Literature review on extrapolation, external validity and transfer

Deliverable D2.2

Milestone ML2

Coordinator of this work: Università di Teramo

Authors: Simone Busetti, Lorenzo Pagni

info@evaluatingfoodpolicy.it

Finanziato dall'Unione europea NextGenerationEU

PRIN 2022-PNRR-SH2-Prof. SIMONE BUSETTI-PI Unite (codice P2022XWBT8_01) Finanziato dall'Unione europea – Next Generation EU – CUP C53D23009980001

1. Literature Review for Extrapolation

External validity and extrapolation are crucial concepts in scientific research, especially in social sciences and policy evaluation. External validity refers to the extent to which the results of a study can be generalized beyond the specific sample and context in which they were obtained (Cook & Campbell, 1979). This is a fundamental aspect of empirical research, as it determines the applicability and relevance of the findings to broader populations and different settings. High external validity means that the study's conclusions can be extended to other groups, settings, and times with confidence. Factors that influence external validity include the representativeness of the sample, the ecological validity of the study environment, and the robustness of the experimental design (Shadish et al., 2002).

Extrapolation, on the other hand, involves applying the findings of a study to populations or settings not directly examined in the study. It is a process that extends the inferences made from the study to broader contexts. Extrapolation is particularly important in fields like medicine, education, and social policy, where direct experimentation on all possible contexts is impractical or unethical. The process of extrapolation requires careful consideration of the similarities and differences between the original study conditions and the new contexts to which the findings are being applied (Berk, 1983).

Understanding external validity and the ability to extrapolate findings are essential for policymakers because they ensure that the results from research studies can be applied to broader populations. This is particularly important when implementing policies that aim to address widespread social issues. For instance, a policy intervention that works well in a small, controlled study setting, may not have the same impact when applied to a larger, more diverse population. Policymakers rely on research findings to design and implement effective interventions, and without high external validity, there is a risk that these interventions may not achieve the desired outcomes in real-world settings (Rossi et al., 2004). Extrapolation allows researchers and policymakers to make informed predictions about how a policy might perform in different contexts, which is critical for effective and efficient policy design and implementation. It helps in predicting the success or failure of policies before they are widely implemented, saving time, resources, and potential negative impacts on the population.

Given the importance of these concepts, a rigorous methodology was employed to identify and select relevant papers for this study. Google Scholar was used as the primary search engine. The following keywords were utilized: "External validity" yielding 1330 results, "Extrapolation" with 12100 results (most of which

pertained to statistical and mathematical modeling, with relevant contributions selected for species extrapolation), "External validity + Policy" with 47 results, "Extrapolation + Policy" with 13 results, "External validity + Policy program" and "Extrapolation + Policy program," both of which returned no results, "External validity + Intervention" with 28 results, and "Extrapolation + Intervention" with 5 results. Citations were preemptively excluded from the search.

Initially, the titles of the first 1000 results were reviewed, reducing the total number of papers to 290. Following a review of the abstracts, the count was further narrowed down to 138, considering that some papers appeared in multiple searches. After consolidating duplicates, 108 unique results were included. These 108 results were then cross-checked using Research Rabbit to identify similar, earlier, and later works, adding 11 more relevant papers to the list. This brought the total to 119 papers.

The final selection comprised 119 papers, categorized as follows: 6 books, 102 journal articles, 5 reports, 3 chapters, and 3 preprints. Following an additional analysis to check for duplicates, 2 papers were removed, resulting in a final total of 117 papers.

Subsequently, to analyze the papers for the purpose of transferring food policies from one context to another, ChatGPT-4 was asked to answer the following questions for each selected paper:

What are the main problems in extrapolation / external validity raised by the article/book?

What are the methods for fixing problems of extrapolation / external validity?

What features of the new context and/or target population can affect extrapolation / external validity?

Does the article/book mention examples of problematic extrapolation/external validity? Which one?

Does the article/book mention examples of successful extrapolation/external validity? Which one?

Consider the case of transferring an urban food security program from one city to another, what would this article suggest to do?

Consider the case of transferring an urban food security program from one city to another, can you rate the relevance of the approach of this article on a five-point scale (1 = not relevant, 5 = very relevant)?

This analysis aimed to identify the papers most relevant to the chosen context. At the end of the process, a list of 28 papers was compiled, each of which received a relevance rating of 5 based on the specified criteria.

The next section includes all the evaluations and detailed insights derived from the papers reviewed, along with the respective relevance ratings.

2. Paper analysis

2.1. A Conceptual Framework for External Validity (Averitt et al., 2021)

The paper by Amelia J. Averitt et al. focuses on the challenges of external validity in randomized controlled trials (RCTs) within the context of evidence-based medicine (EBM). It identifies significant issues that arise when attempting to generalize the results of RCTs to broader, real-world populations and proposes a new framework to address these challenges.

Main Problems in Extrapolation/External Validity

The primary problem highlighted is the poor generalizability of RCT findings due to the controlled settings and selective participant criteria. This leads to a gap between clinical research and practical application, as RCTs often exclude diverse populations like pregnant women, the elderly, and those with comorbidities, resulting in findings that do not reflect real-world diversity.

Methods for Fixing Problems of Extrapolation/External Validity

The authors propose a population-oriented framework to evaluate and improve external validity. This framework involves identifying and analyzing the factors that contribute to poor generalizability, such as participant characteristics and treatment settings. It emphasizes the need for detailed information about these factors to better assess the applicability of RCT results to different populations.

Features of the New Context and/or Target Population Affecting Extrapolation/External Validity

The framework identifies several features that can affect external validity, including demographic differences, healthcare delivery environments, and the specific characteristics of the treatment settings. Understanding these variables is crucial for assessing how well RCT results can be applied to new contexts.

Examples of Problematic Extrapolation/External

Validity The paper discusses the exclusionary criteria of RCTs as a major source of poor external validity. These criteria often result in underrepresentation of significant population segments, leading to findings that do not generalize well to broader, real-world settings.

Examples of Successful Extrapolation/External Validity

While the paper does not provide specific examples of successful extrapolation, it suggests that careful consideration and adaptation of RCT findings to match the new

context and target population can improve generalizability. The proposed framework aims to facilitate this process by providing a structured approach to evaluating external validity.

Transferring an Urban Food Security Program For transferring an urban food security program from one city to another, the article would suggest evaluating the new context's demographic and environmental characteristics. By applying the proposed framework, practitioners can assess the similarity between the original and target populations and settings to ensure effective replication of the intervention.

Overall, the paper emphasizes the importance of understanding and addressing the factors that impact external validity to bridge the gap between clinical research and practical application. The proposed framework offers a comprehensive tool for evaluating and improving the generalizability and applicability of RCT findings in real-world settings.

2.2. A Design-Based Approach to Improve External Validity in Welfare Policy Evaluations (Tipton & Peck, 2016)

The article by Elizabeth Tipton and Laura R. Peck addresses the challenge of ensuring external validity in welfare policy evaluations. It identifies the issue of nonrepresentative site selection in large-scale randomized experiments and proposes a balanced sampling method as a solution. This method involves strategic sample selection, creating strata, and making post hoc adjustments to achieve compositional similarity to a well-defined inference population. The approach aims to improve the generalizability of study findings by accounting for demographic and contextual variations in the target population.

Main Problems in Extrapolation/External Validity

The article identifies several key challenges in ensuring external validity in welfare policy evaluations. One major issue is that traditional random sampling methods are rarely feasible due to high nonresponse rates and logistical constraints, leading to a lack of representativeness in the selected sites. This nonrepresentativeness results in biased estimates that do not generalize well to broader populations.

Methods for Fixing Problems of Extrapolation/External Validity

Tipton and Peck propose a balanced sampling method that does not rely on random sampling. This method involves creating a strategic sample selection plan that focuses on compositional similarity to a well-defined inference population. The approach includes dividing the population frame into strata, ranking units within

these strata, and identifying replacements for non-participating sites. This stratification aims to minimize coverage errors and enable post hoc adjustments to enhance representativeness.

Features of the New Context and/or Target Population Affecting Extrapolation/External Validity

The authors highlight that the success of this design-based approach depends on accurately defining the inference population and the selection of covariates that explain variation in treatment impacts. For welfare policy evaluations, they suggest that relevant features might include state-level policies, local administrative practices, and demographic characteristics of welfare recipients. These features influence the variability in treatment effects across different contexts.

Examples of Problematic Extrapolation/External Validity

The article references the National Evaluation of Welfare to Work Strategies (NEWWS), where the selected sites were not intended to be nationally representative. Despite this, the findings were generalized to predict impacts in broader populations, which may have led to misleading policy conclusions.

Examples of Successful Extrapolation/External Validity

The article does not provide a specific example of successful extrapolation but emphasizes the potential of their proposed method to improve generalizability. By using stratified sampling and targeted recruitment, researchers can create more representative samples that better reflect the diversity of the inference population.

Transferring an Urban Food Security Program

If an urban food security program were to be transferred from one city to another, the article would suggest first defining the inference population and relevant covariates. This involves understanding the specific characteristics of the target population and the context in both cities. The balanced sampling approach would then be used to select sites that are compositionally similar to the target population, ensuring that the findings are generalizable.

Overall, the design-based approach proposed by Tipton and Peck offers a practical solution to improve external validity in welfare policy evaluations. By focusing on strategic sample selection and the use of stratified sampling, the method aims to create more representative samples and enhance the generalizability of study findings.

2.3 A Focus on External Validity (Russell et al., 2007)

The paper by Russell E. Glasgow, Lawrence W. Green, and Alice Ammerman emphasizes the critical importance of external validity in health research. It highlights the need for better reporting and consideration of external validity to ensure that research findings are applicable to real-world settings and diverse populations. The authors provide a detailed framework for evaluating and enhancing external validity in research studies.

Main Problems in Extrapolation/External Validity

The primary problem discussed is the lack of attention to and reporting on external validity in health research. This omission limits the ability of clinicians, public health practitioners, and policymakers to determine the applicability of study results to their specific contexts. The failure to address external validity leads to skepticism about the relevance of research findings and hinders the translation of research into practice.

Methods for Fixing Problems of Extrapolation/External Validity

The authors propose a framework that includes several key elements to improve the reporting and consideration of external validity:

1. Detailed reporting on recruitment and selection procedures, participation rates, and representativeness at multiple levels (individuals, intervention staff, delivery settings).
2. Consistent documentation of the implementation across different program components, settings, staff, and over time.
3. Reporting on a variety of secondary outcomes that are important to different stakeholders, including quality of life, program cost, and adverse consequences.
4. Long-term follow-up reports that address attrition, long-term effects, and program sustainability or modification.

Features of the New Context and/or Target Population Affecting Extrapolation/External Validity

The authors emphasize that various features of the new context and target population can significantly impact the extrapolation of research findings. These include demographic characteristics, local healthcare practices, and the specific needs and conditions of the population. Understanding these features is crucial for determining how well study results can be generalized to different settings.

Examples of Problematic Extrapolation/External Validity

The paper notes that a common issue is the lack of representativeness in study samples, which leads to results that do not generalize well to broader populations. This problem is exacerbated by insufficient reporting on external validity factors, making it difficult to assess the applicability of findings.

Examples of Successful Extrapolation/External Validity

While specific examples of successful extrapolation are not detailed, the paper advocates for systematic reporting and evaluation of external validity to achieve better generalizability. The authors believe that by following their proposed framework, researchers can improve the external validity of their studies.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, the article suggests using the proposed framework to assess the new context's characteristics. This involves evaluating the representativeness of the target population, consistency in implementation, and potential impacts on various outcomes. By systematically addressing these factors, the framework can guide the adaptation of the program to ensure its effectiveness in the new setting.

Overall, the paper underscores the necessity of prioritizing external validity in health research to enhance the translation of evidence into practical applications. By adopting a structured approach to reporting and evaluating external validity, researchers can contribute to more reliable and applicable health interventions.

2.4 A General Algorithm for Deciding Transportability of Experimental Results (Bareinboim & Pearl, 2013)

The article by Elias Bareinboim and Judea Pearl introduces a formal approach to the problem of transportability, which involves transferring causal knowledge from one population (where experimental data is available) to another (where only observational data is available). The authors propose a comprehensive algorithm to determine when and how causal effects can be transported between populations, providing a framework that combines experimental and observational data to yield bias-free estimates of causal relationships.

Main Problems in Extrapolation/External Validity

The primary problem discussed is the difficulty in generalizing causal findings from one population to another due to differences in population characteristics and the

conditions under which data is collected. Traditional methods often fail to account for these differences adequately, leading to biased or invalid conclusions when applied to new populations.

Methods for Fixing Problems of Extrapolation/External Validity

Bareinboim and Pearl propose using causal diagrams and do-calculus to formalize the conditions under which transportability is feasible. They introduce the concept of selection diagrams to represent commonalities and differences between populations. Their algorithm systematically determines whether and how causal effects can be transported by identifying and adjusting for these differences.

Features of the New Context and/or Target Population Affecting Extrapolation/External Validity

Key features include demographic variables, environmental factors, and specific characteristics of the populations involved. The framework requires detailed knowledge of both the source and target populations to identify relevant differences and similarities that affect the transportability of causal effects.

Examples of Problematic Extrapolation/External Validity

The article provides theoretical examples illustrating problematic transportability scenarios, such as when certain demographic or contextual factors differ significantly between populations. These examples highlight the necessity of adjusting for these differences to avoid biased conclusions.

Examples of Successful Extrapolation/External Validity

Successful examples are derived from the application of the proposed algorithm. By using selection diagrams and do-calculus, the authors demonstrate how accurate causal relationships can be inferred even in the presence of population differences, ensuring valid generalizations.

Transferring an Urban Food Security Program

In the context of transferring an urban food security program from one city to another, the article suggests using selection diagrams to identify relevant demographic and contextual differences between the cities. The proposed algorithm would then determine the necessary adjustments to the causal model, ensuring that the program's effectiveness can be accurately predicted and adapted to the new setting.

Overall, the article by Bareinboim and Pearl offers a robust and formalized approach to address the challenges of transportability in causal inference, providing researchers with the tools to make valid generalizations across diverse populations and settings.

2.5 A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results (Kennedy-Martin et al., 2015)

The article by Tessa Kennedy-Martin and colleagues reviews the representativeness of randomized controlled trial (RCT) samples and their implications for the external validity of trial results. The review highlights that RCTs, while the gold standard for determining efficacy, often fail to represent real-world populations due to restrictive eligibility criteria. This lack of representativeness raises concerns about the generalizability of RCT findings to everyday clinical practice.

Main Problems in Extrapolation/External Validity

The primary issue identified is the limited representativeness of RCT samples. RCTs often exclude significant portions of the population, such as the elderly, women, and those with comorbidities, leading to findings that may not apply to real-world settings. This discrepancy can result in biased outcomes that do not reflect the broader patient population.

Methods for Fixing Problems of Extrapolation/External Validity

The authors suggest several methods to improve external validity:

1. Broadening RCT inclusion and exclusion criteria to include a more diverse patient population.
2. Conducting RCTs in more representative settings and populations.
3. Standardizing inclusion/exclusion criteria and diagnostic assessments across studies.
4. Using complementary evidence from observational studies and pragmatic trials to enhance RCT findings.
5. Implementing statistical adjustments and power calculations to ensure subgroup analyses are adequately powered.

Features of the New Context and/or Target Population Affecting Extrapolation/External Validity

Key features impacting external validity include demographic differences, disease severity, comorbid conditions, and variations in treatment practices across different

populations. These factors can significantly influence how well RCT results apply to real-world settings.

Examples of Problematic Extrapolation/External Validity

The review provides multiple examples showing that RCT samples often differ significantly from real-world populations. For instance, in cardiology, patients in RCTs were generally younger, healthier, and less likely to have comorbidities compared to those in routine clinical practice. Similar issues were observed in mental health and oncology trials.

Examples of Successful Extrapolation/External Validity

While the review predominantly highlights issues with external validity, it suggests that broader inclusion criteria and the use of complementary study designs can improve generalizability. Specific successful examples are not detailed, but the recommended strategies are aimed at achieving better representativeness.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, the article would recommend evaluating the demographic and clinical characteristics of the target population. Ensuring that the new population is similar to the original study population in key aspects would help in applying the findings effectively. The use of complementary observational studies to understand the local context better and to validate the program's effectiveness in the new setting would also be advisable.

Overall, the article emphasizes the need for a balanced approach that considers both internal and external validity to ensure that clinical trial results are applicable to real-world settings. This involves methodological adjustments, broader inclusion criteria, and the integration of various types of evidence to provide a comprehensive understanding of treatment effects.

2.6 A New Approach to Argument by Analogy: Extrapolation and Chain Graphs (Steel, 2010)

Daniel Steel (2010) presents a novel approach to extrapolation using analogical reasoning and chain graphs. He proposes this method to address the challenges of making scientific results applicable in different contexts, such as transferring findings from animal models to humans or from experimental settings to real-world applications.

Main Problems in Extrapolation / External Validity

Steel identifies two central challenges in extrapolation: the problem of difference and the extrapolator's circle. The problem of difference arises because there are often causally relevant differences between the model and the target that can affect the outcome of an extrapolation. The extrapolator's circle involves the difficulty of needing evidence to show that the model is similar enough to the target to justify extrapolation, which can seem redundant if the mechanisms in both are already known.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Steel introduces the concept of using chain graphs, which generalize the notion of comparative process tracing. Chain graphs allow for a more nuanced representation of the relationships between variables, using both lines and arrows to indicate connections. This approach emphasizes the role of "fingerprints" or distinctive markers that can help identify whether an analogy is justified without falling into the extrapolator's circle. By focusing on downstream effects that must occur if the mechanism is similar, researchers can more efficiently compare model and target mechanisms.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Steel emphasizes understanding the specific mechanisms and their components that are relevant to the causal claims being extrapolated. By identifying which aspects of the mechanisms need to be similar and which can differ without affecting the outcome, researchers can make more informed decisions about the validity of their extrapolations. This requires detailed knowledge of both the model and the target contexts.

Examples of Problematic Extrapolation / External Validity

Steel discusses how overly simplistic analogies can lead to incorrect extrapolations. For example, the sound of a car engine might lead one to incorrectly assume that two cars have the same type of engine, overlooking other relevant factors such as the type of fuel used. This highlights the need for a deeper understanding of the mechanisms involved to avoid spurious analogies.

Examples of Successful Extrapolation / External Validity

While Steel does not provide extensive examples of successful extrapolations, he illustrates his approach with the case of DNA adducts as markers of exposure to aflatoxin B1. These markers provide a clear link between exposure and effect, allowing for more reliable extrapolations between different organisms.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Using chain graphs to map the mechanisms involved in the food security program and identifying distinctive markers that can be tracked across different urban contexts can help ensure the program's success. By focusing on the specific causal pathways and their downstream effects, policymakers can make more informed decisions about adapting the program to new cities.

Steel's approach emphasizes the importance of understanding the specific mechanisms involved in causal relationships and using chain graphs to represent these mechanisms. By focusing on distinctive markers and the relevant similarities between model and target, researchers can improve the reliability and generalizability of their extrapolations, leading to more effective applications of scientific findings in diverse contexts.

2.7 A Note on Campbell's Distinction Between Internal and External Validity (Hammersley, 1991)

The article by Martyn Hammersley critiques Donald Campbell's distinction between internal and external validity, arguing that the distinction is fundamentally flawed. Hammersley asserts that this separation reflects a misunderstanding of the nature of causal relationships and the generalizability of research findings. He proposes an alternative framework focusing on descriptive and explanatory adequacy.

Main Problems in Extrapolation/External Validity

Hammersley identifies ambiguities in Campbell's definitions of internal and external validity. He argues that the distinction is misleading because it fails to distinguish between relationships among events and relationships among variables. Additionally, Hammersley criticizes the assumption that causal relationships can be discovered independently of their applicability to other cases. This leads to the conclusion that findings cannot be true in one sense (internally valid) but false in another (externally invalid).

Methods for Fixing Problems of Extrapolation/External Validity

Hammersley suggests an alternative conceptualization of validity that avoids the issues he identifies. He proposes focusing on:

1. Descriptive adequacy: This involves ensuring that theoretical variables are accurately represented by the indicators used in research. It includes checking

the accuracy of measurements and the precision needed to detect predicted variations.

2. Explanatory adequacy: This concerns whether the discovered relationships between variables can reasonably be taken to indicate causal relationships as specified in the hypothesis. It involves assessing how well the research design controls for extraneous variables and whether the findings hold in different contexts.

Features of the New Context and/or Target Population Affecting Extrapolation/External Validity

The article suggests that the applicability of findings to new contexts depends on accurately representing theoretical variables and ensuring that the causal relationships hold under different conditions. Factors affecting this include the precision of measurements and the control over extraneous variables.

Examples of Problematic Extrapolation/External Validity

Hammersley discusses the United Nations' information campaign study, which found different results when the original sample was re-interviewed. Campbell interpreted this as internally valid but externally invalid due to an interaction effect. Hammersley argues that the hypothesis itself should be revised to include the sensitizing effect of the initial interview, showing that the original conclusion was not universally true.

Examples of Successful Extrapolation/External Validity

The article does not provide specific examples of successful extrapolation but emphasizes that focusing on descriptive and explanatory adequacy can improve the generalizability of findings.

Transferring an Urban Food Security Program

For transferring an urban food security program, Hammersley's framework would suggest ensuring descriptive adequacy by accurately measuring relevant variables and ensuring explanatory adequacy by controlling for relevant extraneous variables. This approach would help determine if the causal relationships identified in the original context hold true in the new setting.

Overall, Hammersley's article challenges the traditional distinction between internal and external validity, advocating for a more integrated approach that emphasizes accurate measurement and robust causal inference to improve the generalizability of research findings.

2.8 A Protocol for the Extrapolation of 'Best' Practices: How to Draw Lessons from One Experience to Improve Public Management in Another Situation (Ongaro, 2010)

Edoardo Ongaro (2010) proposes a protocol for extrapolating best practices from one context to another, particularly in the field of public management. The paper builds on the concept of "smart practices analysis" by Bardach and integrates recent developments in best practices research to create a systematic approach for transferring successful practices across different settings.

Main Problems in Extrapolation / External Validity

Ongaro identifies significant issues with the transferability of best practices, primarily focusing on the contextual differences that can hinder the effectiveness of extrapolated practices. He emphasizes that the success of practices in their original context does not guarantee success in a new context due to variations in organizational culture, political environment, and specific situational factors. Another major challenge is the lack of a standardized methodology for identifying and adapting best practices to different contexts.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Ongaro presents a five-step extrapolation protocol:

1. **Identify the Function**: Determine the specific function the practice is meant to perform.
2. **Define the Practice**: Clearly describe the practice, including its core components and operational mechanisms.
3. **Describe the System**: Understand how the current system operates and how the practice interacts with the system.
4. **Identify Effects**: Assess the main effects, variations, and possible side effects of the practice.
5. **Define Process Context Factors**: Identify the key context factors under which the practice operates successfully.

This protocol emphasizes a detailed understanding of both the practice and the context in which it will be applied, ensuring that relevant adjustments can be made to fit the new setting.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features affecting the transferability of practices include:

- **Organizational Culture**: Differences in organizational norms and values that can influence the acceptance and effectiveness of new practices.
- **Political Environment**: Variations in political support and stability, which can affect the implementation process.
- **Resource Availability**: Differences in financial, human, and technical resources that can impact the feasibility of transferring practices.
- **Stakeholder Engagement**: Levels of stakeholder involvement and support, which are crucial for successful adaptation and implementation.

Examples of Problematic Extrapolation / External Validity

The paper discusses the complexity of transferring devolution practices in public administration, using the example of the devolution process in Lombardy, Italy. The attempt to transfer similar practices to other regions faced challenges due to differences in local government structures, stakeholder dynamics, and resource availability.

Examples of Successful Extrapolation / External Validity

Ongaro highlights the successful adaptation of devolution practices by leveraging inter-institutional tables and mobilizing structures that facilitated communication and coordination among different levels of government. By adapting these elements to fit the specific context of Lombardy, the devolution process achieved significant momentum and stakeholder support.

Transferring an Urban Food Security Program

While the article does not directly address transferring an urban food security program, the principles discussed can be applied. Using Ongaro's five-step protocol, policymakers can identify the core functions of the food security program, understand how it interacts with the current urban system, and adapt it to the specific context of the new city. By considering organizational culture, political environment, resource availability, and stakeholder engagement, the program can be tailored to ensure its success in the new setting.

Ongaro's protocol offers a structured and detailed approach to the extrapolation of best practices in public management. By focusing on the interplay between practices and their contexts, and by identifying key factors that influence success, researchers and policymakers can enhance the transferability and effectiveness of interventions across different settings.

2.9 "A Review of Generalizability and Transportability (Degtiar & Rose, 2023)

The article by Irina Degtiar and Sherri Rose provides a comprehensive review of methods and frameworks for assessing and addressing the challenges of generalizability and transportability in causal inference. The focus is on improving the external validity of study findings to ensure they can be applied to target populations beyond the study sample.

Main Problems in Extrapolation/External Validity

The primary problems identified include the frequent lack of representativeness of study samples compared to target populations, leading to external validity bias. This bias arises from differences in subject characteristics, settings, treatment implementation, and outcomes between the study and target populations.

Methods for Fixing Problems of Extrapolation/External Validity

The article reviews several methods to address these issues:

1. **Weighting and Matching**: Techniques like inverse probability of participation weighting (IPPW) and propensity score matching are used to adjust for differences in covariate distributions between study and target populations.
2. **Outcome Regression**: Models that predict outcomes based on covariates and treatment assignments, allowing researchers to generalize findings by adjusting for effect modifiers.
3. **Combined Approaches**: Double robust methods that integrate both weighting and outcome regression to improve generalizability and transportability.

