



STRUCTURING A TELEHEALTH EVALUATION

JUNE 2020

► EVALUATION STRATEGY: OVERVIEW¹

Research suggests a four-step process for measuring the efficacy of telehealth

<i>Evaluability</i>	<ul style="list-style-type: none">• Assessment conducted prior to or at the beginning of a program to make explicit the goals and objectives of the program and intended effects or outcomes.• Key components include: (1) framing the research question, (2) determining research design, (3) identifying data collection methods, and (4) determining the analytic method.
<i>Documentation</i>	<ul style="list-style-type: none">• Narrative that describes the implementation of the program.• Documentation should include the procedures used, the difficulties encountered, the steps taken to address barriers, successes/challenges in dealing with those barriers, and how the program can be reproduced in another setting.
<i>Formative evaluation</i>	<ul style="list-style-type: none">• Evaluation focusing on the process of the program being delivered.• Formative evaluations allows for project managers to observe attitudinal changes related to the program, any workforce integration issues, and potential technical issues with program rollout.
<i>Summative evaluation</i>	<ul style="list-style-type: none">• Evaluation providing evidence of the intended effects of the program.• The goal of a summative evaluation is to provide robust evidence of a program, identify the benefits of the program, and provide evidence to policy and decision makers.

1. Agboola, Stephen et al., "Real-World' Practical Evaluation Strategies: A Review of Telehealth Evaluation," JMIR Research Protocols, Vol. 3(4), 2014.

▶ EVALUATION STRATEGY: CODIFYING THE PROGRAM MODEL¹

“Telehealth” can mean many different things, so establishing a clear program model is vital prior to conducting an evaluation

What are the goals of the program?

- As a first step, program staff should **define the primary goals and objectives of the telehealth program**.
 - Objectives can include improving clinical outcomes, increasing engagement, and decreasing costs – **and most likely all three**.
- Project coordinators **should identify and ensure access to the data needed** to measure the goals and the objectives of the telehealth program.
 - Data sources can include **electronic health records, Medi-Cal claims data, and patient surveys**.
- Attention should be paid to the **questions that senior stakeholders and decision-makers will want to have answered**, as these individuals will determine the rate of scale if successful.

How is the program structured?

- Program staff should **clearly define each component** of the program model:
 - Who is **the target population** being served?
 - **Who delivers services**, and what are the required qualifications?
 - Who are the **major stakeholders**, and what is their role?
 - What is the **timeline for services**? When do you expect to be able to measure clinical outcomes?
- ***Performance indicators should be generated for each step in the process.***

1. Agboola, Stephen et al., “Real-World’ Practical Evaluation Strategies: A Review of Telehealth Evaluation,” JMIR Research Protocols, Vol. 3(4), 2014.

▶ OUTCOME METRICS: PROCESS AND SUMMATIVE EVALUATION

Formative evaluation¹

What questions need to be answered to ensure that the program is functioning appropriately and engaging individuals?

- Was the **target audience appropriately reached**?
- Has the **program been adopted** by key stakeholders?
- Are **key stakeholders engaged**?
- Are **engagement patterns different** based on time, demographic group, or location?
- Are some participants **responding differently to engagement techniques** than others?
- Are **ongoing troubleshooting issues** reported and addressed promptly?

Summative evaluation

What questions need to be answered to ensure that the program is improving outcomes for individuals?

Example outcome metrics for different populations may include:

- **Substance use disorder:** changes in ED visits or hospitalizations due to substance use²
- **Veterans with PTSD:** changes in the clinician-administered PTSD scale³
- **Individuals with SMIs:** change in psychiatric symptoms (e.g., psychosis and depression), health outcomes (weight, blood pressure), ED visits and/or hospitalizations due to SMIs⁴

1. Agboola, Stephen et al., "'Real-World' Practical Evaluation Strategies: A Review of Telehealth Evaluation," JMIR Research Protocols, Vol. 3(4), 2014.

2. Huskamp, Haden et al., "How Is Telemedicine Being Used In Opioid and Other Substance Use Disorder Treatment?", Health Affairs, Vol. 37(12), December 2018.

3. Morland, Leslie A et al., "Group Cognitive Processing Therapy Delivered to Veterans via Telehealth: A Pilot Cohort," Journal of Traumatic Stress, Vol. 24(4), August 2011.

4. Pratt, Sarah I et al., "Feasibility and Effectiveness of an Automated Telehealth Intervention to Improve Illness Self-Management in People With Serious Psychiatric and Medical Disorders", Psychiatric Rehabilitation Journal, Vol. 36(4), December 2013.

