PHARMACIST & TECHNICIAN CE:

Questioning Quasi-Experimental Research: An Overview of Quasi-Experimental Research Design

by Cassie Sedgwick, PharmD, Clara Nickel, 2024 PharmD Candidate, Anna Erickson, 2024 PharmD Candidate, Grayson Cooley, 2024 PharmD Candidate, John MacDonald, 2024 PharmD Candidate, Amanda Margolis, PharmD, MS, BCACP

CE FOR **PHARMACISTS & TECHNICIANS**

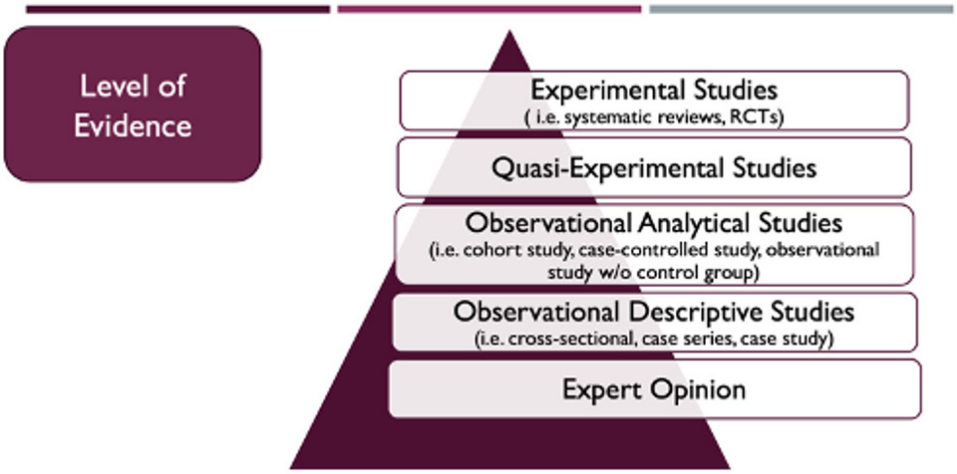
COMPLETE ARTICLE AND CE EXAM AVAILABLE ONLINE: **WWW.PSWI.ORG**

Learning Objectives

- Describe when quasi-experimental research design is used compared to traditional experimental research.
- Differentiate among the major types of quasi-experimental research design.
- Identify the strengths and limitations of propensity score matching and interrupted time series.
- Identify critical appraisal checklists that can be used to evaluate a quasi-experimental study.

Appropriately assessing and using literature is important for informed practice and clinical decision making.[1-3] Pharmacists are well trained in assessing true experimental research, such as systematic reviews, meta-analysis, and randomized controlled trials (RCTs); and observational studies, such as cohort and case-controlled studies. This information is then translated into practice and used to ensure patients are receiving high-quality, evidenced-based care.[4] Quasi-experimental research is being used more often in medical literature, though how to interpret these studies is not always covered in pharmacy curriculum. Quasi-experimental research is similar to traditional research (i.e. RCTs); however, it does not involve randomization. A quasi-experimental design can be used when randomization for an RCT is unethical, such as when the outcome in question may result in harm. It can also be used when a traditional experimental design may be cost- or time-prohibitive or to explore a causal relationship in the early stages of research. Quasi-experimental studies fall below traditional experimental research in level of evidence though are still higher than observational studies or expert opinions (Figure 1).

FIGURE 1. Level of Evidence Pyramid



As quasi-experimental research becomes increasingly prevalent in medical literature, it is important that pharmacists understand how to interpret this study design in order to translate it to clinical practice. A survey was administered to pharmacists across Wisconsin to assess understanding of quasi-experimental research design. The survey included four multiple choice questions and one short answer question to assess general knowledge of quasi-experimental design, including interrupted time series and propensity score matching, and three questions using a 5-point Likert scale (1=not confident at all to 5=very confident) to assess confidence in assessing different research methods including randomized controlled trials, interrupted time series, and propensity score matching. Fifteen pharmacists responded to the survey. Respondents scored between 0% and 40% on the multiple choice and short answer questions assessing knowledge. Mean confidence in assessing RCTs was 4.2 compared to 1.5 and 1.8 for interrupted time series and propensity score matching, respectively. This survey demonstrates that a gap in knowledge exists for pharmacists in how to interpret quasi-experimental design. The purpose of this article is to increase pharmacist knowledge and ability to assess studies using quasi-experimental research design.

When to Use Quasi-Experimental Research

As previously mentioned, quasi-experimental study design is becoming increasingly prevalent in medical literature; therefore, understanding and assessing these types of studies has become essential to clinical practice. The prefix "quasi" means "resembling."[5] Quasi-experimental research resembles experimental research without being truly experimental due to a lack of randomization.[5-7] Despite this difference, the goals of quasi-experimental research remain the same: to establish causal relationships between independent and dependent variables. In quasi-experimental research, the investigator directs the dependent variable and therefore it is a higher level of evidence than observational research.[5] However, quasi-experimental research is still at a higher risk for other limitations compared to RCTs, such as confounding variables.

