AOGS
Acta Obstetricia et Gynecologica Scandinavica

AOGS REVIEW

External validity in perinatal research

ANTHONY O. ODIBO[1] & GANESH ACHARYA[2]

[1]Division of Maternal Fetal Medicine, Department of Obstetrics and Gynecology, Morsani College of Medicine, University of South Florida, Tampa, FL, USA, and [2]Division of Obstetrics and Gynecology, Department of Clinical Science, Intervention and Technology, Karolinska Institute and Center for Fetal Medicine, Karolinska University Hospital, Stockholm, Sweden

Key words
External validity, perinatal research, generalizability

Correspondence
Anthony O. Odibo, Division of Maternal Fetal Medicine, Department of Obstetrics and Gynecology, Morsani College of Medicine, University of South Florida, 2 Tampa General Circle, Suite 6053, Tampa, FL 33606, USA.
E-mail: aodibo@health.usf.edu

Conflict of interest
The authors have stated explicitly that there are no conflicts of interest in connection with this article.

Please cite this article as: Odibo AO, Acharya G. External validity in perinatal research. Acta Obstet Gynecol Scand 2018; 97:424–428.

Received: 4 September 2017
Accepted: 20 November 2017

DOI: 10.1111/aogs.13268

Abstract

Recent studies are beginning to focus on the external validity of well conducted internally valid research. This review gives an overview of external validity, the dimensions involved and suggestions for when future intervention trials are designed, using examples from perinatal research. Finally, we remind the perinatal researcher that it is their duty to provide extensive details beyond those needed to establish internal validity. The latter would help clinicians to determine whether the intervention is applicable to their population.

Introduction

Recent perinatal research has focused a lot on the internal validity of published studies through adherence to rigorous standards when estimating causal effects. However, external validity has received less attention. Internal validity refers to how accurate the effects of an intervention on participants in epidemiological studies are, unlike external validity, which has to do with the effect of the intervention on a target population. External validity tries to answer the question: is the intervention or finding applicable to other patients, with different characteristics, in dissimilar settings, and treated by other clinicians? Unlike study populations in epidemiological studies, which tend to be selected randomly, the target population, which represent the end users of the interventions, is not randomly selected. It is therefore not surprising that the mean or average

Key Message

It is only recently that external validity of research findings started receiving some attention. The review suggests items to be included in perinatal research to improve generalizability to other populations.

 © 2017 Nordic Federation of Societies of Obstetrics and Gynecology,

results obtained from well conducted studies are generally not reproduced in clinical practice in populations that different from that of the study participants. This review focuses on the dimensions of external validity including generalizability, applicability and feasibility, using examples from perinatal research.

Definitions and framework of review

External validity has been defined as the extent to which an internally valid effect measured in a study sample reliably reflects the effect in a population of interest – also described as the target population (1,2). In most clinical trials published in contemporary times, some consideration of generalizability is taken into account, but there are other aspects of external validity such as applicability and feasibility that are usually not well addressed (3). Recently, the concept of transportability of the conclusions regarding causal effects from epidemiological studies to a target population has been added as an important facet of external validity (4).

Unlike internal validity, where there are established standards for evaluation in epidemiological studies, such standards are not at present available for evaluating external validity. This creates a difficult situation for reviewers of epidemiological studies. However, there are some solid epidemiological principles that are covered in this review which can be useful for both investigators, reviewers and clinicians planning to apply study results to their target population.

Generalizability of results of epidemiological studies

The generalizability of the results from a study is how well the conclusions or suggested interventions apply to a different or larger population. This concept has been termed treatment heterogeneity if the effects are expected to differ based on the characteristics of the target population (3). Important questions that need to be addressed by studies include: does the study population represent your target population? Have there been many studies about this intervention and, if so, how consistent have the results been? Has the intervention been tested in several different settings? Finally, the average effect reported across studies should always be considered (3). If the population on which the study was reported is more heterogeneous, the then the likelihood that the results can be generalized to other target populations is higher. There are methods of purposive sampling or recruitment to the study that can be employed in the design phase to enrich the heterogeneity of the population (5).