▶ DATA SOURCES FOR MEASUREMENT

Summative evaluations will require access to administrative physical and behavioral health data

Data source	Primary use	Evaluation type
Patient survey	<ul style="list-style-type: none"> Used to measure client satisfaction with services. Can be particularly helpful for telehealth to measure the setup (e.g., technology) and engagement on the part of the client and clinician. Not useful to measure clinical outcomes. 	<i>Formative</i>
Electronic health records ¹	<ul style="list-style-type: none"> Contain a patient's medical history, diagnoses, treatment plans, immunization dates, etc. Provider notes can provide context on telehealth sessions and scores on different psychiatric rubrics. Will include most of the medical data that is used make clinical decisions, but may not be as easy to code for summative evaluation. 	<i>Formative, summative</i>
Insurance (e.g., Medi-Cal) claims data	<ul style="list-style-type: none"> Will contain each instance that an individual is billed for a medical service. Helpful for understanding both utilization and cost patterns. Will not necessarily provide context on the specific reason for an incident. 	<i>Summative</i>

1. "What is an electronic health record (EHR)?", HealthIT.gov.

▶ POTENTIAL RESEARCH PARTNERS



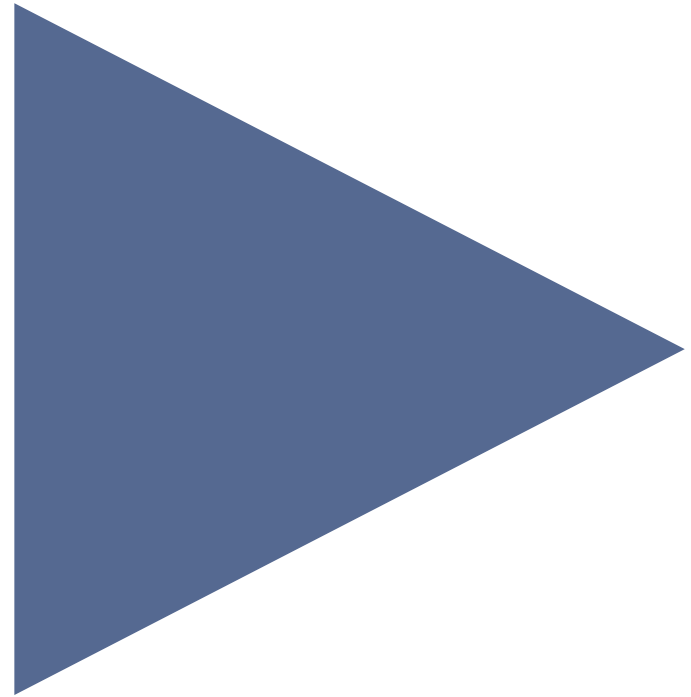
Partner with California's state and local governments to generate scientific evidence that solves California's most urgent problems