Quasi-experimental design can be used in cases where it would be unethical to randomize.[5] An example of this would be if the outcome in question is centered around whether an intervention causes harm. In many cases, patients choose whether they receive a particular intervention, such as a treatment or procedure. Investigators can then follow patients or review the medical record retrospectively to see if a certain intervention caused harm. In these cases, quasi-experimental research may be able to fill gaps in knowledge or answer questions that would otherwise be unethical through an RCT. However, by applying quasi-experimental techniques to the traditional retrospective cohort design, it may produce similar distributions of baseline characteristics and minimize some aspects of selection bias.[7] Quasi-experimental design can also be used to reduce cost or resources to test a hypothesis. The cost of an RCT can be high, especially when considering many involve multi-center approaches. Quasi-experimental design is a less costly way to establish a causal relationship. Cancer screening and prevention trials are generally very long and would be costly to conduct as RCTs, so many of those studies use quasi-experimental approaches.[8-10]

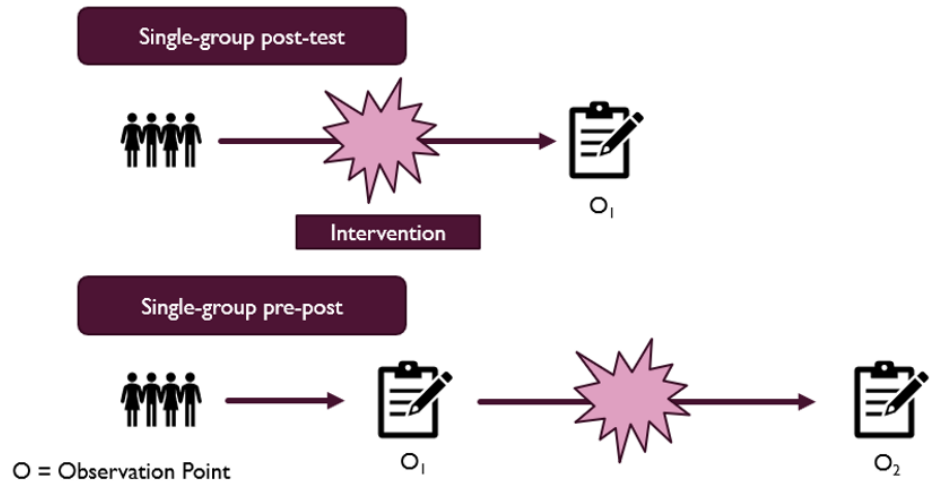
Quasi-experimental research is also used to evaluate new programs, services, educational materials, or workflows.[5,7,11] When a new service or workflow is implemented across a department, randomization, or even prospective data collection, is often not feasible.[11] Because of this, randomization is not possible and a quasi-experimental approach can be used. Additionally, whether a patient is enrolled in a program or receives a specific treatment is often dependent on patient specific factors and is chosen by the patient or clinician. Using quasi-experimental research designs can often include individuals who may be excluded from an RCT and are often considered more pragmatic.[11]

Though quasi-experimental designs are not considered the "gold standard" of research, these methods have many strengths. Since the study population tends to resemble the general population more closely, quasi-experimental studies have higher generalizability.[5,6,11] Additionally, interventions are often assessed in real-world settings rather than a controlled laboratory environment, leading to higher external validity.[6,11] Quasi-experimental research designs meet some criteria for causality which allows for some causal inferences when randomization is not possible.[11,12] A quasi-experimental study can lay the foundation to justify further research when implementing an experimental design may not be feasible due to cost or time constraints.[5,6] Quasi-experimental methods can also be used in the retrospective analysis of interventions that have occurred outside of the investigator's control and can use previously collected data.[5,6,11]

Conversely, quasi-experimental design has many limitations. As subjects are not randomly assigned within the quasi-experimental structure, these study designs have lower internal validity than RCTs.[5-7,11] Less control increases the risk of confounding variables and bias. An example of a potential bias is in the selection of subjects. Without a standard system for randomization, natural human bias can influence who is chosen to be included in the study population, and who may receive the intervention, leading to differences between the intervention and control groups. Additionally, other factors such as age and comorbidities may influence whether a subject receives the intervention. Another threat to the quasi-experimental methodology is historical bias, which occurs when events outside the intervention influence the measured outcomes.[5,11] These unrelated events may precede or coincide with the intervention and may misrepresent trial results. For example, suppose an investigator is evaluating patients seen in a new asthma clinic, but many of those patients are also enrolled in an asthma education course. In that case, it could be challenging to determine if their increase in proper inhaler use was due to the clinic or the education course. Finally, if subjects are followed for extended periods of time, methods of testing may change or evolve. Instrumentation bias, as this phenomenon is called, could complicate data comparisons over time.[11]

Additionally, some quasi-experimental studies may have ambiguous temporal precedence, as the timing of the intervention may not be defined.[6] A vague timeline makes distinguishing between

FIGURE 2. Pretest-Posttest Design



pre-intervention and post-intervention data difficult. For instance, if only a post-intervention test is performed, it can be unclear if the outcome was present before the intervention or if the measured effect was truly due to the intervention.

