CHAPTER SEVENTEEN

Data Management and Evaluation of Translation

Martha Sylvia

What gets measured
gets improved.
—Peter Drucker

The Doctor of Nursing Practice (DNP) Translates evidence into practice on a systems, large-scale level, with the result being an improvement in quality of health-related outcomes (American Association of Colleges of Nursing, 2006; Mundinger, Starck, Hathaway, Shaver, & Woods, 2009). Improvement implies execution of *defining, measuring, analyzing,* and *demonstrating* a positive change in outcomes that is carried out through monitoring and evaluation of the translation project.

EVALUATION APPROACH

Understanding the impact of translation requires a flexible, grounded, iterative, contextualized, and participatory approach and can be referred to as "learning evaluation" (Balasubramanian et al., 2015). While the project is in the implementation phase, *monitoring* of program outcomes is taking place. During the monitoring phase, data is gathered, analyzed, and interpreted to inform project implementation and to improve processes. *Evaluation* of the project takes place when data collection and project implementation are completed and these adhere to a protocol established during the project planning phase. A rigorous monitoring protocol leads to an evaluation that is streamlined and efficiently executed with preconceived expectations of results. Predetermined outcome measures are used for both the monitoring and evaluation phases.

Principles of a learning evaluation include:

- Gathering data to describe the changes made and how the changes are implemented
- Collecting relevant process and outcome data

- Assessing multilevel factors affecting implementation, process, outcome, and transportability
- Assisting organizations in using data for continuous quality improvement
- Operationalizing common measurement strategies to generate transportable results (Balasubramanian et al., 2015)

■ EVALUATION RIGOR

Evaluation of translation projects must be rigorous, scholarly, and most importantly provide convincing evidence that changes in outcomes are a result of the intervention project that translates evidence into practice (Terhaar, Taylor, & Sylvia, in press). Flexible, scientifically rigorous study designs are necessary to evaluate translation projects (Balasubramanian et al., 2015). While it is often not feasible for quality improvement/translation projects to use randomized control trials mainly due to time, financial, and equitability constraints, evaluation of translation can use the tools of research to attain a high level of rigor (Terhaar, 2014).

Internal Validity and Generalizability

In research, internal validity refers to the extent to which a study avoids confounding or the possibility that a factor other than the intervention itself contributed to a change in outcomes. Strong internal validity allows a higher level of confidence that a desired change in outcomes is due to the intervention. Generalizability or external validity is the degree to which the results of a study can be generalized to similar settings and circumstances (Polit & Beck, 2012). In applying these same concepts to translation projects, internal validity is of the utmost importance. Generalizability is not expected; however, transportability, the ability to implement the translation intervention in a different setting, is important. Evaluation of translation projects includes an assessment of internal validity and transportability (Balasubramanian et al., 2015).

■ PREPARING FOR EVALUATION

Evaluation Plan

Planning for evaluation takes place during the project planning phase. The original problem statement informs project aims and subsequent objectives and measures. For instance, if the original problem statement purported a high readmission rate for a population of patients with congestive heart failure, then the aims, objectives, and measures would focus on reducing readmission rates as a result of the translation project.

There may be nuances in the definitions of aims and objectives for a translation project. For instance, in some cases aims may be written in broader terms that describe aspirations for the translation project overall, such as, "improving the health of patients with congestive heart failure (CHF)." Aims may also be written more specifically as in the use of the SMART framework; specific, measurable, achievable, realistic,

and time-phased (Centers for Disease Control and Prevention, 2015). If the aims are in more general, overarching terms, it is usually the objectives that follow each aim that describe more specifics about translation achievements. Regardless of terminology, either the aims or the objectives should describe measurable results. For example, "within 3 months of implementation, patients with CHF who received the intervention will have a 10% lower readmission rate than patients with CHF who did not receive the intervention."

Evaluation takes place in seven phases: (a) planning, (b) data collection, (c) data cleansing, (d) data manipulation, (e) exploratory analysis, (f) outcomes analysis, and (g) dissemination, reporting, and presentation (Sylvia & Terhaar, 2014).

