evaluation findings from the LEAD program evaluation provided evidence of effectiveness that was sufficient to warrant continued funding and program replication. The Federal Bureau of Justice Statistics (BJS) provides answers to commonly asked questions about why you should evaluate your program (see Resources page for the BJS Guide to Program Evaluation).

2 Planning Your Evaluation: Are You Ready?

If you are just starting to explore implementing a program, include evaluation planning now. Your evaluation will be stronger if it's planned concurrently with program planning. In addition, it is important to ask yourself, with whom do you need to partner to achieve success? Program implementation and evaluation are not solitary endeavors. Planning for partnerships and collaboration with external evaluators and across behavioral health, justice, and other systems is critically important for both implementing and evaluating your program. Moreover, partnering with an external evaluator from the start can help ensure that you have a strong and objective, or unbiased evaluation.

As you proceed with implementation of your evaluation plan, it is important to document the challenges you are facing and how you might address each challenge. Decisions regarding evaluations always require trade-offs. The resources needed for evaluation may not be available during the planning stage. But making basic data collection and analysis routine can provide the data you need for an eventual evaluation. One place to start is taking stock of what data your program already collects. Such data might include:

- Diversion encounters routinely logged by type
- Time spent per encounter
- Number of 911 calls related to behavioral health
- Arrests of individuals with behavioral health diagnoses
- Re-arrests by arrestee
- Emergency department drop offs (for police departments and sheriff's offices
- Crisis assessments with location (for mobile crisis teams)
- Crisis center admissions
- Admissions to residential diversion services
- Lengths of stay in residential diversion services
- Fatal and non-fatal overdoses
- Naloxone use by first responders

HOW TO PLAN FOR SUCCESS

- Identify what your program is attempting to accomplish. What problem are you trying to solve?
- Identify important partners who can assist you in addressing the problem.
- Decide upon the intervention(s) will be used to address the problem.
- Identify your theory about how the intervention will address the problem.
- Identify the questions you are trying to answer about the changes you expect to see.
- Map out a logic model with input from stakeholders.
- Decide on the research design that is feasible and will best provide the answers to your research questions.
- Develop and implement your evaluation plan.
- Use the evaluation findings for program improvement, replication if successful, and to identify additional resources needed.

3 Resources Needed for Evaluation

Making the case for funding

Funding your evaluation can be complicated. The National Institute of Justice and the Substance Abuse and Mental Health Services Administration have funded jail diversion demonstration grants that include funds for evaluation. There are a number of funding sources through government agencies and foundations which can be used to support your evaluation. Many state focused foundations may also be interested in funding jail diversion program evaluation.

Your chances of obtaining funding are strengthened by a strong study design and strong stakeholder engagement. Obtaining sustainability funding for a program is often linked to demonstrating positive outcomes, which requires a strong evaluation. Investing in your evaluation from the start will pay off in future program stability. Additional resources to help with evaluation resources can be found in the resources section of this guide.

Partnering with other organizations

By its nature, jail diversion involves multiple service systems, so partnerships are important for both program implementation and evaluation. Your organization may not have all the resources and expertise to conduct research on your own. Universities, medical centers, independent research firms, some advocacy groups, public health organizations, state, or local governments may have some of the resources you need to fund implementation and evaluation of your program. Such organizations can also assist with any requirements for human subjects research, such as working with Institutional Review Boards (IRBs) and how to obtain Federalwide Assurance to conduct human subjects research. Partnering with other organizations can also assist staff in your organization to gain needed research and data collection skills. Partners that have experience in evaluation can assist you in identifying the evaluation design that would work best for your program, taking into account the program you are implementing, your budget, and other factors.

- Universities
- Independent research firms
- Advocacy groups that do evaluation
- Think tanks
- States

Once you have examined the data you already collect, think about what data would be useful to you in the future. If you collect and analyze data on a regular basis to inform programmatic decisions, can those be used for your evaluation? Ask staff and leadership what information would be useful to them. Partnering with other organizations in your community, including universities or public health organizations can help you identify what other data might already be available. Partners can also help identify how best to use your data.

4

Identifying Your Goals for the Evaluation

Research Questions

The first step in planning your evaluation is to identify the questions that you want to answer through the evaluation of your program. The question(s) should be testable –something you can measure – such as an outcome you are trying to achieve. When thinking about outcomes, focus on what your program is trying to change, or accomplish, for whom and how. For example, are you trying to reduce the number of repeat arrests of individuals with serious mental illness for nuisance crimes?

As you will see in the design section, the questions you develop will have an impact on the design of your study. "How" and "why" questions are more suitable for process evaluations and implementation studies. For example, "How do warm handoffs between police and mental health staff take place?" or "Why do some officers use a crisis stabilization center while others do not?" An outcome question might be "Does the crisis diversion center reduce police reliance on emergency departments for individuals in need of mental health or substance use services?"