Variations of Quasi-Experimental Research

Though there are several variations of quasi-experimental research, some of the most common include the pretest-posttest, interrupted time series (ITS), and propensity score matching.[5,6,11] These methods can be used independently; however, many studies combine designs when analyzing casual relationships.

Pretest-Posttest Design

The pretest-posttest method may include a single group or use multiple groups, with both an intervention and control.[5,6,11] The factor that differentiates pretest-posttest from interrupted time series is that the dependent variable is only measured once before (unless the design is post-test only) and once after the intervention. Single group pretest-posttest designs (Figure 2) are considered to be the weakest form of quasi-experimental research. Due to lack of repeated testing, this type of design may be subject to the Hawthorne effect, where individuals act differently or modify their behavior in response to being observed. Also, since there is no control group, it can be difficult to determine if the impact on the dependent variable is due to the intervention, co-occurring events, or if the outcome would have occurred without the intervention.

Interrupted Time Series (ITS)

ITS are similar to the pretest-posttest design; however, with this design, multiple data points are collected before and after the intervention, creating a timeline of outcome measures (Figure 3).[5,6,11] This creates a stronger causal relationship between the independent and dependent variables. ITS can be single-group, or can include a comparator or control group. One example of when a single-group ITS may be used is when a policy or procedural change impacts an entire department or facility and there is not a group that was not impacted by the change. If an institution changes the way antibiotics are ordered in the emergency department to reduce inappropriate antibiotic prescribing, the desired outcome could only be measured prior to the intervention and after. There would not be patients who were not impacted by the change in procedure. In this case, if the number of documented errors or time from ordering to receive medication was previously measured on a consistent schedule, that could be compared to measurements taken post-change to determine the impact. In an ITS, the consistent frequency (i.e. monthly, quarterly, etc.) of observations or outcome measurement should have a clinical or practical significance.[11] Single-group ITS can also be conducted with multiple interventions.[5,6] This can be useful when trying to improve services or workflows with changes being made periodically. The impact of each change could be measured throughout time and each change would be a set timepoint where an intervention occurred.

Use of a comparator or control group (Figure 4) strengthens the design of an ITS. A main difference between an RCT and an ITS is that subjects are not randomized to the intervention and comparator groups.[5,6,11] In an ITS, the clinicians, patients, or other factor (e.g. time of process change, floor admitted to) determines whether a patient receives the intervention. However, the use of a control group in an ITS can help reduce historical and instrumentation bias and distinguish the intervention effect from co-occurring events.[6] Both single and multigroup ITS rely on extrapolation to estimate what the postintervention data would have been without the intervention.[6,11] This helps to determine the impact or the difference that the intervention had on the outcome in question. The difference between the outcome line postintervention and the extrapolated line shows the perceived impact of the intervention (Figure 5).

Figure 5 depicts an example of how data from a comparative ITS may be shown.[5-7,11] The extrapolated data is depicted by the dashed line while the solid line depicts the actual data collected. Use of a comparator group can help to distinguish the impact of the intervention from the impact of co-occurring events. The difference between the extrapolated line of the comparator, or control, group and the collected data postintervention can be considered to be the difference caused by co-occurring events. This difference can then be subtracted from the difference between the extrapolated

FIGURE 3. Single-group Interrupted Time Series

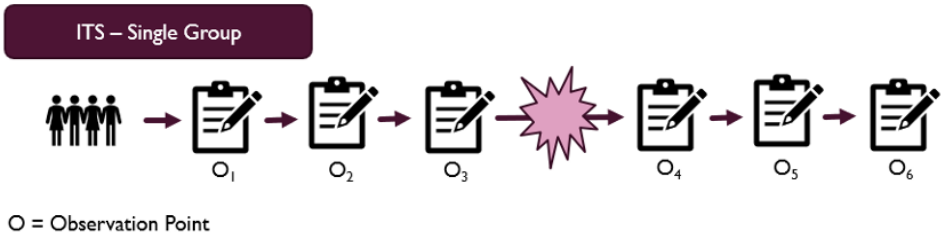
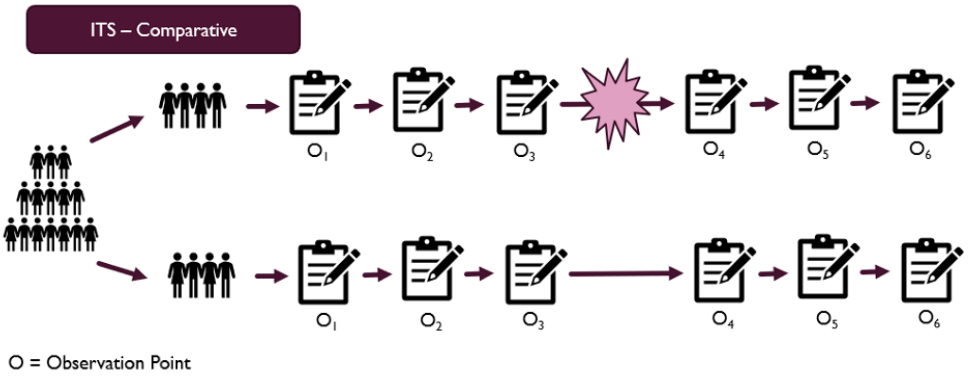