OUTCOMES

Measuring Successful Implementation

Multilevel factors affect successful implementation of evidence-based health innovations and can be measured in constructs that are important for evaluation, derived from the field of dissemination and implementation science (Chaudoir, Dugan, & Barr, 2013). These factors, their description, and measurement examples are described in Table 17.1.

TABLE 17.1 Multilevel Evaluation Factors

Factor	Description	Measurement Example
Structural/Outer Setting	Constructs representing the external structure of the broader sociocultural context or community within which a specific organization is situated	Description of the political climate, social climate, policy, economic climate, or infrastructure (workforce, transportation)
Organizational/Inner Setting	Constructs that represent aspects of the organization in which an innovation is being implemented	Leadership effectiveness, culture, innovation climate (extent to which organization values and rewards evidence-based practice or innovation)
Innovation	Constructs that represent aspects of the innovation that will be implemented	Adaptability, complexity, cost, advantage of the choice of one innovation above another/relative advantage, quality of evidence supporting the innovation's efficacy/evidence strength and quality, process for the innovation/design quality and packaging

(continued)

TABLE 17.1 Multilevel Evaluation Factors (*continued*)

Factor	Description	Measurement Example
Provider Level	Constructs that represent aspects of the individual provider/clinician who implements the innovation with a patient or client	Attitudes toward evidence-based practice, perceived behavioral control for implementing the innovation
Patient Level	Constructs representing patient characteristics or characteristics of the recipient of the intervention	Health-relevant beliefs, motivation, personality traits, health literacy, trust of medical practices
Transferability	Constructs that determine the ability of an intervention to be transferred to another setting and can be used in two situations: (1) prior to deciding on transferring an intervention to a new setting; (2) when evaluating the replica intervention to help explain the effects of the intervention that has been replicated	The ASTAIRE tool, which has four categories of analyzing the transferability with 23 structured criteria • Population • Environment • Implementation • Support for transfer
Patient/Client Outcomes	These are outcomes that the intervention is meant to directly or indirectly impact	Functional status, quality of life, perceived health status

ASTAIR, *AnalySe de la Transférabilité et accompagnement à l'Adaptation des Interventions en pRomotion de la santE.*

Sources: Cambon, Minary, Ridde, and Alla (2013); Chaudoir et al. (2013); Lewis et al. (2015).

Multiple instruments exist for specific measures in each of the factors (Chaudoir et al., 2013; Lewis et al., 2015).

Value

The concept of value in health care varies depending on the perspective of the stakeholder. Patients may describe value in terms of the quality of their relationship with their provider, meeting personal goals, or living normal lives. Providers may suggest that value improvement means developing diagnostic and treatment tools that offer them more confidence in the effectiveness of the services they offer. Employers discuss value improvements in terms of keeping workers and their families healthier and more productive at lower costs, while health insurers perceive value as the delivery of interventions that have a substantial evidence base for effectiveness and efficiency. Finally, health product innovators value improved products for patients that are more profitable, differentiated, and innovative (Institute of Medicine, 2009).

The "triple aim" goals, as outlined by Berwick and colleagues through the Institute of Health Improvement (IHI), offer an established and accepted framework for determining value in health innovation with three goals:

1. Improving the individual experience of care
2. Improving the health of populations
3. Reducing the per capita costs of care for populations (Berwick, Nolan, & Whittington, 2008)

A suggested set of outcome measures designed to address the three domains are as follows:

- Mortality
  - Years of potential life lost; life expectancy; standardized mortality ratio
- Health and functional status
  - Health Related Quality of Life (HRQOL-4) or multidomain assessment such as the Patient Reported Outcomes Measurement Information System (PROMIS)
- Health life expectancy
- Disease burden
  - Incidence and/or prevalence of major chronic conditions
  - Morbidity burden
- Behavioral and psychological factors
  - Behaviors such as smoking, alcohol consumption, physical activity, and diet
  - Physiological factors such as blood pressure, body mass index, cholesterol, and blood glucose
- Experience of care
  - Consumer Assessment of Healthcare Providers and Systems (CAHPS)
  - Likelihood to recommend
  - Measures based on the Institute of Medicine's six aims for improvement – Safe, effective, timely, efficient, equitable, and patient-centered
- Cost
  - Total cost
  - Utilization of services, that is, hospital and emergency department (Stiefel & Nolan, 2012)

DESIGNS FOR MEASURING TRANSLATION INTERVENTION OUTCOMES

Certain analysis designs are commonly used for measuring the impact of evidence translation projects. Each of these designs are described with the appropriate statistics.