Features of the New Context and/or Target Population Affecting Extrapolation/External Validity

Key features include demographic characteristics, baseline health conditions, geographic factors, and variations in healthcare practices. These factors can significantly influence the applicability of study findings to new populations.

Examples of Problematic Extrapolation/External Validity

The article highlights issues such as the underrepresentation of certain demographic groups (e.g., African Americans in cancer RCTs), which can lead to biased conclusions and inappropriate generalizations.

Examples of Successful Extrapolation/External Validity

While specific examples of successful extrapolation are not detailed, the article emphasizes the potential of proposed methods to enhance external validity when properly applied. Techniques such as balancing covariate distributions and robust regression models are advocated to improve the reliability of extrapolated results.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, the article would recommend using weighting and matching methods to ensure the new target population is comparable to the original study population. Additionally, outcome regressions could be employed to adjust for local contextual differences, ensuring that the program's effectiveness is maintained in the new setting.

Overall, the review by Degtiar and Rose underscores the importance of robust methodological approaches to address external validity biases, ensuring that causal inferences drawn from studies can be reliably applied to broader and diverse populations.

2.10 The Tamil Nadu Integrated Nutrition Project: A Review with Special Emphasis on the Monitoring and Information System (Shekar, 1991)

Meera Shekar (1991) reviews the Tamil Nadu Integrated Nutrition Project (TINP), focusing on its components, success factors, and monitoring systems. The review assesses TINP's effectiveness and offers insights into its successful design and implementation features.

Main Problems in Extrapolation / External Validity

Shekar identifies several challenges related to the external validity of TINP. A primary issue is the variability in local conditions and the presence of multiple overlapping nutrition programs, such as the Noon Meal Program (NMP), which complicates attributing improvements solely to TINP. Furthermore, the data's credibility is occasionally questioned due to inconsistencies and the need for rigorous statistical analysis to validate findings.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Shekar emphasizes the need for thorough and rigorous statistical analysis of existing data to better understand TINP's impact. She also advocates for detailed multivariate analyses to identify specific program inputs contributing to improved nutritional status. Additionally, building institutional capacity

for program evaluation is crucial to ensure robust and reliable data collection and analysis.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features affecting the transferability of TINP include the socio-economic status of the target population, the local political and administrative environment, and the existing infrastructure for health and nutrition services. Understanding these factors is essential for adapting TINP to new contexts effectively.

Examples of Problematic Extrapolation / External Validity

Shekar points out that many TINP beneficiaries also participated in the NMP, which provided supplementary nutrition to older children. This overlap complicates the attribution of nutritional improvements to TINP alone, especially in children older than two years. Additionally, the reported impact on different age groups within TINP shows inconsistencies, suggesting that some observed improvements might be due to factors other than TINP.

Examples of Successful Extrapolation / External Validity

The review suggests that despite methodological limitations, the TINP data indicate a significant positive impact on the nutritional status of children, particularly those younger than two years, who were not covered by the NMP. This improvement suggests that TINP's focused interventions were effective in this specific demographic.

Transferring an Urban Food Security Program

While the review does not directly address the transfer of an urban food security program, the principles discussed are applicable. When considering such a transfer, policymakers should ensure detailed statistical analysis and context-specific adaptation. Key steps would include:

- Conducting a thorough needs assessment to understand the specific challenges and resources of the new urban setting.
- Leveraging existing data and conducting pilot studies to identify potential barriers and facilitators to program success.
- Ensuring strong institutional support and building capacity for ongoing monitoring and evaluation.
- Engaging local stakeholders and tailoring program components to fit the socio-economic and cultural context of the new city.

In summary, Shekar's review highlights the importance of rigorous data analysis, contextual adaptation, and strong institutional support for successfully transferring and scaling nutrition programs like TINP. These principles are crucial for ensuring the external validity and effectiveness of such programs in new settings.

2.11 A Simple Approximation for Evaluating External Validity Bias (Andrews & Oster, 2019)

Isaiah Andrews and Emily Oster (2019) develop a method to assess and correct for external validity bias in randomized trials. Their approach focuses on understanding and adjusting for selection biases that arise from both observable and unobservable variables.

Main Problems in Extrapolation / External Validity

The primary issue identified by Andrews and Oster is the difference between the experimental sample and the target population. This difference arises because participants in randomized trials often self-select based on observable and unobservable factors, leading to biased estimates of treatment effects. The paper emphasizes that even if observable differences can be adjusted, unobserved factors may still bias the results.

Methods for Fixing Problems of Extrapolation / External Validity

The authors propose a framework that approximates the bias introduced by unobservable factors. They suggest using reweighting techniques to adjust for observable differences and developing approximations for the impact of unobservable variables on participation decisions. This involves calculating the covariance between the treatment effect and the selection weights, and using these approximations to provide benchmarks for assessing external validity.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features affecting extrapolation include the demographic and socio-economic characteristics of the target population, as well as the nature of participation decisions. The framework developed by Andrews and Oster takes into account how both observable and unobservable factors influence the likelihood of participation in the trial, and how these factors impact the treatment effect.

Examples of Problematic Extrapolation / External Validity

The paper discusses instances where treatment effects from a randomized trial are not generalizable due to significant differences between the trial participants and the broader population. For example, the productivity gains observed in a working-from-home experiment may not apply to all employees if the trial participants were self-selected based on their ability to self-motivate.

Examples of Successful Extrapolation / External Validity

While specific examples of successful extrapolation are not detailed, the framework itself is designed to improve the likelihood of successful extrapolation by systematically addressing both observable and unobservable selection biases.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. The framework proposed by Andrews and Oster suggests the following steps:

1. **Assessing Observables**: Collect and analyze detailed data on the target city's demographic and socio-economic characteristics. Use reweighting techniques to adjust for these observable differences.
2. **Estimating Unobservables**: Develop approximations for the potential biases introduced by unobservable factors. This involves estimating the covariance between the expected treatment effect and the participation decision weights.
3. **Applying Sensitivity Analysis**: Conduct sensitivity analysis to determine how variations in unobservable factors might affect the program's outcomes. This helps in understanding the robustness of the program's effects when transferred to a new city.
4. **Pilot Testing and Iterative Adjustments**: Implement pilot programs in the new city to collect preliminary data and make iterative adjustments based on the observed outcomes and contextual factors.

By systematically addressing both observable and unobservable factors, policymakers can better estimate and mitigate the biases that may affect the program's success in the new urban setting.

In summary, Andrews and Oster's framework provides a robust method for evaluating and correcting external validity bias, offering valuable insights for transferring programs across different contexts. By carefully considering both observable and unobservable factors, their approach enhances the reliability and applicability of experimental findings in new settings.

2.12 A Typology of Useful Evidence: Approaches to Increase the Practical Value of Intervention Research (Hasson et al., 2020)

Henna Hasson, Laura Leviton, and Ulrica von Thiele Schwarz (2020) present a typology aimed at enhancing the practical value of intervention research by proposing approaches that improve the usefulness of evidence-based interventions (EBIs) for practitioners, policymakers, and other end users.

Main Problems in Extrapolation / External Validity

Hasson et al. identify that many studies on EBIs are not sufficiently useful for end users because they often lack detailed information about the intervention's core components, implementation strategies, context, and outcomes. This lack of information hampers the ability of practitioners to adapt and implement these interventions effectively in different settings.

Methods for Fixing Problems of Extrapolation / External Validity

The authors propose a typology with three levels—Description, Analysis, and Design—each addressing different aspects of intervention research:

1. **Description**: Providing detailed information about the intervention, its core components, and how it was implemented. This includes describing the context and the outcomes valued by end users.
2. **Analysis**: Using various analytic methods to understand how and why an intervention works, identifying the impact of core components and implementation strategies, and analyzing the context's moderating effects.
3. **Design**: Developing interventions with usefulness in mind from the outset, incorporating pragmatic trials, and ensuring that interventions can be effectively implemented in real-world settings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The typology emphasizes the need to understand the specific context in which the intervention will be implemented. This includes both the inner context (organizational structure, culture) and the outer context (political, economic, and social factors). Understanding these factors helps tailor the intervention to fit the new setting, increasing the likelihood of success.

Examples of Problematic Extrapolation / External Validity

The authors discuss the failure of the Nurse Family Partnership (NFP) program when transferred from the US to the UK. Despite being one of the best-documented EBIs in public health, the UK trial did not replicate the positive effects seen in the US, likely due to differences in existing services and implementation contexts.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not extensively detailed, the paper emphasizes that following the proposed typology can improve the generalizability and usefulness of EBIs. This involves detailed descriptions, rigorous analysis, and thoughtful design tailored to specific contexts.

Transferring an Urban Food Security Program

Although the article does not specifically address transferring an urban food security program, the principles discussed are highly applicable. The typology suggests:

1. **Detailed Descriptions**: Clearly document the food security program's core components and implementation strategies. Ensure that the description includes context-specific adaptations.
2. **Context Analysis**: Conduct thorough analysis of the new urban setting, including socio-economic conditions, organizational structures, and community needs. Use frameworks like the Consolidated Framework for Implementation Research (CFIR) to guide this analysis.
3. **Pragmatic Trials and Adaptive Design**: Implement the program in the new city through pragmatic trials that test the program's effectiveness in real-world conditions. Adapt the program based on ongoing feedback and outcomes.
4. **Engage End Users**: Involve local stakeholders, including community members and policymakers, in the program's adaptation and implementation. Ensure that the outcomes measured are relevant to the new context.

By following these steps, policymakers can enhance the likelihood of successfully transferring and implementing an urban food security program in a new city. This approach ensures that the program is not only evidence-based but also contextually appropriate and practically useful.

Hasson et al. emphasize the importance of integrating practical considerations into the research and implementation process. By doing so, they aim to bridge the gap between evidence generation and real-world application, ultimately making EBIs more useful and effective across different settings.

2.13 Across the Boundaries: Extrapolation in Biology and Social Science (Steel, 2007)

Daniel P. Steel's book, "Across the Boundaries: Extrapolation in Biology and Social Science," explores the methodologies and challenges of extrapolating causal relationships from one context to another. The work delves into the principles and practices needed to ensure that findings from one population or setting can be reliably applied to another, emphasizing the role of mechanisms in establishing the validity of such extrapolations.

Main Problems in Extrapolation / External Validity

Steel identifies several central problems in extrapolation, primarily the heterogeneity of populations and the extrapolator's circle. The heterogeneity problem highlights the differences in causal mechanisms between populations, making it difficult to generalize findings. The extrapolator's circle refers to the paradox where establishing the validity of an extrapolation requires prior knowledge of the target population, which would make the extrapolation redundant.

Methods for Fixing Problems of Extrapolation / External Validity

Steel proposes a mechanisms-based approach to address these issues. This involves understanding and mapping the causal mechanisms in both the model and target populations. Comparative process tracing is a key method recommended by Steel, which involves identifying stages in the causal process where significant differences are likely to occur and focusing on downstream stages where these differences are minimal. This helps in establishing a basis for reliable extrapolation despite some differences between the model and target.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Understanding the specific causal structures and mechanisms in the new context is crucial. Factors such as genetic, environmental, and social differences can impact the applicability of the findings. Steel emphasizes the need for detailed knowledge about these factors to ensure that the extrapolation does not overlook critical differences that could affect the outcome.

Examples of Problematic Extrapolation / External Validity

Steel discusses the example of animal models used for testing carcinogenic substances. He points out that while these models provide valuable insights, the differences between species can lead to significant errors in extrapolation. The

extrapolator's circle is particularly problematic in such cases, where detailed knowledge of human biology is needed to validate findings from animal studies.

Examples of Successful Extrapolation / External Validity

While the book does not provide extensive examples of successful extrapolations, Steel emphasizes the importance of a rigorous mechanisms-based approach that can lead to more reliable extrapolations. The case of using aflatoxin B1 as a carcinogen in rodents to infer its effects in humans is discussed, illustrating how focusing on conserved mechanisms can aid in successful extrapolation.

Transferring an Urban Food Security Program

Steel's mechanisms-based approach can be applied to the transfer of an urban food security program. Key steps would include:

1. **Mapping Mechanisms**: Identify the causal mechanisms that make the program successful in the original city, such as community engagement, supply chain logistics, and nutritional education.
2. **Comparative Process Tracing**: Analyze the new city's context to identify potential differences in these mechanisms. This could involve understanding local dietary habits, economic conditions, and existing infrastructure.
3. **Pilot Testing and Iterative Adjustments**: Implement the program on a small scale in the new city and make iterative adjustments based on observed outcomes. This helps in adapting the program to fit the new context effectively.

By focusing on the specific mechanisms and their functioning in different contexts, policymakers can enhance the likelihood of successfully transferring and implementing an urban food security program. Steel's approach underscores the importance of understanding underlying causal processes to ensure that interventions are effective across different settings.

2.14 Against External Validity (Reiss, 2019)

Julian Reiss in his article "Against External Validity" challenges the traditional focus on external validity in causal inference, arguing that it encourages poor evidential reasoning. Reiss proposes an alternative approach that emphasizes direct reasoning about target systems and highlights the contextualist methodology over the foundationalist approach that prioritizes certainty over relevance.

Main Problems in Extrapolation/External Validity

The main problem identified is that framing causal inference in terms of external validity leads to poor evidential reasoning. Reiss argues that this approach mistakenly prioritizes establishing analogous claims in model systems before making inferences about target systems, which is often unnecessary and misguiding.

Methods for Fixing Problems of Extrapolation/External Validity

Reiss advocates for a contextualist approach that focuses on directly establishing hypotheses about target systems and identifying the types of evidence needed to support those hypotheses. This approach integrates various sources of evidence and emphasizes the role of direct and indirect support for causal claims without relying heavily on external validity considerations.

Features of the New Context and/or Target Population Affecting Extrapolation/External Validity

Key features include specific contextual factors such as the socio-economic environment, cultural aspects, and the particularities of the population under study. Understanding these contextual elements is crucial for making reliable causal inferences directly about the target system.

Examples of Problematic Extrapolation/External Validity

Reiss critiques the reliance on animal models to infer human carcinogenicity, noting that differences between species can lead to unreliable extrapolations. He also discusses the general issues with applying findings from controlled experimental settings to more complex, real-world environments.

Examples of Successful Extrapolation/External Validity

The article does not provide specific examples of successful extrapolation but emphasizes that when appropriate evidence is directly gathered and analyzed in the context of the target system, more reliable inferences can be made.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, Reiss's approach would suggest directly formulating hypotheses about the new context and gathering relevant evidence to support those hypotheses. This would involve understanding the specific socio-economic and cultural factors of the new city and using this information to adapt and implement the program effectively without over-relying on the findings from the original context.

Overall, Reiss's work argues for a shift from traditional notions of external validity to a more pragmatic, context-based approach to causal inference, aiming for relevance and reliability in real-world applications.

2.15 Assessing External Validity (Bo & Galiani, 2021)

In their paper "Assessing External Validity," Hao Bo and Sebastian Galiani propose a formal, data-driven approach for evaluating the external validity (EV) of randomized controlled trials (RCTs). They argue that while RCTs are effective at addressing internal validity, the generalizability of their results to other populations is often uncertain. The authors develop a methodology to measure the stability of causal relationships across different populations by reweighting samples and assessing the degree to which findings hold under these adjustments.

Main Problems in Extrapolation/External Validity

The main problem identified is the challenge of ensuring that causal relationships established in one population (through RCTs) are applicable to other populations. This issue arises because of differences in population characteristics, leading to instability in the conditional distribution of outcomes given treatment across different populations.

Methods for Fixing Problems of Extrapolation/External Validity

Bo and Galiani propose a method that involves reweighting the original sample to create new samples representing different populations. This method allows for the evaluation of whether the original conclusions hold in these reweighted samples. The approach includes:

1. **Reweighting**: Creating multiple reweighted samples from the original RCT data to simulate different populations.
2. **Assessment**: Calculating the average treatment effect (ATE) and its standard error for these new samples to see if the original findings remain significant and consistent.
3. **Global and Local Measures**: Introducing global measures of EV to evaluate how often the original findings hold across all new populations and local measures to assess EV in populations similar to the original.

Features of the New Context and/or Target Population Affecting Extrapolation/External Validity

Key features include demographic variables, socioeconomic factors, and baseline conditions of the new population. The method requires understanding how these variables influence the causal relationships under study and adjusting for them in the reweighted samples.

Examples of Problematic Extrapolation/External Validity

The paper discusses the general issue of variability in treatment effects across different populations. They cite the difficulty in making generalizations from trials conducted in specific contexts to broader or different populations without proper adjustments.

Examples of Successful Extrapolation/External Validity

While the paper does not provide specific empirical examples, it illustrates the application of their methodology through simulated examples. These simulations show how their measures can help determine the extent to which the original RCT results are likely to be valid in new populations.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, Bo and Galiani's method would suggest creating reweighted samples based on the characteristics of the new city's population. By assessing the stability of the original program's outcomes in these reweighted samples, one can determine whether the program's effects are likely to be replicated in the new setting. This involves analyzing demographic, economic, and contextual variables to ensure the program's effectiveness is maintained.

Overall, the paper provides a robust framework for assessing external validity, emphasizing the importance of data-driven approaches to evaluate the generalizability of causal findings from RCTs.

2.16 Assessing Methods for Generalizing Experimental Impact Estimates to Target Populations (Kern et al., 2016)

Holger L. Kern, Elizabeth A. Stuart, Jennifer Hill, and Donald P. Green (2016) evaluate various statistical methods for generalizing the impact estimates from randomized experiments to broader target populations. Their work addresses the increasing demand to apply experimental findings beyond the specific contexts of the original studies, particularly in fields such as education, sociology, and public health.

Main Problems in Extrapolation / External Validity

The primary challenge discussed by Kern et al. is the discrepancy between the experimental sample and the target population. This issue arises because the experimental participants often do not represent the broader population, leading to potential biases when generalizing findings. The paper emphasizes the difficulties in dealing with both observed and unobserved differences between the experimental sample and the target population.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Kern et al. investigate several statistical methods. They focus on propensity score weighting, which adjusts the experimental data to resemble the target population by reweighting based on propensity scores. They also explore Bayesian Additive Regression Trees (BART), a flexible modeling approach capable of handling complex interactions and non-linear relationships. Another method is Inverse Probability of Treatment Weighting (IPTW), which uses weights derived from propensity scores to adjust for differences between the sample and the target population. Additionally, Double Robust (DR) methods combine outcome modeling with propensity score adjustments to account for both observed and unobserved confounders. The authors find that flexible modeling approaches, particularly BART, generally outperform traditional regression-based methods when the necessary assumptions are met.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features impacting generalization include the socio-economic and demographic characteristics of the target population, which can differ significantly from those of the experimental sample. Contextual factors such as local policies, cultural norms, and infrastructure also influence the effectiveness of interventions when applied in new settings.

Examples of Problematic Extrapolation / External Validity

The complexity of generalizing findings from school-based dropout prevention programs to different educational settings highlights the difficulties in achieving reliable extrapolation. Significant variations in school environments, student populations, and local policies can lead to different outcomes.

Examples of Successful Extrapolation / External Validity

While the paper does not provide specific examples of successful extrapolation, it emphasizes that the use of advanced statistical methods can improve the reliability

of generalizations. Methods like BART and DR approaches, when correctly applied, can help mitigate the biases associated with extrapolation.

Transferring an Urban Food Security Program

Kern et al.'s findings suggest several steps for transferring an urban food security program from one city to another. Policymakers should gather comprehensive data on the target city's socio-economic characteristics and existing food security infrastructure. Applying propensity score weighting can help adjust for observed differences between the original and target cities. Utilizing flexible modeling approaches, such as BART, allows for accounting for complex interactions and non-linear relationships in the data. Conducting pilot studies in the new city to test the program and making necessary adjustments based on initial findings can further enhance the program's success.

By systematically addressing both observed and unobserved factors, policymakers can enhance the likelihood of successfully transferring and implementing the food security program in a new urban context. In summary, Kern et al. provide a robust evaluation of statistical methods to improve the generalizability of experimental findings. Their work underscores the importance of rigorous data analysis and the use of flexible modeling techniques to address the challenges of extrapolation and ensure effective application across diverse settings.

2.17 Assessing the applicability of public health interventions (Burchett et al., 2018)

Burchett et al. (2018) present a detailed examination of the applicability of public health interventions to different contexts. They discuss the increasing interest in generalisability and applicability, highlighting the lack of guidance and consistent criteria for assessing these aspects in existing literature. The paper reviews various tools designed to assess the applicability of interventions, noting the variability in their use, interpretation, and effectiveness.

Main Problems in Extrapolation / External Validity:

The paper highlights the challenges in assessing whether public health interventions evaluated in one context can be generalized or applied to another. These challenges stem from the complexity and context-dependency of such interventions. There is a notable lack of detailed guidance on assessing applicability, leading to inconsistent and often subjective evaluations by different users.

Methods for Fixing Problems of Extrapolation / External Validity:

The authors suggest that tools with set criteria for assessing applicability are not ideal. They argue that a standardized list of criteria is impractical due to the limitless number of potential factors affecting implementation and effectiveness. Instead, they propose that tools should focus on relevant factors specific to the new context, guided by high-level conceptual judgments and informed by empirical evidence.

Features of the New Context and/or Target Population:

The paper emphasizes the importance of understanding the new context, which includes anything external to the intervention that may act as a barrier or facilitator. This encompasses political, economic, and social contexts, as well as organizational characteristics and modifiable factors like culture and climate.

Examples of Problematic Extrapolation/External Validity:

The study references instances where tools led users to judge interventions as not applicable to new contexts. This often resulted from a focus on differences between settings or populations without considering both similarities and differences.

Examples of Successful Extrapolation/External Validity:

The paper notes that some tools supported judgments of reasonable applicability by focusing on high-level conceptual judgments and mechanisms of action. These tools were less structured and allowed for more flexible interpretation.

Transferring an Urban Food Security Program from One City to Another:

For transferring an urban food security program, Burchett et al. would suggest a careful assessment of the new context, including political, economic, and social factors, as well as organizational characteristics. They would recommend using tools that focus on high-level judgments and mechanisms of action to determine applicability. Additionally, considering potential modifications to the program to overcome barriers in the new setting would be crucial.

This approach aligns with the need for a systematic yet flexible evaluation of context-specific factors, ensuring that the intervention can be adapted and effectively implemented in the new urban environment.

2.18 Beyond 'Context Matters': Context and External Validity in Impact Evaluation (Williams, 2020)

Martin J. Williams (2020) examines the significance of context in determining the external validity of impact evaluations. He proposes a framework to help policymakers

effectively integrate evidence from different contexts by focusing on the interaction between policy mechanisms and contextual features.

Main Problems in Extrapolation / External Validity

Williams highlights that the primary challenge in external validity is understanding how context influences the effectiveness of a policy. While experimental methods provide valuable evidence, they often fail to account for contextual differences that can alter outcomes. The key issue is that aspects of context that interact with a policy's mechanism can significantly affect its success when transported to a new setting. Without a thorough understanding of these interactions, policymakers risk implementing interventions that do not perform as expected.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Williams introduces the concept of mechanism mapping. This method involves setting out the policy's mechanism or theory of change and identifying the contextual assumptions required for each causal step to occur. Policymakers then compare these assumptions to the actual characteristics of the target context. This iterative process helps identify which parts of the policy need adaptation and which can remain unchanged. By systematically assessing the interaction between mechanisms and context, policymakers can better predict the success of interventions in new settings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key contextual features include the local organizational structure, cultural norms, existing policies, and socio-economic conditions. These factors can significantly impact how a policy's mechanism operates in the new context. Understanding these interactions is crucial for adapting policies to ensure their effectiveness in different settings.

Examples of Problematic Extrapolation / External Validity

Williams illustrates the challenges of extrapolation with the example of the Tamil Nadu Integrated Nutrition Programme (TINP) and its attempted replication in Bangladesh. Although both regions faced similar malnutrition problems, the differing roles of women in household food allocation led to different outcomes. In Tamil Nadu, mothers who received nutritional advice could directly apply it, while in Bangladesh, where men and mothers-in-law controlled food distribution, the same advice did not translate into improved nutrition.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not detailed, the framework proposed by Williams aims to improve the likelihood of successful extrapolation. By using mechanism mapping and focusing on context-mechanism interactions, policymakers can more accurately predict and enhance the effectiveness of policies in new settings.

Transferring an Urban Food Security Program

Applying Williams' framework to the transfer of an urban food security program involves several steps. Policymakers should start by clearly defining the program's mechanisms, such as food distribution, nutritional education, and community engagement. They must then identify the contextual assumptions necessary for these mechanisms to function effectively, considering factors like local governance, cultural practices, and socio-economic conditions. Comparing these assumptions with the actual characteristics of the new urban setting can highlight potential areas needing adaptation. Conducting pilot studies and iterating based on feedback will further refine the program to fit the new context.

By systematically evaluating the interaction between policy mechanisms and contextual features, Williams' approach helps ensure that interventions are effectively adapted and implemented in new settings. This enhances the external validity and practical applicability of evidence-based policies across diverse contexts.

2.19 Beyond External Validity (Calder et al., 1983)

The article "Beyond External Validity" by Bobby J. Calder, Lynn W. Phillips, and Alice M. Tybout critiques the conventional emphasis on external validity in theoretical research. The authors argue that the applicability of behavioral research should not be pursued through attempts to incorporate "real-world" variables into theory-testing studies. Instead, they propose that applicability is best achieved through rigorous testing of theories and the subsequent evaluation of theory-based interventions.

Main Problems in Extrapolation/External Validity

The primary problem identified is the overemphasis on external validity in individual theory-testing studies. Calder et al. argue that such an emphasis often leads to inappropriate inclusion of atheoretical variables, which detracts from the objective of developing applicable theories through scientific progress.