A good research question takes time to develop. It should be fairly narrow (but not too narrow) and measurable. For example, while you likely want to ask, "Is this program effective?", you need to define what you mean by "effective." A better question might be, "What impact does this program have on reducing arrests among individuals experiencing a mental health crisis?"

A few resources for narrowing down your research questions are provided in the resources section. Once you have identified the "what" and "for whom" of your research questions, the next step is to determine the "how": what system components (such as mental health services, policing, current diversion practices) are you trying to implement to produce changes in your outcomes? How are you trying to implement those components? What resources and training are needed to implement those components? How will you know if you are successful in the short term and in the long term?

Logic Models

A useful tool for thinking through how you will achieve and measure your desired outcomes is a logic model. Logic models are like road maps that can guide you where you want to go. They are the basis for every good evaluation. A logic model is a graphic representation or picture of what resources (the "inputs") and activities are included in your program, what you expect your program to produce (the "outputs"), and short- and long-term "outcomes." One way to think through the logic model for your program is to base the inputs, activities, outputs, and outcomes on your theory of change. A theory of change may sound abstract, but essentially it captures the outcomes you are trying to accomplish and the ways in which your program is designed to influence those outcomes. For example, the LEAD program is based on the assumptions (i.e., theories) that a sufficiently funded public health approach is the best way to address drug crime, and that improving health and social conditions through service delivery will lead to reductions in criminal behavior and recidivism, improve public safety, and reduce costs to the criminal justice system.

Resources to help you develop your theory of change are located in the resources section.

Most logic models are represented as a series of boxes with arrows that show a sequence of steps taken toward achievement of program outcomes. However, the steps through a logic model can be shown in many different ways. They can be simple (as shown below in the LEAD example) or more detailed, naming evaluation questions, partners, and assessment tools. The level of detail you include will be determined by the information you need to guide your program evaluation.

Logic Model of LEAD Program's Evaluation*

Input	Activities	Outputs	Short Term Outcomes	Long Term Outcomes
Community Advisory Board	Referrals from Law Enforcement	Indivdualized Case Plans Developed for all Participants	Decreased Non-drug Related Criminal Activity	Decreased Recidivism Rates
Funding	Intensive Case Management	Rapid Linkages to Legal Services, Employment Housing, and Transportation	Decreased Harm from Drug and Alcohol Use	Decreased Demand for Social Services
Contracted Treatment Provider	Individual Intervention Plans	Case Management Meetings	Savings from Diversion	Improved Relationships Between Police and those Policed
(Case managers Peer Support Staff)	Peer Outreach and Counseling	Peer Counseling for all Participants	Participant Increased Psychosocial Functioning	Business and other Leaders increased Satisfaction with Public Safety
	Comprehensive Treatment and Social Services		Improved Participant Health	Increased Efficiency in Use of Public Safety Resources
	Leadership Development Training		Increase Participant Hope for the Future	
	Legal Advocacy			
	Harm Reduction			
	Engagement of Neighborhood Public Safety Leaders			

*This logic model was developed based on resources related to the LEAD program and its evaluation.

LEAD RESEARCH QUESTIONS

1. What is the impact of the LEAD program on reducing criminal recidivism (i.e., arrests and charges) compared to the 'system-as-usual' comparison condition?
2. What is the impact of the LEAD program on reducing publicly funded legal and criminal justice service utilization and associated costs (i.e., prosecution, public defense, jail, prison) compared to the 'system-as-usual' comparison condition?
3. What is the impact of the LEAD program on psychosocial, educational, employment, and housing variables?

There are a number of resources that can guide you and your team in developing a useful logic model. Many of these resources include advice on engagement of key stakeholders and coming to consensus on the components of your logic model.

Time and resources will be needed to develop partnerships, develop your theory of change, plan program, and develop your logic model. However of these tools is useful for gaining agreement on:

- What outcome(s) you are trying to achieve;
- What needs to happen to achieve the outcome(s); and
- What contextual factors can help you achieve your outcomes and what factors may get in the way.

The time you spend up front will ensure that your program and evaluation are well implemented and that you will be able to accurately measure your outcomes of interest.

5 Evaluation Designs

Once you have established the logic model for your program, the next step is to determine the most appropriate evaluation design to use. Evaluation designs provide the structure for how you will collect the data needed to answer your research questions.

There are a range of evaluation design options, each with different characteristics, requirements, strengths, and weaknesses. Which one you choose will depend upon the goals of the evaluation, available samples, availability and quality of data, budget, and time. Evaluation partners should be engaged early and actively in determining the best option(s).

Evaluation designs vary from experimental designs that can show whether a program causes particular outcomes to non-experimental designs that are more descriptive in nature. Randomized control trials (RCTs), where individuals are randomly assigned to groups, are considered the gold standard method of evaluation. However, such a design may not be desired, or even appropriate, depending on what you are trying to understand through your research questions. Process or implementation evaluations may be a good first step, where you are evaluating how your program is being implemented and identifying whether it is being implemented as intended (i.e., with fidelity).

In this section, we describe several design options, including experimental, quasi-experimental, and non-experimental designs.