Applicability of a study

The applicability of a study refers to how interventions described in the study might work in a different context. Information should be provided about study site characteristics such as urban vs. rural, the general demographics of the population and the distribution age, race or ethnicity, gender, parity, socioeconomic status and other pertinent questions about the similarity of the sample relative to your target population (3). The subtle difference from generalizability is that applicability focuses on the circumstances under which the intervention was rigorously tested. For example, if tested only in urban settings with easy access to public transport and ancillary amenities, the results may not be reproducible in a rural location without similar characteristics. The service delivery context such as the setting in which the study was conducted, for example an in-school setting, is an important consideration when assessing the applicability of a study. Details regarding how race or ethnicity was ascertained, standardization of definitions of education attainment, age and employment status are factors that can introduce bias and consequent difficulty applying a study result to a different population.

Feasibility of interventions

Feasibility refers to whether an intervention can be used in your target population given the available local resources. Elements to be considered when evaluating the feasibility of an intervention include the actual components of the intervention, the cost of the intervention, the education and training required of the personnel who would implement the intervention, and the expected time-frame needed to implement the intervention (3). The dosage of the intervention and the delivery format of the intervention are also important considerations under the domain of feasibility. We must stress that some epidemiologists do not consider feasibility to be a precise element of external validity.

Transportability of study findings

This is a specific study design method for providing external validity for a study and is classically demonstrated in clinical prediction models when a model is developed using a certain population and then validated in a new population (6). This method can also employ post study statistical adjustments to minimize the population differences and potentially improve generalizability to other target populations (7). As an example, a recent study employed a permutation test, focusing on discrimination as quantified by the c-statistic to judge

whether a prediction model was transportable to a new setting (8). Other statistical methods employed to evaluate external validity of trials are beyond the scope of this review.

Examples from perinatal research

Systematic reviews have recently focused on how the population settings from which primary studies are conducted can influence their generalizability to other target populations. As an example, a recent Cochrane database review evaluated studies using incentives to increase prenatal care use by women in an attempt to improve maternal and neonatal care. Of five studies eventually included, the largest two were from low-income, homogeneously Hispanic communities in Central America (9). This setting introduced a number of confounding factors which limit the generalizability of the study findings to other ethnically diverse and more urban communities.

In a well-conducted multi-center clinical trial, the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) and Maternal-Fetal Medicine (MFM) Network showed that 17-OH progesterone caproate prevented recurrent preterm birth in 2003 (10). This led to recommending the use of the medication by professional authorities in the USA. In a recently published observational study, investigators did not find a similar benefit as previously reported (11). It must be highlighted that the latter study only used a historical reference group as a comparison group with the women on 17 alpha-hydroxy progesterone caproate.

Why are study results not always transferable to the clinical setting?

There are several reasons why the results from trials are not always reproducible in the clinical setting. First, the typical participants in clinical trials are generally not representative of the average clinical patient we encounter. Studies have demonstrated that certain personality types are more likely to volunteer to be in a study than others. These personality differences include gender, presence of other co-morbidities, sociodemographic factors and age distribution. Such volunteers may be more compliant with a prescribed intervention compared with average patients. A literature review that was conducted of published English language articles that reported the findings of studies assessing external validity by a comparison of the patient sample included in randomized controlled trials reporting on pharmaceutical interventions with patients from everyday clinical practice, revealed that samples were highly selected and had a lower risk profile than real-world populations, with frequent exclusion of elderly patients and patients with co-morbidities (12). One interesting example from perinatal research is from the Surfactant Positive Airway Pressure and Pulse Oximetry Randomized Trial (SUPPORT) antenatal consent study, which demonstrated that mothers of infants enrolled in the SUPPORT trial had significantly different demographics and exposure to antenatal steroids compared with mothers of eligible, but not enrolled, infants (13). These differences were likely due to enrollment bias resulting from the antenatal consent process. The authors concluded that additional research and regulatory review would need to be conducted to ensure that large trials that require antenatal consent can be conducted so as to ensure the generalizability of results. Some studies offer incentives such as cash, gift vouchers or free transport to subjects in order to increase recruitment to studies. In actual practice, providing such incentives may be impractical and therefore reduce generalizability. Other factors already considered above in this review also contribute significantly to lowering the generalizability of study results to other populations. The rigor applied by investigators in the research setting to retain study subjects and also to encourage compliance may be stricter than in the clinical setting and may consequently affect the external validity. Finally, even when studies are well conducted, certain details about the study may deter users who are considering using the intervention from recommending it. The latter speaks to the transportability of the trial findings.