FIGURE 4. Comparative Interrupted Time Series



and collected data lines of the intervention group to determine the difference caused by the intervention. This is also referred to as a difference-in-differences approach.[6,7]

Though ITS have many strengths, there are still several limitations. Like pretest-posttest design, it can be difficult to determine if the impact on the dependent variable is due to the intervention, some co-occurring events, or if the outcome would have occurred without the intervention in a single-group ITS.[5,6,11] As previously mentioned, use of a comparator group can help to distinguish intervention effect from the effect of co-occurring events.[6] In ITS, the time point of an intervention can be unspecific or difficult to determine. Varied implementation of an intervention can also contribute to ambiguity of the start of the intervention. The impact on the dependent variables may have a delayed or weak impact, which may not be detected by the investigators depending on the length of the ITS. Lastly, given lack of randomization, there is also the risk for selection bias.

There are ways to strengthen an ITS and reduce the impact of bias and limitations. One way would be to control the time point of the intervention. This could involve choosing to intervene when threats to the intervention are less likely. Additionally, collecting more data points both pre- and post-intervention can help to strengthen the study design. Lastly, investigators can match subjects in the treatment and comparison group based on covariates and excluding some outliers to increase similarities between the two groups. This can be done through propensity score matching.

Propensity Score Matching

Propensity score matching can be used to help strengthen quasi-experimental research as it helps to better compare groups that were not established randomly.[5] A propensity score is the conditional probability that a subject belongs to the treatment group based on specified covariates.[6] The use of propensity score matching allows investigators to make the intervention and control groups more similar and improve internal validity of the study. In randomized controlled trials, subjects are assigned randomly to either the intervention or control group and are often stratified by characteristics specified ahead of time by the investigators. As shown in Figure 6, this creates groups that are similar at baseline. This increases the likelihood that any differences seen between groups is due to the intervention and not differing baseline characteristics.

Traditional cohort studies involve following subjects over time and measuring dependent variables in groups of people who were exposed to certain risk factors compared to those who were not exposed to determine the impact of the exposure.[6] However, without randomization, whether a subject receives an intervention depends on various factors. For example, certain subjects may be more or less likely to receive a certain pharmacologic treatment or procedure depending on age, comorbidities, geographic location, subject's ability to perform self-cares, support system, or socioeconomic status.[5,6] This often