Methods for Fixing Problems of Extrapolation/External Validity

The authors propose a two-stage procedure for developing applicable theories:

1. **Rigorous Theory Testing**: Theories should be subjected to rigorous tests in controlled settings to identify and incorporate missing variables through a process of falsification and refinement.
2. **Intervention Testing**: After a theory has been rigorously tested, it should be used to design and test interventions in specific real-world settings. This approach ensures that the interventions are evaluated based on the practical context and specific background factors relevant to the setting.

Features of the New Context and/or Target Population Affecting Extrapolation/External Validity

The authors emphasize that background factors, which include specific contextual elements such as socio-economic conditions and cultural factors, should be considered during the intervention testing stage. These factors are essential for evaluating the applicability of interventions in new contexts.

Examples of Problematic Extrapolation/External Validity

Calder et al. critique the approach suggested by Lynch, which involves blocking on a limited number of background factors during theory testing. They argue that this approach relies on induction and may not provide a reliable basis for inferring external validity.

Examples of Successful Extrapolation/External Validity

The authors do not provide specific examples of successful extrapolation but argue that successful applicability is achieved through the rigorous testing of theories and subsequent focused evaluation of interventions.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, Calder et al.'s approach would recommend first ensuring that the underlying theoretical framework has been rigorously tested and refined. Then, the program should be adapted and tested in the new context, taking into account specific socio-economic and cultural factors. This two-stage process helps ensure that the intervention is both theoretically sound and practically applicable in the new setting.

Overall, Calder et al.'s article advocates for a methodological shift from emphasizing external validity in theory testing to focusing on rigorous theory testing followed by context-specific intervention evaluation. This approach aims to achieve practical applicability through scientific progress and targeted evaluation.

2.20 Against External Validity (Reiss, 2019)

Reiss's (2019) paper critiques the notion of external validity in scientific research, arguing that it is often overemphasized or misunderstood. He asserts that external validity, while important, should not overshadow the need for robust internal validity and should be contextualized based on specific research objectives and practical implications. Reiss also highlights that the practical relevance of scientific findings often depends more on the specific context and less on generalized external validity claims.

Main Problems in Extrapolation/External Validity

Reiss identifies several problems with the concept of external validity. One major issue is that it is often treated as a static property of a study rather than a context-dependent judgment. He argues that the external validity of a study's findings can vary significantly depending on the context in which the findings are applied. Reiss also criticizes the tendency to prioritize external validity over internal validity, suggesting that this can lead to weaker scientific conclusions.

Methods for Fixing Problems of Extrapolation/External Validity

To address these issues, Reiss advocates for a more nuanced approach that considers the specific context and practical objectives of the research. He suggests that researchers should focus on the mechanisms underlying observed phenomena and understand how these mechanisms might operate differently in various contexts. This approach can help to identify the conditions under which findings might be validly extrapolated.

Features of the New Context and/or Target Population Affecting Extrapolation/External Validity

Reiss emphasizes the importance of considering factors such as cultural, social, economic, and institutional differences when extrapolating findings to new contexts. These contextual factors can significantly influence whether and how research findings apply to different populations or settings.

Examples of Problematic Extrapolation/External Validity

Reiss provides examples where research findings have failed to generalize across different contexts, highlighting the limitations of assuming universal applicability. These examples underscore the need for careful consideration of contextual factors in assessing external validity.

Examples of Successful Extrapolation/External Validity

While Reiss is critical of over-reliance on external validity, he acknowledges that successful extrapolation is possible when researchers carefully account for contextual differences and underlying mechanisms. He suggests that studies that explicitly address these factors are more likely to achieve valid extrapolation.

Transferring an Urban Food Security Program from One City to Another

Reiss would likely suggest a careful analysis of the contextual factors unique to each city, such as local food supply chains, economic conditions, and cultural practices related to food security. By understanding these specific factors, policymakers can better adapt the program to ensure its success in the new context.

Overall, Reiss advocates for a more flexible and context-sensitive approach to external validity, emphasizing the importance of understanding the mechanisms at play and the specific conditions under which research findings can be validly extrapolated.

2.21 Can understanding mechanisms solve the problem of extrapolating from study to target populations (the problem of 'external validity')? (Howick et al., 2013)

The article by Howick, Glasziou, and Aronson explores whether a deeper understanding of mechanisms can address the problem of extrapolating findings from study populations to target populations, known as the issue of external validity. The authors discuss the limitations of traditional methods for generalizing study results and propose a framework for using mechanistic knowledge to improve the reliability of extrapolations.

Main Problems in Extrapolation/External Validity

The primary problem identified is that average study results do not always apply to target populations. This issue arises due to differences in contextual factors, populations, and mechanisms between the study and target settings. These differences can lead to misleading or harmful applications of study findings when generalized without considering the underlying mechanisms.

Methods for Fixing Problems of Extrapolation/External Validity

Howick et al. propose using mechanistic knowledge to justify extrapolation. Their approach involves:

1. **Identifying Mechanisms**: Understanding the causal chain linking the intervention with outcomes.

2. **Comparing Mechanisms**: Ensuring that the mechanisms in the study and target populations are similar.
3. **Contextual Analysis**: Evaluating whether the contexts in which the mechanisms operate are sufficiently similar.

Features of the New Context and/or Target Population Affecting Extrapolation/External Validity

Key features include genetic factors, age- and sex-related factors, physiological variants, comorbidities, drug-drug interactions, and sociological factors. These features can influence the applicability of study findings to new populations by affecting the mechanisms of action.

Examples of Problematic Extrapolation/External Validity

The article provides examples such as the use of antiarrhythmic drugs, which were initially thought to reduce mortality based on their mechanism of action in clinical trials but later found to increase mortality due to unanticipated mechanisms in broader populations. Similarly, the Tamil Nadu Integrated Nutrition Program succeeded in its original context but failed when implemented in Bangladesh due to differing social structures.

Examples of Successful Extrapolation/External Validity

The authors mention successful extrapolations where mechanistic knowledge was correctly applied, such as the use of macrogols for treating constipation in children. The mechanism of action was known to be age-independent, justifying its use even in younger children despite the lack of direct trial evidence.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, Howick et al.'s approach would suggest a thorough analysis of the mechanisms underlying the program's success and comparing these mechanisms in the new context. This involves assessing genetic, demographic, and sociological factors that could affect the program's outcomes. Ensuring that the mechanisms and contextual factors are sufficiently similar would help in making a justified and successful extrapolation.

Overall, the article emphasizes the importance of mechanistic knowledge in improving the reliability of extrapolating study findings to new populations, advocating for a systematic approach to understand and compare mechanisms and contexts.

2.22 Causal Interaction and External Validity: Obstacles to the Policy Relevance of Randomized Evaluations (Muller, 2015)

In his 2015 paper, Muller addresses the challenges of extrapolating findings from randomized control trials (RCTs) to broader policy contexts. He critiques the common reliance on atheoretical replication and highlights the importance of understanding interacting factors to achieve credible external validity. Muller argues that better knowledge of these factors and empirical data is crucial for making reliable policy claims based on experimental results.

Main Problems in Extrapolation/External Validity

Muller identifies several issues with extrapolation in RCTs, emphasizing the problem of interaction effects. These interactions, if not properly understood and accounted for, can lead to inaccurate predictions when applying findings to different populations or contexts. Without proper identification of interacting variables, results from one context may not hold true in another.

Methods for Fixing Problems of Extrapolation/External Validity

Muller suggests addressing these problems requires detailed knowledge of the interacting factors and their empirical distributions in both the experimental and target populations. He highlights "conditional external validity," which involves using covariate information to adjust treatment effects accordingly. This approach is similar to methods used in observational studies to account for selection bias.

Features of the New Context and/or Target Population Affecting Extrapolation/External Validity

Muller stresses the importance of contextual factors such as social institutions and historical differences that can fundamentally alter the applicability of results. Variations in social structures can interact with treatment variables in ways that make straightforward extrapolation impossible.

Examples of Problematic Extrapolation/External Validity

The paper cites the example of school class size interventions, where the effect of class size on student outcomes can be influenced by teacher quality. If the average teacher quality differs between the experimental and policy populations, the extrapolated results may be misleading.

Examples of Successful Extrapolation/External Validity

Muller does not provide specific examples of successful extrapolation but emphasizes that achieving this requires stringent empirical and theoretical groundwork to identify and measure relevant interacting factors.

Transferring an Urban Food Security Program

Muller's approach would suggest a thorough assessment of the interacting factors in both cities. Factors such as economic conditions, social institutions, and population demographics should be carefully compared and adjusted for to ensure the intervention's effectiveness in the new context. He advocates for a detailed empirical analysis to identify and measure these factors before extrapolation.

Overall, Muller's work emphasizes the importance of understanding and accounting for interaction effects in achieving credible external validity, advocating for a detailed, context-specific approach to policy extrapolation.

2.23 Causality is Good for Practice: Policy Design and Reverse Engineering (Busetti, 2023)

The paper by Busetti (2023) delves into how causal mechanisms can be utilized to improve policy design and practice. The core argument emphasizes the necessity of understanding and leveraging causal mechanisms to facilitate intelligent replications, program adjustments, and the design of new interventions. The paper discusses the limits of standard program designs across different contexts and how mechanisms can inform modifications to fit target contexts better.

Main Problems in Extrapolation / External Validity

The primary issue discussed is the difficulty in replicating the success of a policy program in different contexts. This stems from contextual discrepancies that can significantly alter the program's effectiveness. Understanding which elements are crucial and which can be disregarded without losing the program's effectiveness is a major challenge.

Methods for Fixing Problems of Extrapolation / External Validity

Busetti suggests using causal mechanisms as a solution. By abstracting the causal powers and understanding how these can be exercised in the target context, designers can create functional equivalents that replicate the desired outcomes. Techniques like "mechanism mapping" can help predict the impact of transporting policies by comparing contextual assumptions.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The paper highlights that new contexts may have unique features, some of which might be unknown initially, that could impair the program's success. These features can range from cultural practices to social dynamics, such as the role of mothers-in-law in household food allocation in Bangladesh, which differed from the Tamil Nadu context.

Examples of Problematic Extrapolation / External Validity

An example given is the Tamil Nadu nutrition program, which was unsuccessful when replicated in Bangladesh. This failure was due to the different social structures regarding household food control, demonstrating how critical contextual differences can lead to unsuccessful extrapolations.

Examples of Successful Extrapolation / External Validity

The paper does not explicitly mention successful examples but implies that successful extrapolation is possible when causal mechanisms are properly understood and adapted to the new context.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, Busetti would likely suggest a detailed assessment of the target context using causal mechanisms. This involves identifying the key causal powers that led to success in the original context and ensuring that these can be replicated in the new context, possibly by adjusting certain elements of the program or incorporating additional design features to address new contextual challenges.

Overall, Busetti's work underlines the importance of tailoring interventions to fit specific contexts rather than blindly replicating programs, thereby enhancing the likelihood of successful policy implementation across different environments.

2.24 Checklist for the Qualitative Evaluation of Clinical Studies with Particular Focus on External Validity and Model Validity (Bornhöft et al., 2006)

Bornhöft et al. (2006) present a comprehensive checklist for evaluating clinical studies, emphasizing external validity and model validity. The authors argue that external validity is often overlooked in clinical trials, which can distort conclusions about the applicability of study findings to broader populations. Their checklist aims

to systematize and operationalize the assessment of external and model validity to improve the generalizability and practical relevance of clinical research.

Main Problems in Extrapolation / External Validity

The main problem identified is the frequent neglect of external validity in clinical studies. This oversight can lead to biased conclusions about the generalizability of study results. The authors highlight that the representativeness of the study population, intervention realism, and relevance of outcome measures are critical factors often inadequately addressed.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Bornhöft et al. propose a checklist that includes:

1. **Assessment of Selection Bias**: Evaluating whether the study population represents the target population in terms of demographic and clinical characteristics.
2. **Assessment of Performance Bias**: Ensuring that the intervention reflects typical clinical practice, including dosage, administration, and duration.
3. **Assessment of Detection and Attrition Bias**: Checking the relevance of outcome measures to everyday clinical practice and monitoring dropout rates and reasons.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features include the similarity between the study and target populations in terms of severity and duration of illness, socio-demographic characteristics, and therapy preferences. The clinical setting and the qualifications and experience of those administering the intervention are also crucial.

Examples of Problematic Extrapolation / External Validity

The paper discusses how differences in consent processes and patient preferences can lead to non-representative study populations. For example, patients who consent to participate in studies often differ significantly from those who do not, which can limit the generalizability of findings.

Examples of Successful Extrapolation / External Validity

While the paper focuses more on the theoretical framework and checklist rather than specific examples, it implies that successful extrapolation is more likely when studies rigorously apply the proposed checklist criteria.

Transferring an Urban Food Security Program

To transfer an urban food security program from one city to another, Bornhöft et al.'s approach would suggest using their checklist to assess the external validity of the original program. This involves evaluating the representativeness of the original study population, ensuring the intervention methods are consistent with everyday practice in the new context, and verifying that the outcome measures are relevant to the new setting. By systematically applying these criteria, policymakers can better understand and adapt the program to fit the new urban environment.

Overall, Bornhöft et al. emphasize the importance of systematically assessing external and model validity to enhance the practical applicability and reliability of clinical research findings.

2.25 Checklists for External Validity: A Systematic Review (Dyrvig et al., 2014)

Dyrvig et al. (2014) conduct a systematic review of existing checklists for assessing external validity in clinical studies. The authors identify the strengths and weaknesses of these checklists and propose the need for a new, empirically validated checklist that can better support decision-makers in healthcare by ensuring the generalizability of study findings.

Main Problems in Extrapolation / External Validity

The main problem identified is the lack of empirical support for the items included in existing checklists for external validity. This absence of evidence undermines the reliability of these checklists in assessing the generalizability of clinical study findings. Additionally, the variability in the items included across different checklists creates inconsistencies in how external validity is evaluated.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Dyrvig et al. suggest the development of a new checklist that incorporates items with strong empirical support. They propose using a systematic approach to identify and validate the items to be included in the checklist. This involves:

1. **Systematic Literature Review**: Identifying existing checklist items and assessing the methodological justification for their inclusion.
2. **Empirical Validation**: Conducting studies to provide empirical evidence for the relevance and importance of each item.

3. **Consensus Building**: Engaging experts to reach consensus on the most critical items for assessing external validity.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features include demographic characteristics, disease prevalence, healthcare infrastructure, and cultural factors that may differ between the study population and the target population. Understanding these contextual factors is crucial for determining whether the study findings can be generalized to the new setting.

Examples of Problematic Extrapolation / External Validity

The review highlights the frequent inclusion of items in checklists without empirical justification, which can lead to incorrect assessments of external validity. For example, demographic characteristics such as age and gender may significantly influence treatment outcomes, yet are often inadequately considered in many checklists.

Examples of Successful Extrapolation / External Validity

While specific examples of successful extrapolation are not provided, the authors imply that a checklist with empirically supported items would enhance the likelihood of successful generalizations by ensuring all relevant factors are considered systematically.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, Dyrvig et al.'s approach would suggest developing a comprehensive checklist that includes empirically validated items relevant to the new context. This involves assessing demographic, socio-economic, and cultural factors that could affect the program's outcomes. Ensuring the checklist items are based on empirical evidence will help accurately evaluate the program's potential success in the new setting.

Overall, Dyrvig et al. emphasize the need for a more robust and empirically validated approach to assessing external validity, which can significantly improve the generalizability and applicability of clinical research findings to broader populations and different settings.

2.26 Conceptual Tools for Assessing Experiments: Some Well-Entrenched Confusions Regarding the Internal/External Validity Distinction (Jiménez-Buedo, 2011)

María Jiménez-Buedo examines the conceptual issues surrounding the internal and external validity distinction in experimental research. The paper critiques the classical categories of validity as formulated by Donald Campbell and highlights the misunderstandings and confusions in their application, especially in the field of behavioral economics. The author argues that these categories, though central to discussions about experimental methodology, often lead to problematic and inconsistent methodological recommendations.

Main Problems in Extrapolation / External Validity

The primary problem identified is the ambiguity and conceptual flaws in the distinction between internal and external validity. Jiménez-Buedo argues that this distinction is often misapplied or misunderstood, leading to confusion about what can be considered internally or externally valid. This confusion undermines the ability to make reliable generalizations from experimental findings.

Methods for Fixing Problems of Extrapolation / External Validity

The paper suggests re-evaluating the use of internal and external validity distinctions and possibly abandoning them in favor of more precise and context-specific terms. Jiménez-Buedo advocates for a clearer conceptual framework that better aligns with the practical and theoretical goals of experiments, focusing on the specific inferences and their contexts rather than broad and often misleading validity categories.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features include the relevance and similarity of the experimental conditions to the new context, the underlying mechanisms at play, and the specific attributes of the target population. Understanding these factors is crucial for assessing whether findings can be generalized to different settings.

Examples of Problematic Extrapolation / External Validity

The paper critiques the simplistic application of internal/external validity distinctions in behavioral economics, where experiments often aim to establish the existence of particular phenomena rather than measure the effects of well-established variables. This can lead to misleading conclusions about the generalizability of findings.

Examples of Successful Extrapolation / External Validity

Jiménez-Buedo does not provide specific examples of successful extrapolation but emphasizes that a more nuanced and context-aware approach to validity could improve the reliability of generalizations.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, Jiménez-Buedo's approach would suggest moving beyond simplistic validity distinctions and focusing on the specific contextual factors that could influence the program's effectiveness. This includes understanding the socio-economic environment, cultural differences, and specific needs of the new target population. By focusing on these detailed contextual factors, policymakers can make more informed and reliable decisions about adapting and implementing the program.

Overall, Jiménez-Buedo's work calls for a rethinking of traditional validity distinctions to better address the complexities of generalizing experimental findings to new contexts, advocating for a more detailed and context-sensitive approach.2.27

2.27 Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix (Pritchett and Sandefur, 2014)

Pritchett and Sandefur's (2014) paper delves into the complexities of external validity in the context of development practices. The authors emphasize that development policies often face significant challenges when applied across different contexts due to varying local conditions. They argue that achieving external validity requires a thorough understanding of the local context and the factors that influence policy outcomes.

Main Problems in Extrapolation/External Validity

The paper highlights several key issues with extrapolation and external validity. One major problem is the variability of contextual factors that affect policy implementation and outcomes. The authors argue that policies successful in one context may not produce the same results in another due to differences in social, economic, and political conditions.

Methods for Fixing Problems of Extrapolation/External Validity

To address these issues, Pritchett and Sandefur advocate for a more nuanced approach to policy design and implementation. They suggest that policymakers should not only replicate successful interventions but also adapt them to the specific contextual factors of the new environment. This involves conducting thorough local research and engaging with local stakeholders to understand the unique challenges and opportunities present.

Features of the New Context and/or Target Population Affecting Extrapolation/External Validity

The authors point out that factors such as local governance structures, cultural norms, economic conditions, and existing infrastructure can significantly influence the success of policy interventions. Understanding these features is crucial for adapting policies to new contexts.

Examples of Problematic Extrapolation/External Validity

The paper provides examples where policies that were successful in one context failed in another due to a lack of consideration of local conditions. These cases underscore the importance of context-specific adaptation in policy design.

Examples of Successful Extrapolation/External Validity

While the paper primarily focuses on the challenges of external validity, it implies that successful extrapolation requires careful adaptation and contextualization. Policies that are thoughtfully modified to fit local conditions have a higher likelihood of success.

Transferring an Urban Food Security Program

In the case of transferring an urban food security program, the authors would likely suggest a detailed assessment of the new city's contextual factors. This would involve understanding local food systems, economic conditions, and cultural attitudes towards food security. Engaging with local communities and stakeholders would be essential to tailor the program effectively.

Overall, Pritchett and Sandefur's approach emphasizes the critical role of understanding and adapting to local contexts to achieve successful policy implementation and external validity in development practices.

2.28 Cross-Sample Comparisons and External Validity (Krupnikov and Levine, 2014)

Krupnikov and Levine (2014) address the challenges of cross-sample comparisons and their implications for external validity. The paper discusses the complexities involved in generalizing findings across different samples and highlights the methodological considerations necessary to improve the robustness and applicability of experimental results.

Main Problems in Extrapolation / External Validity

The main problem identified is the difficulty in ensuring that findings from one sample can be generalized to other populations. Differences in sample characteristics, contextual factors, and experimental settings can lead to variations in outcomes, thereby challenging the external validity of the results.

Methods for Fixing Problems of Extrapolation / External Validity

Krupnikov and Levine suggest several methods to address these issues:

1. **Replication Studies**: Conducting replication studies in different contexts to assess the consistency of findings across diverse samples.
2. **Meta-Analysis**: Using meta-analytic techniques to aggregate results from multiple studies and evaluate the overall effect size and its variability.
3. **Contextualized Sampling**: Designing experiments with samples that closely resemble the target population to enhance the relevance and applicability of the findings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features include demographic differences, socio-economic factors, cultural norms, and the specific context in which the study is conducted. These factors can significantly influence how findings from one sample translate to another population.

Examples of Problematic Extrapolation / External Validity

The paper discusses examples where findings from one sample did not hold true in another due to differences in contextual factors. These instances highlight the risks of assuming that results are universally applicable without considering sample and context-specific variations.

Examples of Successful Extrapolation / External Validity

While specific examples of successful extrapolation are not detailed, the authors imply that rigorous methodological approaches, such as replication and meta-analysis, can improve the likelihood of successful generalization by systematically addressing variability across samples.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, Krupnikov and Levine would likely suggest conducting preliminary studies in the new context to replicate the findings and assess their applicability. This involves using samples that are representative of the new city's population and considering contextual factors such as local economic conditions, cultural practices, and social structures. By

doing so, policymakers can better understand how the program might perform in the new setting and make necessary adjustments to ensure its effectiveness.

Overall, Krupnikov and Levine's work emphasizes the importance of careful methodological planning and the use of robust techniques to enhance the external validity of experimental findings, thereby improving their applicability across different populations and contexts.

2.29 Designing Multi-Actor Implementation: A Mechanism-Based Approach (Busetti & Dente, 2018)

Busetti and Dente (2018) explore the complexities of policy implementation involving multiple actors. The authors propose a mechanism-based approach to policy design, emphasizing the importance of understanding and leveraging causal mechanisms to ensure successful implementation. They argue that traditional top-down approaches are often insufficient and that a more nuanced understanding of the interactions among various actors is necessary for effective policy design.

Main Problems in Extrapolation / External Validity

The primary problem identified is the difficulty in ensuring cooperation among diverse actors involved in policy implementation. Differences in goals, resources, and strategies among these actors can hinder the effectiveness of policy interventions. The authors highlight the complexity of multi-actor interactions and the challenge of designing policies that can be successfully implemented across different contexts.

Methods for Fixing Problems of Extrapolation / External Validity

Busetti and Dente propose a mechanism-based approach to policy design, which involves:

1. **Identifying Causal Mechanisms**: Understanding the specific causal processes that lead to successful implementation.
2. **Designing to Support Mechanisms**: Creating policy designs that trigger and support these mechanisms, ensuring that the necessary conditions for cooperation are met.
3. **Reverse Engineering Successful Cases**: Analyzing successful implementations to identify the underlying causal mechanisms and adapt them to new contexts.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features include the organizational structure, existing power dynamics, cultural norms, and specific incentives or disincentives for cooperation. Understanding these contextual factors is crucial for designing policies that can effectively engage multiple actors and ensure successful implementation.

Examples of Problematic Extrapolation / External Validity

The paper discusses instances where policies failed due to a lack of consideration of the complex interactions among implementing actors. For example, rigid top-down designs that did not account for local variations and the autonomy of local actors often resulted in poor implementation outcomes.

Examples of Successful Extrapolation / External Validity

Busetti and Dente provide several examples of successful policy implementations where a mechanism-based approach was applied. These cases demonstrate how understanding and leveraging causal mechanisms can lead to more effective and adaptable policy designs.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, the authors would suggest:

1. **Analyzing Successful Cases**: Identifying successful food security programs and understanding the causal mechanisms that made them work.
2. **Adapting Mechanisms to the New Context**: Ensuring that the key mechanisms, such as local stakeholder engagement, incentive structures, and resource distribution methods, are adapted to fit the new city's context.
3. **Engaging Local Actors**: Involving local stakeholders in the design process to ensure that the program is tailored to the specific needs and conditions of the new urban environment.

Overall, Busetti and Dente emphasize the importance of a mechanism-based approach in designing policies that are adaptable and effective across different contexts, highlighting the need for a deep understanding of the interactions among implementing actors.

2.30 Designing Randomized Controlled Trials with External Validity in Mind (Chassang and Kapon, 2022)

Chassang and Kapon (2022) discuss strategies for designing randomized controlled trials (RCTs) that enhance external validity. The paper emphasizes the dynamic

nature of research processes and the importance of considering research externalities. The authors argue that improving external validity requires systematic steps, including measuring rich covariates, using diverse contexts, and leveraging structured speculation and pilot studies.

Main Problems in Extrapolation / External Validity

The primary problem identified is the difficulty of extrapolating findings to new contexts due to varying covariates and contextual factors. The authors highlight that many research practices prioritize internal validity but do not sufficiently address external validity, leading to challenges in generalizing results.

Methods for Fixing Problems of Extrapolation / External Validity

Chassang and Kapon suggest several methods:

1. **Pre-Registration and Timely Reporting**: Ensuring transparency and reducing biases in research findings.
2. **Rich Covariate Measurement**: Collecting detailed covariate data to better understand treatment effects across different contexts.
3. **Diverse Contexts**: Conducting studies in varied contexts to capture a broader range of covariates and improve the generalizability of findings.
4. **Structured Speculation**: Including sections in research reports where researchers speculate about the applicability of findings to other contexts in a clear and testable manner.
5. **Pilot Studies**: Using pilot data to predict long-term effects and refine treatment approaches based on context-specific findings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features include demographic characteristics, socio-economic conditions, cultural norms, and institutional environments. Understanding these features helps tailor the extrapolation of study findings to new contexts.