Qualitative Analysis

Qualitative analysis can be used to describe the conditions in which the translation project takes place, the types of changes that are taking place, how the changes are made, facilitators and barriers to the process, and key stakeholders' experiences with the translation. To

carry out this type of analysis, key documents are collected, observations and interviews are performed, and diaries can be implemented (Balasubramanian et al., 2015).

Qualitative analysis in the form of focus groups is often used to determine the results of translation with participants in whom the intervention is planned to impact. For instance, in an intervention designed to improve bonding between parents and newborns, focus groups may be formed to obtain more detailed responses about the experience of participating in the intervention from the parents. Information gleaned from the focus group could be used to describe the context for certain observed newborn behaviors.

Descriptive Design

Descriptive evaluation designs are used most often when the cohort receiving the intervention is small and/or when there is no comparison group. This is most often used in feasibility and pilot translation projects. Descriptive designs are the sole way to present the results of a project that does not use any type of comparisons; but the same type of descriptive information is used as the initial way to present the results of a project that does use a comparison group.

A descriptive design provides information about a group receiving the translation intervention and about the outcome metrics. Information about a group might include characteristics that are demographic such as age, gender, and ethnicity; geographic such as zip code, city, county, and state; psychosocial such as marital status or social support; or financial such as income level. The outcome metric would be presented without any comparisons within the project itself but may be compared to similar outcomes in the literature. When a translation intervention is meant to impact an event, for example, when trying to decrease the wait time for an emergency department visit, the descriptive information is about that certain event as opposed to describing the characteristics of a group. Descriptive information for an event might include the date and time; staff involved in the event; and the patient involved in the event.

Univariate statistics are used in descriptive evaluation designs; mean, median, mode, and proportions are used to present aggregated information about the group or set of events exposed to the intervention. Additionally, distributions of the values for an outcome may be presented using histograms and box plots. Because there is no comparison group, there is no independent or dependent variable.

Pre–Post Measurement in One Cohort

Paired grouping evaluation designs measure the same outcome in one cohort over two or more points in time. The comparison is within the same cohort with one measurement representing a time period where the cohort was not exposed to the intervention and the comparison measurement representing a time period where the cohort was exposed to the intervention. The independent variable is the exposure to the intervention and is represented by the time period (pre or post). The dependent variable is the outcome.

In the case where the dependent outcome variable is continuous (e.g., weight, diastolic blood pressure), the mean, standard deviation, and distributions are used to describe the outcome. A paired *t*-test is used to determine whether the differences in means are statistically significant. When a dichotomous outcome variable is used (e.g., weight loss achieved: yes/no, diastolic blood pressure less than 80: yes/no), then proportions are used to describe the outcome and McNemar's test of proportions is

the appropriate test to determine statistically significant differences. In the case of an ordinal outcome variable (e.g., Likert scales), proportions are still used to describe the outcome and the Wilcoxon signed-rank test is used to determine statistically significant differences.

Comparison of Two Different Cohorts

Different cohorts can be compared either in the same time frame of intervention delivery or in different time periods. When comparing cohorts in the same time period, a cohort is selected, using inclusion and exclusion criteria, to receive the intervention and is referred to as the "intervention cohort." A second cohort is selected using the same inclusion and exclusion criteria but does not receive the intervention and is referred to as the "comparison cohort."

Alternatively, a comparison cohort may be selected from a time period previous to the time of the delivery of the intervention, sometimes referred to as "causal comparative design" (Brewer & Kubn, 2013). When selecting a cohort from a previous time period, the same consideration must be given to inclusion and exclusion criteria as when selecting a comparison cohort from the same time period. Additionally, special consideration should be given to the impact of time on the outcomes. For instance, when measuring costs as an outcome, changes in policy or inflation rates could account for changes in cost aside from the intervention itself and could have a negative or positive impact on the outcome.