Experimental Designs

In evaluations that use an experimental design, individuals from the same pool of clients or program participants are randomly assigned to one of two study groups: the treatment group or the comparison group. Outcomes are measured for both groups and then compared to each other to determine whether the program made a difference.

RANDOMIZED CONTROL TRIAL (RCT): The RCT is widely considered the "gold standard" evaluation design because people are randomly sorted into one of the two study groups. This randomization process makes the two groups more balanced or similar prior to receiving a program (for example, men and women have an equal chance of being assigned to each group using randomization) so that it is more feasible to conclude that any differences in the groups' outcomes were produced by the program.

CONSIDERATIONS: Is it ethically feasible to randomly assign program participants to either receive programming/services (i.e., treatment) or not (i.e., comparison)? In other words, does the creation of a comparison group mean that some clients would be denied needed services?

POTENTIAL RANDOMIZATION STRATEGIES:

- There is an expected waiting list or oversubscription of services that would allow for randomization without denying services to clients who would otherwise be served absent the evaluation, and there is no prioritization of those who receive the service, i.e., moved to the top of the waiting list.
- Services could be expanded to other locales or clients that are not currently being served, thereby allowing the creation of a comparison group from a currently unserved population and the context between the two groups can be taken into account.
- Components of the program could be randomized to compare the effectiveness of receiving one service (or set of services) to another, rather than comparing the effect of receiving all services to receiving no services.

Quasi-Experimental Designs

In quasi-experimental evaluations, methods other than random assignment, such as systematic matching, are used to create the study conditions. Groups are matched based on characteristics of people, rather than random assignment (for example, white men in their 30s are matched to other white men in their 30s and the total number of white men in their 30s is divided into the treatment or control groups). Outcomes are measured for both groups and compared to each other.

PROPENSITY SCORE MATCHING (PSM):
PSM is a statistical technique where participants are assigned to the treatment and control conditions based on their chances for a particular group membership, known as a propensity score.

- The technique still requires the presence of a similar but unserved group of people available for comparison to the group of people receiving the program or intervention.
- Because the groups are not randomly assigned using this method, there is potential for bias in the outcomes. In other words, outcomes could be the result of an unobserved factor and not the treatment, or the treatment combined with an unobserved factor.

DIFFERENCE IN DIFFERENCES (DD):
DD evaluations capitalize on natural experiments. This design measures changes, or differences, over time within two groups: one that experience an event or "treatment" and one that did not (the comparison group). First, within each group, the evaluator measures the difference between that group's outcomes after the event occurred compared to prior to the event. These differences within each group are then compared to each other to see what the difference is in the change over time between the group that experienced the event and the one that did not. In other words, the change over time in the

treatment group is compared to the change over time in the comparison. Hence, the name difference in differences.

- DD evaluation designs work well when researchers have access to good administrative data. For example, suppose there are two counties that are collecting data on opioid use and overdose, and both counties are observing an increase in the rate of opioid overdose. County A implements a law requiring police officers to carry naloxone, but County B does not. The effect of that law can be evaluated using a DD approach. First, you would calculate the difference in the rate of opioid overdose six months before the law was implemented to six months after the law was implemented, in each county. Then you could compare those differences between the counties to see if County A, the county that implemented the new naloxone law, had a different outcome than the one that did not, County B.

REGRESSION DISCONTINUITY DESIGN (RDD): A key feature of RDD is to apply a cut-off point or threshold to a continuum of risks or needs as a way to assign participants to a treatment or comparison group. RDD is based on the assumption that participants who are very close to the cut-off point are similar enough to use for comparison.

- RDD is appropriate when random assignment to a treatment and control group is not feasible, but when the evaluator can collaborate with program personnel to assign participants to the program systematically on the basis of need, merit, or some other qualifying condition.
- RDD may use administrative data already collected, and does not change service delivery or access for program or study participants.
- RDD is not as statistically powerful as RCT designs and cannot be implemented if there are an insufficient number of program participants who have been screened or too few cases of each outcome. Power calculation will determine sample size required.

Non-Experimental Designs

In non-experimental evaluations, there is no control condition and participants' outcomes are measured before and after intervention or program exposure.

PRE-POST DESIGN: This design does not incorporate a control group, or counterfactual (i.e., a measure of how participants would have fared without the program). Without a counterfactual, evaluators are unable to determine whether measured outcomes are a result of exposure to the program or historical changes in the context.

QUALITATIVE, OR DESCRIPTIVE, DESIGN: Qualitative methods are an effective way to analyze process. Qualitative methods of data collection include observations of activities, interviews, focus groups, and descriptive checklists or surveys. One consideration for conducting qualitative designs is the importance of an objective perspective. Partnering with an external evaluator is a great way to ensure objectivity in the evaluation. Since they are external to the program and not directly invested in it, they can evaluate process and program fidelity in an objective manner.

6 Sampling

Different evaluation designs require different sampling methods. Sampling is the process of selecting the individuals from the population who will be part of the treatment group and part of the control or comparison group.

