International Journal of Public Health and Safety

Mobley et al., Int J Pub Health Safe 2016, 1:2

Commentary Open Access

Epidemiological Principles in Public Health Program Evaluation

Sandra Mobley, Suzanne Dixson Thomas*, Jodi L Hudgins, Donald E Sutherland and Sandra B Inglett

Department of Biobehavioral Nursing, Augusta University, Georgia, USA

***Corresponding author:** Suzanne Dixson Thomas, Department of Biobehavioral Nursing, Augusta University, 300 Gardners Mill Court, Augusta, Georgia 30907, USA, Tel: 7066316140; Fax: 7066510250; E-mail: SUTHOMAS@augusta.edu

Received: July 02, 2016; **Accepted:** August 23, 2016; **Published:** August 31, 2016

Commentary

Public health program evaluation has certain limitations that are addressed with principles of epidemiology. Evaluation of service programs is limited to non-randomized designs and research techniques [1]. TREND (Transparent Reporting of Evaluations with Nonrandomized Designs) checklist [2] provides a useful outline based upon epidemiological principles to structure studies with nonrandomized designs. Service populations are defined within a geopolitical population; they are targeted for beneficial health changes but measurements rarely show direct impact upon population indices. Impact estimates are strengthened by use of internal and external controls that support validity [3].

Stakeholders must be identified prior to program evaluation and must be informed and consulted at every stage of the process. They include funding agency staff, local community leaders, and program leaders, and most of all persons from the population served. Funding and policy decisions are based upon program evaluations. It is ideal to have a program evaluator who is not a stakeholder to prevent any question of conflict of interest. The evaluator's commitment to ethics and integrity in the research process must underlie any research performed [4].

The purpose of program evaluation is to measure the impact of services upon the target population. The study sample may be the entire target population or a logical sub-set. Definition of the study sample begins with an understanding of the type, level, and structures of service. It is likely that many sub-population groups will be represented in the sample and their intermediate and final outcomes will be described with data [5].

Bias is defined as a systematic tendency that prejudices outcomes [6]. Biases are important detriments to evaluation that can undermine reasonable attempts to measure outcomes of interventions. Some potential sources of bias and how they may be controlled are included below. The long list of potential sources of bias and how their effects could be diminished is beyond the scope of this brief commentary. The reader is directed instead to references listed herein as starting points in the management of bias to protect validity [3,6-9].

Avoiding the impact of history

The study period must be determined when there is little evidence of historical change. Changes such as new strategies for service, new measurements, new data collection forms, and new leadership staff can cloud the measurement of impact unless accommodations are made. For example, when a new measurement instrument is introduced to staff and a significant amount of training and development takes place, begin the study period, (the period when data collection will occur) after the new instrument is understood and is used in service by the majority of staff. History also refers to the historical context in which the study is conducted. For example, the study data were collected during a time when staff were able to contact most patients by cell phone but before the advent of the Affordable Care Act [5].

Avoiding the impact on service staff

It is important to define an endpoint to the study period. That is, data will be collected through a pre-determined time point when no further data will be added to the data set for evaluation. That endpoint is not shared with service staff. Service data constitutes a living, on-going record of peoples' responses to services provided. Thus, the program evaluator must establish an arbitrary endpoint so that on-going data collection will not be affected by certain types of bias and selection. If staff know that the data they are collecting will become part of the program evaluation up to a certain date they cannot help but be influenced, and objectivity is compromised.

Objectivity of scoring

If staff has been trained in correct measurement principles, they will not look at the previous measurement of a service participant's progress while they record the present measurement. Objectivity in scoring is destroyed when staff are allowed or encouraged to look back at previous measurements and to show "progress." It goes without saying that accurate scoring depends upon on-going training and supervisory conferences to continue to build staff members' understanding of the phenomena they measure in the extremely complex, non-clinical settings where public health programs take place.

Sample selection bias

Selection bias is seen when staff want to be sure to include certain cases in the sample-those they most proudly proclaim or those they use as cautionary tales. Individual cases should not be considered when determining the study period that determines the data set. All participants with the defined service characteristics in the defined study period are included.

Protecting the data set

Once a study period is set, all the data for that period of time should be collected and separated from the on-going collection of data. The program evaluator takes control of the data set and maintains control until all evaluation work has been completed. The data set can and must, of course, be shared with others on the leadership team. But the program evaluator must allow no changes to be made to the data set once it is established and checked for accuracy and completeness.

