August 2023

Using Single-Case Designs to Build Evidence

Introduction

Generating evidence of program, practice, and policy effectiveness is a high priority for Teen Pregnancy Prevention (TPP) program developers, implementing organizations, and other stakeholders. To produce evidence of program effectiveness, evaluators of TPP programs have historically employed traditional impact evaluation designs, such as randomized controlled trials (RCT) or quasi-experimental designs (QEDs). Although these designs can produce credible evidence of effectiveness, they often require large samples to be sufficiently well powered and can take several years to provide evidence of effectiveness.

Traditional approaches to designing evaluations might not be well suited for all TPP programs and with all populations. These approaches can be particularly challenging for programs serving populations that are small or difficult to recruit or retain in studies—for example, youth in foster care, tribal populations, and expectant and parenting teens. Although some evaluations can overcome small sample sizes by holding a lengthy enrollment period (for example, over multiple years), this is not always feasible. In general, policymakers and practitioners would prefer to understand program effectiveness on an expedited schedule.

As an alternative to group design evaluations, evaluators can use single-case designs (SCDs) to build evidence of the effectiveness of TPP programs, practices, and policies. Well-executed SCDs can provide evidence of effectiveness that some audiences feel is as compelling as the evidence from a well-implemented group design study. In addition, SCDs do not require large samples or random assignment, and SCDs can often be completed in less time than traditional group design studies (particularly, those that involve multiple cohorts or long-term follow-up data collection). SCD research and evidence is prevalent in psychology, medicine, substance abuse, and education, and bringing this approach to the TPP field might address a methodological need.

This brief describes SCDs and provides examples of how these designs could be implemented in the TPP field. The approaches described in this brief might be particularly useful to those who are ready for small-scale evaluations of innovations, or who want to develop initial evidence of effectiveness for interventions with a small study to set the stage for a future rigorous evaluation.

The brief is intended to illustrate these topics for a broad group of audiences, including policymakers, project directors, implementers, and evaluators. It contains numerous links to other resources and references for researchers and evaluators interested in learning more about SCD approaches.

The first section of this brief introduces SCD research principles, focusing on two designs that might lend themselves particularly well to future TPP research. The next two sections build on this foundation by describing how to apply SCDs to the TPP field and providing several examples of hypothetical studies

HHS Office of Population Affairs
Web: opa.hhs.gov | Email: opa@hhs.gov
LinkedIn: HHS Office of Population Affairs | YouTube: HHSOfficeofPopulationAffairs

and suggestions for designing a strong SCD study. The brief concludes with descriptions of the limitations of SCD research and recommendations for the future.

Introduction to SCDs

SCDs are studies that examine outcomes over time during a period before and immediately following the introduction of an intervention. Kratochwill and colleagues (2010) highlight the following key features of an SCD:

- An individual case is the unit of data for intervention and analysis. A case can be an individual or a group/cluster, such as a classroom of students or collection of patients at a clinic.
- A case serves as its own control when defining the comparison condition (a condition without an intervention). In an SCD, outcome data measured for a case during the intervention condition are compared with outcome data measured for the case during the non-intervention condition.
- Evaluators repeatedly measure the outcome variable targeted by the program, both within and across intervention and comparison conditions. These different conditions are called “phases” in SCD research.
- Within a typical SCD, researchers determine when the intervention is introduced to or withdrawn from a case, as a means to identify a causal relationship between the intervention and the outcome of interest.
- SCDs require continuous or discrete outcomes (with multiple values) that can be measured frequently over time, and that are expected to change across repeated measurements.

In addition, in nearly all SCDs, researchers present a graphical display of the outcome data to help assess the effect (or absence of an effect) of an intervention. Evaluators interpret the effectiveness of an intervention in an SCD by examining trends in outcomes in intervention and comparison phases, where changes in the level and slope of the trend indicate intervention effect. The visual presentation of data and analysis of outcomes across various study conditions to show program effectiveness differs markedly from traditional group design research, in which researchers compare differences in means across intervention and comparison groups to estimate a program effect. For additional information on SCD principles, see Kratochwill and Levin (2014), Kratochwill and colleagues (2010), and Dallery and Raiff (2014).

There are several types of SCDs, just as there are several types of group design approaches (for example, RCTs, QEDs, and regression discontinuity designs). Different SCD approaches lend themselves to different types of programs and outcomes, as discussed in the next part of this brief. The remainder of this brief focuses on two commonly used designs that lend themselves well to TPP research: (1) reversal/withdrawal designs and (2) multiple baseline designs.[1]

[1] This brief does not discuss other types of SCD approaches, including changing criterion and alternating treatments. See Kratochwill and colleagues (2010) and Dallery and Raiff (2014) for details about these approaches.