Probability or random sampling refers to strategies for selecting a sample where individuals have equal chance of being part of the treatment group or the comparison group. There are four main types of probability sampling:

- Simple random sampling – individuals have an equal chance of being selected for the treatment or comparison groups.
- Stratified random sampling – individuals are stratified, or grouped, based on certain characteristics and then the researcher randomly assigns individuals from each group proportionally to the treatment group or comparison group. This strategy ensures the sample reflects the population based on those characteristics.
- Systematic sampling – individuals in the population are given a number and researchers select every nth (e.g., 5th) person in the population to be in the sample, using a random starting point on the list to begin selection. As they are selected, individuals alternately are assigned to either the treatment or comparison group.
- Cluster sampling – the population is divided into a certain number of groups, and then the researchers randomly select a certain number of clusters to be included in the sample. Once the sample is identified, clusters are randomly assigned to either the treatment or comparison group.

If you are using a quasi-experimental or non-experimental evaluation design and are unable to randomly assign individuals to the treatment or comparison group, you can use convenience sampling to identify your sample. Convenience sampling is just as it sounds; individuals are chosen to be part of the sample because they are easily accessible, or convenient. An example of convenience sampling is conducting interviews with individuals who are receiving services on a particular day – the day you are collecting the data.

LEAD SAMPLE

Adults suspected of low-level drug or prostitution offenses (N=318)

LEAD (N = 203)
Comparison (N = 115)

Law enforcement officers randomized to be in either LEAD or control shifts

Individuals arrested by officers in LEAD shifts offered entry to LEAD; individuals arrested by officers in control shift were not

Eligible individuals also referred to LEAD group by social contacts (officers)

Comparison group areas, in addition to control shifts, were added to account for increase in LEAD participants from social referrals

7 Data Elements

Once you have identified the appropriate evaluation design for your program, you must choose the measures, or data elements that will be collected to help answer your research questions. Identifying the proper measures that will demonstrate program effectiveness and will also be feasible to collect may be your biggest challenge.

One reason that the identification of measures is so difficult is because you cannot measure something that didn't happen, or a non-event. Especially for broad community-based prevention efforts, measuring the impact on arrests, for instance, is difficult because you would need population level data to determine if the program is reducing overall arrests in the jurisdiction, and be able to attribute it to the program and not another factor. Depending on available data, it can be easier to measure prevention of re-arrest for individuals who have been in contact with a program.

When considering what measures you might use, consider what data you are already collecting (e.g., in arrest records or client records) so you do not have to collect a lot of new data. Additionally, your partners, such as treatment organizations, justice organizations, and state agencies are also likely already collecting data that might be useful for your evaluation. In order to use partners' data for your evaluation, you may need to establish a Data Use Agreement (DUA) or Memorandum of Understanding (MOU) that outlines how data will be used and exchanged in order to ensure protections for the individuals from whom the data were collected.

In this section, we'll cover some the types of measures you might want to collect and provide examples of jail diversion-related measures in each category. We'll also provide potential data sources that might yield those measures.

COMMON MEASURES TO CONSIDER

INDICATORS	POSSIBLE DATA SOURCE
Baseline Equivalence Measures: **These are measures that you may use to test to determine if your treatment/intervention group and your control/comparison group are similar enough on related variables to demonstrate that the program is causing differences in outcomes rather than difference in the two groups.**	
Age	Client record; arrest record
Gender	Client record; arrest record
Previous Contact with Police	Arrest/court records
Number of Previous Arrests	Arrest/court records
Previous psychiatric hospitalization or detox/rehab	Client record; insurance record
Current connection with treatment	Behavioral health authority; Medicaid authority
Housing status	Client record; participant survey
Employment status	Client record; participant survey
Population Measures: **Rates of prevalence of a certain illness, condition, risk factor, protective factor, or demographic factor among the population as defined by a geographic area.**	
Percentage of individuals with MI/SUD diagnoses	National Survey on Drug Use and Health; state or local public health agency
Percentage of individuals living in poverty	Census
Distribution by race/ethnicity	Census
Encounter Measures: **Encounter measures count the number of times an activity or intervention is completed. This may include activities a program is trying to increase (e.g., contacts involving diversion) or activities a program is trying to decrease (e.g., use of force).**	
Police contacts with individuals with MI/SUDs	Police logs, calls for service
Narcan use	Police logs, EMS records, emergency department data
Contacts involving diversion	Police logs, crisis center admissions data
Connections to services	Police logs, diversion program data, behavioral health provider client data
Use of force	Police logs
Process Measures: **Process measures track what has been done as part of program implementation. These may include number of individuals served or trained, fidelity to the model, or satisfaction with the program.**	
Police satisfaction with model	Program/partner survey
Individual/service recipient satisfaction with model	Program/client survey
Number of officers trained	Program data
Implementation fidelity (for existing models with fidelity standards)	Program data
Collaboration between behavioral health organizations and law enforcement/corrections (e.g., new memoranda of agreements)	Program data
Changes in communications and resource ties (i.e., social network analysis)	Qualitative analysis
Outcome Measures: **Outcome measures are used to determine if the program is having the desired effect. They are indicators that can be changed or improved as a result of the program.**	
Percentage of arrests involving SU intoxication and/or mental illness	Police logs
Percentage of re-arrests involving individuals with SUD and/or MI	Police logs; public records; diversion program data
Percentage of re-arrests of individuals diverted	Police logs; public records; diversion program data
Time from diversion to next police contact	Police logs; public records; diversion program data
Treatment Outcomes, like hospitalizations, clinical measures of substance use and mental health, overdoses and overdose deaths	Hospital data; client records; emergency department data; CDC WONDER (overdose deaths)
Homelessness	Homelessness point in time surveys (a standardized survey of individuals, youth, and families experiencing homelessness at one point in time)
Costs and Cost Offsets: **The decline in costs to other parts of the organization or system due to the program.**	
Money saved at the programmatic/precinct level as well as longer-term costs to the community in reduced health care costs, increased tax revenues (from those entering and staying in the workforce), etc.	Budgets for the program, organization (if the organization also provides other services), law enforcement entity, jails, and other costs such as medical and behavioral health services.