FIGURE 5. Single-group Interrupted Time Series

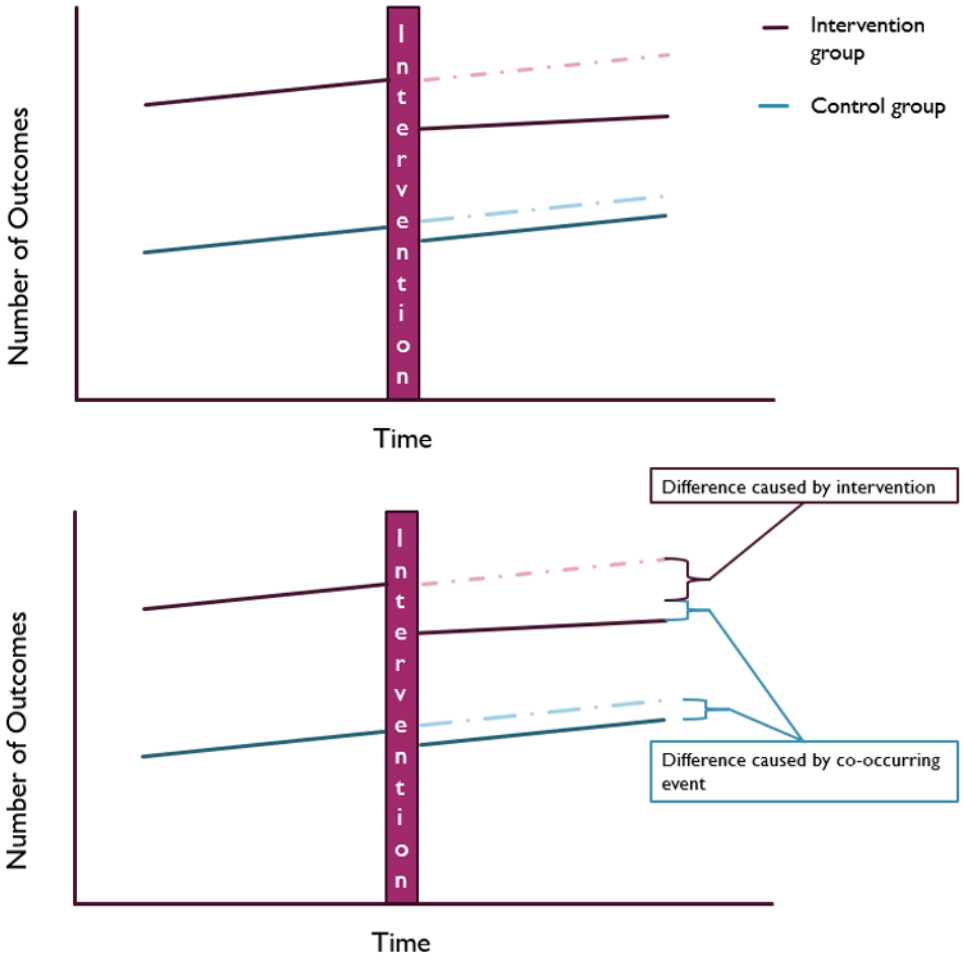
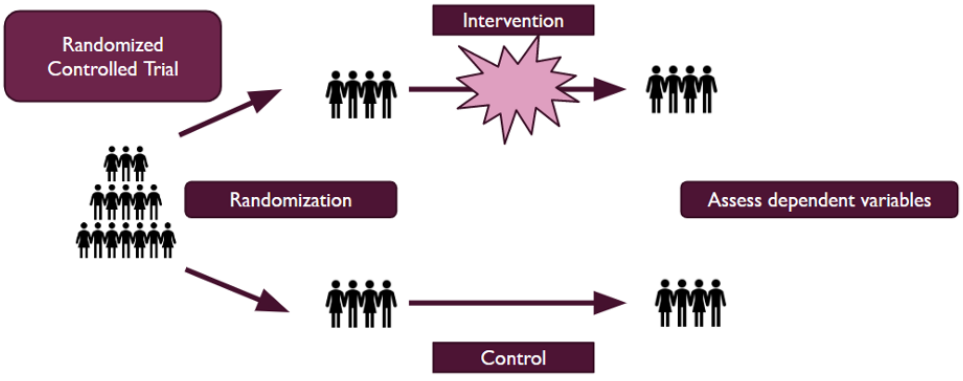


FIGURE 6. Randomized Controlled Trial Design



causes group to differ at baseline (Figure 7). This can confound the results and make it unclear whether the effect of the intervention is due to the intervention itself or the difference in baseline characteristics. Propensity score matching is one quasi-experimental technique that can be used to create more similar groups at baseline without the use of randomization.[6,7]

The first step to conducting a propensity score match is to determine the dependent variable. The dependent variable is selected by investigators and is the primary outcome of interest. Next, investigators select what covariates should be used to calculate the propensity score.[7] This includes anything that the investigator may suspect would impact whether a subject receives the intervention, as long as it was measured, but may also include factors that may impact the outcome of interest. These are often baseline characteristics seen in "Table 1" of research papers.

Next, a statistician builds a model to estimate a subject's likelihood, or propensity, to be in the intervention group.[6] This is the propensity score. Subjects in the intervention and control groups are then matched based on their propensity scores. The unmatched subjects tend to be excluded from the analysis if there is no one in the other group that has the same propensity to be in the intervention group as they do. For example, someone in the intervention group who has a very high propensity score may not have a match to a subject in the control group —there may not be someone in the control group who has that high a likelihood based on covariates that would have been in the intervention group. Figure 9 illustrates this with the size of the person correlating to a higher propensity score. The subjects who are excluded have a very high or very low propensity score. There are a number of techniques statisticians can use to match subjects by propensity. Once the propensity score matching has been completed, and baseline characteristics between groups are more similar, the treatment effect can be determined.

One limitation of propensity score matching is that, though it creates more similar groups, by removing subjects from the study, generalizability of the results decreases. Additionally, the best methods for balancing and matching based on propensity score are still being determined.

