

**CBEVSS:
What's the Difference?
Constructing Meaningful Comparison Groups**

[00:00:00 silent]

[Introduction Graphic – Child Welfare Evaluation Virtual Summit Series]

[Title Graphic – What's the Difference? Constructing Meaningful Comparison Groups.

Image shows colorful doodles of diverse stick people of many colors and sizes holding hands and smiling.]

[Image fades to doodle graphics of a stick man between two projector screen type presentation boards. The boards illustrate the points being spoken by the female narrator as she talks.]

Female Narrator: [00:00:21] As child welfare professionals, we're constantly comparing. It's how we understand, learn, solve problems, and improve. We may often find ourselves asking questions like...

[Graphic – Photo of a man, with text of the following question spoken by the male narrator]

Male Narrator: [00:00:30] Are more children in my community victims of maltreatment than in other communities?

[Graphic – Photo of a woman, with text of the following question spoken by the Female Narrator]

Female Narrator: [00:00:34] Is my private agency's adoption program outperforming other agencies?

[Graphic – Photo of a different man, with text of the following question spoken by the male narrator]

Male Narrator: [00:00:38] Is a child doing better now, perhaps behaviorally or emotionally, than prior to my child welfare agency's involvement?

[Graphic – Doodle people carrying puzzle pieces come together as narrator speaks, puzzle reads "partnership" when complete.]

Female Narrator: [00:00:45] We're constantly partnering with our internal and external stakeholders to make our services better, to strengthen our collaborations, and to improve outcomes for the children and families we are serving. [Graphics of happy families with children] And we always want to know if what we are doing is working.

[Graphics fade in and out of adults working closely with children.]

Male Narrator: [00:00:57] And that's particularly true when we implement a new program or policy and we don't know yet whether it's going to be a better option than the programs or services our children and families are currently receiving.

[Graphic of Doodle person shooting a bow and arrow at a target, hits the bull's-eye.]

Female Narrator: [00:01:07] Right. But how do we know if the interventions we're implementing are doing what they're supposed to? [Graphics of happy children fade in and out.] Or more precisely, how do we really know if our reform efforts are responsible for any changes in outcomes we are observing or if they're due to other factors like growth and development over time or other services children and families may be receiving?

[Video - young adult's working together with laptops and folders, no additional audio]

Male Narrator: [00:01:24] Well, that question is the driving force behind child welfare outcome evaluation efforts, and it's exactly what researchers and evaluators are getting at when they're talking about developing rigorous program evaluations.

[Video – Hand writing with pen and paper]

Female Narrator: [00:01:35] I know we see the requirement for rigorous evaluations when applying for grants or other funding opportunities, but what does rigorous mean?

[Graphic Text – What does rigorous evaluation design mean?]

And how do we go about developing a rigorous evaluation design?

[Graphic – Doodles of graphs, charts, people, and information, demonstrating the points of rigorous evaluation design]

Male Narrator: [00:01:44] Well, first it's important to know that there are many factors that contribute to a rigorous outcome evaluation. One is a strong evaluation design that includes a meaningful comparison group. Additional components of a rigorous evaluation effort include the selection of valid and reliable tools such as surveys and assessment instruments to measure change in children and families as well as systematic procedures for collecting data.

[Graphic Text – Did it work? Doodle woman uses pencil marking “yes or no” checkbox]

Female Narrator: [00:02:07] And while a good comparison group design and quality measures may help us determine whether our intervention was successful, those elements by themselves cannot explain why it was successful. [Graphic Text – Why did it work?] Answering those questions requires additional methods that should be part of a thorough process evaluation.

[Graphic – Doodles demonstrate comparison factors]

Male Narrator: [00:02:22] For now let's focus on the development of meaningful comparison groups and how they can help us figure out whether our child welfare programs are impacting the outcomes we're trying to measure. In short, comparison

groups can help us answer the big question, did it work? A comparison group can help us answer this question because it gives us a systematic way of predicting what would have happened to children and families if they had not received the program or service in question.

[Graphic Text – What does a comparison group look like?]

Female Narrator: [00:02:46] So, what does a comparison group look like?

Male Narrator: [00:02:48] Well, no matter what we're evaluating, a comparison group will always consist of units. That's a researchy word, which in child welfare most often refers to children, parents, families, case workers, or agencies. [Graphic – Comparison groups illustrated on doodle presentation boards, on left "Received Service" on right Comparison Group] As child welfare leaders committed to evaluating our programs, we want to identify groups of children or families or whatever the unit is who did not receive our new intervention so that we can compare them with those people who did receive it, and we want those two groups to be as equivalent as possible.