Examples of Measures Used in Jail Diversion

Process

- Number Trained
- Satisfaction with Services Police Satisfaction with Mode
- Satisfaction with Training
- Fidelity to Model*
- Number Enrolled
- Percent Completed Treatment
- Number of Encounters
- Linkage to Treatment
- Police Referrals to Program

Descriptive

- Age
- Gender
- Race
- Ethnicity
- Prior Arrest
- Percent of Population in Poverty
- Number or Percentage of Population Without Housing
- Percentage of Population with SUD or SMI

Justice Outcomes

- Arrests
- Non-Warrant Arrests
- Total Charges
- Felony Charges
- Jail Bookings
- Jail Days
- Use of Force
- Calls to Scene for Behavioral Health
- Co-Responder Interventions
- Use of Protective Custody

Clinical Outcomes

- Naloxone Use For Opioid Overdose Reversal
- Depression Symptoms
- Abstinence from Opioids

Cost, Cost-Benefit

- Cost of Programming
- Staff Costs
- Dollars Saved Arrests Deflected
- Dollars Saved Hospitalization
- Justice and Legal System Costs
- Jail Bed Costs

LEAD OUTCOME MEASURES

Research Question 1:
recidivism (shorter-term: 6 months pre/post; longer-term: entirety of evaluation time frame)
Arrests
Non-warrant arrests
Total charges
Felony charges

Research Question 2:
reduction in service utilization
Jail bookings
Jail days
Prison incarceration
Misdemeanor and felony cases
Costs associated with justice and legal system utilization

8

Data Collection

Now that you have determined the appropriate evaluation design for your research question(s) and identified measures and their data sources, you next need to prepare for and begin collecting data. Factors to consider for data collection include:

a. What is ethical? When dealing with vulnerable populations, researchers must consider what and how they are collecting sensitive information, and be sure they have proper approval from an Institutional Review Board (IRB) and the participants. [See the resources page for additional ethics resources.]
b. What is logically feasible? Are data already available? Have you chosen a research design and sample that fit into your budget? Are the chosen outcome measures too difficult to collect and result in unacceptably large amount of missing data?
c. Establish the timeline and elements needed to establish a baseline. It's crucial to have baseline measures of the population of interest prior to the start of the program. That's why it is so important to plan your evaluation while you are planning your program to ensure the availability of data related to the outcomes of interest prior to program or service receipt.
d. What is the timing and frequency of data collection? How often and at what time points do data need to be collected to be able to answer your research questions?
e. How will you find a comparison group? How will you identify and measure that group? Is there a waiting list for services? Would a quasi-experimental matching technique or non-experimental design be more appropriate or feasible?
f. What quality assurance processes are in place, or will need to be put in place, to ensure data are being collected accurately and on time?

LEAD ETHICAL CONSIDERATIONS

When LEAD conducted semi-structured interviews of program participants as part of their evaluation, they used an informed consent procedure to inform participants:

- about the purpose and procedures of the interview
- that their participation in the interview will not affect their services
- that their comments would be kept confidential and will be aggregated so program staff would not be able to attribute comments to any one participant
- the offered incentive ($5 gift card for fast food) would be provided regardless of the comments they provide

9

Data Analysis

Once you have collected all of the data required for your evaluation, the data will need to be analyzed. The methods and techniques you will use to analyze your data will depend on the evaluation design that you used, although some techniques will be consistent regardless of your approach. For instance, you will likely want to run basic descriptive statistics on your data, such as frequencies (or counts), means (or averages), and ranges of the values on each variable. You will then likely want to run more sophisticated analyses, but how sophisticated will depend on your design. For example, in non-experimental designs, you may only need to conduct a comparison of means (or averages) at two time points (before and after). Other analytic techniques, such as linear and logistic regression, analysis of variance (ANOVA), cluster analysis, or hierarchical linear modeling (HLM), may be required depending on the structure of your data and the design of your evaluation.