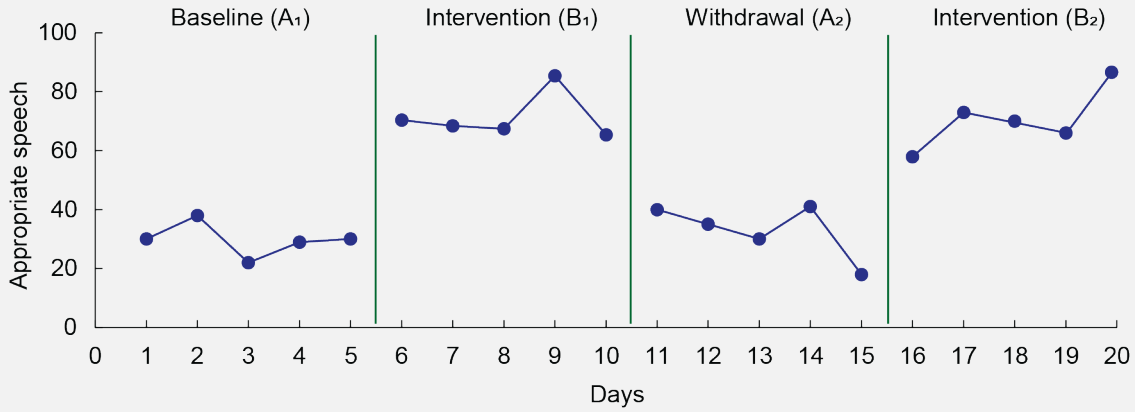
Reversal/withdrawal designs

A reversal/withdrawal design involves collecting data for a case across multiple phases, during which the intervention is systematically introduced and removed several times during the study. For example, an ABAB design is a common type of reversal/withdrawal study design that includes four phases (one for each letter): a baseline nonintervention period (A_1), an intervention period (B_1), a second nonintervention period (A_2), and a final intervention period (B_2). Reversal/withdrawal designs are particularly useful and appropriate for studies focusing on outcomes that are expected to change with the introduction of an intervention but return to a pre-intervention level when the intervention level is removed. Reversal/withdrawal designs require only one case (for example, one individual).

In an ABAB design, there are three opportunities for an intervention's effect to be observed on the outcome of interest within a single case: (1) during the transition from the initial baseline phase to the first intervention phase (A_1 to B_1), (2) during the transition from the first intervention phase to the second nonintervention phase (B_1 to A_2), and (3) during the transition from the second nonintervention phase to the second intervention phase (A_2 to B_2). If the outcome changes sufficiently during adjacent phases, the intervention can be viewed as having an effect on the outcome. See Figure 1 for an example of this design.

Consider the study of an intervention that aims to increase appropriate speech and that uses an ABAB design (Lancaster 2004). At the start of the study, the researcher measures appropriate speech (one of the outcome variables of interest) for five days. The researcher continues to measure appropriate speech for individuals during each session in all phases for the duration of the study. The first phase, during which the individuals do not receive any intervention, serves as a baseline phase (A_1). After the initial baseline phase, researchers introduced the intervention and proceed with the first intervention phase (B_1). After the first intervention phase, the intervention is withdrawn (A_2), and the individuals return to a nonintervention phase. Finally, the researcher reintroduces the intervention and continues with a final intervention phase (B_2). See Figure 1 for an illustrative presentation of the information from this study.

Figure 1. Illustrative ABAB graphical presentation



HHS Office of Population Affairs
Web: opa.hhs.gov | Email: opa@hhs.gov
LinkedIn: HHS Office of Population Affairs | YouTube: HHSOfficeofPopulationAffairs

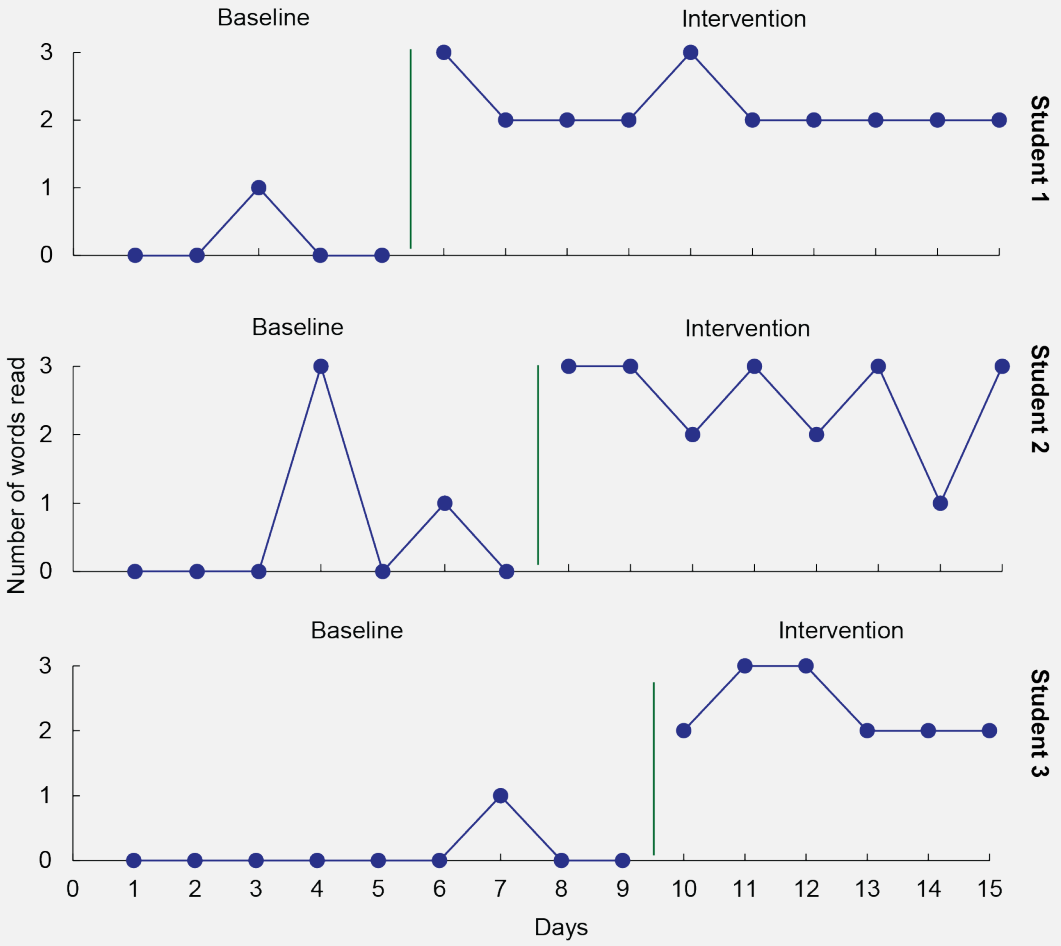
Multiple baseline designs

A multiple baseline is a different type of SCD, in which the collection of data occurs in two phases per case: a baseline phase and an intervention phase. Multiple baselines require more than one case (for example, multiple individuals). With this design, each case provides an opportunity for a single demonstration of an effect of a program. To obtain replications of a demonstration of an effect, data collection occurs across multiple cases, in which the timing of the introduction of the intervention is staggered across cases. For example, a multiple baseline with three people will provide an opportunity for three demonstrations of a program's effect on the outcome of interest. Multiple baseline designs might be more appropriate as a design when the intervention is expected to have a lasting effect on the outcomes of interest (unlike a reversal/withdrawal design, in which outcomes are expected to return to a baseline level in the absence of the intervention).