UC DAVIS
HEALTH

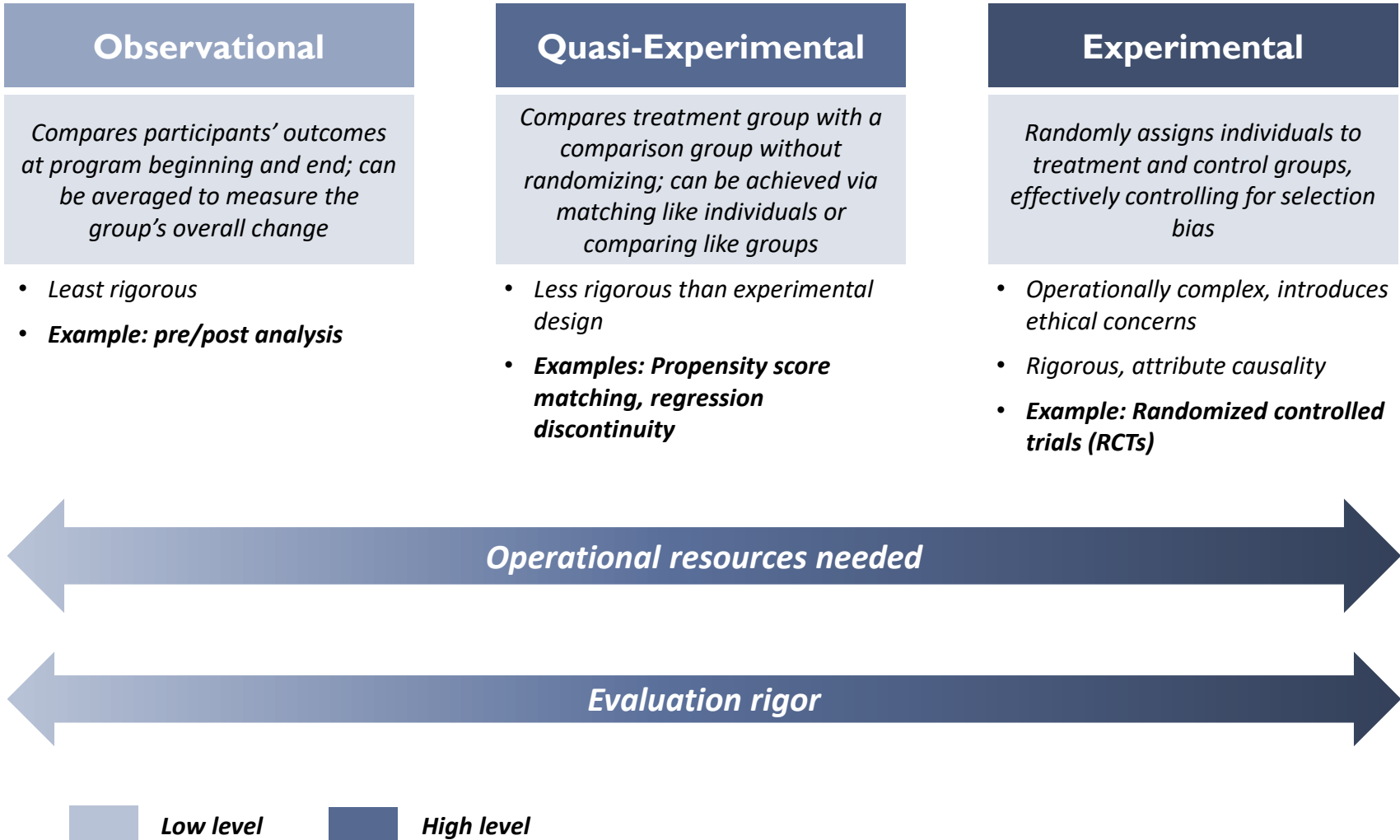
*Clinical and Translational
Science Center*

Serves as a catalyst to support biomedical research projects by providing services and resources to investigators, staff, scholars and trainees, and community partners.

▶ APPENDIX: EVALUATION DESIGNS



▶ TYPES OF EVALUATION DESIGNS



OVERVIEW OF EVALUATION METHODS

Evaluation method	Definition	Strengths	Weaknesses
Randomized Controlled Trial (RCT)	Randomly assign target population to treatment and control groups to isolate intervention effects	<ul style="list-style-type: none"> • Considered most credible • Clear distinction between treatment and control groups • Simple to interpret 	<ul style="list-style-type: none"> • Potential ethical concerns • Operationally complex to set-up • Potential strict research protocols
Quasi-experimental design (QED): Regression discontinuity	Use a threshold to assign contemporaneous treatment and comparison groups and estimate the marginal impact of being near a cutoff	<ul style="list-style-type: none"> • Closest substitute to randomization • Assumes that those who just barely received treatment are comparable to those who just barely did not receive treatment 	<ul style="list-style-type: none"> • Can't extrapolate findings beyond narrow bandwidth (less precise further away from cutoff) • Difficult to find large enough sample near the threshold to estimate precise results
QED: Matching	Match program participants with non-treated, contemporaneous "control" group using demographic variables	<ul style="list-style-type: none"> • Can build on existing evaluations and assumptions to inform evaluation design • Can be conducted retrospectively • Easier to obtain stakeholder buy-in 	<ul style="list-style-type: none"> • Potential for selection bias -- cannot control for unobservable characteristics • Difficult to achieve baseline equivalence between groups (may need large comparison sample)
QED: Difference-in-differences	Compare before-and-after effects for a group or region of intervention participants vs non-treated group or region	<ul style="list-style-type: none"> • Relatively simple to calculate; can be calculated with a single, basic regression 	<ul style="list-style-type: none"> • Relies heavily on assumption that absent the program groups would have had "parallel trends"
Historical baseline	Compare outcomes of intervention participants to incidence of outcome in a historical group	<ul style="list-style-type: none"> • Less operationally complex and costly 	<ul style="list-style-type: none"> • High potential for selection bias • Policy changes that have occurred may skew observed results • Not very rigorous