Resources for commonly used statistical techniques can be found in the resources section.

10

Conclusion

Data are powerful. Data on the number of opioid overdoses can be used to leverage funds from state or local government for program development and evaluation. Data on officer time spent on mental health encounters could be useful in justifying community diversion services or co-response models.
With planning, engagement of stakeholders and other partners, and a feasible design given the resources at hand, evaluation of your jail diversion program can inform both ongoing program operations and policy related to jail diversion for individuals with behavioral health diagnoses. You can use your evaluation to make programmatic decisions, and the data can also be useful to your partner organizations, policy makers, service recipients, and community members. The resources provided in this Resource Guide can get you started on making vital decisions about evaluation design, data collection, and analysis.

Glossary

Business Associates Agreement (BAA)	A contract between a health care entity/organization and another entity/organization (business associate) that does work for or on behalf of the health care entity/organization. Business associates are required to follow all confidentiality requirements of the primary entity/organization. Behavioral health care providers are almost always "covered entities" under the Health Insurance Portability and Accountability Act. More information and a sample contract are located at https://www.hhs.gov/hipaa/for-professionals/covered-entities/sample-business-associate-agreement-provisions/index.html.[1]
Comparison Group	Research subjects who receive no treatment or alternative treatment.
Contextual Factor	Outside influences which can impact the implementation or outcome of a program.
Counterfactual	What the outcomes would be if the program did not take place.
Crisis	A state of emotional instability and risk due to spiritual, relationship, emotional, or increase in psychiatric symptoms that lead to functional impairment, suicidality or risk of physical harm to self or others.
Crisis Respite	Specific definitions vary by state but crisis respites are non-clinical residential crisis settings with or without peer support.
Crisis Stabilization Service	Typically a non-hospital based residential unit or center where individuals can people can stay up to two weeks and receive counseling, help with medication stabilization, and support to resolve a psychiatric or emotional crisis. Specific definitions vary by state.
Data Sharing Agreement	A legal agreement that agreement establishes the terms and conditions under which one organizations/party agrees to share, acquire, and use data from the other party. Either party may be a provider of data to the other, or a recipient of data from the other.
Experimental Design	A design in which individuals from the same study participant pool are randomized into one of two study conditions: treatment or control. Outcomes are measured for both groups and compared to each other.
Evidence-based	A treatment practice or approach that is backed by a strong body of research evidence.
Federalwide Assurance	An agreement obtained from Health and Human Services that states the research conducted protects the rights and welfare of human subjects.
Forensic Peer Support	Peer support specialists trained to work with individuals with behavioral health disorders involved in the justice system from pre-booking jail diversion to community re-entry.[2]
Human Subjects Research	Any research project, study, or evaluation that involves a human being, and the researcher/evaluator collects data through an program or interaction that contains identifiable private information.[3] There are additional protections for incarcerated individuals.
Implementation Study	Research dedicated to the promotion of best practices to improve quality and effectiveness in care.

[1]https://www.hhs.gov/hipaa/for-professionals/covered-entities/sample-business-associate-agreement-provisions/index.html

[2]Barron, R. (2011). Forensic Peer Specialists: An Emerging Workforce. New Brunswick, NJ: Center for Behavioral Health Services and Criminal Justice Research, Institute for Health, Health Care policy, and Aging Research.

[3]https://www.hhs.gov/ohrp/regulations-and-policy/guidance/guidance-on-engagement-of-institutions/index.html

Informed Consent	The informed consent process involves three key features: (1) disclosing to potential research subjects information needed to make an informed decision; (2) facilitating the understanding of what has been disclosed; and (3) promoting the voluntariness of the decision about whether or not to participate in the research. Informed consent must be legally effective and prospectively obtained. HHS regulations at 45 CFR 46.116 and 45 CFR 46.117 describe the informed consent requirements.[4]
Input	The resources that will go into your program/intervention, as represented in your logic model.
Institutional Review Board (IRB)	The administrative body that protects the rights and welfare of human subjects in research.
Measure	An indicator that will be used to assess whether a program is reaching the intended outcome. A measure will assist you to know if change occurred and in what direction was the change. Measures are usually broken down into process and outcome measures. An example of a process measure might be, number of officers trained in Crisis Intervention Teams, implementation of a new behavioral health assessment procedure, number of diversions, or establishment of a memorandum of understanding between police and a behavioral health authority. An example of outcome measure might be opioid overdose deaths or recidivism/repeated police encounter rates. It is difficult to measure if a non-event occurred---if someone was not arrested or the person does not encounter police but need some kind of intervention.
Mobile Crisis Intervention / Mobile Crisis Team	Mobile crisis interventions are services delivered in the field…i.e. in someone's home, school, work or other locale, generally provided by a licensed mental health clinician and often a peer support specialist who assess the adult/child/youth's needs and recommend services to stabilize the crisis. The original aim of these interventions was to avoid hospitalization and to help stabilize the person's crisis in the least restrictive setting possible. Specific names of these teams/programs vary by state and locale.
Natural Experiments	Studies in which individuals receive either the experimental and control conditions due to factors outside the control of evaluators.
Overdose	An overdose occurs when a person uses enough of a drug to produce a life-threatening reaction or death.
Peer Support	The process of giving and receiving non-clinical assistance to achieve long-term recovery from severe psychiatric, traumatic or addiction challenges. This support is provided by peer supporters - people who have "lived experience" and have been trained to assist others in initiating and maintaining long-term recovery and enhancing the quality of life for individuals and their families. Peer support services are inherently designed, developed, delivered, evaluated and supervised by peers in long-term recovery.[5,6]
Population	A collection of individuals who are a part of the group being studied.
Population Level Data	A data set that is representative of an entire population.
Protected Health Information	Health information created, stored, and used by healthcare providers, insurers, and their business associates covered by the Health Insurance Portability and Accountability Act Privacy Rule.[7] Information on behavioral health services, dates of enrollment, case managements services, data on opioid use and overdoses, etc. might be protected health information included in an evaluation of a jail diversion program.