Consider a multiple baseline study that focuses on an intervention serving English language learners (youth who do not speak English at home and who have limited English proficiency) to improve their reading fluency (see Klingbeil et al. 2017 as an example of this approach).[2] At the start of this multiple baseline study, researchers measure the number of words that all English language learner students can read from a list (the outcome variable of interest). Researchers continue to measure the number of words each student can read during each class session for the duration of the study. The first several measurements, which are taken before researchers introduce the intervention to any students, serve as a common baseline phase. The researchers then introduce the intervention to each student at a different point in time. First, they introduce the intervention to Student A, then a few days later, they introduce it to Student B, and so on. Staggering the introduction of the intervention over time allows the effect of the intervention to occur at different times, decreasing the internal validity threat of an external event occurring that simultaneously causes a change in outcomes (the common concern of a traditional pre-post design without a counterfactual condition). See Figure 2 for an illustrative presentation of the information from this study.

[2] Briefly, the intervention being tested involves peer tutors working individually with English language learners to introduce unknown words, provide feedback, and intersperse known words to the English language learners as a refresher to improve word knowledge and maintain existing word fluency.

Figure 2. Illustrative multiple baseline graphical presentation



Multiple probe design

A multiple probe design is a variant of the multiple baseline design that relaxes some of the data collection requirements of a multiple baseline. In a typical multiple baseline design, repeated measurements of the outcome are necessary within each of the phases—ideally, with at least five measurements per phase to reliably document trends in outcomes. However, in some situations, it might be infeasible or too burdensome to collect outcome data this often. With a multiple probe design, there are scenarios where it is feasible to produce credible effects with fewer observations, potentially using as few as three observations across cases at baseline and additional assessments immediately before and after the introduction of the intervention across cases. See Ledford and Gast (2018) for more details about approaches for multiple probe designs.

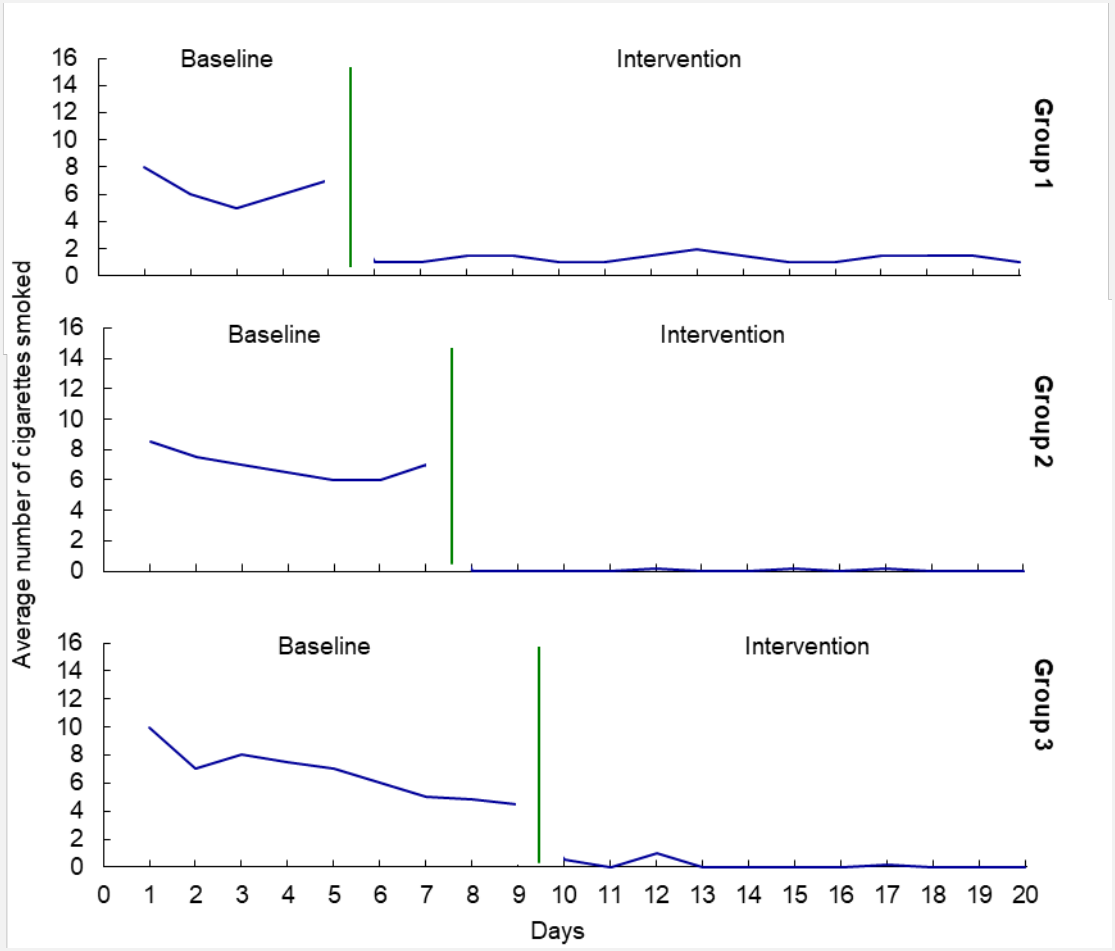
Note on unit of assignment

In the examples presented above, individuals served as the unit of assignment and analysis. However, SCDs can also be conducted where groups of individuals (for example, youth within a classroom or patients attending a clinic) can be the unit of assignment and analysis. This is true for all SCDs, including the specific designs described above previously. The same principles still apply, with data collected from all the individuals (or a representative sample) within the group, and the aggregated version of the outcome across individuals is used to show the effectiveness of the intervention.