FIGURE 7. Traditional Cohort Design

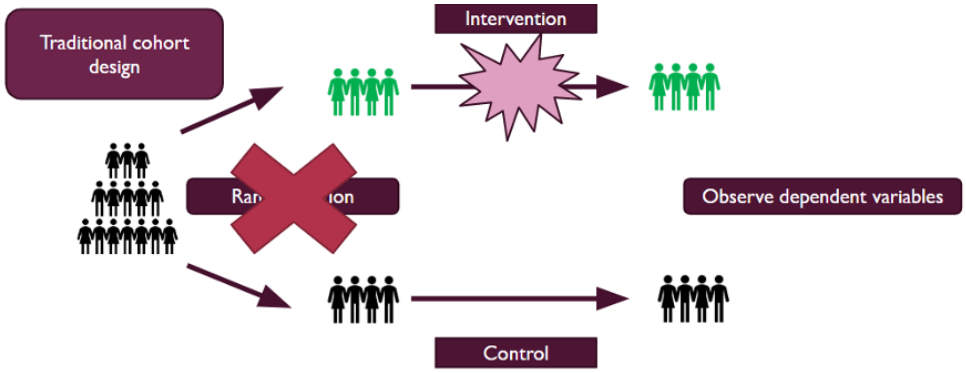


FIGURE 8. Propensity Score Matching

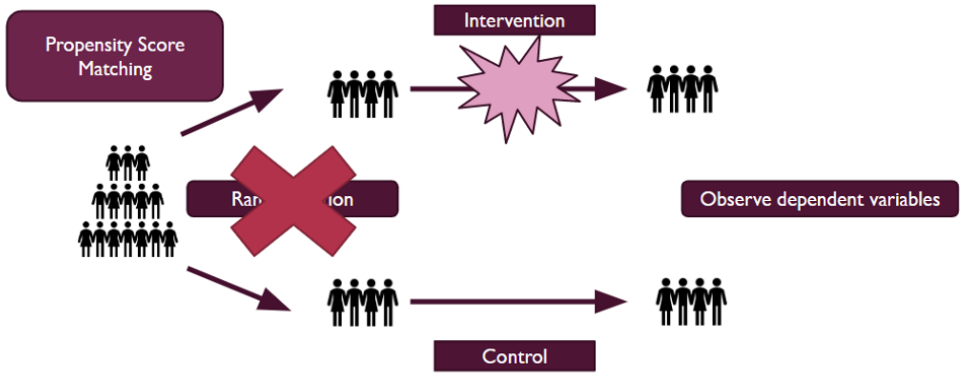
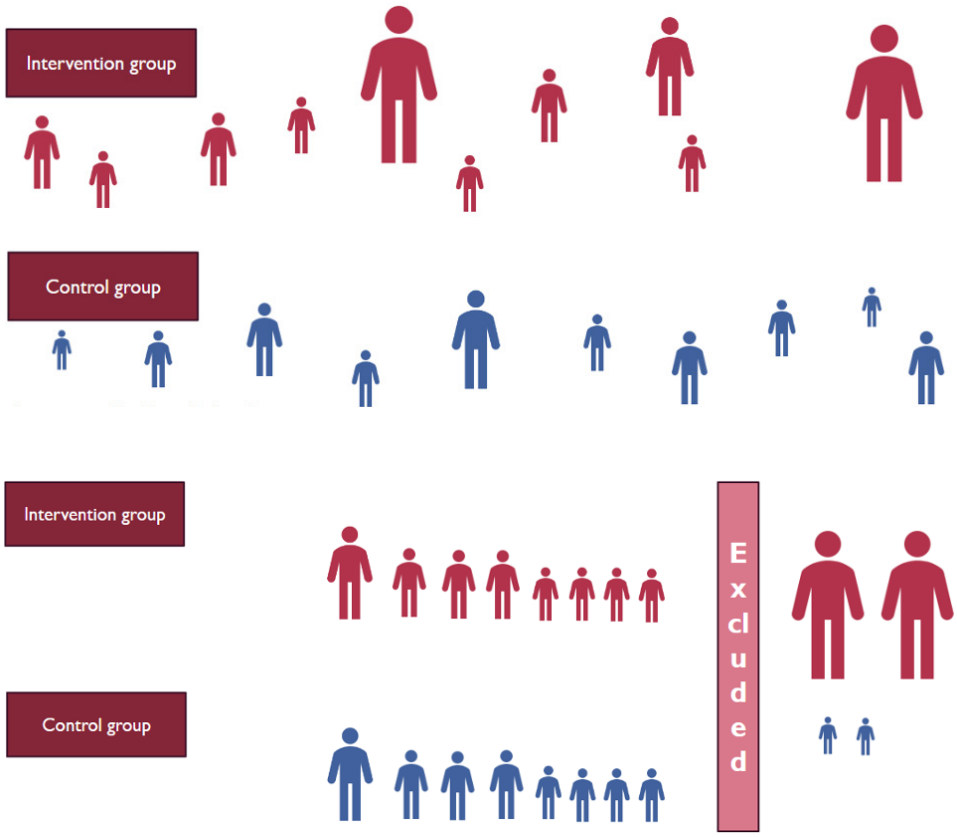


FIGURE 9. Propensity Score Matching Example



Overall, using propensity score matching in quasi-experimental research can help to minimize selection bias and improve the strength and quality of the study.

Critical Appraisal Tools

Being able to assess the quality of scientific literature is an important skill all healthcare professionals should have in order to provide optimal, evidence-based care to patients.[13] A critical appraisal tool is a checklist of prompts created to evaluate the quality of a study.[14,15] The prompts challenge the reader to question the study's design, conduct, and analysis to consider inconsistencies and potential biases thatmay result in misleading conclusions. Proven critical appraisal tools have been formulated for different types of study designs, as each has unique capacities for those inconsistencies and biases. There are critical appraisal tools specific for quasi-experimental design, which pharmacists are encouraged to seek out when reading quasi-experimental studies (Table 1).