[4]https://www.hhs.gov/ohrp/regulations-and-policy/guidance/faq/informed-consent/index.html

[5]White, W. (2009). Peer-Based Addiction Recovery Support: History, Theory, Practice and Scientific Evaluation.

[6]https://www.psychrehabassociation.org/newsletters/psyr/developing-role-peer-specialists-within-jail-diversion-program

Pre-post Study Design (Also called a Before and After Study)	In a pre-post study, you measure your outcomes of interest before and after the program. However, you will not know whether changes in outcomes are a result of the program or other factors because you won't have data on people with similar problems (comparison group) who are not getting the intervention. Pre-post study designs only provide a snap shot of changes in outcomes, so it's they are weaker designs.[8,9] There are different variations to help strengthen pre-post designs but they are not as robust as more expensive quasi-experimental or experimental designs.[10]
Prescription Drug Misuse	The use of a medication in ways or amounts other than intended by a doctor, by someone other than for whom the medication is prescribed, or for the experience or feeling the medication causes. This term is used interchangeably with "nonmedical" use, a term employed by many national drug use surveys.
Process Evaluation	Evaluations that assess how programs are implemented and if they are executed according to plan.
Quasi-Experimental Design	A design where outcomes for a group of people who receive an intervention are compared with outcomes for group who do not receive the intervention. Individuals are not randomly assigned to the program and comparison groups. Comparison groups can be created through propensity score matching or by assigning groups by thresholds of risk.
Regression to the Mean	In the course of time, some people will recover on their own from a condition (or at least improve) and some people who are doing fine may develop worsening symptoms. If you just collect measures of interest before and after an intervention, some people will do better and some worse. In terms of jail diversion, some individuals with behavioral health conditions will come into contact with the justice system again or for the first time and some will appear to be doing better. This is known as regression to the mean. In this case, the study will not be able to clearly show if the program achieved the outcome of interest. One way to address the problem is by having a comparison group of individuals with similar characteristics to those receiving the intervention. Or even better, by randomly assigning people to receive the intervention or not.
Reliability	The reliability of a measure refers to whether or not the instrument or tool used to measure outcomes dependably gives you the same outcome or score. For example, the tool you might use to measure opioid use collects consistent information not matter who is collecting the data, each time it is used, and between questions on opioid use. Basically, the measure would give you consistent data as it is used by different people over time.[11] See also validity.
Sequential Intercept Model	A framework that outlines a series of points, or intercepts, along the criminal justice continuum where intervention could occur to divert individuals to community resources, and away from the criminal justice system. Intercept 0 encompasses services in the community designed to prevent the involvement of vulnerable populations in the justice system.
Theory of Change	A theory of change first defines intended outcomes, then works backwards to determine necessary steps to take in order to reach those outcomes.

[7]http://www.hhs.gov/ocr/privacy/hipaa/understanding/index.html

[8]http://coalition4evidence.org/wp-content/uploads/2009/05/study-design-hierarchy-6-4-09.pdf

[9]http://158.132.155.107/posh97/private/research/evaluation/Chp_3.pdf

[10]http://www.excellenceforchildandyouth.ca/sites/default/files/olm/evaluation/5/01-07-08.pdf