Examples of Problematic Extrapolation / External Validity

The paper discusses instances where lack of consideration for local variations and the omission of relevant covariates led to poor extrapolation of study findings.

Examples of Successful Extrapolation / External Validity

While the paper primarily provides theoretical insights, it implies that successful extrapolation involves rigorous methodology and a detailed understanding of contextual factors.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, Chassang and Kapon would recommend:

1. **Conducting Pilot Studies**: Implementing the program on a small scale in the new city to gather context-specific data.
2. **Measuring Rich Covariates**: Collecting detailed demographic and socio-economic data to understand how the program might perform in the new environment.
3. **Engaging in Structured Speculation**: Speculating about the potential challenges and adaptations needed for the program to succeed in the new context.
4. **Using Diverse Contexts**: Drawing lessons from similar programs implemented in various contexts to inform the adaptation process.

Overall, Chassang and Kapon emphasize a dynamic, context-aware approach to designing RCTs and policy interventions, highlighting the importance of continuous learning and adaptation to achieve external validity.

2.31 Elements of External Validity: Framework, Design, and Analysis (Egami & Hartman, 2023)

Egami and Hartman (2023) present a comprehensive framework for understanding and improving external validity in social science research. The authors formalize the concept of external validity by breaking it down into four dimensions: X-validity (populations), T-validity (treatments), Y-validity (outcomes), and C-validity (contexts). They propose various methodological approaches to enhance the external validity of empirical studies.

Main Problems in Extrapolation / External Validity

The primary problem identified is the challenge of generalizing causal findings across different populations, treatments, outcomes, and contexts. Traditional studies often focus on internal validity, neglecting the factors that affect whether results can be generalized to broader or different settings.

Methods for Fixing Problems of Extrapolation / External Validity

Egami and Hartman propose several methods to address these issues:

1. **Formal Framework**: A systematic framework that considers all four dimensions of external validity.
2. **Effect-Generalization**: Three classes of estimators (weighting-based, outcome-based, and doubly robust) to adjust for selection into experiments and treatment effect heterogeneity.
3. **Sign-Generalization**: A multiple-testing procedure to generalize the direction of causal effects under weaker assumptions, focusing on purposive variations.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features include demographic characteristics, socio-economic conditions, cultural norms, and institutional environments. Understanding these factors helps tailor the extrapolation of study findings to new contexts.

Examples of Problematic Extrapolation / External Validity

The paper discusses various scenarios where extrapolation fails due to differences in populations, treatments, or contexts that were not adequately accounted for in the original study design.

Examples of Successful Extrapolation / External Validity

Egami and Hartman illustrate their methods using empirical examples from field, survey, and lab experiments, showing how their framework can be applied to achieve successful generalizations.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, Egami and Hartman would recommend:

1. **Using the Formal Framework**: Evaluating all four dimensions of external validity.
2. **Conducting Pilot Studies**: Gathering preliminary data in the new context to assess applicability.
3. **Adjusting for Contextual Differences**: Using estimators to account for differences in populations and contexts.
4. **Engaging Local Stakeholders**: Involving local communities to ensure the program is tailored to the new environment.

Overall, Egami and Hartman emphasize a detailed, systematic approach to enhancing external validity, providing practical tools and methodologies for researchers to improve the generalizability of their findings.

2.32 Establishing the Internal and External Validity of Experimental Studies (Slack & Draugalis, 2001)

This paper by Slack and Draugalis (2001) delves into the intricacies of establishing both internal and external validity in experimental studies, specifically within the context of pharmaceutical education research. It highlights the importance of robust experimental design and thorough evaluation to ensure that study findings are not only accurate but also generalizable to broader populations and settings.

Main Problems in Extrapolation / External Validity

The paper identifies several key issues in extrapolation and external validity. These include the challenge of ensuring that the study sample is representative of the broader population, the difficulty in controlling for all variables that might influence outcomes in different settings, and the potential for interaction effects that may not be apparent in the original study context but become significant when applied elsewhere.

Methods for Fixing Problems of Extrapolation / External Validity

To address these problems, the authors suggest a variety of methods. These include the use of stratified sampling to ensure diverse representation, conducting pilot studies in different settings to identify potential interaction effects, and employing statistical techniques such as meta-analysis to synthesize findings from multiple studies and enhance generalizability.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features that can impact extrapolation include demographic variables (e.g., age, gender, socio-economic status), cultural factors, and specific characteristics of the target population (e.g., health status, prior exposure to interventions). The authors emphasize the need to consider these variables when designing studies and interpreting their findings.

Examples of Problematic Extrapolation / External Validity

The paper mentions instances where findings from clinical trials failed to generalize to real-world settings due to differences in patient populations, treatment protocols, and

healthcare systems. These examples underscore the necessity of rigorous external validity assessments.

Examples of Successful Extrapolation / External Validity

Conversely, the authors provide examples where careful consideration of external validity factors led to successful extrapolation of study results. For instance, studies that included diverse samples and multiple sites were more likely to produce findings that could be generalized across different populations and settings.

Transferring an Urban Food Security Program from One City to Another

If the paper were to suggest strategies for transferring an urban food security program from one city to another, it would likely recommend a comprehensive evaluation of the target city's specific needs and context. This might include pilot testing the program in the new location, gathering data on local demographics and socio-economic factors, and adjusting the program components based on these insights to ensure they are culturally and contextually appropriate.

In summary, Slack and Draugalis (2001) provide a thorough examination of the challenges and solutions related to establishing external validity in experimental studies. Their insights are invaluable for researchers aiming to ensure that their findings are both accurate and applicable to broader contexts.

2.33 Estimates of External Validity Bias When Impact Evaluations Select Sites Nonrandomly (Bell et al., 2016)

Bell et al. (2016) address the concern of external validity in impact evaluations, particularly when sites are chosen nonrandomly. The authors argue that nonrandom site selection can lead to biases in estimating the impact of programs on the broader population. The study combines nonrandomly selected school districts from 11 educational impact studies with population data from the Reading First program to estimate the bias in impact evaluations. The findings suggest that nonrandom site selection can substantially bias impact estimates, usually downwards, compared to the broader population.

Main Problems in Extrapolation / External Validity

The primary problem discussed is the bias introduced by nonrandom site selection in impact evaluations, which can lead to incorrect generalizations of program effectiveness to the broader population of interest. The bias is typically downward,

meaning that the estimated impact in the selected sites is lower than what it would be in a more representative sample of the broader population.

Methods for Fixing Problems of Extrapolation / External Validity

The paper suggests that ideally, impact evaluations should include a representative sample of sites selected at random to ensure statistical equivalence to the population. This would help in obtaining unbiased estimates of the program's impact on the broader population. When random selection is not possible, the paper highlights the importance of understanding and correcting for the factors that influence site selection to adjust the impact estimates accordingly.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The features that can affect extrapolation include the specific criteria used for site selection, such as the capacity to support the evaluation, oversubscription to support random assignment, and the mix of urban and rural sites. Additionally, the paper mentions that sites often opt out of evaluations, which can further skew the representativeness of the sample and affect the external validity.

Examples of Problematic Extrapolation / External Validity

The paper provides examples from the 11 educational impact studies, showing how nonrandom site selection leads to biased impact estimates when generalized to the broader population. These examples illustrate that the estimates from these selected sites are, on average, 0.10 standard deviations lower than the impact estimates in the broader population.

Examples of Successful Extrapolation / External Validity

The paper does not provide specific examples of successful extrapolation but implies that achieving external validity requires rigorous site selection processes and adjustments for selection biases.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, the article would suggest ensuring that the sites for the urban food security program are selected randomly or representatively to the broader population of interest. If random selection is not feasible, it would recommend understanding the criteria used for site selection and adjusting the impact estimates to account for these factors to ensure that the program's effectiveness can be accurately generalized to the new city.

2.34 Evidence-Based Policy: A Practical Guide to Doing It Better (Cartwright & Hardie, 2012)

Cartwright and Hardie (2012) delve into the intricacies of evidence-based policy, focusing on the challenges and methodologies for ensuring that a policy proven effective in one context can be successfully implemented in another. They argue that the success of a policy in one setting does not guarantee its effectiveness in another due to variations in causal roles and support factors between contexts.

Main Problems in Extrapolation / External Validity

The primary issue discussed is the difficulty of ensuring that a policy which worked in one context will have the same effect in a different context. This is largely due to differences in the causal roles a policy may play and the necessary support factors that may not be present in the new context. The authors highlight the failure of the Bangladesh Integrated Nutrition Project (BINP) as a case where external validity was problematic due to contextual differences between Bangladesh and Tamil Nadu.

Methods for Fixing Problems of Extrapolation / External Validity T

The authors suggest a thorough analysis of the causal roles and support factors necessary for a policy to succeed in the new context. They propose a "vertical search" to identify the appropriate level of abstraction for the principles underpinning the policy and a "horizontal search" to ensure all necessary support factors are identified and secured. The book emphasizes the importance of detailed context-specific adjustments rather than direct replication of policies.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features that can affect extrapolation include local socio-cultural dynamics, existing infrastructure, and the distribution of support factors. For example, in the BINP case, the social structure in Bangladesh, such as the "mother-in-law factor" and the "male-shopper factor," played significant roles in the failure of the policy that had succeeded in Tamil Nadu.

Examples of Problematic Extrapolation / External Validity

Cartwright and Hardie illustrate problematic extrapolation with the BINP, where the policy failed due to differences in food distribution dynamics and decision-making hierarchies within families compared to Tamil Nadu. This failure underscores the need to understand and adapt to local conditions rather than assuming direct transferability of policy success.

Examples of Successful Extrapolation / External Validity

While the book primarily focuses on challenges, it does highlight the theoretical framework needed for successful extrapolation. Successful cases require rigorous analysis and adaptation of the policy to fit the new context's specific causal structures and support factors. However, specific successful examples are not detailed as prominently as the challenges.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, Cartwright and Hardie would recommend conducting a detailed analysis of the causal mechanisms and support factors in both the original and new contexts. They advise against assuming that the same methods will work without modification. Instead, they suggest tailoring the program to the local conditions, ensuring all necessary support factors are present, and adjusting the implementation strategy to fit the new environment.

In summary, Cartwright and Hardie emphasize the importance of understanding the specific context and causal mechanisms when transferring policies between different settings. They advocate for detailed, context-specific adaptations to ensure that policies proven effective elsewhere can be successfully implemented in new environments.

2.35 Expanding the Framework of Internal and External Validity in Quantitative Research (Onwuegbuzie, 2000)

Onwuegbuzie (2000) addresses the complexity of internal and external validity in quantitative research, highlighting the need for a more comprehensive framework beyond the traditional threats identified by Campbell and Stanley. The paper argues for a thorough assessment of both internal and external validity across all quantitative research designs.

Main Problems in Extrapolation / External Validity

The main issue discussed is the generalizability of research findings to broader contexts outside the experimental settings. Traditional frameworks often fall short in addressing threats to validity in non-experimental research designs, which can lead to incorrect assumptions about the applicability of findings to different populations, settings, and times.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, the paper proposes expanding the traditional framework to include additional dimensions and subdimensions of validity. It suggests a systematic approach to identifying and mitigating threats through rigorous research design, data collection, and data analysis processes. The importance of replicating studies in different contexts to confirm external validity is also emphasized.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features that can affect extrapolation include the demographic characteristics of the population, environmental conditions, and the specific settings in which the research is conducted. The paper stresses the importance of considering these variables when designing studies and interpreting results.

Examples of Problematic Extrapolation / External Validity

The paper highlights the frequent omission of discussions on threats to internal and external validity in empirical research reports, particularly in non-experimental studies. This omission can lead to a misunderstanding of the extent to which findings can be generalized, as illustrated by various examples from educational research where population and ecological validity were significant concerns.

Examples of Successful Extrapolation / External Validity

While the focus is primarily on the challenges, the paper underscores the necessity of addressing validity threats to achieve successful extrapolation. It advocates for detailed discussions of these threats in research reports to improve the generalizability of findings and guide future research.

Transferring an Urban Food Security Program

For transferring an urban food security program from one city to another, Onwuegbuzie would recommend a careful analysis of the population and environmental differences between the two cities. The program should be adjusted to account for these differences, and additional pilot studies should be conducted in the new city to ensure that the intervention will be effective in the new context.

In conclusion, Onwuegbuzie emphasizes the critical importance of thoroughly addressing internal and external validity in all quantitative research designs. By expanding the traditional frameworks and rigorously assessing validity threats, researchers can improve the reliability and generalizability of their findings.

2.36 Experimental and Quasi-experimental Designs for Generalized Causal Inference (Shadish et al., 2002)

Shadish et al. (2002) delve into the complexities of external validity, especially in the context of experimental and quasi-experimental designs. They emphasize that external validity pertains to the extent to which the causal relationships observed in a study can be generalized to other contexts, populations, and times. The authors argue that achieving high external validity is challenging but essential for the applicability of research findings.

Main Problems in Extrapolation / External Validity

The primary issue highlighted is the difficulty in generalizing findings from one context to another. This is compounded by the problem of short extrapolations being more justifiable than long ones, due to the likelihood of qualitative changes that may not be accounted for. There is also the inherent uncertainty when a significant gap exists between the sample and the target population.

Methods for Fixing Problems of Extrapolation / External Validity

The authors suggest that shorter extrapolations are preferable and emphasize the importance of having a well-identified functional form over the sampled range. They advocate for robust theoretical frameworks to guide the identification of inflection points that could affect extrapolation. Moreover, using multiple operations of constructs and ensuring that these constructs are not biased by constant sources can aid in mitigating extrapolation issues.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Features such as the specific characteristics of the population, the setting, and the variables involved can significantly impact the external validity of a study. The authors note that differences in settings, such as urban versus rural, or cultural differences, can lead to variations in how interventions work. Additionally, differences in time periods and the presence of additional, unmeasured variables in the new context can affect the validity of extrapolations.

Examples of Problematic Extrapolation / External Validity

Shadish et al. provide examples of how different constructs and operational definitions can lead to issues. For instance, they discuss how generalizing findings from one demographic group to another (e.g., from male job applicants in one city to female job applicants in another) can be problematic if the constructs are not

accurately represented across contexts. They also highlight issues in extrapolating data from animal studies to human populations, which often involves significant risks.

Examples of Successful Extrapolation / External Validity T

he text does not provide explicit examples of successful extrapolations but implies that success hinges on careful and incremental extensions of both theory and experiment. The authors argue that reasonable extrapolations are often those that involve small, manageable variations in study features rather than wholesale changes.

Transferring an Urban Food Security Program

In the context of transferring an urban food security program from one city to another, Shadish et al. would suggest a meticulous approach. They would recommend ensuring that the fundamental causal mechanisms identified in the original program are applicable to the new setting. This would involve a detailed analysis of the new population and setting, adjusting for any contextual differences, and conducting pilot studies to test the applicability of the program before full-scale implementation.

In conclusion, Shadish et al. emphasize the importance of understanding and addressing the unique features of new contexts when attempting to generalize research findings. They advocate for rigorous theoretical and methodological approaches to ensure that extrapolations are valid, cautioning against the assumption that findings from one context can be easily applied to another without significant consideration of potential differences.

2.37 Experimental Localism and External Validity (Guala, 2003)

Guala (2003) explores the concept of "experimental localism," emphasizing the significance of context-specific knowledge and the limitations of universal theories in scientific research. The paper critiques Bruno Latour's radical localism and discusses its implications for external validity, proposing solutions through Mayo's error-probabilistic approach.

Main Problems in Extrapolation / External Validity

The primary issue highlighted is the challenge of generalizing experimental results beyond the controlled settings of a laboratory. Latour's perspective suggests that without replicating laboratory conditions in the real world, scientific generalizations often fail. This skepticism towards broad applicability of experimental findings is seen as a significant hurdle in achieving external validity.

Methods for Fixing Problems of Extrapolation / External Validity

Guala suggests that to address these challenges, researchers should employ rigorous error-probabilistic methods as proposed by Mayo. This involves designing experiments that can withstand severe tests of hypotheses, thereby reducing the likelihood of errors. Additionally, creating analogies between experimental and real-world settings can help, provided these analogies are robust and account for all critical variables.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The specific characteristics of the target population and the new context, such as environmental factors, demographic differences, and cultural elements, can significantly impact the validity of extrapolated findings. Ensuring that these features are adequately considered and matched to the experimental conditions is crucial for successful generalization.

Examples of Problematic Extrapolation / External Validity

Guala discusses instances where experimental results fail to translate to real-world settings, using the example of Pasteur's antianthrax vaccine, which did not work effectively outside controlled laboratory conditions. Such examples underscore the need for careful consideration of context when attempting to generalize findings.

Examples of Successful Extrapolation / External Validity

While the paper focuses more on challenges, it hints at the possibility of successful extrapolation through careful design and rigorous testing. The discussion implies that successful cases often involve incremental extensions of experimental conditions to the real world, ensuring all critical variables are accounted for.

Transferring an Urban Food Security Program

In the context of transferring an urban food security program from one city to another, Guala would advocate for a detailed analysis of the new context, ensuring that all relevant variables and support mechanisms from the original setting are present or adapted appropriately. This might involve pilot studies and adjustments to the program to fit the specific needs and conditions of the new urban environment.

In conclusion, Guala emphasizes the need for rigorous methodologies and careful consideration of context-specific factors when attempting to generalize experimental findings. By addressing the unique challenges of external validity, researchers can enhance the applicability and reliability of their results in diverse settings.

3.38 Experimental Practices and Objectivity in the Social Sciences (Jiménez-Buedo & Russo, 2021)

Jiménez-Buedo and Russo (2021) explore the complexities of experimental practices in the social sciences, particularly focusing on the notion of objectivity and its implications for external validity. The authors argue that the concept of external validity is often misunderstood and misapplied, leading to simplistic views about the generalizability of experimental findings. They emphasize the importance of construct validity and the role of background assumptions in shaping the interpretation of experimental results.

Main Problems in Extrapolation / External Validity

The primary issue highlighted is the misconception that experiments inherently possess high internal validity but low external validity. This simplistic dichotomy overlooks the fact that the same experimental intervention can represent different constructs based on varying background assumptions. Consequently, this can lead to misunderstandings about the generalizability of experimental results and the conditions under which they can be applied.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, the authors suggest moving beyond the binary internal/external validity framework. They propose a more nuanced approach that recognizes the role of background assumptions in determining the constructs represented by an experiment. By fostering a shared set of assumptions among researchers, it is possible to improve the objectivity and generalizability of experimental findings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The success of extrapolation depends significantly on the alignment of constructs between the original experimental setting and the new context. Differences in background assumptions, cultural contexts, and specific conditions of the target population can all influence the validity of extrapolating findings from one setting to another.

Examples of Problematic Extrapolation / External Validity

The paper discusses the common critique that randomized controlled trials (RCTs) are "good in internal validity but bad in external validity." This critique often fails to account for the complexities of construct validity and the importance of background assumptions. The authors illustrate how different interpretations of constructs can

lead to varied conclusions about the effectiveness of an intervention, thus complicating the process of generalizing results.

Examples of Successful Extrapolation / External Validity

While the paper does not provide specific examples of successful extrapolation, it implies that success is more likely when there is a pre-existing consensus on the validity of the constructs involved. Shared background assumptions among researchers can facilitate more accurate generalizations and applications of experimental findings to new contexts.

Transferring an Urban Food Security Program

In the context of transferring an urban food security program from one city to another, this paper would suggest ensuring that the constructs underlying the program are clearly understood and agreed upon by all stakeholders. It would be crucial to align the background assumptions about food security, cultural practices, and socioeconomic conditions between the two cities to enhance the likelihood of successful implementation and generalization of results.

In conclusion, Jiménez-Buedo and Russo (2021) advocate for a deeper understanding of construct validity and the role of background assumptions in experimental practices. By moving beyond a simplistic internal/external validity dichotomy, researchers can improve the objectivity and generalizability of their findings, ultimately enhancing the impact of social science research.

2.39 Experimentation in the 21st century: The importance of external validity (Winer, 1999)

In his 1999 article, Winer discusses the significance of external validity in contemporary experimentation, particularly in the context of the 21st century. He emphasizes the challenges and necessities of ensuring that experimental findings are not only valid within the confines of a controlled environment but also applicable to real-world scenarios. The paper explores various facets of external validity and provides insights into how researchers can enhance the generalizability of their experimental results.

Main Problems in Extrapolation / External Validity

Winer identifies several key issues that arise when attempting to extrapolate findings from a controlled experimental setting to a broader context. One significant problem is the potential for results to be too context-specific, thereby limiting their applicability

to different populations or settings. He also notes the risk of overgeneralization, where findings from a limited sample are incorrectly assumed to be universally applicable.

Methods for Fixing Problems of Extrapolation / External Validity

To address these problems, Winer suggests the implementation of more robust and varied sampling techniques. He advocates for the inclusion of diverse demographic groups in the experimental design to ensure that the findings are not biased toward a particular segment of the population. Additionally, he recommends conducting replication studies across different settings to verify the consistency of the results.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Winer highlights that factors such as cultural differences, environmental variables, and socio-economic conditions can significantly influence the external validity of experimental results. Understanding these contextual factors is crucial for accurately extrapolating findings from one setting to another.

Examples of Problematic Extrapolation / External Validity

Winer provides instances where extrapolation has led to misleading conclusions. For example, he discusses cases in educational research where findings from studies conducted in urban schools were inappropriately generalized to rural schools, resulting in ineffective policy recommendations.

Examples of Successful Extrapolation / External Validity

Although Winer does not provide detailed examples of successful extrapolation, he implies that studies which carefully consider and incorporate contextual variables tend to have more reliable and generalizable results. Successful extrapolation often involves iterative testing and validation across diverse settings.

Transferring an Urban Food Security Program

In the case of transferring an urban food security program from one city to another, Winer's article would suggest a thorough assessment of the new context's specific characteristics. This includes understanding the local dietary habits, economic conditions, and community structures. He would likely recommend pilot testing the program in the new setting and making necessary adjustments based on the initial outcomes to ensure the program's effectiveness and relevance.

In summary, Winer's work underscores the critical importance of external validity in experimentation. By addressing the challenges associated with extrapolation and

emphasizing the need for context-aware methodologies, he provides a framework for enhancing the applicability of experimental findings to real-world scenarios.

2.40 External Validity (Findley et al., 2021)

Findley et al. (2021) examine the concept of external validity within the social sciences, highlighting its crucial role in ensuring that research findings are applicable to broader contexts. The paper critiques the common neglect of external validity in favor of internal validity and provides a detailed framework for enhancing the generalizability and transportability of research results.

Main Problems in Extrapolation / External Validity

The authors identify a primary issue: the overemphasis on internal validity, which often leads to a neglect of external validity. This imbalance results in findings that are highly credible within the specific study settings but may not be applicable to broader or different populations. They argue that without addressing external validity, the applicability and usefulness of social science research are severely limited.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Findley et al. propose several methodological improvements. They advocate for the use of the M–STOUT framework, which includes mechanisms, settings, treatments, outcomes, units, and time. This framework helps ensure that all relevant dimensions are considered when assessing the external validity of a study. They also recommend employing robust sampling techniques, such as random or stratified random sampling, to enhance representativeness.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features that can impact the extrapolation of findings include differences in the demographic characteristics of the population, cultural contexts, environmental factors, and the specific settings in which the research was conducted. The authors stress the importance of clearly defining the scope and populations at the theoretical and design stages to ensure that findings can be accurately generalized or transported.

Examples of Problematic Extrapolation / External Validity

The paper provides examples of how internal validity-focused studies often fail to address external validity concerns adequately. For instance, randomized controlled

trials (RCTs) with unrepresentative samples may produce biased estimates that do not apply to the broader population, leading to misleading policy recommendations.

Examples of Successful Extrapolation / External Validity

While the paper focuses more on the challenges, it implies that successful extrapolation is achievable through careful design and rigorous methodological practices. Studies that employ the M–STOUT framework and robust sampling methods are more likely to produce findings that are generalizable to other contexts.

Transferring an Urban Food Security Program

In the context of transferring an urban food security program from one city to another, Findley et al. would suggest a thorough evaluation of the new context's characteristics, including demographic, cultural, and environmental factors. They would recommend using stratified random sampling to ensure that the sample is representative of the new city's population and conducting pilot studies to test the program's applicability and effectiveness in the new setting.

In conclusion, Findley et al. emphasize the necessity of balancing internal and external validity to enhance the applicability of social science research. By adopting comprehensive frameworks and rigorous sampling techniques, researchers can ensure that their findings are not only internally valid but also generalizable and transportable to other contexts.

2.41 External Validity and Evaluation Research: A Codification of Problems (Bernstein et al., 1975)

Ilene N. Bernstein, George W. Bohrnstedt, and Edgar F. Borgatta (1975) provide a comprehensive examination of the factors threatening the external validity of evaluation research. Their work codifies these factors into distinct categories and discusses their implications for generalizing research findings to broader populations.

Main Problems in Extrapolation / External Validity

Bernstein et al. identify five primary categories of factors that threaten external validity: selection effects, measurement effects, confounded treatment effects, situational effects, and effects due to differential mortality. Selection effects occur when the sample does not accurately represent the target population, often due to self-selection or convenience sampling. Measurement effects arise from unreliable or invalid measurement tools and the reactivity of subjects to being measured. Confounded treatment effects include the presence of multiple treatments or

variations in treatment implementation. Situational effects are related to the specific context of the experiment, such as the staff administering the program or the geographic setting. Finally, differential mortality refers to the attrition of subjects or programs, which can skew results if not randomly distributed.