Consider an example of a multiple baseline design adapted from Wen et al. (2019) in which groups of people are the units of assignment and analysis. [3] In this example, there were three groups, each with 10 pregnant women and each group received the intervention at different times. One group of women received the intervention first (after a short baseline period), a second group of women received the intervention next, and a final group of women received the intervention last. Women reported the number of cigarettes they smoked each day throughout the study, and the study reported the average number of cigarettes smoked each day across each group as the outcome of interest. The study showed the effectiveness of the intervention by examining the changes in the average number of cigarettes smoked between the baseline period and the intervention period for each group of women. See Figure 3 for an illustration of this study design.

[3] Wen et al. (2019) used a nonconcurrent multiple baseline design that potentially introduces other internal validity threats. See Slocum et al. (2022) for guidance on ways to avoid internal validity threats in multiple baseline studies, generally. The adapted example presents a concurrent multiple baseline design.

Figure 3. Illustrative multiple baseline graphical presentation of a study using groups as the unit of assignment



Bringing SCDs to TPP

Developing a well-designed SCD study of a TPP program requires careful planning of the types of outcomes that are appropriate for an SCD and design approaches that are most feasible and well aligned with the intervention.

SCDs and outcomes

It is important to consider the outcome or outcomes to examine whether they are appropriate for an SCD study. Like group design studies, SCD studies of TPP programs should focus on outcomes that are likely to be affected by the intervention. That is, studies should focus on outcomes that are in the programs' theory of change or logic model and can be measured reliably. In addition, SCDs require continuous or discrete outcomes (with multiple values) that can be measured frequently over time, and that are expected to change across repeated measurements.

For youth programs, there are several types of outcomes that could be appropriate for an SCD study.[4] Outcomes that might align with content for a TPP program could include knowledge, attitudes, intentions, and measures of engagement and self-efficacy that can be measured using Likert and other scales. Behavioral outcomes are also potentially relevant and well aligned with program content, and might include sexual behaviors, such as number of recent sexual partners, number of recent condom-protected sex acts, condom use skills, or minutes of app use. Given some of the nuances of SCD research, consider the following when identifying and selecting outcomes for evaluations of TPP programs:

- **Recall window.** Some outcomes will only be viable for an SCD when measured during a short recall window so they can be measured repeatedly over time without overlapping recall windows. For example, when collecting data from respondents about recent sexual activity, respondents could report on the number of sexual encounters in the past week once a week over the course of several weeks. In this example, it might not be feasible to measure the number of sexual encounters over a longer recall window (for example, a month) if the goal is to document outcomes repeatedly, unless the study data collection period allows for several months of assessments.
- **Relative permanence of outcomes.** Some outcomes are relatively permanent once acquired or achieved, such as knowledge or self-efficacy. These types of outcomes might only be appropriate for multiple baseline designs and not a reversal/withdrawal design because they will not revert to a pre-intervention level once the intervention is removed. Other common outcomes in TPP research that are relatively permanent and unchanging, such as sexual initiation or pregnancy, are also not ideal outcomes for nearly any SCD.

[4] The examples presented in this brief include a mix of outcomes currently used by Teen Pregnancy Prevention Evidence Review (for example, sexual behaviors and their consequences) and other outcomes (for example, knowledge and intentions) that are not currently eligible for review under version 7.0 of the standards (Mathematica, 2023).

- **Dichotomous outcomes.** Many TPP studies examine behavioral outcomes that can take a value of yes or no (for example, whether a respondent used a condom or birth control during a recent sexual encounter). These outcomes are often not sensitive enough to show programmatic effects in repeated measurements in an SCD in which individuals are the unit of assignment and analysis. There are a couple of exceptions to this rule. The first exception occurs when the case in an SCD is a group of people. With data from a group, the dichotomous variables can be converted into a percentage (for example, a proportion of the group that had sex without a condom in the past week). The second exception occurs when researchers measure outcomes for people repeatedly during one session, and then present the count or percentage per session. These types of continuous outcomes might be more sensitive to showing program effects.

If the intervention being studied is a change in policy or practice—for example, at the clinic or community level—consider other types of outcomes beyond those that are typically self-reported in surveys. At a clinic, it could be feasible to measure the number of referrals, number of people taking up services, number of appointments attended, and patient satisfaction rates. Evaluators could collect these data as part of the study or from existing administrative data collected as part of regular practice. Regardless of how the outcome data are obtained, they can be used to assess the effect of a policy change.

SCDs as a design approach for TPP studies

Like group designs, certain programs or contexts might lend themselves to certain SCD approaches. The most appropriate study design will depend on the program and the extent to which the TPP program, policy, and practice is expected to have a lasting effect for a given outcome of interest.

To help demonstrate this, we present a variety of examples of hypothetical TPP interventions and associated SCDs to assess their effectiveness. The examples that follow use straightforward designs and focus on interventions and outcomes for which we would expect to see an immediate effect with the introduction of the intervention. These illustrations are intended to showcase a variety of interventions, outcomes, units of assignment, and demonstrations of program effectiveness.