Because translation projects do not commonly use randomization to assign individuals to intervention and comparison cohorts, careful attention is given to determining whether the groups differ on known characteristics other than receipt of the intervention. An evaluation of the differences in known characteristics such as age, gender, ethnicity, and so on between groups is necessary to determine if there are reasons other than the intervention that may account for the realized changes in outcomes. This is accomplished by testing for statistical significance between groups on these characteristics. For instance, if the mean difference in age between the intervention group and the comparison group is statistically significant, then consideration of age as a confounder is necessary when statistically testing the differences in the outcome between groups (Sylvia & Murphy, 2014).

When describing the characteristics of each cohort and the outcomes, univariate statistics are used. Proportions and means are used for categorical and continuous variables, respectively, as previously described in the descriptive design section. Additionally, bivariate statistics are used to determine whether the differences in cohort characteristics and in the outcomes are statistically significantly different between groups. The independent *t*-test for continuous variables and the chi-square test for dichotomous and categorical variables are most commonly used. If confounding variables are discovered in the evaluation of the differences in characteristics between groups, then multivariate analysis, most commonly using logistic or linear regression, is used to determine statistically significant differences in outcomes between groups due to the translation intervention.

Ongoing Monitoring Using Run Charts and Statistical Process Control

Run charts and/or statistical process control are designs used to understand the outcomes of implementation in close to real time in order to continuously inform decisions and improve processes (Sherry, 2014). These designs allow for the distinction between *natural*

cause variation, which occurs naturally due to individual differences, temporal factors (e.g., seasonality), or other factors and *special cause variation*, which occurs because of the intervention (Carey, 2003).

The run chart is used to plot data points of a process over time and displays each point chronologically, with a median line delineated on the graph. A "run" is defined as one or more consecutive points that are on the same side of the median line. Statistical process control is a more sensitive and powerful way of displaying data over time that involves taking various methods for monitoring processes and combining them to determine whether changes seen in the data over time are statistically significantly different from normal variation (indicative of special cause variation). The charts are useful for monitoring processes, identifying early signs of correlation between processes and outcomes, identifying differences across groups, or aiding in self-management interventions (Marsteller, Huizinga, & Cooper, 2013).

Once control charts are plotted, there is typically a set of tests to determine significance. The following tests are usually used to test for significance on statistical process control (SPC) charts (Benneyan, Lloyd, & Plsek, 2003):

- One point outside the upper or lower control limits
- Two of three successive points more than 2 sigma from the mean on the same side of the center line
- Four out of five successive points more than 1 sigma from the mean on the same side of the center line
- Eight successive points on the same side of the center line
- Six successive points increasing or decreasing (a trend)
- Obvious cyclical behavior

See Exemplar 21.3, "Rapid Access to Tertiary Care: Mitigating Barriers Impacting Clinical, Financial, and Operational Outcomes," by Scott Newton in which a run chart to report the outcomes of a translation project is used.

Other Designs

Other design elements can be incorporated with more sophistication to adjust for biases that may be inherent due to the comparison and intervention cohort selection process; the correlation of multiple time-dependent measurements of the same outcome in the same cohort; and non-normal distributions of outcome data. A statistician should be consulted in these instances.

Power Analysis for Determining Adequate Numbers

Power analysis can be a great tool for determining, during the translation project planning phase, the number of individuals (or events) needed in each of the comparison groups to find statistically significant differences in outcomes if they do exist. Using power analysis for sample size determination adds to the internal validity of the overall translation project. Information needed to use power analysis includes (a) the statistical significance level or the percentage chance that the intervention is determined to be successful when it actually is not (usually 0.05% or 5%); (b) the level of statistical

power or the chance that the intervention is determined not to be successful when it actually is (usually 0.80% or 80%); (c) the degree of difference in the outcome between the intervention and comparison groupings (i.e., mean or proportion expected in the outcome metric for each group); and (d) the appropriate statistical test (Sylvia, 2014). Many sample size calculators are available online, with one favorite by Russ Lenth (www.stat.uiowa.edu/~rlenth/Power).

CONCLUSION

The evaluation of translation is an integral part of translating evidence into practice. A sound, rigorous evaluation plan developed during the project planning phase will pave the way to solid outcomes that describe success of the intervention on multiple levels for multiple stakeholders, of which the patient is the utmost important one.