Methods for Fixing Problems of Extrapolation / External Validity

To mitigate these issues, the authors suggest several strategies. For selection effects, they recommend random sampling and the use of control groups to ensure the sample is representative of the broader population. Measurement effects can be addressed by using reliable and valid instruments and employing unobtrusive measures where possible. For confounded treatment effects, they propose clearly defining and standardizing treatments across different settings. Situational effects can be minimized by carefully considering the impact of context and replicating studies in diverse settings. Addressing differential mortality involves tracking attrition rates and using statistical techniques to adjust for non-random dropout.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The authors emphasize that the characteristics of the new context and target population, such as demographic, cultural, and environmental factors, can significantly affect the generalizability of research findings. These features must be carefully analyzed and accounted for when attempting to apply findings to new settings.

Examples of Problematic Extrapolation / External Validity

The paper provides examples where generalizing findings proved problematic due to contextual differences. One instance involves the Head Start program, where variations in local implementation and context led to inconsistent results across different settings. Another example is the evaluation of income maintenance experiments, where differences in local economies and social structures impacted the applicability of findings.

Examples of Successful Extrapolation / External Validity

While the authors do not provide specific examples of successful extrapolation, they underscore the importance of rigorous methodological approaches to enhance the reliability of generalizations. By systematically addressing the identified threats to external validity, researchers can improve the applicability of their findings across diverse contexts.

Transferring an Urban Food Security Program

Applying the insights from Bernstein et al. to transferring an urban food security program involves several steps. Policymakers should begin by ensuring that the sample for the initial evaluation accurately represents the target population of the new city. This includes using random sampling techniques and control groups to minimize selection biases. Reliable and valid measurement tools should be employed, and efforts should be made to ensure that the new context's characteristics are comparable to those of the original setting.

Standardizing the treatment protocol and ensuring consistency in implementation across different sites can help mitigate confounded treatment effects. Additionally, the program should be piloted in the new city to identify any situational factors that may influence outcomes. Monitoring and adjusting for differential attrition rates is crucial to maintaining the validity of the results.

By systematically addressing these factors, policymakers can enhance the likelihood of successfully transferring and implementing the urban food security program in a new setting, ensuring that it achieves the desired outcomes.

Bernstein et al. provide a detailed framework for understanding and addressing the challenges of external validity in evaluation research. Their codification of problems and suggested solutions offer valuable guidance for researchers and policymakers seeking to generalize findings across diverse contexts.

2.42 External Validity and Meta-Analysis (Slough & Tyson, 2023)

Slough and Tyson (2023) explore the theoretical foundations of meta-analysis, emphasizing the importance of external validity in combining results from multiple studies. They provide a framework to assess the theoretical conditions necessary for meta-analyses to achieve meaningful and interpretable conclusions, focusing on the concepts of target-equivalence and harmonization.

Main Problems in Extrapolation / External Validity

The primary issue highlighted is the lack of external validity in individual studies, which can lead to incorrect generalizations when these studies are combined in a meta-analysis. The authors argue that many meta-analyses fail to account for differences in study settings, populations, and measurement strategies, which undermines the validity of the combined results.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, the authors propose ensuring that constituent studies in a meta-analysis are target-equivalent, meaning they aim at the same empirical target. They introduce the concepts of contrast harmonization and measurement harmonization, which are necessary for achieving target-equivalence. By aligning the comparisons made and the outcomes measured across studies, researchers can improve the external validity of meta-analyses.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features that affect the extrapolation include differences in demographic characteristics, environmental conditions, and cultural contexts between the original study settings and the new context. These factors must be carefully considered and aligned to ensure that the findings are applicable to the new setting.

Examples of Problematic Extrapolation / External Validity

The authors provide examples of how failure to achieve harmonization can lead to misleading conclusions. For instance, studies that measure outcomes differently or use different treatment contrasts can produce inconsistent results, making it difficult to draw meaningful inferences from the meta-analysis.

Examples of Successful Extrapolation / External Validity

While specific examples of successful extrapolation are not detailed, the framework suggests that successful meta-analyses are those that rigorously ensure harmonization of measurement strategies and contrasts. This alignment helps in achieving consistent and interpretable results across different studies.

Transferring an Urban Food Security Program

In the context of transferring an urban food security program from one city to another, Slough and Tyson would recommend ensuring that the program's components are harmonized with the new context. This involves aligning the demographic and environmental factors, as well as ensuring that the measurement strategies for assessing program outcomes are consistent with those used in the original context. Pilot studies in the new setting can help in making necessary adjustments to enhance the program’s effectiveness and generalizability.

In conclusion, Slough and Tyson emphasize the critical importance of external validity and harmonization in meta-analyses. By carefully aligning study settings, measurement strategies, and treatment contrasts, researchers can ensure that their findings are robust, generalizable, and applicable to new contexts.

2.43 External Validity and Model Validity: A Conceptual Approach (Khorsan & Crawford, 2014)

Khorsan and Crawford (2014) delve into the complexities of external and model validity within healthcare RCTs, emphasizing the challenges posed by varying motivations for participation and the issue of sample attrition. They highlight the need for a representative sample ideally drawn randomly from the target population to ensure statistical inferences can be made accurately.

Primary Problems in Extrapolation / External Validity

The main issues identified involve the difficulty in generalizing results due to non-representative samples, which often arise from nonrandom participation, recruitment, and retention rates. The text points out that these factors lead to biases that limit the generalizability of study results.

Methods for Addressing Problems of Extrapolation / External Validity

To mitigate these issues, the authors suggest a more rigorous approach to participant selection, emphasizing the importance of random sampling and the need for detailed analyses to understand the effects of sample attrition. They also propose the development of methodological checklists to assess external validity more systematically.

Features of New Contexts and/or Target Populations Affecting Extrapolation / External Validity

Key features that can impact extrapolation include demographic variables such as gender, ethnicity, and socio-economic status, as well as the geographical and temporal context of the study. The authors stress that these variables must be considered to ensure the results are applicable to different populations.

Examples of Problematic Extrapolation / External Validity

The article discusses how many studies fail to adequately report the necessary information for assessing external validity, leading to limited generalizability of their findings. The absence of detailed descriptions of how demographic and contextual factors might influence results is a common problem.

Examples of Successful Extrapolation / External Validity

While the article does not provide specific examples of successful extrapolation, it implies that studies which incorporate rigorous sampling methods and thorough reporting of external validity criteria are more likely to achieve successful generalization.

Transferring an Urban Food Security Program

For an urban food security program, the authors would likely recommend ensuring that the program's implementation in the new city involves random sampling or, at the very least, a representative sample of the target population. They would also advise conducting a thorough analysis of local contextual factors and addressing potential sample attrition to maintain the program's effectiveness.

In conclusion, Khorsan and Crawford (2014) emphasize the importance of random sampling and detailed contextual analysis to enhance the external validity of healthcare RCTs, providing a robust framework for generalizing results across different populations and settings.

2.44 External Validity and Policy Adaptation: From Impact Evaluation to Policy Design (Williams, 2020)

Williams (2020) investigates the challenges of applying impact evaluation evidence to policy design, particularly focusing on the external validity of such evidence. He introduces the concept of "mechanism mapping" to bridge the gap between generalizability and applicability of policy interventions across different contexts.

Primary Problems in Extrapolation / External Validity

Williams identifies the main issue as the interaction between a policy's theory of change and the contextual variables in which it is implemented. Failures in external validity often arise when these contextual factors are not aligned, leading to different outcomes than those observed in the original setting. He also highlights the problem of relying solely on impact evaluations without considering local contextual information.

Methods for Addressing Problems of Extrapolation / External Validity

To mitigate these problems, Williams proposes "mechanism mapping," which involves mapping a policy's theory of change against the contextual assumptions required for each step of the mechanism to function. This approach helps identify potential external validity issues and suggests appropriate policy adaptations. He also emphasizes the importance of combining strong evidence from other contexts with detailed local information.

Features of New Contexts and/or Target Populations Affecting Extrapolation / External Validity

Factors such as location, target group characteristics, existing related policy interventions, and the specific time period can significantly impact the external validity of a policy. These dimensions of context must be carefully analyzed to ensure that the policy will work similarly in the new setting.

Examples of Problematic Extrapolation / External Validity

Williams provides the example of the Bangladesh Integrated Nutrition Programme (BINP), which failed to replicate the success of a similar program in Tamil Nadu due to differences in household food allocation practices. Another example is the Tools of the Mind early childhood education program, which showed negative impacts when scaled up due to the complexities of integrating the program into regular school activities.

Examples of Successful Extrapolation / External Validity

While the paper does not provide specific examples of successful extrapolation, it implies that policies which undergo rigorous mechanism mapping and careful contextual adaptation are more likely to succeed when transported to new settings.

Transferring an Urban Food Security Program

In the case of transferring an urban food security program, Williams would recommend using mechanism mapping to align the policy's theory of change with the contextual realities of the new city. This involves assessing the local demographic, socio-economic, and political conditions to ensure that the program components will function as intended.

In conclusion, Williams emphasizes the necessity of understanding and adapting to local contexts when applying evidence from impact evaluations to new policy settings. By using mechanism mapping, policymakers can enhance the external validity and effectiveness of their interventions.

2.45 External Validity and the Research Process: A Comment on the Calder/Lynch Dialogue (McGrath & Brinberg, 1983)

Joseph E. McGrath and David Brinberg (1983) provide an insightful commentary on the dialogue between Calder, Phillips, Tybout, and Lynch concerning external validity in research. They discuss points of agreement and disagreement among the scholars and introduce a schema for analyzing validity within the research process.

Main Problems in Extrapolation / External Validity

McGrath and Brinberg identify the oversimplification of external validity as a primary problem. They argue that equating external validity with realism or representativeness fails to capture the complexity of generalizing findings across contexts. External validity should consider the fit between theoretical constructs and empirical findings and the robustness of findings across different settings.

Methods for Fixing Problems of Extrapolation / External Validity

The authors propose a three-stage schema:

1. **Prior Validities**: This stage involves developing and selecting research elements within their conceptual, methodological, and substantive domains.
2. **Internal Validities**: Here, elements from two domains are combined to form an intermediate structure, which is tested by integrating the third domain, ensuring the alignment of research design, hypotheses, and observations.
3. **External Validities**: In this stage, findings from Stage Two are verified, extended, and delimited through replication and robustness analysis. Researchers assess the generalizability of findings by exploring their scope and limits across various facets of the research domains.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The authors emphasize understanding differences in conceptual models, methodological tools, and substantive events between contexts. This comprehensive approach involves considering the theoretical, methodological, and practical aspects of both the original and target settings.

Examples of Problematic Extrapolation / External Validity

They discuss issues such as applying findings from laboratory settings to real-world environments, noting that increasing realism in lab experiments often fails to address underlying contextual differences, leading to misleading conclusions about generalizability.

Examples of Successful Extrapolation / External Validity

While not providing specific examples, McGrath and Brinberg argue that their systematic approach enhances reliability and applicability. By thoroughly assessing the conceptual, methodological, and substantive elements, researchers can ensure the robustness and generalizability of their findings.

Transferring an Urban Food Security Program

Applying McGrath and Brinberg's framework involves several steps. First, thoroughly understand the conceptual, methodological, and substantive domains of the original program, ensuring that core components and mechanisms are adapted to the new setting. Second, test the program's design and implementation in the new urban environment through pilot studies, making necessary adjustments based on outcomes and feedback. Third, conduct robustness analyses and replication studies to assess generalizability and identify limits. This multi-stage approach helps tailor the program to the new context, enhancing its external validity and practical applicability.

McGrath and Brinberg's approach provides valuable insights for transferring and adapting interventions like urban food security programs. Their emphasis on a thorough, multi-stage analysis ensures that programs are effectively tailored to new contexts, enhancing their external validity and practical applicability.

2.46 External Validity and Translation from Research to Implementation (Prohaska & Etkin, 2010)

Prohaska and Etkin (2010) focus on the significant challenges associated with translating scientifically validated research findings into community-based programs. They emphasize the hurdles of ensuring that health promotion and disease prevention programs for older adults are effectively disseminated, implemented, and maintained in real-world settings.

Primary Problems in Extrapolation / External Validity

The main issues discussed involve the slow and fragmented translation of research findings into practice, often due to the divergence in priorities between researchers and practitioners. Additionally, the stringent criteria for internal validity in initial trials can limit the external validity of these programs, making it difficult to generalize findings to broader, more diverse populations.

Methods for Addressing Problems of Extrapolation / External Validity

To improve external validity, the authors advocate for using the RE-AIM framework, which stands for Reach, Efficacy, Adoption, Implementation, and Maintenance. This framework helps in systematically evaluating the impact of health interventions across different settings and populations. They also stress the importance of involving community stakeholders early in the research process to ensure that the programs are relevant and practical for the intended populations.

Features of New Contexts and/or Target Populations Affecting Extrapolation / External Validity

Several key features can influence the generalizability of research findings, including the demographic characteristics of the population, cultural and socio-economic factors, and the availability of resources in the new setting. Prohaska and Etkin emphasize the necessity of adapting programs to fit the local context to enhance their effectiveness.

Examples of Problematic Extrapolation / External Validity

The article discusses issues such as the failure of programs to be adopted widely due to a lack of attention to external validity. For instance, health promotion interventions that worked well in controlled clinical settings often failed when applied in community settings due to differences in population characteristics and available resources.

Examples of Successful Extrapolation / External Validity

The authors highlight the Chronic Disease Self-Management Program (CDSMP) as a successful example, noting its widespread adoption and effectiveness across various community settings. The program's success is attributed to its adaptability and inclusion of culturally appropriate content, which enhanced its appeal and effectiveness among diverse older adult populations.

Transferring an Urban Food Security Program

In the context of transferring an urban food security program, Prohaska and Etkin would likely recommend using the RE-AIM framework to evaluate and adapt the program to fit the new city's specific context. This includes understanding local needs, involving community stakeholders, and ensuring the program is adaptable to the new demographic and socio-economic conditions.

In summary, Prohaska and Etkin highlight the importance of external validity in translating research into practice. By utilizing frameworks like RE-AIM and engaging with community stakeholders, researchers can enhance the applicability and impact of their programs in diverse settings.

2.47 External Validity in IS Survey Research (King & He, 2005)

King and He (2005) delve into the issue of external validity within the realm of IS survey research, emphasizing the importance of generalizability of study findings beyond the immediate study sample. They argue that while internal validity ensures the causal relationships within the study are sound, external validity concerns whether these findings can be applied to broader populations and settings.

Primary Problems in Extrapolation / External Validity

The article identifies several key problems in extrapolation and external validity, primarily focusing on the challenges of generalizing findings from a study's specific sample to a wider population. One significant issue is the representativeness of the sample, which often fails to capture the diversity of the target population, leading to potential biases and limited generalizability of the study results.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, King and He propose multiple methodologies to enhance external validity. They emphasize the necessity of using random sampling techniques to ensure the sample accurately represents the broader population. Additionally, they suggest the use of statistical adjustments and modeling to account for any sample biases that might arise, thereby improving the robustness and applicability of the findings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The article discusses various features that can influence the generalizability of research findings. These include demographic differences, such as age, gender, and socio-economic status, as well as contextual factors like cultural norms and organizational environments. These variables can significantly affect the external validity of IS survey research, necessitating careful consideration and adjustment during study design and analysis.

Examples of Problematic Extrapolation / External Validity

King and He provide examples illustrating the pitfalls of poor external validity. They highlight instances where survey results from a homogenous group, such as university students or employees from a single organization, failed to apply to a more diverse population. Such cases underscore the need for more inclusive and representative sampling strategies to ensure broader applicability of research findings.

Examples of Successful Extrapolation / External Validity

The article does not explicitly provide examples of successful extrapolation but implies that studies employing rigorous sampling techniques and thorough contextual analysis are more likely to achieve high external validity. These practices ensure that the study's conclusions are more robust and applicable to various settings and populations.

Transferring an Urban Food Security Program

In considering the case of transferring an urban food security program from one city to another, the article suggests that ensuring external validity would require a

comprehensive understanding of the target population's characteristics and contextual factors in the new city. By employing representative sampling and accounting for local demographic and contextual differences, policymakers can better gauge the program's potential effectiveness and adaptability to the new setting.

In summary, King and He's examination of external validity in IS survey research underscores the critical need for representative sampling and contextual analysis to ensure that research findings are generalizable and applicable across different populations and settings. This approach is essential for translating research into practical, effective policies and programs.

2.48 External Validity in Policy Evaluations That Choose Sites Purposively (Olsen et al., 2013)

Olsen et al. (2013) investigate the challenges of achieving external validity in policy evaluations that use purposive site selection rather than random sampling. They develop a conceptual model to understand and quantify the bias introduced by purposive site selection, highlighting its impact on the generalizability of evaluation findings.

Primary Problems in Extrapolation / External Validity

The authors identify a significant issue with purposive site selection: the resulting sample of sites may not be representative of the broader population. This selection bias can lead to incorrect estimates of the program's impact when generalized to the entire population of interest. They point out that this bias is exacerbated when there is variation in program impacts across different sites.

Methods for Addressing Problems of Extrapolation / External Validity

To address these challenges, the authors propose several strategies. They suggest using a formal model to understand the biases introduced by purposive site selection. Additionally, they recommend employing statistical methods to adjust for these biases, such as reweighting the sample to better reflect the population or using stratified random sampling where feasible.

Features of New Contexts and/or Target Populations Affecting Extrapolation / External Validity

Key factors that can influence the external validity include demographic differences, the specific needs of the target population, and varying local conditions. The authors

emphasize the need to account for these factors in both the design and analysis phases to ensure that findings are more broadly applicable.

Examples of Problematic Extrapolation / External Validity

The paper discusses instances where purposive site selection has led to significant biases. For example, evaluations of welfare-to-work programs showed considerable variation in impacts across different sites, leading to misleading conclusions when generalized. Such examples underscore the importance of considering site variability in the analysis.

Examples of Successful Extrapolation / External Validity

While the paper primarily focuses on the challenges, it implies that evaluations which employ rigorous methodologies to adjust for selection biases are more likely to achieve successful extrapolation. These approaches can help mitigate the impact of non-representative sampling.

Transferring an Urban Food Security Program

In the context of transferring an urban food security program, Olsen et al. would advise ensuring that the site selection process considers the representativeness of the new city's population. Using methods like reweighting or stratified sampling could help in adjusting the program's design to better fit the new context, thereby enhancing its effectiveness and generalizability.

In summary, Olsen et al. emphasize the importance of understanding and adjusting for the biases introduced by purposive site selection to improve the external validity of policy evaluations. By employing appropriate statistical methods and considering contextual factors, researchers can enhance the generalizability and applicability of their findings.

2.49 External Validity Is More Than Skin Deep: Some Answers to Criticisms of Laboratory Experiments (Berkowitz & Donnerstein, 1982)

Berkowitz and Donnerstein (1982) respond to criticisms of laboratory experiments in psychology, specifically addressing concerns about their external validity. They argue that the criticisms often stem from a misunderstanding of the goals of laboratory experiments and the concept of external validity.

Main Problems in Extrapolation / External Validity

The authors identify a primary problem in the criticism of laboratory experiments: the assumption that ecological validity (the resemblance of the laboratory setting to the real world) is necessary for external validity. They argue that this is not necessarily the case and that the meaning assigned by subjects to the laboratory setting and their actions is more important for generalizability.

Methods for Fixing Problems of Extrapolation / External Validity

Berkowitz and Donnerstein suggest focusing on experimental realism rather than mundane realism. This means ensuring that the psychological processes elicited in the laboratory are the same as those in real-world settings, rather than simply making the physical setting of the experiment resemble real-world situations.

Features of New Contexts and/or Target Populations Affecting Extrapolation / External Validity

Key features affecting external validity include the participants' interpretation of the experimental situation and their motivations. The authors stress that understanding these psychological aspects is crucial for determining whether laboratory findings can be generalized to other settings.

Examples of Problematic Extrapolation / External Validity

The authors provide examples from aggression research where critics argue that laboratory settings fail to capture the complexities of real-world aggression. They counter that the laboratory setting can still yield valid findings if the psychological processes involved are correctly interpreted and replicated.

Examples of Successful Extrapolation / External Validity

While specific examples are not detailed, Berkowitz and Donnerstein imply that successful extrapolation occurs when laboratory studies focus on experimental realism and ensure that the psychological processes are similar to those in real-world contexts.

Transferring an Urban Food Security Program

In the context of transferring an urban food security program, Berkowitz and Donnerstein's approach would suggest ensuring that the psychological processes and motivations involved in the program are understood and maintained in the new setting. This involves focusing on how participants interpret the program and its components rather than merely replicating the physical aspects of the original program.

In summary, Berkowitz and Donnerstein (1982) emphasize the importance of experimental realism and the psychological interpretation of experimental settings in achieving external validity. By focusing on these aspects, researchers can enhance the generalizability of laboratory findings to real-world situations.

2.50 External Validity of Trivial Experiments: The Case of Laboratory Aggression (Anderson & Bushman, 1997)

Anderson and Bushman (1997) examine the external validity of laboratory experiments, specifically in the context of aggression research. They address the criticism that laboratory studies are too artificial to be applicable to real-world scenarios, arguing instead that the theoretical relations identified in laboratory settings can indeed generalize to real-world contexts.

Main Problems in Extrapolation / External Validity

The primary problem discussed is the perceived lack of external validity in laboratory experiments due to their artificial nature. Critics argue that findings from such controlled settings do not apply to real-world situations where numerous uncontrolled variables are at play.

Methods for Fixing Problems of Extrapolation / External Validity

To counter these criticisms, the authors propose that the key to external validity lies in ensuring that the psychological processes observed in the laboratory are the same as those in real-world settings. They suggest using meta-analytic techniques to compare findings from laboratory and field studies, thereby validating the generalizability of laboratory results.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key factors affecting external validity include the similarity between the study sample and the broader population, the ecological validity of the experimental setup, and the psychological realism of the tasks and stimuli used. The authors emphasize that ensuring conceptual similarity rather than physical similarity is crucial for generalization.

Examples of Problematic Extrapolation / External Validity

The paper discusses how laboratory aggression studies have been criticized for not capturing the complexities of real-world aggression. Critics argue that such studies

oversimplify the phenomena and fail to account for the myriad of factors influencing aggressive behavior outside the lab.

Examples of Successful Extrapolation / External Validity

Anderson and Bushman provide meta-analytic evidence showing that the effects of various situational and individual difference variables on aggression are similar in both laboratory and real-world settings. This evidence supports their argument that laboratory findings can be generalized to broader contexts.

Transferring an Urban Food Security Program

In the context of transferring an urban food security program, the authors would likely recommend ensuring that the key psychological processes and motivations involved in the original program are maintained in the new setting. This involves focusing on the underlying mechanisms rather than the superficial aspects of the program, ensuring that the new implementation context matches the conceptual framework of the original.

In summary, Anderson and Bushman argue that laboratory experiments can achieve external validity if they focus on the generalizability of the underlying theoretical processes rather than the superficial details of the experimental setup. By using meta-analytic techniques and ensuring psychological realism, researchers can enhance the applicability of their findings to real-world situations.

2.51 External Validity of a Framed Field Experiment (Lusk et al., 2006)

Jayson L. Lusk, J.R. Pruitt, and Bailey Norwood (2006) examine the external validity of framed field experiments by comparing the results of a value elicitation experiment conducted in a grocery store with actual retail sales of a new product. Their study explores whether consumer behavior observed in an experimental setting reflects behavior in a naturally occurring market.

Main Problems in Extrapolation / External Validity

The authors highlight that a significant concern in experimental economics is whether the behavior of experimental subjects mirrors the behavior of real buyers and sellers in the field. They note that pro-social behavior, such as a preference for environmentally friendly products, tends to be more pronounced in experimental settings where subjects know they are being observed. This discrepancy raises questions about the external validity of such experiments.

Methods for Fixing Problems of Extrapolation / External Validity

To address these concerns, Lusk et al. propose conducting framed field experiments, which are carried out in a field context with a field commodity, but where participants are aware they are part of an experiment. This approach aims to bridge the gap between laboratory experiments and real-world settings. They conducted a framed field experiment in a grocery store where participants were asked to choose between different types of pork chops, including antibiotic-friendly options. The subsequent sales data of these pork chops in the grocery store provided a natural field experiment for comparison.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study emphasizes the need to consider how subjects' knowledge of being observed can influence their behavior. It also points out that the socio-economic and demographic characteristics of the experimental sample can differ from those of the broader market, potentially affecting the generalizability of findings.

Examples of Problematic Extrapolation / External Validity

The authors refer to previous studies, such as List (2006), which found that subjects exhibited significant pro-social behavior in laboratory settings that disappeared in a naturally occurring market. This discrepancy highlights the difficulty of ensuring that experimental results accurately reflect real-world behavior.

Examples of Successful Extrapolation / External Validity

Lusk et al. find that the results of their framed field experiment corresponded reasonably well with actual retail sales, suggesting that framed field experiments can provide a close approximation of consumer behavior in the marketplace. However, they also note a tendency for subjects to exhibit more pro-social behavior in the experiment than in the natural market, consistent with the findings of Levitt and List (2005).

Transferring an Urban Food Security Program

The principles discussed by Lusk et al. can be applied to transferring an urban food security program. Policymakers should consider conducting framed field experiments in the target city to simulate the program's implementation and observe potential consumer behavior in a controlled yet realistic setting. This involves setting up experimental booths or pilot projects within the community to gather data on how residents respond to different aspects of the food security program.

By comparing these experimental results with actual outcomes from similar programs in other cities, policymakers can better understand how contextual factors such as

socio-economic conditions and local cultural practices might influence the program's success. Conducting natural field experiments by implementing small-scale versions of the program and monitoring their performance can provide further insights into how well the program will scale in the new urban setting.

In summary, Lusk et al.'s study underscores the value of framed field experiments in evaluating external validity and suggests that this approach can help policymakers predict the success of programs like urban food security initiatives in new contexts. By systematically testing and adjusting the program based on experimental and real-world data, policymakers can enhance its effectiveness and applicability across different urban settings.