Illustrative examples of a reversal/withdrawal design for TPP studies

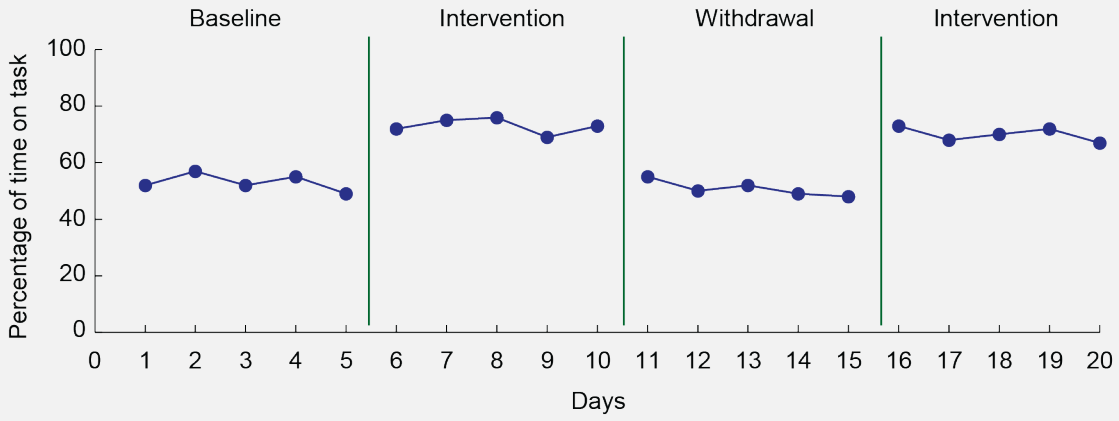
Reversal/withdrawal designs can be used to evaluate whole interventions, or components of an intervention, not expected to have a lasting effect after the intervention is withdrawn. The following two examples illustrate how a reversal/withdrawal design could be used for a TPP study.

Example 1 of a TPP reversal/withdrawal study: A new facilitation technique designed to improve student engagement

Design: Reversal/withdrawal
Intervention: New facilitation technique
Unit of analysis: Class
Measurement: Observer-recorded data
Outcome: Amount of time students are on task

A school district wants to evaluate a new facilitation technique designed to improve the amount of time students spend on task. Researchers work with a health teacher who delivers sexual health education in the district high school to test the new technique. The researchers identify a 10th-grade class to participate in the study. To start, there is an initial five-day baseline phase during which the health teacher uses their typical facilitation technique. The first intervention phase starts on the sixth day, and during this phase, the health teacher uses the new facilitation technique for five days. In the third phase, the intervention is withdrawn, and for five days, the teacher does not use the new technique. The final intervention phase includes five days during which the teacher uses the new technique again. During all phases, independent observers observe youth in the classroom daily and record their time on task. The outcome of interest for the study is the average percentage of time on task across all students in each classroom session. The data patterns in the intervention and non-intervention phases presented in Figure 4 suggest the new facilitation technique is effective. The level of time on task is markedly higher during the first intervention phase compared with the two adjacent phases with no intervention, and the amount of time on task is also higher during the second intervention phase compared with the adjacent withdrawal phase.

Figure 4. Visual analysis of reversal/withdrawal example focused on facilitation techniques to improve student achievement



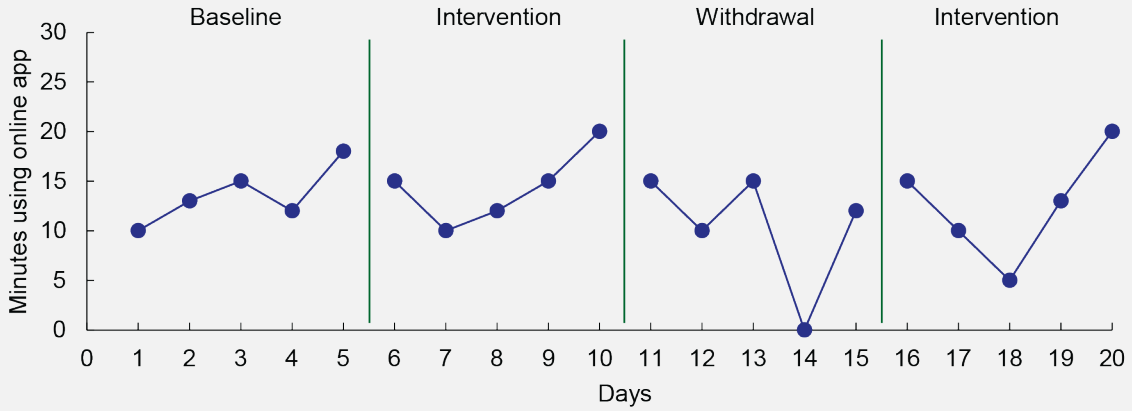
HHS Office of Population Affairs
Web: opa.hhs.gov | Email: opa@hhs.gov
LinkedIn: HHS Office of Population Affairs | YouTube: HHSOfficeofPopulationAffairs

Example 2 of a TPP reversal/withdrawal study: Text message reminders

Design: Reversal/withdrawal
Intervention: Text message reminders
Unit of analysis: Individual
Measurement: Daily app data
Outcome: Minutes of app use

A program developer wants to test the effectiveness of text message reminders, a program component designed to encourage youth to engage with an online sexual health program. Researchers identify a teenager to participate in a study of the program component. To start, there is an initial five-day baseline phase during which the teenager has access to the online program but does not receive any text message reminders. The first intervention phase starts on the sixth day, and during this phase, the teenager receives daily text message reminders for five days. In the third phase, the intervention is withdrawn, and for five days, the teenager does not receive any text message reminders. The final intervention phase includes five days during which the teenager receives text messages reminders again. Researchers rely on daily data on minutes of app use collected administratively using the app as the outcome of interest for the study. The data patterns in the intervention and non-intervention phases presented in Figure 5 do not suggest the text message reminders had an effect on the number of minutes of app use.