REFERENCES

American Association of Colleges of Nursing. (2006). *The essentials of doctoral education for advanced nursing practice*. Washington, DC: Author.

Balasubramanian, B. A., Cohen, D. J., Davis, M. M., Gunn, R., Dickinson, L. M., Miller, W. L., . . . Stange, K. C. (2015). Learning evaluation: Blending quality improvement and implementation research methods to study healthcare innovations. *Implementation Science, 10*(31), 1–11.

Benneyan, J., Lloyd, R., & Plsek, P. (2003). Statistical process control as a tool for research and healthcare improvement. *Quality and Safety in Health Care, 12*, 458–464.

Berwick, D., Nolan, T., & Whittington, J. (2008). The triple aim: Care, health, and cost. *Health Affairs, 27*(3), 759–769.

Brewer, E. W., & Kubn, J. (2013). *Causal-comparative design*. Retrieved from Encyclopedia of Research Design: http://smo.sagepub.com/view/encyc-of-research-design/n42.xml

Cambon, L., Minary, L., Ridde, V., & Alla, F. (2013). A tool to analyze the transferability of health promotion interventions. *BMC Public Health, 13*(1184), 1–10.

Carey, R. G. (2003). *Improving healthcare with control charts: Basic and advanced SPC methods and case studies*. Milwaukee, WI: ASQ.

Centers for Disease Control and Prevention. (2015, March 2). *Communities of practice*. Retrieved from Develop SMART Objectives: http://www.cdc.gov/phcommunities/resourcekit/evaluate/smart_objectives.html

Chaudoir, S. R., Dugan, A. G., & Barr, C. H. (2013). Measuring factors affecting implementation of health innovations: A systematic review of structural, organizational, provider, patient, and innovation level measures. *Implementation Science, 8*(22), 1–20.

Institute of Medicine. (2009). *Value in health care: Accounting for cost, quality, safety, outcomes, and innovation*. Washingdon, DC: National Academy of Sciences.

Lewis, C. C., Stanick, C. F., Martinez, R. G., Weiner, R. G., Kim, M., Barwick, M., & Comtois, K. A. (2015). The society for implementation research collaboration instrument review project: A methodology to promote rigorous evaluation. *Implementation Science, 10*(1), 2.

Marsteller, J., Huizinga, M., & Cooper, L. (2013). *Statistical process control: Possible uses to monitor and evaluate patient-centered medical home models*. Rockville, MD: Agency for Healthcare Research and Quality.

Mundinger, M., Starck, P., Hathaway, D., Shaver, J., & Woods, N. (2009). The ABCs of the doctor of nursing practice: Assessing resources, building a culture of clinical scholarship, curricular models. *Journal of Professional Nursing*, 25(2), 69–74.

Polit, D. F., & Beck, C. T. (2012). *Generating and assessing evidence for nursing practice* (9th ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Sherry, M. (2014). Ongoing monitoring. In M. Sylvia & M. Terhaar (Eds.), *Clinical analytics and data management for the DNP* (pp. 189–214). New York, NY: Springer Publishing Company.

Stiefel, M. S., & Nolan, K. (2012). *A guide to measuring the triple aim: Population health, experience of care, and per capita cost*. Cambridge, MA: Institute for Health Improvement.

Sylvia, M. (2014). Basic statistical concepts and power analysis. In M. Sylvia & M. Terhaar (Eds.), *Clinical analytics and data management for the DNP* (pp. 8–18). New York, NY: Springer Publishing Company.

Sylvia, M. L., & Murphy, S. (2014). Outcomes data analysis. In M. Sylvia & M. Terhaar (Eds.), *Clinical analytics and data management for the DNP* (pp. 141–166). New York, NY: Springer Publishing Company.

Sylvia, M., & Terhaar, M. (2014). An approach to clinical data management for the doctor of nursing practice curriculum. *Journal of Professional Nursing, 30*(1), 56–62.

Terhaar, M. (2014). Introduction to clinical data management. In M. Sylvia & M. Terhaar (Eds.), *Clinical analytics and data management for the DNP* (pp. 1–6). New York, NY: Springer Publishing Company.

Terhaar, M., Taylor, L., & Sylvia, M. (in press). The doctor of nursing practice: Shifting from start-up to impact. *Education Perspectives*.