TABLE 1. Critical Appraisal Tools

Organization	*Types of Critical Appraisal Tools*	*Website*
Joanna Briggs Institute (JBI)[14]	Analytical Cross Sectional Studies Case Control Studies Case Reports Case Series Cohort Studies Diagnostic Test Accuracy Studies Economic Evaluations Prevalence Studies Qualitative Research Quasi-Experimental Studies Randomized Controlled Trials Systematic Reviews Text and Opinion	jbi.global/critical-appraisal-tools
Centre for Evidence-Based Medicine (CEBM)[15]	Systematic Reviews Diagnostics Prognosis Randomized Controlled Trials Critical Appraisal of Qualitative Studies IPD Review	https://www.cebm.ox.ac.uk/resources/ebm-tools/critical-appraisal-tools
Critical Appraisal Skills Programme (CASP)[16]	Randomized Controlled Trial Systematic Review Qualitative Studies Cohort Study Diagnostic Study Case Control Economic Evaluation Clinical Prediction Rule	casp-uk.net/casp-tools-checklists/

Conclusion

Using literature is fundamental in informing evidence-based practice.[1-3] Meta-analyses and randomized controlled trials have the highest level of evidence and should be used whenever possible. However, a quasi-experimental study design is an alternative that can be used to evaluate potentially causal relationships in cases where traditional research design cannot be used. With this design becoming more and more common, it is necessary for healthcare providers to understand how to interpret these types of studies as well as recognize their strengths and limitations.

Cassie Sedgwick is a clinical pharmacy practitioner at the William S. Middleton Memorial Veterans Hospital in Madison, WI. Clara Nickel, Anna Erickson, Grayson Cooley, and John MacDonald are 2024 Doctor of Pharmacy Candidates at the University of Wisconsin-Madison School of Pharmacy in Madison Wisconsin. Amanda Margolis is an Associate Professor at the University of Wisconsin-Madison School of Pharmacy in Madison Wisconsin..

Disclosure: The authors declare no real or potential conflicts or financial interest in any product or service mentioned in the manuscript, including grants, equipment, medications, employment, gifts, and honoraria.

Corresponding Author: *Cassie Sedgwick - Cassie.Sedgwick@va.gov*

References

1. Gentry CK, Parker RP, Ketel C, et al. Integration of clinical pharmacist services into an underserved primary care clinic utilizing an interprofessional collaborative practice model. *J Health Care Poor Underserved.* 2019;27(615):1-7. doi: 10.1353/hpu.2016.0005.
2. Pottie K, Farrell B, Haydt S, et al. Integrating pharmacists into family practice teams: physicians' perspectives on collaborative care. *Can Fam Physician.* 2008;54(12).
3. Truong H, Kroehl ME, Lewis C, et al. Clinical pharmacists in primary care: provider satisfaction and perceived impact on quality of care provided. *SAGE Open Med.* 2017;5:205031211771391. doi:10.1177/2050312117713911
4. Sackett D, Rosenberg W, Gray J, Haynes R, Richardson W. Evidence based medicine: what it is and what it isn't. *Br Med J.* 1996;312(7023):71-72. doi: 10.1136/bmj.312.7023.71
5. Research Methods in Psychology: 7.3 Quasi-Experimental Research. University of Minnesota Libraries Publishing. Published 2016. Accessed March 6, 2023. https://open.lib.umn.edu/psychologyresearchmethods/chapter/7-3-quasi-experimental-research/
6. Kim Y, Steiner P. Quasi-experimental designs for causal inference. *Edu Psychol.* 2016;51(3-4):395-405. doi:10.1080/00461520.2016.1207177
7. White H, Sabarwal S. Quasi-experimental design and methods. UNICEF Office of Research. Published September 2014. Accessed June 23, 2023. https://www.unicef-irc.org/KM/IE/img/downloads/Quasi-Experimental_Design_and_Methods_ENG.pdf
8. Young B-R, Gwede CK, Thomas B, et al. A systematic review of U.S.-based colorectal cancer screening uptake intervention systematic reviews: available evidence and lessons learned for research and practice. F*ront Public Health.* 2019;7:145. doi: 10.3389/fpubh.2019.00145
9. Flood T, Wilson IM, Prue G, McLaughlin M, Hughes CM. Impact of school-based educational interventions in middle adolescent populations (15-17yrs) on human papillomavirus (HPV) vaccination uptake and perceptions/knowledge of HPV and its associated cancers: a systematic review. P*rev Med.* 2020;139:106168. doi: 10.1016/j.ypmed.2020.106168
10. Saei Ghare Naz M, Kariman N, Ebadi A, Ozgoli G, Ghasemi V, Fakari FR. Educational interventions for cervical cancer screening behavior of women: a systematic review. *Asian Pac J Cancer Prev.* 2018;19(4):875-884. doi: 10.22034/APJCP.2018.19.4.875
11. Schweizer ML, Braun BI, Milstone AM. Research methods in healthcare epidemiology and antimicrobial stewardship – quasi-experimental designs. *Infect Control Hosp Epidemiol.* 2016;37(10):1135-1140. doi: 10.1017/ice.2016.117.
12. Fedak KM, Bernal A, Capshaw ZA, Gross S. Applying the Bradford Hill criteria in the 21st century:

how data integration has changed causal inference in molecular epidemiology. *Emerg Themes Epidemiol.* 2015;12(1):1-9. doi:10.1186/s12982-015-0037-4
13. Burkiewicz JS, Zgarrick DP. Evidence-based practice by pharmacists: utilization and barriers. *Ann Pharmacother.* 2005;39(7-8):1214-1219. doi:10.1345/aph.1E663
14. Johannes Briggs Institute. Checklist for Quasi-Experimental Studies (Non-randomized Experimental Studies). 2020. Accessed January 15, 2023. https://jbi.global/sites/default/files/2020-08/Checklist_for_Quasi-Experimental_Appraisal_Tool.pdf
15. Centre for Evidence-Based Medicine. Critical appraisal tools. Accessed May 5, 2023. https://www.cebm.ox.ac.uk/resources/ebm-tools/critical-appraisal-tools
16. Nuffield Department of Primary Care Health Sciences. Centre for Evidence-Based Medicine. Published 2023. Accessed May 5, 2023. www.cebm.ox.ac.uk/resources/ebm-tools/critical-appraisal-tools

Assessment Questions

1. Which study design has the highest internal validity?
   a. Cohort with propensity score matching
   b. Interrupted time series
   c. Randomized controlled trial
   d. Case-control study

2. Which of the following scenarios is most likely to require a quasi-experimenal approach as an RCT is less feasible?
   a. Determining efficacy and safety for a new medication for use in diabetes
   b. Evaluation of a new FDA approved antibiotic used to treat urinary tract infections compared to nitrofurantoin
   c. Determining the titers developed following a new vaccine for respiratory syncytial virus (RSV)
   d. Evaluation of the efficacy and safety of change in antibiotic order process in the emergency department

3. Which of the following study questions is most appropriate for the quasi-experimental design?
   a. Determining the efficacy and safety of a new anticoagulant compared to apixaban in patients with atrial fibrillation
   b. Determining the change in prescribing habits after implementation of a diabetes education program
   c. Evaluating the safety of a new SGLT1/SGLT2 inhibitor in patients with heart failure with midrange ejection fraction (HFmrEF)
   d. Assessing an association between physical activity and cognitive function in older adults

4. **True or False:** Quasi-experimental research design has a lower level of evidence than a case-controlled study.
   a. True
   b. False

5. Which study design does the following study describe?

   A checklist for checking chemotherapy infusions is implemented in an inpatient pharmacy. Error rates and near misses are reported monthly and data is available over the previous three years. One month after the implementation of the checklist, error rates and near misses are collected again and a month after implementation the effects of using the checklist are evaluated.
   a. Pretest-Posttest
   b. Randomized controlled trial
   c. Interrupted time series
   d. Cohort with propensity score matching

6. Which statement regarding bias is specific to an interrupted time series?
   a. The time point of the intervention is not always clear
   b. Interrupted time series have low generalizability as specific to time of intervention
   c. Interrupted time series have greater power than randomized controlled trials
   d. In an interrupted time series, investigators are unable to use a control arm, which reduces selection bias

7. What is an advantage of using propensity score matching?
   a. There are clear methods available for balancing propensity scores
   b. Propensity score matching increases generalizability of cohort studies
   c. Readers do not need to be concerned about other potential forms of bias such as historical bias or attrition
   d. Propensity score matching reduces selection bias

8. **True or False:** Critical appraisal checklists can be used to evaluate strength of a study using quasi-experimental design
   a. True
   b. False

CE FOR **PHARMACISTS & TECHNICIANS**

Continuing Education Credit Information

The Pharmacy Society of Wisconsin is accredited by the Accreditation Council for Pharmacy Education as a provider of continuing pharmacy education. Continuing education credit can be earned by completing the self assessment questions. Questions may be completed online. Participants receiving a score of 70% or better will be granted 0.5 hour (0.05 CEU) credit through CPE Monitor. Accurate birth date (MMDD) and CPE Monitor ID must be provided in order to receive this credit as required by ACPE. This CE offering is offered free-of-charge to all PSW members. Nonmembers are charged $25.

July/August 2023
Questioning Quasi-Experimental Research: An Overview of Quasi-Experimental Research Design
ACPE Universal Activity Number: 0175-0000-23-085-H99-P,T
Target Audience: Pharmacists
Activity Type: Knowledge-based
Release Date: July 1, 2023
(No longer valid for CE credit after July 1, 2026)

Submit Your CE Online
www.pswi.org/Education/Journal-CE