Figure 5. Visual analysis of reversal/withdrawal example focused on text message reminders



HHS Office of Population Affairs
Web: opa.hhs.gov | Email: opa@hhs.gov
LinkedIn: HHS Office of Population Affairs | YouTube: HHSOfficeofPopulationAffairs

Illustrative examples of a multiple baseline design for TPP studies

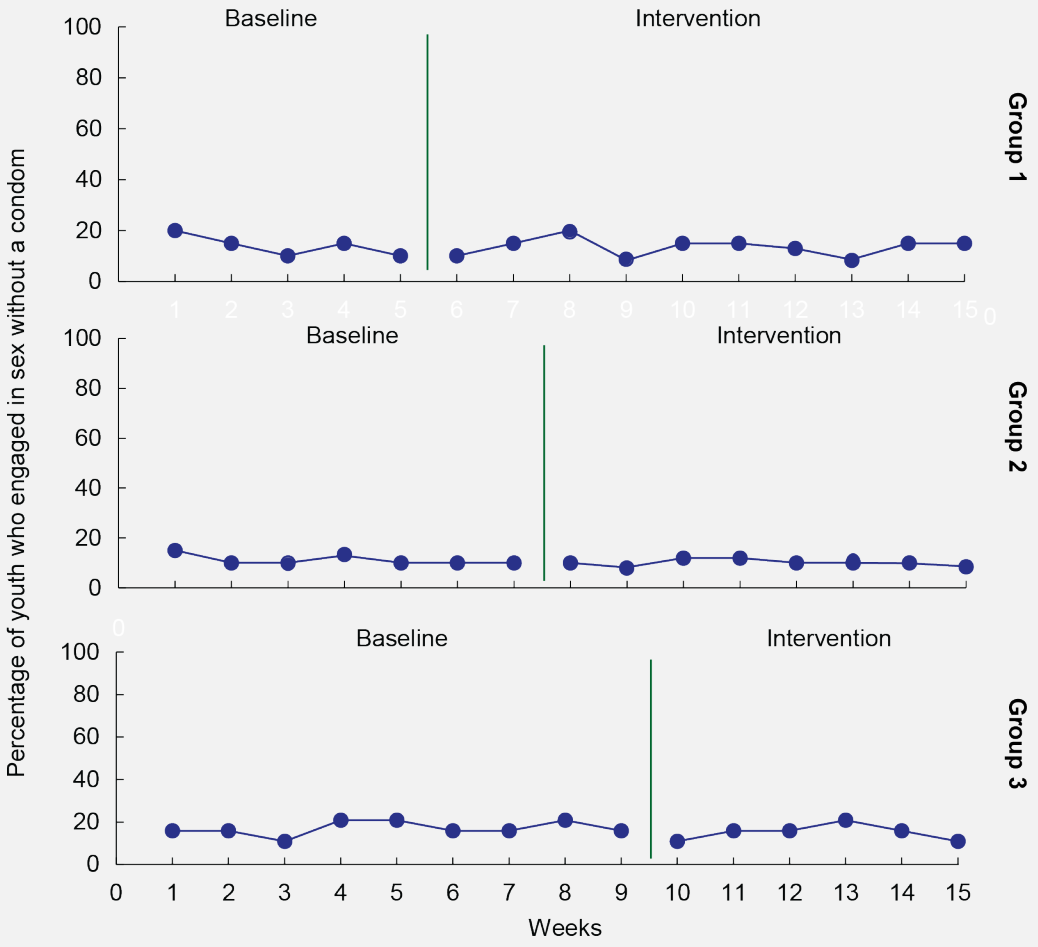
Multiple baseline designs work well to assess the effect of a whole intervention or components of an intervention expected to have a lasting effect. This could be a program, practice, or policy that is implemented with individuals, classrooms, or clinics and for which the introduction of the intervention can be purposefully staggered. The following examples show how a multiple baseline design could be used in the TPP field.

Example 1 of a TPP multiple baseline study: Sexual health online application

Design: Multiple baseline across three groups
Intervention: Sexual health online app
Unit of analysis: Groups
Measurement: Self-reported survey data
Outcome: Condom use

A program developer is interested in assessing the effectiveness of a sexual health online application on condom use. Researchers recruit three groups of teenagers to participate. All three groups start with a baseline phase, during which none of the teenagers have access to the online application. Researchers then intentionally introduce the online application to the groups of teenagers at different times. After five weeks in the baseline phase, Group 1 receives access to the online application. Two weeks later, after seven weeks in the baseline phase, Group 2 receives access to the online application. Finally, two more weeks later, after nine weeks in the baseline phase, Group 3 receives access to the online application. Starting with the first week in the baseline phase, researchers use a brief text messaging survey to collect data on condom use from all teenagers once a week throughout the 15-week study period. Condom use is the outcome of interest for this study. Researchers examine the data patterns before and after the intervention is introduced for each group. The data patterns in the baseline and intervention phases for each group of teenagers shown in Figure 6 do not suggest there is an intervention effect.

Figure 6. Visual analysis of a multiple baseline example focused on a sexual health online application



Example 2 of a TPP multiple baseline study: Community-wide initiative to encourage use of clinics and sexually transmitted infection (STI) testing