Challenges to external validity in perinatal research

One important challenge to be overcome if studies are to focus more attention on external validity, is for the Societies or governing bodies in charge of women's health research to emphasize the importance of generalizability in their guidelines. The problem with perinatal health research is that most of our colleges, for example the American College of Obstetrics and Gynecology, Royal College of Obstetricians and Gynecologists and the Nordic Federation for Obstetrics and Gynecology, do not publish guidelines for reporting research. We are therefore left with guidelines produced by the Oxford Center for Evidence Based Medicine (14) and the revised CONSORT statement (15) which touch on external validity only briefly. Routinely collected health data, obtained for administrative and clinical purposes without specific a priori research goals, are increasingly used for research. This is particularly common in the Nordic countries,

 © 2017 Nordic Federation of Societies of Obstetrics and Gynecology,

Table 1. Checklist for perinatal research to improve external validity (modified from Dijkers 2011) (14).

Category	Item needing full description	Check
1. Rationale	Perinatal problem addressed by the intervention; alternative interventions	
2. Target population	Characteristics of all women for whom the intervention might be applicable: clinical, ethnic and sociodemographic characteristics	
3. Subject recruitment	Referral sources; advertising; duration of recruitment period; patient incentives offered	
4. Inclusion criteria	Gestational age range, body mass index, singleton or multifetal gestations; severity or stage of target disorder: screening or diagnostic tests used	
5. Exclusion criteria	Maternal age and other demographic factors; twins and higher-order multiples; patient skills and abilities (for example literacy level): likelihood of compliance, non-response and adverse events (for example as determined by use of running periods); current/recent participation in another trial; type of co-morbidities for example hypertension, pre-gestational diabetes	
	Justification of choices	
6. Number of women/subjects	Sample size calculations; total number screened; number excluded per inclusion criteria; number refusing informed consent; number studied, by center and/or by interventionist; number of drop-outs; patient flow diagram	
7. Characteristics of subjects	Clinical (for example severity and duration of disorder, co-morbidities, medications) and demographic factors (for example age, gestational age, race/ethnicity, education, socioeconomic status, gravidity or parity); baseline risks of good/poor outcomes in the treatment/control group; representativeness of patients of target population	
8. Intervention and method of intervention delivery	Were investigators blinded to patient's study group assignment? Number, timing and duration of intervention; contents of intervention; individual vs. group delivery; local modifications of the protocol used at participating sites; treatment acceptability (other than refusal and drop-out): cultural and financial issues, patient preferences	
	Any safety concerns for intervention -maternal or neonatal	
9. Comparator	Same items as described for intervention (#8)	
	Similarity of intervention and comparator (if blinding was not possible)	
10. Data analysis	Degree of contamination between treatment and comparator; impact of non-compliance on outcomes; blinding success; sensitivity analyses for alternative assumptions; short- and long-term neonatal outcomes	
11. Discussion	Clinical significance of results; treatment feasibility: contexts, professional and other workload, costs, equipment, treater/patient-client compliance; choice of comparator; acceptability of treatment: fit with current practice, fit with patient culture; groups (clinical, demographic) for whom the treatment may be particularly suitable/unsuitable; blinding failure, risk for potential bias	
	Tertiary care vs. community health centers, level of nursery care	