[11]https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3184912/

Time Series Design	A more robust version of a pre-post test design.[12] The difference between a simple pre-post test design and a time series design is the number of times you measure outcomes. Ideally, in a time series study, measures of interest are collected at least four times. For example, data are collected before the program, during the program, at the end of the program, and at some time point(s) after the program has lasted. Robust experimental and quasi-experimental designs collect measures over time as well but with a randomly assigned control group or comparison group respectively.
Trauma-informed Approach	Term used by SAMHSA to mean "A program, organization, or system that is trauma-informed: Realizes the widespread impact of trauma and understands potential paths for recovery; Recognizes the signs and symptoms of trauma in clients, families, staff, and others involved with the system; Responds by fully integrating knowledge about trauma into policies, procedures, and practices; and Seeks to actively resist re-traumatization."[13]
Treatment Group	Research subjects who receive the treatment or participate in the program being evaluated.
Validity	When we talk about the validity of research and evaluation, we are talking about whether or not the study/evaluation whether or not the study used strong enough methods to truly answer the question posed by your study: did the program improve the outcomes you expected it to improve? When we think about validity in terms of measuring an outcome, it's important to know if a measure actually measures what we think it measures. The outcome for rearrests is pretty straight forward to measure. Engagement in treatment is trickier to measure (See measures). Making the first appointment for treatment might not be a valid measure of engagement in treatment.[14,15,16,17]
Variable	An element of interest that is expected to vary over time and across study participants. A variable is measured and entered into a statistical equation to determine if the participant's/organization's outcomes improver or deteriorate over time. An example might be the cost of services for an individual. Do costs go up, stay the same, or go down, over the course of and after a program.

[12]http://eknygos.lsmuni.lt/springer/587/63-76.pdf

[13]https://store.samhsa.gov/system/files/sma14-4884.pdf

[14]https://cirt.gcu.edu/research/developmentresources/research_ready/experimental/validity

[15]https://www.aqr.org.uk/glossary/validity

[16]https://opentextbc.ca/researchmethods/chapter/reliability-and-validity-of-measurement/

[17]https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3184912/

Resources

General Evaluation Resources

Agency for Healthcare Research and Quality, A Guide to Real-World Evaluations of Primary Care Interventions: Some Agency for Healthcare Research and Quality, A Guide to Real-World Evaluations of Primary Care Interventions: Some Practical Advice: https://pcmh.ahrq.gov/sites/default/files/attachments/PCMH_Evaluation_Guide.pdf

American Evaluation Association: https://www.eval.org/

Bureau of Justice Statistics, Guide to Program Evaluation: https://www.bja.gov/programs/crppe/research-eval.html

Centers for Disease Control and Prevention, A Framework for Program Evaluation: https://www.cdc.gov/eval/framework/index.htm

Centers for Disease Control and Prevention, Introduction to Program Evaluation for Public Health Programs: A Self-Study Guide: https://www.cdc.gov/eval/guide/index.htm

Children, Youth, and Families At-Risk (CYFAR), Interactive Learning Modules: https://cyfar.org/learning-mods-home

LEAD Program Evaluation

LEAD Evaluations: https://www.leadbureau.org/evaluations

LEAD Program Overview: https://docs.wixstatic.com/ugd/6f124f_535679d78c2541fdaf433d3983cb2a31.pdf

Research Questions

BetterEvaluation, Key Evaluation Questions: https://www.betterevaluation.org/en/rainbow_framework/frame/specify_key_evaluation_questions

Duke University Writing Studio, What Makes a Good Research Question?: https://sites.duke.edu/urgws/files/2014/02/Research-Questions_WS-handout.pdf

Grand Canyon University Center for Innovation in Research and Teaching, Research Questions and Hypotheses: https://cirt.gcu.edu/research/developmentresources/research_ready/quantresearch/question_hypoth

Logic Models

The Aspen Institute, The Community Builder's Approach to Theory of Change: A Practical Guide to Theory Development: http://www.theoryofchange.org/pdf/TOC_fac_guide.pdf

Bureau of Justice Statistics, Logic Models: https://www.bja.gov/programs/crppe/logic-models.html

Innovations for Poverty Action Goldilocks Toolkit, Theory of Change: Laying the Foundation for Right-Fit Data Collection: https://www.poverty-action.org/sites/default/files/publications/Goldilocks-Toolkit-Theory-of-Change.pdf

University of Kansas Center for Community Health and Development, Community Tool Box, Developing a Logic Model of Theory of Change: https://ctb.ku.edu/en/table-of-contents/overview/models-for-community-health-and-development/logic-model-development/main

Ethics

Health and Human Services Office for Human Research Protections: https://www.hhs.gov/ohrp/

Health and Human Services, Federalwide Assurance (FWA) for the Protection of Human Subjects: https://www.hhs.gov/ohrp/register-irbs-and-obtain-fwas/fwas/fwa-protection-of-human-subjecct/index.html

National Institutes of Health, Protection of Human Subjects: https://grants.nih.gov/policy/humansubjects.htm

Data Analysis

BetterEvaluation, Analyse Data: https://www.betterevaluation.org/en/rainbow_framework/describe/analyse_data

Grand Canyon University Center for Innovation in Research and Teaching, Analyzing Quantitative Research: https://cirt.gcu.edu/research/developmentresources/research_ready/quantresearch/analyze_data

Resources Needed for Evaluation

Michigan State University and Bureau of Justice Assistance, Identifying and Working with a Research Partner: https://psn.cj.msu.edu/tta/researchpartnerqa_version-2_june2017.pdf

September, 2019