2.52 External Validity of Randomised Controlled Trials: To Whom Do the Results of This Trial Apply? (Rothwell, 2005)

Rothwell (2005) discusses the often-overlooked issue of external validity in randomized controlled trials (RCTs) and its implications for clinical practice. The article highlights how RCTs, despite being the gold standard for determining treatment effects, frequently lack consideration of generalizability, leading to underuse of effective treatments in routine practice.

Primary Problems in Extrapolation / External Validity

The main problems identified include the non-representativeness of trial participants compared to the broader patient population and the controlled conditions of RCTs, which may not reflect real-world clinical settings. This discrepancy can lead to findings that are not applicable to everyday medical practice.

Methods for Fixing Problems of Extrapolation / External Validity

Rothwell proposes several methods to address these issues, such as ensuring the trial population is representative of the broader patient population by including diverse participants and settings. He also suggests better reporting of the context and conditions of trials, along with guidelines and checklists to assess and enhance external validity.

Features of New Contexts and/or Target Populations Affecting Extrapolation / External Validity

Key factors affecting extrapolation include differences in healthcare systems, patient demographics, and clinical practices across different regions and populations. These variations can significantly impact the applicability of trial results to different settings.

Examples of Problematic Extrapolation / External Validity

The article provides examples from various medical fields where RCT findings did not translate well into practice due to differences in patient populations and treatment settings. For instance, certain cardiovascular treatments that were effective in trials showed less benefit in broader clinical practice due to differences in patient characteristics and treatment adherence.

Examples of Successful Extrapolation / External Validity

While specific examples of successful extrapolation are not detailed, Rothwell emphasizes that trials with well-reported contextual details and diverse participant groups are more likely to have findings that are generalizable to broader clinical settings.

Transferring an Urban Food Security Program

In the context of transferring an urban food security program, Rothwell's insights suggest ensuring that the program's evaluation includes a representative sample of the target population and considers the local healthcare and socio-economic conditions. Detailed reporting and contextual analysis would be crucial to adapt the program effectively to the new setting.

In summary, Rothwell underscores the critical need for greater attention to external validity in RCTs to ensure that their findings are applicable to real-world clinical practice. By enhancing the representativeness of trial populations and improving the reporting of trial contexts, researchers can bridge the gap between research findings and everyday medical care.

2.53 External Validity, Generalisability, Applicability, and Directness: A Brief Primer (Murad et al., 2018)

Murad et al. (2018) provide a comprehensive overview of the concepts of external validity, generalisability, and applicability in clinical research. They differentiate between these terms and discuss their relevance in translating research findings into clinical practice.

Primary Problems in Extrapolation / External Validity

The authors identify that a major issue is the confusion between generalisability and applicability. Generalisability pertains to extending results from a sample to the broader population from which the sample was drawn, while applicability concerns using study findings in the care of patients from any population. A common problem is that many clinical trials do not use random sampling, which limits generalisability. Additionally, narrow eligibility criteria in trials can reduce the applicability of findings to broader, real-world patient populations.

Methods for Fixing Problems of Extrapolation / External Validity

To improve external validity, the authors suggest employing random sampling where possible and ensuring that study populations are representative of the broader patient population. They also recommend assessing the similarity between trial populations and target patient groups in terms of clinical, socio-economic, and healthcare factors to enhance applicability.

Features of New Contexts and/or Target Populations Affecting Extrapolation / External Validity

Key features include clinical characteristics, such as disease severity and comorbidities, socio-economic status, and healthcare system factors. Differences in these areas between the study and the target population can significantly impact the external validity of research findings.

Examples of Problematic Extrapolation / External Validity

The paper highlights that trials often invite convenient samples and rely on patient volunteers, which can result in non-representative samples. This limits the ability to generalize findings to the broader population. For example, trials with narrow eligibility criteria might only apply to a small subset of the population.

Examples of Successful Extrapolation / External Validity

While specific examples are not provided, the authors imply that studies with broader eligibility criteria and more representative samples tend to have higher applicability.

They also suggest that using frameworks like GRADE, which considers indirectness, can help in evaluating and improving the applicability of evidence.

Transferring an Urban Food Security Program

For transferring an urban food security program, the principles outlined by Murad et al. would suggest ensuring that the target population in the new city closely matches the characteristics of the original study population. This includes considering socio-economic, cultural, and healthcare factors to ensure the program's applicability and effectiveness.

In summary, Murad et al. emphasize the distinction between generalisability and applicability and highlight the importance of ensuring study populations are representative and relevant to target patient groups. By addressing these factors, researchers can improve the external validity and applicability of their findings in clinical practice.

2.54 External Validity, Generalizability, and Knowledge Utilization (Ferguson, 2004)

Ferguson (2004) delves into the essential concepts of external validity and generalizability, emphasizing their significance in research utilization and knowledge application, particularly within the context of evidence-based practice. The article explores strategies to enhance the generalizability of research findings, addressing the inherent conflicts between internal and external validity in research designs.

Main Problems in Extrapolation / External Validity

One primary issue highlighted is the inverse relationship between internal validity and external validity. When researchers focus on controlling extraneous variables to ensure internal validity, they may inadvertently compromise the generalizability of their findings. Additionally, threats such as interaction effects of selection biases with experimental variables, reactive effects of experimental arrangements, and multiple treatment interference are noted as significant challenges.

Methods for Fixing Problems of Extrapolation / External Validity

Ferguson suggests several methods to mitigate these problems, including the use of random selection and random assignment of participants to ensure representative samples. The inclusion of multiple sites and varied settings in research designs can also help address threats related to interaction effects of treatment and setting. Meta-

analysis and meta-synthesis are recommended to increase sample size and validate findings across different settings and populations.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features affecting extrapolation include the representativeness of the sample and the setting of the study. Ferguson stresses the importance of ensuring that the sample accurately reflects the target population and that the study's context is relevant to other settings. The article also highlights the need for researchers to consider and report on the effects of historical events and other contextual factors that may influence the generalizability of their findings.

Examples of Problematic Extrapolation / External Validity

The article discusses how strict controls to ensure internal validity can limit the external validity of findings, making them less applicable to broader populations or different settings. Specific examples of these issues are not provided, but the general concept is thoroughly examined.

Examples of Successful Extrapolation / External Validity

While the article does not provide specific examples of successful extrapolation, it emphasizes the importance of methodological rigor and transparency in reporting research findings. The use of comprehensive meta-analyses and the establishment of research partnerships with practitioners and policymakers are suggested as ways to enhance the applicability of research results.

Transferring an Urban Food Security Program

If transferring an urban food security program from one city to another, Ferguson would likely recommend ensuring the representativeness of the sample and the relevance of the research context. The use of random selection, multiple sites, and meta-analytic approaches to validate findings across different settings would be essential. Additionally, fostering partnerships with local stakeholders to address specific contextual factors would enhance the program's generalizability and effectiveness in the new setting.

In summary, Ferguson (2004) provides a comprehensive framework for understanding and enhancing the external validity and generalizability of research findings, emphasizing the need for methodological rigor and practical strategies to bridge the gap between research and practice.

2.55 External Validity: From Do-Calculus to Transportability (Pearl & Bareinboim, 2022)

Pearl and Bareinboim (2022) discuss the challenges and methodologies for addressing external validity in the context of generalizing causal inferences from one population to another. They introduce formal tools such as selection diagrams and transport formulas to provide a structured approach to transportability analysis. These tools help in understanding the conditions under which causal effects estimated in one setting can be validly transported to another.

Main Problems in Extrapolation / External Validity

The authors identify the primary issue of extrapolation as the differences in populations that can affect the transportability of causal effects. They emphasize that without formal tools to account for these differences, generalizations can be invalid.

Methods for Fixing Problems of Extrapolation / External Validity

The paper proposes the use of selection diagrams and theorems to create transport formulas that adjust for population differences. These formulas recalibrate learned relationships to account for these differences, thus enabling accurate transport of causal inferences.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features include demographic differences, such as age distribution, and latent variables that might differ between the original and target populations. Understanding these differences is crucial for applying the correct transport formulas.

Examples of Problematic Extrapolation / External Validity

An example discussed involves estimating causal effects from a study in Los Angeles and attempting to generalize them to New York City, considering significant demographic differences like age distribution.

Examples of Successful Extrapolation / External Validity

The paper provides a theoretical framework showing that successful extrapolation is possible when the correct transport formulas are applied, demonstrating this through various mathematical proofs and theoretical examples.

Transferring an Urban Food Security Program

When considering the transfer of an urban food security program from one city to another, the article suggests using selection diagrams to identify and adjust for key demographic and contextual differences. This ensures that the causal effects

observed in the original city can be validly applied to the new city by recalibrating the learned relationships.

In summary, Pearl and Bareinboim's work offers a comprehensive theoretical framework for understanding and addressing the challenges of external validity, providing practical tools for researchers to ensure their findings can be accurately generalized across different populations and contexts.

2.56 External Validity: Is There Still a Problem? (Marcellesi, 2015)

Marcellesi (2015) addresses the issue of external validity, proposing a distinction between two types of external validity inferences: predictive and explanatory. He argues that while predictive external validity inferences have been adequately addressed, explanatory inferences remain problematic.

Main Problems in Extrapolation / External Validity

Marcellesi identifies a significant issue in the understanding and application of predictive and explanatory external validity inferences. Predictive inferences concern the ability to generalize findings to new populations, while explanatory inferences deal with explaining phenomena observed in different contexts. He argues that there is a satisfactory framework for predictive inferences, but explanatory inferences are still inadequately understood.

Methods for Fixing Problems of Extrapolation / External Validity

For predictive external validity, Marcellesi references the work of Cartwright and Hardie, as well as Bareinboim and Pearl, who have developed robust methods to ensure that causal inferences are valid across different populations. These methods involve identifying and matching relevant factors that affect the outcome in both the original and target populations.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features affecting the external validity of predictive inferences include the presence of similar causal mechanisms and contextual factors in both the original and new populations. For explanatory inferences, understanding the underlying mechanisms and how they interact with different contextual variables is crucial.

Examples of Problematic Extrapolation / External Validity

Marcellesi discusses how failing to account for differences in causal mechanisms and contextual factors can lead to poor external validity. For example, a policy

proven effective in one school district may not yield the same results in another if the contextual factors differ significantly.

Examples of Successful Extrapolation / External Validity

He highlights that successful predictive inferences often involve rigorous methodological approaches that account for and adjust to the differences between populations. Theoretical advancements by Cartwright and Hardie, and Bareinboim and Pearl, provide frameworks for such successful generalizations.

Transferring an Urban Food Security Program

Although the article does not specifically address transferring an urban food security program, the principles discussed by Marcellesi could be applied to analyze the validity of such a transfer. By ensuring that the causal mechanisms and contextual factors affecting food security are similar in both cities, and by using robust predictive inference methods, policymakers could enhance the likelihood of successful program implementation.

In summary, Marcellesi argues that while predictive external validity has been effectively addressed through advanced methodologies, explanatory external validity remains a complex issue. The principles discussed can potentially be applied to various contexts, including the transfer of urban food security programs, by ensuring methodological rigor and contextual similarity.

2.57 External Validity: The Neglected Dimension in Evidence Ranking (Persaud & Mamdani, 2006)

Persaud and Mamdani (2006) argue that external validity is often neglected in the ranking of evidence, which traditionally emphasizes internal validity. They propose that both internal and external validity should be considered equally to ensure that clinical evidence is both accurate and relevant.

Main Problems in Extrapolation / External Validity

The authors highlight that many randomized controlled trials (RCTs) focus heavily on internal validity, often at the expense of external validity. This focus can lead to findings that do not generalize well to broader, more diverse patient populations. They point out that differences in study populations, settings, and clinician decision-making processes can all impact the generalizability of results.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Persaud and Mamdani suggest incorporating external validity considerations into evidence ranking systems. They propose a two-dimensional ranking system that evaluates both internal and external validity. This includes assessing how closely study populations and settings match those in actual clinical practice and considering patient preferences and behaviors.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features affecting external validity include the representativeness of the study population, the types of institutions and physicians involved, and the role of clinician and patient decision-making. The authors stress the importance of ensuring that these features align with those in the target population to enhance the applicability of the findings.

Examples of Problematic Extrapolation / External Validity

The article provides the example of the Randomized Aldactone Evaluation Study (RALES) for heart failure treatment, where subsequent non-experimental studies found a higher incidence of hyperkalemia in real-world settings compared to the controlled trial. This discrepancy highlights the dangers of relying solely on internally valid evidence without considering external validity.

Examples of Successful Extrapolation / External Validity

While the paper does not provide specific examples of successful extrapolation, it suggests that combining high internal validity (from RCTs and meta-analyses) with high external validity (from non-experimental studies) can lead to more reliable and applicable clinical evidence.

Transferring an Urban Food Security Program

Although the article does not explicitly address transferring an urban food security program, the principles discussed could be applied to analyze the validity of such a transfer. Ensuring that the program's evaluation includes both internally and externally valid evidence and that the target population in the new city closely matches the original study population would be essential steps.

In summary, Persaud and Mamdani emphasize the importance of considering both internal and external validity in evidence ranking to ensure that clinical research findings are both accurate and applicable to real-world settings. This approach can help bridge the gap between research and practice, enhancing the relevance and impact of clinical evidence.

2.58 External Validity: The Next Step for Systematic Reviews (Avellar et al., 2017)

Avellar et al. (2017) emphasize the importance of considering external validity in systematic reviews to enhance the applicability of research findings to diverse populations and settings. They argue that while systematic reviews are crucial for synthesizing evidence, their usefulness is limited if the included studies lack external validity.

Main Problems in Extrapolation / External Validity

The authors identify a significant issue in the generalizability of findings from systematic reviews, primarily due to the inclusion of studies with narrow participant criteria and controlled environments that do not reflect real-world conditions. This limits the applicability of the synthesized evidence to broader populations.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Avellar et al. propose incorporating external validity assessments into systematic reviews. They suggest using frameworks and tools to evaluate the generalizability of individual studies and the overall body of evidence. This includes assessing the similarity between study populations and target populations, and considering contextual factors that may influence the applicability of the findings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features include demographic differences, cultural contexts, and variations in healthcare systems and practices. The authors stress the importance of understanding these factors to enhance the external validity of systematic review findings and ensure they are relevant to diverse populations.

Examples of Problematic Extrapolation / External Validity

The paper highlights instances where systematic reviews have included studies with limited external validity, leading to recommendations that do not apply well to broader populations. For example, reviews of interventions tested predominantly on young, healthy individuals may not be applicable to older adults with multiple comorbidities.

Examples of Successful Extrapolation / External Validity

While specific examples of successful extrapolation are not detailed, the authors imply that systematic reviews which assess and report on external validity are more likely to produce generalizable and applicable findings. Integrating studies with

diverse populations and settings can enhance the relevance of the synthesized evidence.

Transferring an Urban Food Security Program

Although the article does not specifically address transferring an urban food security program, the principles discussed could be applied to analyze the validity of such a transfer. Ensuring that the program's evaluation includes studies with high external validity and considering the demographic and contextual differences between the original and new cities would be essential steps.

In summary, Avellar et al. highlight the necessity of incorporating external validity assessments in systematic reviews to ensure that their findings are applicable to diverse populations and real-world settings. This approach can enhance the relevance and impact of the synthesized evidence in practice.

2.59 External Validity: We Need to Do More (Glasgow et al., 2006)

Glasgow et al. (2006) discuss the crucial importance of external validity in health research, emphasizing the need to improve the design, evaluation, and reporting of applied research to enhance its applicability in real-world settings. They highlight the gap between research findings and their implementation in practice, stressing the importance of addressing external validity to bridge this gap.

Main Problems in Extrapolation / External Validity

The authors identify a significant problem in the generalizability of research findings due to the controlled and often idealized settings of many studies. This limits the applicability of these findings to more diverse and less controlled real-world environments. They argue that current reporting standards, like the CONSORT criteria, focus primarily on internal validity, often neglecting external validity.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Glasgow et al. propose adopting new reporting standards that emphasize external validity. They suggest including detailed information on the context, settings, and populations of studies to enable better assessment of their applicability. The authors advocate for the use of practical clinical trials and implementation research methods that prioritize generalizability and feasibility in diverse settings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features that impact external validity include the representativeness of study samples, the relevance of the settings to real-world conditions, and the adaptability of interventions to different populations and contexts. The authors emphasize the need to report on factors such as participation rates, representativeness, implementation consistency, and outcomes relevant to decision-makers.

Examples of Problematic Extrapolation / External Validity

The article highlights the lack of information on external validity in many intervention studies, which makes it difficult for practitioners and policymakers to judge the relevance of research findings to their specific contexts. For instance, studies conducted in highly controlled environments may not account for variations in practitioner expertise or patient characteristics in typical clinical settings.

Examples of Successful Extrapolation / External Validity

While the article does not provide specific examples of successful extrapolation, it implies that studies which report extensively on contextual and implementation factors are more likely to produce findings that are applicable to broader populations and settings.

Transferring an Urban Food Security Program

Although the article does not specifically address transferring an urban food security program, the principles discussed by Glasgow et al. could be applied to analyze the validity of such a transfer. Ensuring detailed reporting on the context, participant characteristics, and implementation methods would help in assessing the program's applicability to a new city.

In summary, Glasgow et al. call for a greater focus on external validity in health research to enhance the applicability and impact of findings in real-world settings. By adopting comprehensive reporting standards and prioritizing practical clinical trials, researchers can improve the relevance and utility of their work for practitioners and policymakers.

2.60 ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework (Angrist & Fernandez-Val, 2010)

Angrist and Fernandez-Val (2010) develop a covariate-based approach to enhancing the external validity of instrumental variables (IV) estimates within the Local Average Treatment Effect (LATE) framework. They explore how differences in the

characteristics of compliers can influence IV estimates and propose methods for constructing externally valid estimates for new subpopulations.

Main Problems in Extrapolation / External Validity

The primary issue identified is the instrument-specific nature of IV estimates. Each instrument generates its own compliant subpopulation, and the causal effects estimated for one group may not generalize to another. This lack of generalizability raises concerns about the external validity of IV estimates.

Methods for Fixing Problems of Extrapolation / External Validity

The authors suggest constructing covariate-specific LATEs and using reweighting procedures to create estimates for new subpopulations. They introduce a method that employs the traditional overidentification test statistic to define a population for which a pair of IV estimates holds external validity. This method involves comparing and reconciling the observed characteristics of compliers.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features include the demographic and socio-economic characteristics of the compliant subpopulations. Differences in these characteristics can significantly impact the generalizability of IV estimates. The authors emphasize the need to account for these differences when extrapolating causal effects to new populations.

Examples of Problematic Extrapolation / External Validity

The paper discusses how different IV estimates, such as those derived from twins versus sex-composition instruments for studying the effects of childbearing on labor supply, can produce significantly different results due to variations in compliant subpopulations. These differences highlight the challenges in achieving external validity when the characteristics of compliers differ across instruments.

Examples of Successful Extrapolation / External Validity

While the paper does not provide specific successful examples, it illustrates that by using covariate-specific reweighting techniques, researchers can improve the external validity of IV estimates. This approach allows for the construction of more generalizable causal effect estimates across different subpopulations.

Transferring an Urban Food Security Program

Although the article does not specifically address transferring an urban food security program from one city to another, the principles discussed could potentially be applied. By using covariate-specific reweighting techniques, policymakers could

ensure that the characteristics of the original program's target population are matched with those of the new city's population, thereby enhancing the program's applicability and effectiveness.

In summary, Angrist and Fernandez-Val (2010) propose a methodological approach to improve the external validity of IV estimates by focusing on the characteristics of compliant subpopulations. Their framework offers a robust solution to the challenge of generalizing causal effects across different populations, providing valuable insights for researchers and policymakers alike.

2.61 Extrapolation of Causal Effects – Hopes, Assumptions, and the Extrapolator's Circle (Khosrowi, 2019)

Khosrowi (2019) discusses the challenges associated with extrapolating causal effects from experimental to target populations. He critically examines various strategies proposed in the econometrics literature and introduces the concept of the "extrapolator's circle," which highlights the paradoxical requirement for extensive knowledge about the target population that can render the need for extrapolation redundant.

Main Problems in Extrapolation / External Validity

The primary issue discussed is the extrapolator's circle, where the knowledge required about the target population for successful extrapolation is so extensive that one could derive the causal effects from the target population alone, making extrapolation unnecessary. This challenge is prevalent in both econometrics and evidence-based policy.

Methods for Fixing Problems of Extrapolation / External Validity

Khosrowi reviews several strategies, including interactive covariate-based approaches and comparative process tracing (CPT). He suggests that while CPT is designed to evade the extrapolator's circle, it may not be effective in typical econometrics and evidence-based policy applications. The author advocates for the consideration of qualitative evidence alongside quantitative data to address these issues more comprehensively.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features include the distribution of interactive covariates, the parameters governing causal mechanisms, and the structural makeup of these mechanisms in

both experimental and target populations. Differences in these areas can significantly impact the validity of extrapolating causal effects.

Examples of Problematic Extrapolation / External Validity

The paper provides examples from the econometrics literature, such as the difficulty in ensuring that variables like teacher quality (an interactive covariate) are similarly distributed across populations when extrapolating the effects of educational interventions.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not detailed, the author implies that successful extrapolation requires robust methodologies that account for differences in covariate distributions and mechanisms. Integrating qualitative evidence is suggested as a promising approach to achieving better extrapolation outcomes.

Transferring an Urban Food Security Program

Although the article does not specifically address transferring an urban food security program, the principles discussed by Khosrowi could potentially be applied. By using both qualitative and quantitative evidence to understand the differences and similarities between the original and new populations, policymakers could better assess the validity of transferring such a program to a different city.

In summary, Khosrowi (2019) highlights the significant challenges and paradoxes involved in extrapolating causal effects. He suggests that incorporating qualitative evidence can help address these issues and improve the external validity of such extrapolations.

2.62 Factors That Can Affect the External Validity of Randomised Controlled Trials (Rothwell, 2006)

Rothwell (2006) examines various factors that influence the external validity of randomized controlled trials (RCTs). He emphasizes the importance of designing and reporting trials in a manner that allows clinicians to assess their applicability to diverse patient populations and clinical settings.

Main Problems in Extrapolation / External Validity

Rothwell identifies several key issues, including the selection of trial settings, participating centers and clinicians, and patient populations. He points out that the stringent control over internal validity often results in trial conditions that do not reflect real-world clinical environments, limiting the generalizability of the findings.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Rothwell suggests enhancing the reporting of trial settings and the selection processes for centers, clinicians, and patients. He advocates for the use of pragmatic trials that mirror real-world conditions more closely. Additionally, he recommends including detailed information on trial eligibility and exclusion criteria, run-in periods, and the treatment protocols used.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features affecting external validity include differences in healthcare systems, the demographic and clinical characteristics of patient populations, and variations in diagnostic and treatment practices. Understanding these factors is essential for assessing the generalizability of trial results.

Examples of Problematic Extrapolation / External Validity

Rothwell discusses the variability in trial outcomes due to differences in healthcare systems and patient management practices. For instance, he highlights how the European Carotid Surgery Trial showed different treatment effects in countries with varying healthcare system efficiencies, illustrating the impact of contextual differences on trial results.

Examples of Successful Extrapolation / External Validity

While the article does not provide specific examples of successful extrapolation, it implies that trials designed with broad eligibility criteria and inclusive of diverse settings are more likely to produce generalizable results. The GISSI-1 trial of thrombolysis for acute myocardial infarction is noted for its high recruitment rates, which enhanced its external validity.

Transferring an Urban Food Security Program

Although the article does not explicitly discuss transferring an urban food security program, the principles discussed by Rothwell could be applied to analyze the validity of such a transfer. Ensuring that the program's design and evaluation consider the specific healthcare systems, demographic characteristics, and local practices of the new city would be crucial for effective implementation.

In summary, Rothwell (2006) emphasizes the importance of considering external validity in the design and reporting of RCTs. By addressing the factors that influence generalizability, researchers can improve the applicability of trial findings to diverse patient populations and clinical settings.

2.63 Fidelity and Adaptation of Programs: Does Adaptation Undermine Fidelity or Strengthen It? (Nolt & Leviton, 2023)

Nolt and Leviton (2023) explore the balance between fidelity and adaptation in program implementation, particularly focusing on how adaptation can impact the fidelity and effectiveness of programs. They investigate whether changes made to programs during implementation help or hinder their intended outcomes, emphasizing the importance of context in these modifications.

Main Problems in Extrapolation / External Validity

The authors highlight the challenge of maintaining program fidelity while allowing for necessary adaptations to fit new contexts. They argue that strict adherence to original program designs may not always be feasible or desirable when transferring programs to different settings, leading to concerns about whether these adaptations compromise the program's effectiveness.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Nolt and Leviton suggest using a framework that balances fidelity and adaptation. This involves identifying core components of the program that must be maintained to ensure its integrity while allowing flexibility in other areas to accommodate local needs and contexts. They recommend ongoing evaluation and feedback mechanisms to monitor the impact of adaptations on program outcomes.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features include cultural, socio-economic, and environmental differences between the original and new settings. The authors emphasize the importance of understanding these contextual factors and engaging local stakeholders in the adaptation process to ensure the program remains relevant and effective.

Examples of Problematic Extrapolation / External Validity

The article discusses instances where rigid adherence to original program designs has led to poor outcomes in new settings. For example, a health intervention program developed in a high-resource urban area may not perform well in a low-resource rural setting without appropriate adaptations.

Examples of Successful Extrapolation / External Validity

Nolt and Leviton provide examples where thoughtful adaptations have enhanced the effectiveness of programs in new contexts. They highlight cases where

engaging local communities and modifying program components to fit local needs have resulted in successful implementation and positive outcomes.

Transferring an Urban Food Security Program

Although the article does not specifically address transferring an urban food security program, the principles discussed by Nolt and Leviton could be applied. Ensuring a balance between fidelity and adaptation by identifying essential program components and allowing for context-specific modifications would be crucial. Engaging local stakeholders and continuously evaluating the adapted program's effectiveness would help in successfully transferring the program to a new city.