Design: Multiple baseline across four clinics

Intervention: Community-wide initiative to encourage clinic use and STI testing

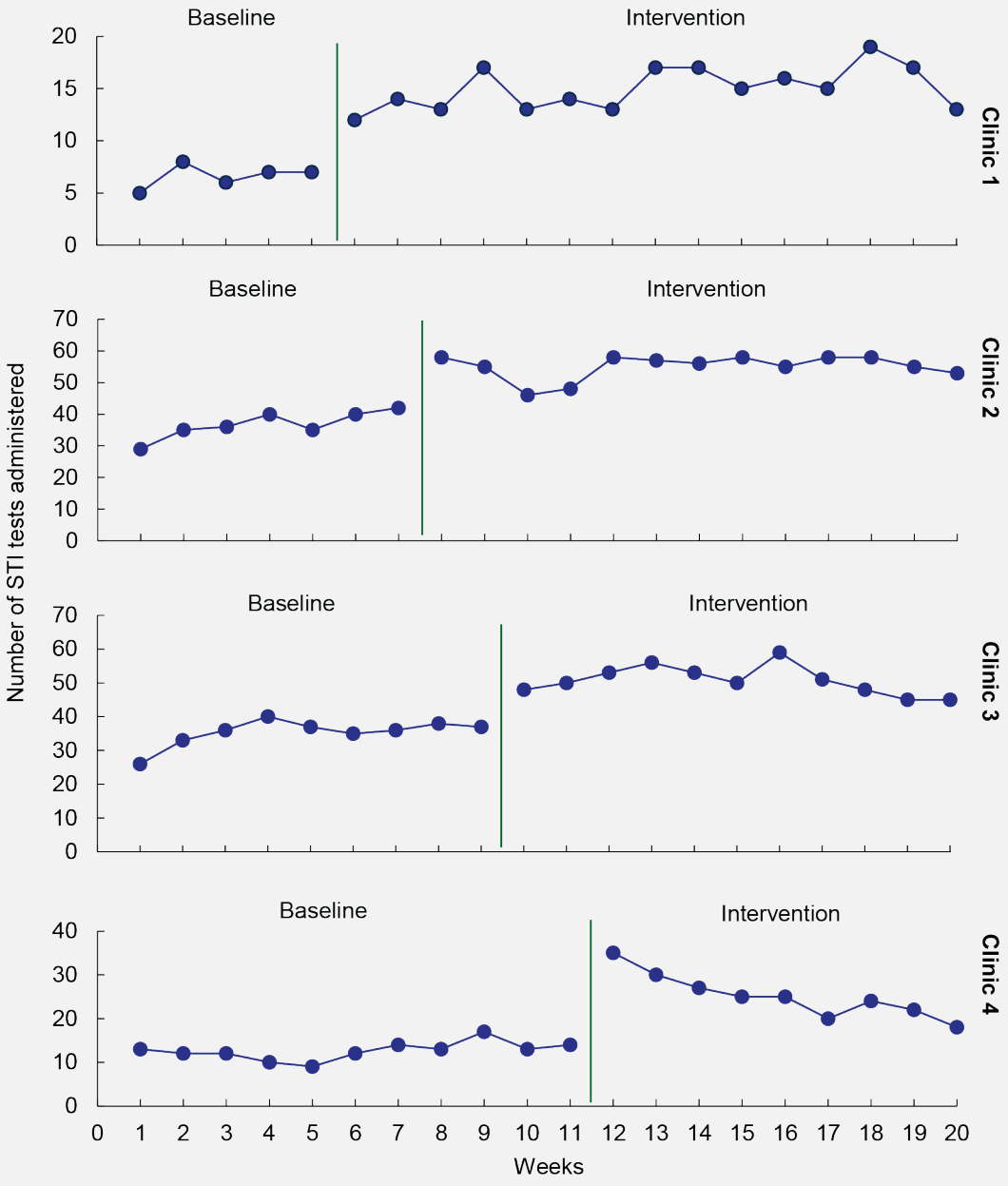
Unit of analysis: Clinics

Measurement: Administrative data

Outcome: STI tests administered

A county is interested in assessing the effectiveness of a community-wide initiative to encourage use of clinics and STI testing. Researchers identify four clinics in four different communities to participate in a study. During the baseline phase, all four clinics continue with business-as-usual practices. Researchers then intentionally introduce the intervention to each clinic at different times. After five weeks in the baseline phase, Clinic 1 starts a campaign to encourage clinic visits and STI testing. The intervention is offered to Clinic 2 in Week 7, to Clinic 3 in Week 9, and to Clinic 4 in Week 11. Researchers rely on clinic testing data records as the outcomes of interest to examine daily data on the number of STI tests administered during the 20-week study period, which includes data from the intervention and baseline phases for all participating clinics. Researchers examine the data patterns before and after each phase change. The data patterns in the intervention and baseline phases for each clinic suggest that the community-wide initiative increased the number of STI tests administered at the clinics (Figure 7). For all clinics, the number of STI tests administered is notably higher in the intervention phase than in the baseline phase.

Figure 7. Visual analysis of a multiple baseline example focused on a county-wide initiative



As noted above, the appropriate SCD design depends on the intervention being tested and the outcomes of interest. These examples above are meant to be illustrative and help showcase approaches for using SCDs to generate evidence about TPP approaches. The sections that follow provide additional information and suggestions for designing a strong SCD study.

Considerations, suggestions, and best practices

SCDs require outcome data collected repeatedly over time, so it is important to be thoughtful about how and when to measure outcomes. Here are some suggestions:

- Limit data collection during the repeated assessments to only the outcomes necessary to answer primary research question(s), such as a brief survey with a few questions.
- Before beginning the baseline sessions, use a separate survey or data collection mechanism to collect more data about the people participating in the study.
- Use two independent observers when feasible for outcomes that require observation, such as student engagement or skill assessment, and measure inter-assessor agreement to document the reliability of their assessment. Consider whether existing administrative data could be used as a potential outcome measure for the study.

As noted earlier, the Teen Pregnancy Prevention Evidence Review does not have standards for SCDs, but the What Works Clearinghouse (WWC) has standards that can be a useful resource for planning prospective SCDs. This includes requirements for the number of phases and data points needed to produce credible estimates of program effectiveness. For example, for the most compelling evidence from an SCD design, the WWC requires that reversal/withdrawal designs have at least four phases (three phase changes) and at least five data points in each phase. Similarly, for the most compelling evidence from a multiple baseline design, the WWC requires at least six phases (three phase changes) and at least five data points in each phase. The WWC standards provide requirements for design features that can be helpful to consider, including reliability standards for measuring outcomes, confounding factors, and attrition.[5]

Beyond the design and reporting recommendations from the WWC mentioned above, when reporting on SCDs, researchers should be transparent and provide sufficient information for the reader. Here are some suggestions:

- Present data graphically and report on any visual analysis in a clear and transparent way.
- Calculate and report SCD effects sizes, as this is becoming increasingly important and common (see sidebar).

