Common Evaluation Designs in Medical Education III
Dario M. Torre, MD, MPH, PhD, associate professor of medicine and vice chair of education, Drexel University College of Medicine; Allison Ferris, MD, associate program director, Internal Medicine, Drexel University College of Medicine; Barbara Daley, PhD, interim associate dean, College of Nursing, and professor, Adult and Continuing Education Program, University of Wisconsin–Milwaukee; and Steven J. Durning, MD, PhD, professor of medicine and pathology and director, Intro to Clinical Reasoning Course, The Uniformed Services University of the Health Sciences

The goal of this Last Page, a follow-up of two previous Last Pages,1,2 is to describe one more evaluation design—known as interrupted time series—and to explain its advantages and challenges. We also provide an example and mention the internal threats to the design’s validity, such as testing, maturation, and reactive testing.

Interrupted Time Series Design (Longitudinal and Successive)

**Interrupted** defines the time when the program is implemented and administered. The interrupted time series design can include:

1. **Longitudinal**: same group measured multiple times at regular intervals
2. **Successive**: different yet equivalent groups (Ea and Eb) measured multiple times at regular intervals

**Pros:**
- Very helpful to study the longitudinal effects of a program
- Useful if the effect of the intervention is occurring across time
- Helpful when treatment has to be given to all students and cannot be withheld for periods of time
- The successive time series design allows longitudinal evaluation of a program when courses are of short duration and new students start the rotation every 4 to 6 weeks (e.g., for a subinternship) as part of their curriculum

**Cons:**
- Requires time
- Need for multiple equivalent measurements across time
- Reactive testing (i.e., the content of the pretest may alert students to focus on specific topics) is a threat as a result of the repeated testing
- More complex statistical procedures may be needed to correctly interpret results
- History (occurrence of other events during the time of intervention that may affect outcome) and maturation (natural growth of learners independent of the intervention) may be difficult to assess
- Randomization may be needed in the successive time series design to ensure group equivalency

**Examples:**

**Longitudinal:** A group of medical students enrolled in a 14-week rotation is tested with an MCQ knowledge-based test every 2 weeks for 6 weeks on the topic of CHF. At 6 weeks, a Web-based interventional instructional program is administered to the group of students for 2 weeks; after the program is no longer administered, a similar MCQ knowledge-based test is administered every 2 weeks to the same group on the topic of CHF for an additional 6 weeks. Test results are then compared and analyzed for trends pre and post program implementation.

**Successive:** A group of medical students enrolled in a 6-week rotation is tested with an MCQ knowledge-based test every 2 weeks on the topic of CHF. At 6 weeks, a Web-based interventional instructional program is administered to the group of students for 2 weeks; after the program is no longer administered, a different yet equivalent group (randomization may be needed to ensure the two groups are equivalent) begins a new 6-week rotation, and a similar MCQ knowledge-based test is administered every 2 weeks to the new group on the topic of CHF for an additional 6 weeks. Test results of the two equivalent groups of students are then compared and analyzed for trends before and after the intervention.

**Legend**
- E = experiment; C = control group; I = instructional program; T = test, measurement, or observation

**References:**

**Additional resources:**

**Author contact:** dario.torre@usuhs.edu

CHF indicates congestive heart failure; MCQ, multiple-choice question.