In summary, Nolt and Leviton (2023) emphasize the importance of balancing program fidelity with necessary adaptations to ensure successful implementation in new contexts. By understanding and accommodating local differences, programs can maintain their effectiveness while being adapted to meet the needs of diverse populations.

2.64 From Local to Global: External Validity in a Fertility Experiment (Dehejia et al., 2021)

Dehejia et al. (2021) explore the challenges of establishing external validity when scaling up experimental findings from local contexts to broader, more diverse populations. Their study focuses on a fertility experiment conducted in developing countries and examines the factors that can influence the generalizability of the results.

Main Problems in Extrapolation / External Validity

The authors identify several key issues with external validity. The primary problem is the variability in local contexts, which can lead to different outcomes when the same intervention is applied in different settings. This includes differences in cultural norms, economic conditions, and access to resources, which can all impact the effectiveness of the intervention.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Dehejia et al. propose a framework that involves collecting detailed data on local conditions and using this information to adjust the intervention accordingly. This includes using statistical techniques such as meta-analysis to combine results from multiple studies and identify common patterns, as well as conducting pilot studies in new settings before full-scale implementation.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The authors emphasize the importance of understanding the specific characteristics of the new context and target population. This includes demographic factors such as age and income, as well as cultural attitudes towards fertility and family planning. By taking these factors into account, researchers can better assess whether the intervention is likely to be effective in the new setting.

Examples of Problematic Extrapolation / External Validity

The paper provides examples of cases where interventions that were successful in one context failed in another due to differences in local conditions. For instance, a fertility program that worked well in urban areas might not be as effective in rural areas with different social and economic structures.

Examples of Successful Extrapolation / External Validity

While the authors do not provide specific examples of successful extrapolation, they highlight the importance of using rigorous methods to test and adapt interventions to new contexts. This includes conducting pilot studies and using statistical techniques to account for differences between the original and new populations.

Transferring an Urban Food Security Program

Although this article does not specifically address transferring an urban food security program, the principles discussed could potentially be applied. By collecting detailed data on the new city's context and using this information to adjust the program, policymakers could enhance the program's effectiveness and ensure its successful implementation in the new setting.

In summary, Dehejia et al. emphasize the importance of understanding and adapting to local conditions when scaling up experimental findings. Their framework provides practical tools for researchers and policymakers to improve the external validity of their interventions, ensuring that they are effective across different contexts and populations.

2.65 Generalization in the Tropics – Development Policy, Randomized Controlled Trials, and External Validity (Peters et al., 2018)

Peters et al. (2018) investigate the complexities of establishing external validity in randomized controlled trials (RCTs) conducted in developing countries. They review

RCTs published between 2009 and 2014, emphasizing the necessity for transparent reporting to prevent overgeneralization and to aid policymakers in accurately interpreting results.

Main Problems in Extrapolation / External Validity

The authors highlight several critical issues affecting external validity, including variability in local contexts, which can result in different outcomes when interventions are applied in new settings. Key problems include Hawthorne and John Henry effects, general equilibrium effects, specific sample problems, and differences in how treatments are provided.

Methods for Fixing Problems of Extrapolation / External Validity

To tackle these challenges, Peters et al. propose detailed reporting and transparency, conducting pilot studies, using meta-analysis to combine results from multiple studies, and systematic reviews. They also recommend adaptation frameworks, such as those suggested by Banerjee et al. (2017), to facilitate the scaling of interventions from proof of concept to broader applications.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study underscores the importance of understanding the specific characteristics of the new context and target population, including demographic factors, economic conditions, cultural norms, and logistical capabilities. Recognizing these differences is essential for effectively adapting interventions to new settings.

Examples of Problematic Extrapolation / External Validity

The paper provides examples of RCTs encountering issues when scaled up, such as an education program that failed to replicate its positive results when implemented by a government body instead of an NGO, due to differences in implementation capacity and political pressures.

Examples of Successful Extrapolation / External Validity

While specific examples of successful extrapolation are not detailed, the authors emphasize the importance of rigorous methods and transparent reporting to enhance the likelihood of successful generalization.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. By collecting detailed data on the new city's context and adjusting the program accordingly, policymakers can improve

its effectiveness. Conducting pilot studies and utilizing meta-analyses can help identify potential issues before full-scale implementation.

Peters et al. emphasize the crucial role of external validity in RCTs for development policy. They call for more systematic and transparent reporting on external validity issues to ensure that findings are genuinely transferable to other contexts, thus providing policymakers with the necessary information to make informed decisions.

2.66 Generalizing about Public Health Interventions: A Mixed-Methods Approach to External Validity (Leviton, 2017)

Leviton (2017) examines the challenges of achieving external validity for public health interventions. The paper emphasizes the need for a mixed-methods approach to generalize findings from evidence-based interventions (EBIs) across different populations and settings. The author highlights the urgency for public health professionals to ensure that interventions are effective in real-world scenarios and discusses methods to improve the generalizability of these interventions.

Main Problems in Extrapolation / External Validity

The main issues identified include the variability in populations, settings, treatments, measurements, and historical periods. There is often insufficient attention to the details of implementation and context in which EBIs are tested, leading to challenges in generalizing the findings. The paper also points out that many registries and systematic reviews do not adequately address external validity, making it difficult for practitioners to apply EBIs effectively in diverse settings.

Methods for Fixing Problems of Extrapolation / External Validity

Leviton proposes several methods to address these challenges, including better description of EBIs and their contexts, combining statistical tools with logical inferences about study samples, defining the theory behind the intervention more sharply, and systematic consultation with practitioners. The use of mixed-methods research, which integrates quantitative and qualitative data, is recommended to expand causal generalizations and improve the applicability of interventions.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The paper stresses the importance of understanding context features likely to impact the intervention's success. These features include demographic variables, cultural norms, and logistical considerations such as resources and infrastructure. By

focusing on these elements, researchers can better tailor interventions to fit new settings.

Examples of Problematic Extrapolation / External Validity

Leviton discusses several cases where interventions failed to generalize successfully. For example, the Nurse Family Partnership program, which showed positive results in the United States, did not replicate these results in Britain due to differences in the support systems available for new parents. This highlights how contextual differences can significantly impact the effectiveness of an intervention.

Examples of Successful Extrapolation / External Validity

The article does not provide specific examples of successful extrapolation but emphasizes the importance of using robust methods and detailed reporting to improve generalization. By systematically addressing the factors that influence external validity, it is possible to increase the likelihood of successful implementation across different settings.

Transferring an Urban Food Security Program

While the article does not directly discuss transferring an urban food security program, the principles outlined can be applied. Collecting detailed data on the new city's context and adjusting the program based on this information is crucial. Pilot studies and meta-analyses can help identify potential issues before full-scale implementation, ensuring that the program is tailored to meet the specific needs of the new setting.

Leviton advocates for a systematic and thorough approach to assessing external validity in public health interventions. By combining detailed descriptions, mixed-methods research, and practical insights from practitioners, the likelihood of successful generalization and implementation of EBIs can be significantly improved.

2.67 Generalizing Causal Knowledge in the Policy Sciences: External Validity as a Task of Both Multi-Attribute Representation and Multi-Attribute Extrapolation (Cook, 2014)

Cook (2014) addresses the complexities of achieving external validity in the policy sciences by emphasizing the need for both multi-attribute representation and multi-attribute extrapolation. The paper discusses the importance of ensuring that causal findings from research can be generalized to broader, real-world contexts, which is crucial for informing effective policy decisions.

Main Problems in Extrapolation / External Validity

Cook identifies several key challenges in achieving external validity. These include the difficulty of generalizing findings from specific samples to broader populations and the complexities of extrapolating results to different settings and times. The variability in units, treatments, outcomes, settings, and times (utosti) poses significant obstacles, as does the lack of attention to these dimensions in many studies. The paper highlights that traditional sampling methods often fall short in adequately representing the populations and contexts to which researchers wish to generalize.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Cook advocates for a more systematic approach to both representation and extrapolation. He suggests using propensity score matching to enhance representativeness and response surface modeling to better understand the relationships between different factors and outcomes. Additionally, meta-analysis is recommended as a tool to identify the conditions under which causal relationships hold and to generalize findings across different contexts. Cook also emphasizes the importance of conducting scale-up studies to test interventions in broader, more varied settings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The paper underscores the necessity of understanding the specific attributes of new contexts and target populations that can affect the generalizability of causal findings. These features include demographic variables, cultural norms, economic conditions, and logistical factors that can influence the implementation and effectiveness of interventions.

Examples of Problematic Extrapolation / External Validity

Cook discusses cases where interventions that were successful in controlled settings failed when applied to broader contexts. For example, the New Jersey Negative Income Tax experiment used an incomplete factorial design that did not fully capture the variability needed for generalization, leading to challenges in applying the findings to other settings.

Examples of Successful Extrapolation / External Validity

While specific examples of successful extrapolation are not provided in detail, Cook highlights the potential of methods like meta-analysis and response surface modeling to improve the generalizability of findings. These methods help to identify the conditions under which causal relationships are likely to hold, thereby enhancing the likelihood of successful extrapolation.

Transferring an Urban Food Security Program

Although the article does not specifically address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context and adjusting the program based on this information is essential. Conducting pilot studies and using meta-analysis can help identify potential issues and ensure the program is effectively tailored to the new setting.

Cook's paper emphasizes the importance of both representation and extrapolation in achieving external validity in policy research. By adopting systematic methods and focusing on the specific attributes of new contexts, researchers can enhance the generalizability of their findings and ensure that policies are effective across diverse settings and populations.

2.68 Generalizing Treatment Effect Estimates From Sample to Population: A Case Study in the Difficulties of Finding Sufficient Data (Stuart & Rhodes, 2017)

Stuart and Rhodes (2017) investigate the challenges of generalizing treatment effect estimates from randomized trials to broader populations. Their focus is on the practical difficulties of finding sufficient data to make reliable generalizations. Using a case study of a supplemental program for Head Start centers, they illustrate the complexities involved in assessing external validity.

Main Problems in Extrapolation / External Validity

The authors identify several significant challenges in achieving external validity. A major issue is the lack of common data elements between trial samples and the target population, which makes reliable generalization difficult. They emphasize the problems posed by differences in demographic variables, socioeconomic status, and other contextual factors that can influence the effectiveness of interventions.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Stuart and Rhodes propose several methods. They suggest using propensity score matching to enhance representativeness and response surface modeling to understand the relationships between different factors and outcomes better. Meta-analysis is recommended to identify conditions under which causal relationships hold and to generalize findings across different contexts. The authors also stress the importance of conducting scale-up studies to test interventions in broader settings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Understanding the specific characteristics of the new context and target population is crucial. This includes demographic variables, cultural norms, economic conditions, and logistical factors that can influence the implementation and effectiveness of interventions. These features must be carefully considered to ensure the intervention's success in new settings.

Examples of Problematic Extrapolation / External Validity

The paper discusses cases where interventions that were successful in controlled settings faced issues when scaled up. For example, the Head Start REDI program, which showed positive results in a limited trial, encountered difficulties when generalizing findings to the national Head Start population due to differences in data elements and contextual factors.

Examples of Successful Extrapolation / External Validity

While specific examples of successful extrapolation are not provided, the authors highlight the potential of methods like meta-analysis and response surface modeling to improve generalizability. These methods help identify conditions under which causal relationships are likely to hold, thus enhancing the likelihood of successful extrapolation.

Transferring an Urban Food Security Program

Although the article does not specifically address transferring an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context and adjusting the program based on this information is essential. Conducting pilot studies and using meta-analysis can help identify potential issues and ensure the program is effectively tailored to the new setting.

Stuart and Rhodes emphasize the importance of systematic methods and detailed data collection to enhance external validity in policy research. By adopting these approaches, researchers can improve the generalizability of their findings and ensure that interventions are effective across diverse settings and populations.

2.69 How Do We Know When Research From One Setting Can Be Useful in Another? A Review of External Validity, Applicability and Transferability Frameworks (Burchett et al., 2011)

Burchett et al. (2011) review various frameworks used to assess the external validity, applicability, and transferability of health research across different settings. They highlight the need for reliable methods to determine when and how research findings can be applied to new contexts, particularly in public health.

Main Problems in Extrapolation / External Validity

The main issues identified include the lack of empirical data supporting many frameworks, insufficient detail in criteria for applicability and transferability, and a general neglect of external validity in health services literature. The authors also note the subjective and context-specific nature of these assessments, which complicates their generalization.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Burchett et al. propose developing validated frameworks that include comprehensive criteria for assessing applicability and transferability. They suggest using empirical data to support the development of these frameworks and emphasize the need for clearer reporting on contextual factors. The paper also recommends integrating applicability and transferability assessments into more general decision-making frameworks to streamline the process for decision-makers.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Understanding the specific characteristics of the new context and target population is essential. Factors such as the local need for the intervention, setting characteristics (including resources and existing services), and population demographics are crucial for determining whether research findings can be effectively applied to a new setting.

Examples of Problematic Extrapolation / External Validity

The paper highlights the general issue of research findings not being adequately assessed for their applicability and transferability, leading to challenges in their implementation in different settings. Specific problematic examples are not provided, but the emphasis is on the general lack of systematic assessments.

Examples of Successful Extrapolation / External Validity

While specific examples of successful extrapolation are not detailed, the authors discuss the potential benefits of using well-developed frameworks to improve the generalizability and practical use of research findings. They stress that better assessment methods can enhance the relevance and impact of research across various contexts.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context and adjusting the program based on this information is crucial. Conducting pilot studies and utilizing comprehensive frameworks for applicability and transferability can help identify potential issues and ensure the program's success in the new environment.

Burchett et al. emphasize the importance of developing and using validated frameworks to assess the external validity, applicability, and transferability of research. By focusing on empirical support and detailed contextual reporting, researchers and policymakers can improve the generalizability of findings and ensure that interventions are effectively adapted to new settings.

2.70 How Much Can We Generalize From Impact Evaluations? (Vivalt, 2020)

Eva Vivalt (2020) explores the extent to which impact evaluations can be generalized across different contexts, using a comprehensive dataset of impact evaluation results. This paper examines the heterogeneity of treatment effects and how various study characteristics influence these effects. By analyzing 15,024 estimates from 635 papers on 20 types of interventions in international development, Vivalt provides insights into the factors that affect the generalizability of impact evaluation findings.

Main Problems in Extrapolation / External Validity

The primary issue highlighted by Vivalt is the large amount of effect heterogeneity observed in impact evaluations. This heterogeneity poses significant challenges for generalizing findings across different contexts. Key problems include the variability in effect sizes due to differences in implementation by governments versus NGOs or academic institutions, and the influence of study characteristics such as sample size and context-specific factors.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Vivalt suggests accounting for study characteristics when analyzing treatment effects. This includes using Bayesian hierarchical models to better understand the variability and to predict the likely effectiveness of interventions in new contexts. Additionally, she advocates for more detailed reporting on study design, context, and implementation characteristics to enhance the transparency and applicability of findings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Understanding the specific features of the new context and target population is crucial. These features include the type of implementing organization, local socioeconomic conditions, and the specific nature of the intervention. For instance, government-implemented programs tend to have smaller effect sizes compared to those run by NGOs or academic researchers, highlighting the importance of considering organizational differences.

Examples of Problematic Extrapolation / External Validity

Vivalt provides examples where the heterogeneity in treatment effects has led to challenges in generalizing results. For instance, studies with smaller sample sizes often report larger effect sizes, which may not hold true in broader implementations. This indicates that findings from small, controlled studies may not always be applicable to larger, more diverse populations.

Examples of Successful Extrapolation / External Validity

While specific examples of successful extrapolation are not extensively detailed, Vivalt's analysis suggests that understanding and adjusting for study characteristics can significantly reduce heterogeneity and improve the generalizability of results. Her use of Bayesian hierarchical models demonstrates a method for achieving more reliable generalizations.

Transferring an Urban Food Security Program

Although the paper does not directly address the transfer of an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context and adjusting the program based on this information is essential. By accounting for organizational differences and other study characteristics, policymakers can better predict the effectiveness of the program in the new setting. Conducting pilot studies and utilizing Bayesian models can help identify potential issues before full-scale implementation.

Vivalt's paper underscores the importance of considering study and context-specific characteristics to enhance the generalizability of impact evaluations. By employing rigorous methods and transparent reporting, researchers and policymakers can improve the applicability of findings across different contexts, thereby making more informed decisions in the field of international development.

2.71 How to Assess the External Validity of Therapeutic Trials: A Conceptual Approach (Dekkers et al., 2010)

Dekkers et al. (2010) explore the complexities of assessing the external validity of therapeutic trials. They argue that external validity, which pertains to the generalizability of study results to different populations, is more intricate than often perceived. The paper proposes a conceptual framework for evaluating external validity, emphasizing the need for a nuanced approach that considers various factors influencing outcomes.

Main Problems in Extrapolation / External Validity

The primary challenges identified include the variability in patient populations, geographical and temporal differences, and the specificity of treatment settings. The authors note that the study population often differs from the target population in significant ways, such as in demographic characteristics, disease prevalence, and healthcare infrastructure. These differences can significantly impact the generalizability of study results.

Methods for Fixing Problems of Extrapolation / External Validity

Dekkers et al. propose a three-step approach to assess external validity:

1. **Evaluating the Representativeness of the Study Population:** This involves comparing the study population with the target population to identify any significant differences in characteristics that could affect outcomes.
2. **Considering Temporal, Ethnical, and Geographical Differences:** Researchers should examine how these factors might influence the generalizability of the study results.
3. **Assessing Generalizability Beyond Eligibility Criteria:** This involves determining whether the study findings can be applied to populations that do not meet all the original eligibility criteria but are still relevant in clinical practice.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The authors highlight the importance of understanding the new context and target population's specific features, such as demographic variables, cultural norms, and healthcare settings. These factors can significantly influence the applicability of study findings to new populations. For instance, treatment effects observed in a clinical trial may differ when applied in a different geographical location with distinct healthcare practices and patient characteristics.

Examples of Problematic Extrapolation / External Validity

The paper discusses the issue of generalizing results from trials conducted in highly controlled environments to routine clinical practice. For example, the outcomes of a trial conducted in tertiary care centers may not be applicable to patients treated in primary care settings due to differences in patient management and available resources.

Examples of Successful Extrapolation / External Validity

While specific examples of successful extrapolation are not detailed, the authors stress the importance of a systematic approach to assessing external validity. By thoroughly evaluating the factors that influence generalizability, researchers can improve the likelihood of successfully applying study findings to broader populations.

Transferring an Urban Food Security Program

Although the article does not directly address the transfer of an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context and adjusting the program based on this information is crucial. Conducting pilot studies and utilizing a systematic framework to assess external validity can help identify potential issues and ensure the program's success in the new environment.

Dekkers et al. underscore the complexity of assessing external validity in therapeutic trials. Their proposed framework provides a structured approach to evaluating the generalizability of study findings, highlighting the need for careful consideration of various factors that influence outcomes. By adopting this approach, researchers and policymakers can improve the applicability of clinical trial results to diverse populations and settings.

2.72 Interaction of Theory and Practice to Assess External Validity (Leviton & Trujillo, 2017)

Leviton and Trujillo (2017) address the intricate relationship between theory and practice in assessing the external validity of evidence-based interventions (EBIs). The authors argue for a more systematic integration of theoretical frameworks and practical insights to enhance the generalizability of research findings across various contexts.

Main Problems in Extrapolation / External Validity

The primary challenges discussed include the variability in local contexts and the difficulty of maintaining intervention fidelity while allowing necessary adaptations. The authors highlight the issue of different settings and populations, which can significantly impact the effectiveness of an intervention if not properly accounted for. They stress that weak external validity inferences hinder the widespread adoption of EBIs.

Methods for Fixing Problems of Extrapolation / External Validity

Leviton and Trujillo propose using program theory as a foundation for assessing external validity. They advocate for a systematic interaction between researchers and practitioners to refine theories and understand the variations in treatment. This approach includes rigorous methodological frameworks that combine empirical data with theoretical insights to address the complexities of different contexts.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Understanding the new context and target population's characteristics is essential. This includes demographic factors, cultural norms, and the local infrastructure, which all play crucial roles in the success of an intervention. The authors emphasize the need for localized intelligence and adaptability to ensure that the core components of an EBI are relevant and effective in new settings.

Examples of Problematic Extrapolation / External Validity

The paper discusses general issues rather than specific examples, highlighting the challenges of applying findings from controlled settings to broader real-world contexts. It underscores the need for careful consideration of contextual differences that can affect intervention outcomes.

Examples of Successful Extrapolation / External Validity

While specific successful cases are not extensively detailed, the authors illustrate the potential benefits of their proposed approach through various examples of improved theoretical and practical integration. These include enhanced program theories and better identification of core components that allow for effective adaptations.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context and adjusting the program based on this information is crucial. By engaging practitioners who are familiar with the local environment and integrating their insights with robust theoretical frameworks, policymakers can enhance the program's effectiveness. Conducting pilot studies and systematically evaluating treatment variations can help ensure the program is well-suited to the new setting.

Leviton and Trujillo emphasize the importance of combining theoretical rigor with practical insights to assess and enhance the external validity of EBIs. Their approach advocates for a collaborative effort between researchers and practitioners to adapt interventions thoughtfully, ensuring their effectiveness across diverse contexts.

2.73 Internal and External Validity of the Comparative Interrupted Time-Series Design: A Meta-Analysis (Coopersmith et al., 2022)

Coopersmith et al. (2022) conduct a meta-analysis of 12 studies examining the internal and external validity of the Comparative Interrupted Time-Series (CITS) design, which is often used to evaluate the effects of social policy interventions. This paper evaluates the bias in CITS by comparing its impact estimates with those derived from theoretically unbiased benchmark studies, such as randomized controlled trials (RCTs) and regression discontinuity (RD) studies.

Main Problems in Extrapolation / External Validity

The primary challenge identified is the potential bias in CITS studies compared to more robust experimental designs like RCTs. The authors highlight issues such as differential historical events, changes in instrumentation, and selection biases that might affect the treatment and comparison groups differently, potentially leading to inaccurate generalizations.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Coopersmith et al. suggest several methods. These include using multiple pre-intervention time points to model trends accurately, matching

treatment and comparison groups on key characteristics, and employing robust statistical methods to adjust for observed and unobserved biases. The use of meta-analysis itself is a recommended approach, as it synthesizes results across various studies to provide more reliable estimates of bias and generalizability.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study emphasizes the importance of considering the specific features of new contexts and target populations. Factors such as the similarity of baseline trends, the stability of the pre-intervention outcome trends, and the comparability of treatment and comparison groups are crucial. The authors stress that differences in these features can significantly impact the validity of extrapolating findings to new settings.

Examples of Problematic Extrapolation / External Validity

The paper discusses instances where CITS results might diverge from those of RCTs due to biases introduced by different historical events or changes in the composition of the treatment and comparison groups. For example, studies where local historical events influenced the treatment group more than the comparison group show the potential for biased estimates.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not detailed, the meta-analysis shows that, on average, the bias in CITS studies is minimal. The tight distribution of bias estimates around zero suggests that CITS can be a reliable method for causal inference in social policy research when implemented carefully.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context, ensuring comparability of baseline trends, and using robust statistical methods to adjust for any biases are crucial steps. Conducting pilot studies and utilizing meta-analyses to validate the generalizability of findings can help ensure the program's effectiveness in the new environment.

Coopersmith et al. underscore the potential of the CITS design for evaluating social policies, provided that careful attention is given to controlling for biases and accurately modeling pre-intervention trends. Their meta-analysis supports the view that, with rigorous implementation, CITS can offer high internal and external validity, making it a valuable tool for policy analysis.

2.74 Internal, External, and Ecological Validity in Research Design, Conduct, and Evaluation (Andrade, 2018)

Chittaranjan Andrade (2018) discusses the concepts of internal, external, and ecological validity in the context of research design, conduct, and evaluation. The paper highlights the importance of these types of validity in ensuring that research findings are robust, applicable, and relevant to real-world settings.

Main Problems in Extrapolation / External Validity

Andrade identifies several challenges in achieving external validity. These include differences between the study sample and the broader population, the influence of study settings that differ from real-world environments, and the limited applicability of findings from short-term studies to long-term scenarios. These issues can lead to difficulties in generalizing study results to different populations and settings.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Andrade suggests rigorous methods for improving external validity. These include ensuring that study samples are representative of the target population, considering the ecological validity of study settings, and conducting longer-term studies when necessary. The use of detailed reporting and transparency in study design and implementation is also emphasized to help other researchers and practitioners assess the applicability of findings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Understanding the characteristics of the new context and target population is crucial for external validity. Factors such as demographic variables, cultural norms, healthcare infrastructure, and socioeconomic conditions can significantly impact the generalizability of study findings. Andrade highlights the need to consider these factors when designing studies and when applying findings to new settings.

Examples of Problematic Extrapolation / External Validity

The paper discusses the issue of generalizing results from controlled settings to everyday life, exemplified by the CATIE study. The CATIE findings, while relevant to clinical practice in the USA, may not be applicable to countries like India, where different healthcare delivery systems and family dynamics influence treatment outcomes. This highlights the importance of considering local contexts in assessing external validity.

Examples of Successful Extrapolation / External Validity

Specific examples of successful extrapolation are not detailed in the text. However, Andrade emphasizes that a good understanding of internal, external, and ecological validity is essential for designing and conducting studies that produce generalizable findings. By systematically addressing these types of validity, researchers can improve the applicability of their results to broader populations and settings.

Transferring an Urban Food Security Program

Although the article does not directly address the transfer of an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context, considering the ecological validity of the program, and ensuring that the study sample is representative of the target population are crucial steps. Conducting pilot studies and using robust methods to assess the applicability of findings can help ensure the program's effectiveness in the new environment.