[Graphic – Doodle graphs illustrate female narrator's points on comparison.]

Female Narrator: [00:03:17] So, the actual effect of our program is measured by the difference between the outcomes achieved by the group that is receiving services and those outcomes that are achieved by our comparison group.

[Graphic – Doodles illustrate male narrator’s points on comparison.]

Male Narrator: [00:03:26] Exactly. If we don’t have a comparison group, it could be difficult to tell whether any changes in outcomes we are seeing are due to the new program we implemented or whether they’re due to other factors that play within our community.

[Graphic – Chalkboard graphic, list of graphic chalk text illustrates the narrator’s points.]

Female Narrator: [00:03:38] There are several approaches to developing good comparison groups. Here is four of them.

Male Narrator: [00:03:42] The first option is a Randomized Controlled Trial Design or RCT for short. This is generally regarded as the strongest approach to building a comparison group. With a Randomized Controlled Trial, units – let’s say parents eligible for a new parenting education program – are assigned randomly to either an experimental group, the new parenting program, or to a comparison group, which is referred to as a control group in the context of RCTs.

[Graphic – Doodles of a comparison. On left, Experimental group, on right, Control group.]

[00:04:09] Units assigned to the control group are offered services as usual. In this case whatever parenting education resources were available to parents in the target population before the new intervention became available.

Female Narrator: [00:04:20] So, random assignment means that each eligible unit has an equal chance of being assigned either to the new program or to the usual program.

Male Narrator: [00:04:28] RCTs are often regarded as the gold standard when it comes to conducting research in child welfare and other human service settings because they're very good at isolating the impact of our intervention from other possible explanations for outcome changes. Because the experimental and control groups look almost exactly the same after randomization except that one group receives the new intervention and the other group does not, we can say with more confidence that any differences in the outcomes we're seeing between the two groups are in fact due to the new program and not due to something else.

[Graphic – 3 doodle notepads, text reads – Technical expertise, Resources, Large sample size]

Female Narrator: [00:04:59] But sometimes RCTs aren't possible, right? I mean they require certain technical expertise, resources, or sample sizes that may not be available for all evaluations.

Male Narrator: [00:05:07] Yes. And in those situations and others there are other methods we can consider.

[Graphic – Chalkboard image with chalk writing]

Female Narrator: [00:05:12] One alternative is called a Waitlist Comparison or Overflow Design. This option could be used, for example, to test a new group therapy model for adolescents. [Graphic – Doodle demonstrates waiting list] With the Waitlist Design, all the teenagers eligible for the new therapy model are served on a first-come, first-served basis depending on the service capacity of the therapist. When the first therapy groups fill up, the next teenagers that come along are placed on a waitlist, and their outcomes are watched and compared with outcomes among teenagers that received the new group therapy model.

[00:05:41] As new therapy groups open up, the teenagers on the waitlist are reassigned to the experimental group on a rolling basis and offered the intervention. As with RCTs, Waitlist Designs are useful when we have more children or families that are eligible for a new intervention than we have the organizational capacity or resources to provide it, at least at a given time. However, unlike RCTs, everyone on a Waitlist Design gets the new intervention eventually.

[Graphic Text – Doodle of hanging sign, reads - “10 week support therapy program”]

Male Narrator: [00:06:03] But Waitlist Designs can be tricky to implement. First, they generally only work when we have a very discreet and clearly defined service that is only provided for a relatively short and specific period of time like in the group therapy example we just described so that we can complete our observation of outcomes among the waitlisted group before space opens up in the experimental group.

[Graphic – Doodle of a school bus, side of bus reads “Waitlist Design”]

[00:06:24] For this reason Waitlist Designs are often used with short-term classroom-based educational programs or brief therapeutic interventions in the medical, mental health, and public health fields.

Female Narrator: [00:06:35] It is also really important to maintain strict control over the wait-listing process so that group assignment truly occurs on a first-come, first-served basis rather than depending on staff’s personal assessments of clients’ needs.

[Graphic – Doodle man looks at signpost for information]

Male Narrator: [00:06:45] That’s right. Without those safeguards, our findings can be biased if staff decision-making or other factors besides time of enrollment influence whether units are assigned to the experimental or waitlist comparison group.

[Graphic – Chalkboard image with chalk writing.]

Female Narrator: [00:06:57] So, what other options are there?

Male Narrator: [00:06:59] Well, another alternative to RCTs is referred to as a Matched Case Comparison Design.

[Graphic – Doodle demonstrates “matched case comparison design.”]