where huge administrative databases are available. The rapid availability of these databases has highlighted issues not addressed by existing reporting guidelines, such as Strengthening the Reporting of Observational Studies in Epidemiology (STROBE). Recently, the Reporting of Studies Conducted Using Observational Routinely Collected Health Data (RECORD) statement was created as an extension to the STROBE statement. It addresses reporting items specific to observational studies using routinely collected health data. RECORD consists of a checklist of 13 items related to the title, abstract, introduction, methods, results, and discussion section of articles, and other information required for inclusion in such research reports (16). We also provide a checklist for perinatal research modified from one designed for rehabilitation medicine research (Table 1) (17). The checklist is not exhaustive and is provided here to encourage other researchers to expand on the items included, specifically as related to perinatal research.

Conclusion

This review detailed the need for external validation of studies using examples from current literature. However, only recently has this subject started attracting the attention of investigators. New constrcts are being defined, particularly in statistics, and interested readers should consult more advanced material for these. Future investigators are encouraged to bear the above principles in mind when designing future intervention studies or prediction models. Specifically, we remind perinatal researchers that it is their duty to provide extensive details beyond those needed to establish internal validity. The latter would help clinicians to determine whether the intervention is applicable to their population.

Funding

Professor Odibo is supported by Grant U01HD087213 from the Eunice Kennedy Shriver National Institute of

Child Health and Human Development, which partially funds this work.

References

1. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. Am J Epidemiol. 2010;172:107–15.
2. Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing study results: a potential outcomes perspective. Epidemiology. 2017;28:553–61.
3. Avellar SA, Thomas J, Kleinman R, Sama-Miller E, Woodruff SE, Coughlin R, et al. External validity: the next step for systematic reviews? Eval Rev. 2016;41:283–325.
4. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of trial results using inverse odds of sampling weights. Am J Epidemiol. 2017;186:1010–4.
5. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. New York: Houghton Mifflin Company, 2002.
6. Thangaratinam S, Allotey J, Marlin N, Mol BW, Von Dadelszen P, Ganzevoort W, et al. Development and validation of Prediction Models for Risks of Complications in Early-onset Pre-eclampsia (PREP): a prospective cohort study. Health Technol Assess. 2017;21:1–100.
7. Stuart EA, Rhodes A. Generalizing treatment effect estimates from sample to population: a case study in the difficulties of finding sufficient data. Eval Rev. 2016;41:357–88.
8. Nieboer D, van der Ploeg T, Steyerberg EW. Assessing discriminative performance at external validation of clinical prediction models. PLoS ONE. 2016;11:e0148820.
9. Till SR, Everetts D, Haas DM. Incentives for increasing prenatal care use by women in order to improve maternal and neonatal outcomes. Cochrane Database Syst Rev. 2015;12:CD009916.
10. Meis PJ, Klebanoff M, Thom E, Dombrowski MP, Sibai B, Moawad AH, et al. Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate. N Engl J Med. 2003;348:2379–85.
11. Nelson DB, McIntire DD, McDonald J, Gard J, Turrichi P, Leveno KJ. 17-Alpha hydroxyprogesterone caproate did not reduce the rate of recurrent preterm birth in a prospective cohort study. Am J Obstet Gynecol. 2017;216:600.e1–9.
12. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015;16:495.
13. Rich W, Finer NN, Gantz MG, Newman NS, Hensman AM, Hale EC, et al. Enrollment of extremely low birth weight infants in a clinical research study may not be representative. Pediatrics. 2012;129:480–4.
14. Oxford Center for Evidence-based Medicine. Levels of Evidence (March 2009). Available online at: http://www.cebm.net/?o=1025 (accessed August 15, 2017).
15. Moher D, Schulz KF, Altman D; for the CONSORT Group (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA. 2001;285:1987–91.
16. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected Health Data (RECORD) statement. PLoS Med. 2015;12:e1001885.
17. Dijkers MPJM. External validity in research on rehabilitative interventions: issues for knowledge translation. FOCUS Technical Brief No. 33. Austin: SEDL, National Center for the Dissemination of Disability Research, 2011.