Additional resources for calculating effect sizes

1. Procedural Sensitivities of Effect Sizes for Single-Case Designs with Directly Observed Behavioral Outcome Measures (journal article)
2. A web-based calculator for between-case standardized mean differences (Version 0.5.2) (web application)
3. Design-Comparable Effect Sizes in Multiple Baseline Designs: A General Modeling Framework (journal article)

[5] See the following resources for additional information on internal validity threats for SCDs:

- What Works Clearinghouse. "Key Criteria Used in WWC Reviews of Single-Case Design Research." Institute of Education Sciences, 2017. https://ies.ed.gov/ncee/wwc/Document/264.
- What Works Clearinghouse. "What Works Clearinghouse Procedures and Standards Handbook, Version 5.0." National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2022. https://ies.ed.gov/ncee/wwc/Handbooks.

- Report reliability for measures, including inter-assessor agreement data for observations, and other reliability measures for survey data.
- Provide information on implementation context and study participants. This could include age or grade, biological sex, and race and ethnicity; location; the type of clinic or school involved in the evaluation; and training for staff implementing the intervention.

Limitations

SCDs can provide credible evidence of effectiveness, but there are some caveats and limitations to this approach, relative to the more traditional group design.

- **There is a lack of awareness of the approach in the TPP field.** Historically, the TPP field has relied solely on RCTs and QEDs to show effectiveness of programs and practices. Since 2010, when the Teen Pregnancy Prevention Evidence Review was first funded, the standards have focused on these two study designs because they were considered the only approaches that could generate credible evidence of effectiveness. Increased awareness and understanding of the value and credibility of the findings of SCDs could lead to increased use and acceptance.
- **Replication is necessary to ensure confidence in the findings.** SCDs often involve small samples and lack generalizability. A single SCD can show credible evidence of effectiveness for a program; however, findings from a single SCD might only be relevant for the particular sample and population that was studied. Replicating the findings across multiple studies is seen as best practice for increasing the confidence in the finding (Horner and Spaulding 2010).
- **Certain outcomes relevant to the TPP field are not ideal for SCD research.** TPP research often focuses on sexual behavior outcomes and their consequences (pregnancy and sexually transmitted infections). Although some behavioral outcomes can be appropriate for an SCD study (for example, number of recent sex acts or the proportion of recent sex acts that were condom-protected), many traditional sexual behavior measures, such as “ever sexually active” or “ever pregnant” might not be appropriate outcome measures under SCD research because they are permanent once achieved, and these dichotomous outcomes may not be sensitive enough to capture intervention effects.
- **Data collection can be burdensome for participants.** SCDs require frequent, repeated measurement of or responses from participants to generate the outcome data required for showing the effects of a program. This might be a particular burden to people participating in the study and the organizations collecting the data. The amount of time and resources required for data collection across the many assessment points in an SCD study could be similar to the amount of time and resources spent completing a small number of surveys for a traditional group design study.

Conclusion

SCDs can be useful for TPP evaluators to draw on in certain contexts. These designs can generate evidence of effectiveness for TPP programs, practices, and policies. This is particularly true when group designs are not feasible or appropriate because of small sample sizes or concerns about withholding services from a comparison group. For anyone who wants to create foundational information about the effectiveness of a program or practice, and who wants to use a more rigorous design than a pre-post outcome study (without a counterfactual), an SCD might be a good approach to consider. As with any study, SCDs require careful planning that aligns well with the program being evaluated.

References

Dallery, J., and B.R. Raiff. "Optimizing Behavioral Health Interventions with Single-Case Designs: From Development to Dissemination." *Translational Behavioral Medicine,* vol. 4, no. 3, 2014, pp. 290–303. doi:10.1007/s13142-014-0258-z.

Horner R., and S. Spaulding. "Single-Case Research Designs." In *Encyclopedia of Research Design,* edited by N.J. Salkind. Thousand Oaks, CA: SAGE, 2010.

Klingbeil, David A., Mariola Moeyaert, Christopher T. Archer, Tatenda M. Chimboza, and Scott A. Zwolski, Jr. "Efficacy of Peer-Mediated Incremental Rehearsal for English Language Learners." *School Psychology Review,* vol. 46, no. 1, March 2017, pp. 122–140.

Kratochwill, T.R., J. Hitchcock, R.H. Horner, J.R. Levin, S.L. Odom, D.M. Rindskopf, and W.R. Shadish. "Single-Case Designs Technical Documentation." What Works Clearinghouse, 2010.

Kratochwill, T.R., and J.R. Levin (eds.). *Single-Case Intervention Research: Methodological and Statistical Advances.* American Psychological Association, 2014. https://doi.org/10.1037/14376-000.

Lancaster, B.M., L.A. LeBlanc, J.E. Carr, S. Brenske, M.M. Peet, and S.J. Culver. "Functional Analysis and Treatment of the Bizarre Speech of Dually Diagnosed Adults." *Journal of Applied Behavior Analysis,* vol. 37, no. 3, 2004, pp. 395–399. https://doi.org/10.1901/jaba.2004.37-395.

Ledford, J., and D. Gast. *Single Case Research Methodology: Applications in Special Education and Behavioral Sciences 3rd Edition*. Routledge, 2018.

Mathematica. (2023). "Identifying Programs That Impact Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors Review Protocol Version 7.0."

Slocum, T.A., S.E. Pinkelman, P.R. Joslyn, and B. Nichols. "Threats to Internal Validity in Multiple-Baseline Design Variations." *Perspectives on Behavior Science,* 2022, pp. 1–20.

Wen, X., R.D. Eiden, F.E. Justicia-Linde, Y. Wang, S.T. Higgins, N. Thor, A. Haghdel, A.R. Peters, and L.H. Epstein. "A Multicomponent Behavioral Intervention for Smoking Cessation During Pregnancy: A Nonconcurrent Multiple-Baseline Design." *Translational Behavioral Medicine,* vol. 9, no. 2, 2019, pp. 308–318. doi:10.1093/tbm/iby027.