Design considerations

Having established the purpose, sub-population to be studied, a study period, and a verified data set, the program evaluator and the leadership team discuss the most feasible research design available to them. Research designs are best described by others [8], including the non-randomized designs consistent with a service environment as opposed to an experimental research environment. The TREND guidelines help the program evaluator structure the steps of the study and the report [2].

Public health program evaluation strategies can be applied retrospectively or prospectively with small, geo-political service populations. This scope of program evaluation is consistent with a rural clinic. Alternatively, public health program evaluation strategies can be applied prospectively with on-going data collection, constantly tuning the intervention and controls as subsets are studied, using a dashboard methodology that encompasses various time periods and generating reports for funding agencies. This scope of program evaluation is consistent with the needs of the U. S. Centres for Community Health or large population health agencies. The basic principles of epidemiology apply equally, regardless of scope or size of the study population.

The most frequently used and the weakest design is the pre-post comparison. It is weak because there can be many alternative reasons in addition to the program's impact that could explain the outcomes found [3]. In order to address those alternative explanations and lay them to rest, it is useful to have a comparison group of similar people who did not receive the services of the program, such as demographically similar perinatal women who received home visitation versus those who did not in the same geographical area. Since it is not ethically possible in most cases to randomly assign some participants to the program and some to placebo or "usual treatment" groups for comparison, the program evaluator must consider other forms of comparison groups. The aim of the comparison is to strengthen the argument that outcomes may be associated with the program intervention.

Internal comparison groups

The aim of internal validity control is to demonstrate that there is clear evidence to support the impact claim. An internal comparison group compares a sub-set of those who received the public health intervention and who had significantly (clinically and/or statistically) different outcomes compared to the other participants in the intervention. There can be little argument about differences in service characteristics, intervention, history, data collection, or analysis [9]. A subgroup that is divergent in some measured aspect of response to the program provides information about how the program is working for its participants. Once the main program evaluation is completed, examine carefully why that sub-group is different. Therein lays important information about what is needed to lift services to the next level and perhaps address issues in health equity.

External comparison groups

The aim of external validity control is to demonstrate generalizability to a greater population. In addition to considerations of internal validity, external validity challenges are from potential lack of multi-dimensional similarities, representation, and applicability of findings to the broader population. If one considers a comparison to national, state or regional data, or to data in another state or province, the comparison is fraught with concerns about the vast volume of data versus the smaller local data set, with geopolitical differences, and with the problems of verifying correspondence of data collection methods and accuracy for comparison. Despite the challenges, today's big data sets that have been held in virtual storage can be linked, opening new vistas for exploration with external comparison groups never before possible [9]. In theory, the limitations of public health program evaluation can never be overcome. In practice, evaluators select the strongest designs available, set up internal and external controls with comparison groups, recognize the limitations, and proceed with necessary evaluation.

Acknowledgements

This research was supported in part by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS) under Grant number H40MC00129 from the Healthy Start Initiative, Maternal Child Health Bureau, Health Resources and Services Administration. This information or content and conclusions are those of the authors and should not be construed as the official position or policy of, nor should any endorsements be inferred by HRSA, HHS or the U.S. Government. Original empirical research was conducted under Research Protocol #1010080 for evaluation of participant service. This protocol was approved by the Augusta University Institutional Review Board.

References

1. Thomas SD, Hudgins JL, Sutherland DE, Ange BL, Mobley SC (2015) Perinatal Program Evaluations: Methods, Impacts, and Future Goals. Matern Child Health J 19: 1440-1446.
2. Des Jarlais DC, Lyles C, Crepaz N; TREND Group (2004) Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. Am J Public Health 94: 361-366.
3. Campbell DT, Stanley JC (1963) Experimental and quasi-experimental designs for research. Chicago, Rand-McNally, USA.
4. Patton MQ (2008) Utilization-Focused Evaluation. Thousand Oaks, SAGE Publications, USA.
5. Mobley SC, Thomas SD, Sutherland DE, Hudgins J, Ange BL, et al. (2014) Maternal health literacy progression among rural perinatal women. Matern Child Health J 18: 1881-1892.
6. Pannucci CJ, Wilkins EG (2010) Identifying and avoiding bias in research. Plast Reconstru Surg 126: 619-625.
7. Brinberg D, McGrath JE (1985) Validity and the Research Process. Beverly Hills, SAGE Publications, USA.
8. Shadish WR, Cook TD, Campbell DT (2002) Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston, Houghton-Mifflin, USA.
9. Mobley SC (2015) Pregnancy Risk Assessment Monitoring System National Meeting. October 19-20, Atlanta.