With this approach, each unit assigned to receive an intervention we are testing is matched on a case-by-case basis with a comparison unit that looks as much like that intervention unit as possible. Matching occurs using a variety of previously selected demographic and case-related variables such as gender, age, maltreatment type, and placement history that are relevant to the intervention we’re testing.

[Graphic – Doodle map of the USA used to demonstrate state by state / county by county example]

Female Narrator: [00:07:26] So, an example of a program for which case matching may be a good option is a new in-home case management model directed at a well-defined population in one or a few counties in a state. Families living in other counties in the state that have characteristics similar to those families in the target counties, well, they can be matched and compared with one another over time.

Male Narrator: [00:07:44] That’s correct. But perhaps the biggest challenge to implementing Matched Case Designs involves identifying and collecting detailed data

on presenting problems, demographics, and other case characteristics that are available for both the intervention and matched comparison groups and that ensure that these two groups are in fact as much alike as possible.

[Graphic – Chalkboard with chalk text]

[Graphics – Photos of different places fade in and out to demonstrate comparison site design.]

Female Narrator: [00:08:04] One more option is a Comparison Site Design. With this approach a geographic community such as a neighborhood, city, or county, is identified that has characteristics that are similar to those of the target community in which a new program or policy is being implemented. [Graphic – two graphs show experimental results] Differences in outcomes between the experimental and comparison counties are then tracked and compared over time. So, a situation in which a comparison site design might work could involve a county that is experimenting with the new Child Welfare Reform Initiative like a new caseworker visit policy, while counties in the rest of the state continue to operate under their existing requirements.

Male Narrator: [00:08:34] However, Comparison Site Designs are generally regarded as one of the weaker approaches to developing a comparison group for a number of reasons, but mainly because it's very difficult to identify a geographic unit that's similar enough to the target community to serve as a truly valid unit of comparison. [Graphic demonstrates elimination of different sites that fail to qualify.]

[Video - Shows quick series of urban, sub-urban, and rural environments]

When we consider just about any level of geography, whether it's a neighborhood, a county, or a larger metropolitan area, so much happens that it's nearly impossible to account and control for in an evaluation.

[Graphic – Four Doodle graphs on four doodle notes, text reads - social, economic, political, and demographic.]

[00:09:02] Examples might include the social, economic, political, and demographic forces shaping that area as well as differences in the mix and types of services and resources that are available.

Female Narrator: [00:09:13] Nonetheless, there are circumstances in which a Comparison Site Design may be appropriate, and the successful implementation of a Comparison Site Design depends on several factors beyond finding the right comparison community. [Video – Man and woman shake hands.] We also need to secure the cooperation and assistance of authorities in comparison communities around data sharing and collection. [Video – Man conducting survey.] And we need to make sure that we have the ability to collect data from both communities at baseline and throughout the duration of the evaluation.

Male Narrator: [00:09:37] All right. So, where does that leave us?

[Graphics – Photographs of families with children fade in and out.]

Female Narrator: [00:09:39] Well, we know that rigorous evaluation designs matter when it comes to learning whether our child welfare practices, programs, and policies are really helping the children and families in the way we want them to. And a good comparison group can be an effective way of increasing the rigor of our design.

[Graphic – Chalkboard with chalk text]

[00:09:52] And remember that the approach we take to creating a comparison group affects how confident we can be in the results of our evaluation. We have the greatest confidence when we select the strongest design possible within the scope of our programs and within the resource constraints of our child welfare agencies.

Female Narrator: [00:10:08] Though a comparison group, in whatever form that takes, is just one part of a comprehensive and practical evaluation effort.

[Graphics – Professional teams looking happy and ready to be productive, doodle gears spell out “collaboration.”]

Male Narrator: [00:10:14] While conducting research and evaluation in child welfare settings can be challenging, it can be done successfully with the commitment of agency leadership, collaboration with other stakeholders, and by working closely with evaluation partners from the very beginning of our planning efforts.

[Graphic – Happy doodle family.]

Female Narrator: [00:10:28] And by working together we can use our evaluations to help us figure out what works and for whom so we can make smart decisions about investing in programs and services that make the biggest difference for the children and families we serve.

[On screen text- Special thanks – Elliott Graham, James Bell Associates – Krista Thomas, Children’s Bureau – Brian Deakins, Children’s Bureau.]

[On screen text – This video was created by Pal-Tech, Inc. under contract No. HHSP23337003T funded by the Children’s Bureau, Administration for Children and Families, U.S. Department of Health and Human Services. Its content does not necessarily reflect the official views of the Children’s Bureau.]

[Graphic – Child welfare evaluation virtual summit series.]

[END OF FILE]