Andrade's paper underscores the importance of considering internal, external, and ecological validity in research. By focusing on these aspects, researchers can design studies that produce findings applicable to diverse populations and settings, ultimately enhancing the impact and relevance of their work.

2.75 Introduction to Special Issue: External Validity and Policy (Westbrook, 2017)

T'Pring R. Westbrook (2017) introduces a special issue focused on the importance of external validity in policy research. The article discusses the challenges and significance of generalizing research findings to broader contexts, particularly in the realm of evidence-based policy-making. It highlights the need for systematic approaches to assess and enhance external validity to ensure that interventions are effective across diverse settings and populations.

Main Problems in Extrapolation / External Validity

The primary challenges identified include the lack of consensus on how to conceptualize and assess external validity, and the need for standardized methods to evaluate it. Westbrook emphasizes that while internal validity is often prioritized, external validity is crucial for ensuring that research findings are applicable to real-world settings. The variability in sample characteristics, treatment conditions, and contextual factors poses significant obstacles to generalization.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Westbrook discusses both prospective and retrospective approaches. Prospective approaches involve making design decisions before conducting a study, such as using stratified sampling methods to improve generalization. Retrospective approaches include statistical adjustments made after a study is completed, such as propensity score subclassification. The article also highlights the importance of systematic evidence reviews and frameworks that help practitioners and policymakers choose evidence-based strategies best suited for their context.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The article underscores the necessity of understanding the specific features of new contexts and target populations, such as demographic variables, cultural norms, and local infrastructure. These factors are critical in determining whether an intervention that was effective in one setting will be equally effective in another. Westbrook calls for more detailed contextual information to be included in evidence assessments to aid decision-makers.

Examples of Problematic Extrapolation / External Validity

While specific problematic examples are not extensively detailed, the article highlights general issues such as the challenges of generalizing findings from controlled environments to real-world settings. It underscores the importance of considering local contexts and the potential differences in treatment implementation and population characteristics.

Examples of Successful Extrapolation / External Validity

The article does not provide specific examples of successful extrapolation but emphasizes the importance of incorporating principles of external validity into evidence-based policy-making. By systematically addressing external validity, researchers can enhance the applicability of their findings to diverse settings.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context, using both prospective and retrospective approaches to assess external validity, and ensuring that the program is tailored to meet the specific needs of the new setting are crucial steps. Systematic evidence reviews and contextual frameworks can help identify potential issues and enhance the program's effectiveness.

Westbrook's article highlights the critical role of external validity in policy research and evidence-based decision-making. By focusing on systematic approaches to assess and enhance external validity, researchers and policymakers can ensure that interventions are effective across various contexts, ultimately improving the impact and relevance of their work.

2.76 Is it possible to overcome issues of external validity in preclinical animal research? Why most animal models are bound to fail (Pound & Ritskes-Hoitinga, 2018)

Pound and Ritskes-Hoitinga (2018) examine the challenges of achieving external validity in preclinical animal research and argue that many animal models fail to translate effectively to human clinical settings. They explore various factors that impact the external validity of animal studies and suggest that species differences inherently limit the reliability of translating findings from animals to humans.

Main Problems in Extrapolation / External Validity

The authors identify significant issues with external validity in preclinical animal research. Key problems include unrepresentative animal samples, the inability of animal models to replicate the complexity of human conditions, and the inherent biological differences between species. These factors contribute to the failure of animal models to predict clinical efficacy and safety accurately.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Pound and Ritskes-Hoitinga suggest improving the representativeness of animal samples and enhancing the clinical relevance of animal models. However, they emphasize that species differences pose an insurmountable problem that cannot be fully resolved. The authors argue that focusing on human-relevant research methods and technologies, such as human induced pluripotent stem cells (iPSC), organs-on-chips, and in silico approaches, is a more promising strategy for improving clinical translation.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study highlights the importance of considering the specific features of new contexts and target populations. Factors such as the genetic, biochemical, and physiological differences between animals and humans must be accounted for. The authors stress that these differences often lead to significant disparities in how interventions affect animal models versus human patients.

Examples of Problematic Extrapolation / External Validity

The paper provides several examples where animal models have failed to predict human outcomes accurately. For instance, the SOD1 transgenic mouse model, used to study motor neuron disease, mimics some characteristics of the disease but does not accurately represent the underlying mechanisms in humans. Similarly, treatments that appeared effective in animal models of stroke failed in human clinical trials due to differences in disease progression and treatment response.

Examples of Successful Extrapolation / External Validity

Specific examples of successful extrapolation are not detailed, as the authors focus on the inherent limitations of animal models. They argue that while some improvements can be made, the fundamental issue of species differences will always limit the external validity of animal research.

Transferring an Urban Food Security Program

Although the article does not directly address the transfer of an urban food security program, the principles discussed can be applied. Collecting detailed data on the new context and focusing on human-relevant methods can enhance the program's effectiveness. Pilot studies using human-specific models and technologies can help identify potential issues and ensure the program is well-suited to the new setting.

Pound and Ritskes-Hoitinga emphasize the significant limitations of animal models in achieving external validity due to species differences. They advocate for a shift towards human-relevant research methods to improve the translation of findings from bench to bedside, ultimately enhancing the development of safe and effective treatments for humans.

2.77 Learning from Experiments When Context Matters (Pritchett & Sandefur, 2015)

Pritchett and Sandefur (2015) address the complexities of generalizing findings from experimental studies when context plays a significant role. They discuss the trade-offs between internal and external validity and the challenges policymakers face when interpreting evidence from different contexts. The authors emphasize the importance of considering context-specific factors in the generalization of experimental results.

Main Problems in Extrapolation / External Validity

The main issues identified include the variability in treatment effects across different contexts and the difficulties in applying findings from one setting to another. The

authors point out that even well-conducted experiments can produce results that are not easily transferable due to differences in local conditions, implementation details, and population characteristics. This context-dependence poses a significant challenge to the external validity of experimental findings.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Pritchett and Sandefur suggest combining experimental and non-experimental data to provide a more comprehensive understanding of treatment effects. They advocate for the use of Root Mean Squared Error (RMSE) to measure the reliability of estimates and emphasize the importance of meta-analysis to synthesize findings from multiple studies. Additionally, they recommend considering both the internal and external validity of studies and using robust statistical methods to adjust for context-specific factors.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The paper highlights the necessity of understanding the specific characteristics of the new context and target population. Factors such as demographic variables, cultural norms, economic conditions, and local infrastructure can significantly influence the applicability of study findings. The authors stress that policymakers must carefully consider these factors when attempting to generalize experimental results to new settings.

Examples of Problematic Extrapolation / External Validity

The paper provides examples from microcredit studies where treatment effects varied significantly across different contexts. For instance, the impact of microcredit on business profits and household consumption differed widely among studies conducted in Bosnia, Ethiopia, India, Mexico, Mongolia, and Morocco. These discrepancies illustrate the challenges of generalizing findings from one context to another and highlight the importance of considering local conditions.

Examples of Successful Extrapolation / External Validity

While the paper does not provide specific examples of successful extrapolation, it emphasizes that a better understanding of context-specific factors and the use of comprehensive data analysis methods can enhance the reliability of generalized findings. By systematically evaluating the contextual variations and employing robust statistical techniques, researchers can improve the external validity of their results.

Transferring an Urban Food Security Program

Although the article does not directly address the transfer of an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context and adjusting the program based on this information is crucial. Conducting pilot studies, using meta-analysis to synthesize findings from different contexts, and considering both internal and external validity can help ensure the program's effectiveness in the new environment.

Pritchett and Sandefur underscore the critical role of context in determining the external validity of experimental findings. They advocate for a nuanced approach that combines experimental and non-experimental data, robust statistical methods, and careful consideration of local conditions to improve the generalizability of research results. This approach can help policymakers make more informed decisions and implement interventions that are effective across diverse settings.

2.78 Learning from Second-Hand Experience: Methodology for Extrapolation-Oriented Case Research (Barzelay, 2007)

Michael Barzelay (2007) delves into the methodology for extrapolation-oriented case research, emphasizing how policymakers can learn from the experiences of other contexts. The focus is on the systematic study of successful practices in one context to inform and improve policy decisions in another, using case studies as the primary research method.

Main Problems in Extrapolation / External Validity

Barzelay identifies the central problem of vicarious learning, which is the reliance on second-hand accounts that often lack critical contextual details. These accounts are edited and may omit significant factors that affect the implementation and outcomes of policies in different settings. The challenge lies in accurately capturing and transferring these nuanced details to ensure successful policy adaptation.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Barzelay proposes a structured approach to case study research that includes thorough investigation of source sites (where the practices originated) and careful analysis of how these practices can be adapted to target sites (where the practices will be implemented). This involves understanding the causal mechanisms behind successful practices and designing new interventions that can activate similar mechanisms in different contexts. He also suggests using a heuristic of extrapolation rather than replication, meaning that practices should be adapted to fit the new context rather than copied exactly.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The new context and target population's characteristics, such as institutional structures, cultural norms, and socio-economic conditions, are critical. Barzelay stresses the importance of considering these factors when designing new interventions based on extrapolated practices. This ensures that the adapted practices are feasible and effective in the new setting.

Examples of Problematic Extrapolation / External Validity

The paper discusses the general difficulties of vicarious learning and the potential pitfalls of relying on second-hand accounts. Specific problematic examples are not extensively detailed, but the emphasis is on the risk of oversimplifying or misinterpreting the contextual factors that contribute to the success of a practice in its original setting.

Examples of Successful Extrapolation / External Validity

Barzelay provides a case study of the "Brazil in Action" program as an example of successful extrapolation. The program, which aimed to implement a series of priority projects under President Fernando Henrique Cardoso, demonstrates how careful adaptation of practices to the local context, supported by detailed case study research, can lead to successful policy outcomes. The key was understanding the mechanisms behind the original practice and thoughtfully adapting them to fit the new context.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context, understanding the causal mechanisms behind successful food security programs, and thoughtfully adapting these mechanisms to fit the local conditions are essential steps. Pilot studies and heuristic adaptations can help ensure the program's effectiveness in the new environment.

Barzelay's paper underscores the importance of a rigorous and systematic approach to learning from second-hand experiences. By focusing on the mechanisms behind successful practices and adapting them to new contexts, policymakers can enhance the external validity of their interventions and achieve better outcomes in diverse settings.

2.79 On the External Validity of Laboratory Tax Compliance Experiments (Alm et al., 2015)

Alm, Bloomquist, and McKee (2015) investigate the external validity of laboratory experiments in the context of tax compliance. The paper examines whether the behaviors observed in laboratory settings can be generalized to naturally occurring environments and if the results from student subjects can be extrapolated to the broader population.

Main Problems in Extrapolation / External Validity

The main issues highlighted include the representativeness of student subjects compared to the general taxpayer population and the differences between the controlled laboratory environment and real-world tax compliance settings. These differences can lead to concerns about whether the findings from laboratory experiments are applicable to broader, naturally occurring contexts.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Alm et al. compare the behavior of laboratory participants with actual taxpayer behavior using data from the U.S. Internal Revenue Service's National Research Program. They also compare the responses of student subjects with nonstudent subjects in identical laboratory experiments. By doing so, they assess the similarities and differences in behavior across different subject pools and settings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study emphasizes the importance of demographic factors, such as age, education level, and experience with tax filing, in determining the external validity of laboratory experiments. These factors can significantly impact the applicability of laboratory findings to the general population.

Examples of Problematic Extrapolation / External Validity

The paper discusses general concerns about the applicability of laboratory results to real-world settings due to the controlled nature of the experiments and the non-representativeness of student subjects. However, specific problematic examples are not extensively detailed.

Examples of Successful Extrapolation / External Validity

The authors find that the behavior of laboratory participants closely aligns with that of actual taxpayers, suggesting that laboratory experiments can provide valid insights

into tax compliance behavior. They also find that student and nonstudent subjects exhibit broadly similar behavior in laboratory settings, further supporting the external validity of these experiments.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context, ensuring the representativeness of the study sample, and using robust statistical methods to compare behaviors across different settings are crucial steps. Conducting pilot studies and validating findings with real-world data can help ensure the program's effectiveness in the new environment.

Alm et al. highlight the potential for laboratory experiments to provide valuable insights into policy issues, provided that careful attention is given to ensuring external validity. By systematically comparing laboratory and real-world behaviors, researchers can enhance the generalizability of their findings and improve the applicability of experimental results to broader contexts.

2.80 Evidence, External Validity, and Explanatory Relevance (Cartwright, 2011)

Nancy Cartwright (2011) explores the concepts of evidence, external validity, and explanatory relevance, particularly in the context of evidence-based policy and practice (EBPP). The paper delves into how evidence from randomized controlled trials (RCTs) can be relevant and useful for predicting policy effectiveness in different contexts. Cartwright emphasizes the importance of explanatory relevance for ensuring that evidence from studies is applicable to real-world scenarios.

Main Problems in Extrapolation / External Validity

Cartwright identifies several challenges in achieving external validity, particularly in the context of RCTs. One of the main issues is that RCTs often fail to account for the specific conditions and causal structures of different contexts. This can lead to problems when trying to generalize findings from one setting to another. Additionally, the paper highlights the difficulty of ensuring that the explanatory factors in one context are relevant and applicable in another.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Cartwright suggests focusing on explanatory relevance, where the evidence from RCTs is considered relevant if it can be shown to be part of a correct explanation for the targeted outcome in the new context. This involves

conducting both horizontal and vertical searches to identify common explanatory elements across different contexts. Horizontal searches look for shared factors in similar settings, while vertical searches climb up and down the ladder of abstraction to find higher-level principles that can apply more broadly.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Understanding the specific features of the new context and target population is crucial. Factors such as local causal structures, socio-economic conditions, and institutional arrangements can significantly affect the generalizability of study findings. Cartwright emphasizes the need to gather detailed evidence about these factors to ensure that the conditions necessary for the effectiveness of an intervention are present in the new context.

Examples of Problematic Extrapolation / External Validity

Cartwright provides the example of the California class-size reduction program, which did not yield the expected improvements in educational outcomes due to the rapid implementation and lack of trained teachers. This illustrates how failing to account for local conditions and causal factors can lead to problematic extrapolation and ineffective policy outcomes.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not extensively detailed, the paper highlights the importance of thorough explanatory analysis and evidence gathering. By identifying and confirming the relevant causal structures and conditions, researchers can improve the likelihood of successful extrapolation of findings.

Transferring an Urban Food Security Program

Although the article does not directly address the transfer of an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context, conducting horizontal and vertical searches to identify relevant causal factors, and ensuring that these factors are present in the new setting are essential steps. Pilot studies and systematic evaluation of the explanatory relevance of evidence can help ensure the program's effectiveness.

Cartwright's paper underscores the importance of explanatory relevance and detailed contextual analysis in achieving external validity. By focusing on these aspects, researchers and policymakers can enhance the generalizability of study findings and make more informed decisions about policy interventions in diverse settings.

2.81 Policy Evaluation, Randomized Controlled Trials, and External Validity: A Systematic Review (Peters et al., 2016)

Peters, Langbein, and Roberts (2016) conduct a systematic review of randomized controlled trials (RCTs) published in top economic journals between 2009 and 2014. Their focus is on assessing the external validity of these studies, particularly how well the findings from these trials can be generalized to broader populations and different contexts.

Main Problems in Extrapolation / External Validity

The primary issues identified include the presence of Hawthorne and John Henry effects, general equilibrium effects, specific sample problems, and the special care with which treatments are provided in RCTs. These factors can significantly compromise the generalizability of RCT findings. For example, participants' awareness of being part of an experiment can alter their behavior, and results obtained under highly controlled conditions may not hold when scaled up or applied in less controlled environments.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, the authors propose a more systematic approach to reporting the results of RCTs. This includes providing detailed information on the implementation of experiments, discussing potential hazards to external validity, and considering how the findings might change if the intervention is scaled up or applied in a different context. They advocate for the inclusion of "credibility-enhancing arguments" to support the external validity of RCT findings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study highlights the importance of understanding the characteristics of the new context and target population. Factors such as demographic variables, cultural norms, and institutional settings can significantly influence the applicability of RCT findings. The authors stress the need for detailed contextual information to assess whether the conditions necessary for the effectiveness of an intervention are present in the new setting.

Examples of Problematic Extrapolation / External Validity

The review discusses several issues related to the external validity of RCTs but does not provide specific case studies of problematic extrapolation. It emphasizes the general difficulty of transferring findings from controlled experiments to real-world

settings, where conditions and populations may differ significantly from those in the original study.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not detailed, the authors highlight the importance of addressing external validity issues to improve the generalizability of RCT findings. They suggest that a systematic approach to reporting and discussing external validity can enhance the reliability and applicability of study results.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, considering the potential for Hawthorne and John Henry effects, and ensuring that the program is implemented with the same rigor as in the original study are crucial steps. Conducting pilot studies and using robust methods to assess the generalizability of findings can help ensure the program's effectiveness in the new environment.

Peters et al. emphasize the need for a systematic approach to addressing external validity in RCTs. By focusing on detailed reporting, discussing potential hazards, and considering the context of new implementations, researchers and policymakers can improve the generalizability and practical relevance of their findings.

2.82 Predicting the Efficacy of Future Training Programs Using Past Experiences at Other Locations (Hotz et al., 2005)

Joseph Hotz, Guido Imbens, and Julie Mortimer (2005) investigate the challenges and methodologies for predicting the efficacy of new job training programs based on data from previously implemented programs. The study specifically examines the variability in population characteristics and program components, emphasizing the need for rigorous adjustments to enhance the generalizability of findings.

Main Problems in Extrapolation / External Validity

The authors identify two principal complications: differences in the populations across locations and differences in the program components. Populations may vary in demographics, socio-economic conditions, and prior work experience, while programs may differ in their implementation and components, such as the balance between classroom training and job search assistance. These variations can significantly affect the ability to generalize findings from one location to another.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Hotz et al. suggest several methodological approaches. They emphasize the importance of adjusting for individual characteristics and pre-training variables to control for population differences. The use of experimental data from multiple locations allows for better control of these variables. Additionally, they recommend employing matching methods and propensity score techniques to align the treatment and control groups more accurately across different settings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study highlights the significance of demographic characteristics, such as age, education, and previous work experience, in affecting the generalizability of training program results. Local economic conditions and labor market characteristics are also critical factors. The authors stress that understanding and adjusting for these contextual differences is essential for accurate predictions of program efficacy.

Examples of Problematic Extrapolation / External Validity

The paper notes that traditional econometric methods often fail to replicate experimental estimates, as demonstrated by previous studies. For example, differences in pre-training earnings and employment histories across locations can lead to significant biases if not properly adjusted. The failure to account for such differences can result in misleading conclusions about the effectiveness of training programs.

Examples of Successful Extrapolation / External Validity

Hotz et al. find that adjusting for individual characteristics and pre-training variables significantly reduces biases, making it possible to predict outcomes more accurately across different locations. They show that for sub-populations with previous work experience, such adjustments lead to comparable outcomes across various settings, suggesting a successful generalization.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, adjusting for demographic and economic differences, and using matching methods to align the program components with the local conditions are crucial steps. Conducting pilot studies and validating findings with real-world data can help ensure the program's effectiveness in the new environment.

Hotz et al. underscore the importance of rigorous adjustments and methodological rigor in predicting the efficacy of new programs based on past experiences. By focusing on individual characteristics and local conditions, researchers and policymakers can improve the external validity of their findings and make more informed decisions about implementing social programs in diverse settings.

Problems with Using Mechanisms to Solve the Problem of Extrapolation (Howick et al., 2013)

Jeremy Howick, Paul Glasziou, and Jeffrey K. Aronson (2013) address the challenges of using mechanistic knowledge to solve the problem of extrapolating results from controlled studies to target populations. They argue that while mechanisms can offer some insights, they often face significant limitations and challenges.

Main Problems in Extrapolation / External Validity

The primary issues identified include the incomplete understanding of mechanisms, the limited applicability of mechanistic knowledge outside controlled laboratory conditions, and the paradoxical behavior of mechanisms. The authors emphasize that mechanisms might not behave consistently across different populations or settings, which complicates the process of extrapolation. They also mention the "extrapolator's circle," a problem where mechanistic knowledge in the target population is required to justify extrapolation, but obtaining this knowledge makes the initial study redundant.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, the authors suggest a cautious and critical approach to using mechanistic reasoning. They recommend combining mechanistic knowledge with empirical data from similar contexts and using robust methodologies to test the validity of mechanisms in new settings. This includes conducting comparative process tracing to identify relevant differences between study and target populations. However, they caution that even with these methods, fully solving the problem of extrapolation is unlikely.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Understanding the new context and target population's specific features, such as demographic variables, cultural norms, and environmental conditions, is crucial. Howick et al. stress that these factors can significantly impact the behavior of mechanisms and the generalizability of study findings. They argue that detailed

contextual analysis is necessary to identify and adjust for differences that might affect the applicability of mechanistic knowledge.

Examples of Problematic Extrapolation / External Validity

The authors provide several examples where mechanistic reasoning failed to predict outcomes accurately. For instance, they discuss how antiarrhythmic drugs, which were expected to reduce mortality based on mechanistic understanding, actually increased mortality in clinical trials. They also mention the failure of the Bangladesh Integrated Nutrition Project (BINP), which was modeled after a successful nutrition program in Tamil Nadu, India. The differing social contexts led to the failure of BINP, highlighting the challenges of extrapolation.

Examples of Successful Extrapolation / External Validity

While the paper focuses more on the limitations and problems of mechanistic reasoning, it does acknowledge that in some well-defined cases, mechanistic knowledge can help justify extrapolation. For example, the understanding of the mechanisms behind stroke treatments has allowed for better-targeted interventions, reducing the risk of harm when applying study results to new populations.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context, combining mechanistic insights with empirical data, and conducting rigorous comparative analyses are essential steps. Pilot studies and careful monitoring can help identify potential issues and ensure the program's effectiveness in the new environment.

Howick et al. emphasize the limitations and complexities of using mechanistic knowledge to solve the problem of extrapolation. They advocate for a cautious approach, combining mechanistic reasoning with empirical data and detailed contextual analysis to improve the generalizability and applicability of research findings.

2.84 Randomised Trials for Policy: A Review of the External Validity of Treatment Effects (Muller, 2014)

Sean Muller (2014) reviews the external validity of treatment effects in randomized trials, with a particular focus on their use in policy evaluation. This paper synthesizes contributions from various subdisciplines within economics, as well as philosophy and

medicine, to provide a comprehensive understanding of the challenges and potential solutions for generalizing findings from randomized controlled trials (RCTs) to broader populations.

Main Problems in Extrapolation / External Validity

Muller identifies several key issues affecting the external validity of RCTs. These include the variability in populations and settings, the interactive functional forms of causal relationships, and the inherent limitations of RCTs to address the broader applicability of findings. He emphasizes that the assumptions required for external validity are conceptually similar to those needed for internal validity but are often more complex and less well-understood.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Muller suggests a more rigorous integration of theoretical frameworks to guide the extrapolation process. This involves using structural econometric models to understand the mechanisms behind treatment effects and applying these insights to new contexts. Additionally, Muller advocates for the use of meta-analysis and replication studies across different settings to assess the robustness of findings. He also highlights the importance of explicitly considering external validity during the design and implementation phases of RCTs.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study stresses the importance of understanding the specific characteristics of the new context and target population. Factors such as demographic differences, institutional environments, and economic conditions can significantly impact the generalizability of RCT findings. Muller argues that detailed data on these contextual factors should be collected and analyzed to enhance the applicability of results.

Examples of Problematic Extrapolation / External Validity

Muller discusses general concerns about the external validity of RCTs but does not provide specific problematic examples. He notes that many empirical studies, especially in development economics, fail to include formal analyses of external validity, leading to potential misapplications of findings in different contexts.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not extensively detailed, Muller emphasizes that rigorous methodological approaches, including structural modeling and meta-analysis, can improve the likelihood of successful extrapolation. By systematically

addressing external validity, researchers can enhance the reliability and relevance of their findings.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, using theoretical frameworks to guide adaptation, and conducting pilot studies can help ensure the program's effectiveness. Systematic evaluation and replication across different settings can also contribute to a more robust understanding of the program's impact.

Muller's paper highlights the critical role of theoretical and methodological rigor in addressing the challenges of external validity in RCTs. By integrating structural models, conducting meta-analyses, and explicitly considering contextual factors, researchers and policymakers can improve the generalizability and practical relevance of their findings.

2.85 Regression Discontinuity and Beyond: Options for Studying External Validity in an Internally Valid Design (Wing & Bello-Gomez, 2018)

Coady Wing and Ricardo A. Bello-Gomez (2018) explore the external validity of treatment effect estimates from regression discontinuity designs (RDDs). While RDDs have high internal validity due to their ability to identify causal effects under clear assumptions, their external validity is often limited. This paper reviews four methodologies to enhance the external validity of RDDs.

Main Problems in Extrapolation / External Validity

The main issue with RDDs is that the causal effect identified applies to a narrow subpopulation, typically those near the cutoff point. This limits the ability to generalize findings to the broader population of interest. Additionally, the assumptions that underpin RDDs, such as the non-manipulability of the assignment variable, may not hold in all contexts, further complicating extrapolation.

Methods for Fixing Problems of Extrapolation / External Validity

Wing and Bello-Gomez discuss four techniques to improve external validity in RDDs:

1. **Comparative RDD**: Combines RDD with a comparison group to facilitate extrapolation away from the cutoff by assuming parallel functional forms between the RDD and comparison groups.

2. **Covariate Matching RDD**: Uses covariate matching to validate and extrapolate the treatment effect beyond the cutoff under the assumption of conditional independence.
3. **Treatment Effect Derivatives (TED)**: Estimates the local slope of the treatment effect near the cutoff to understand how the treatment effect varies.
4. **Statistical Tests for Local Selection Bias**: Tests for selection bias at the cutoff to evaluate the validity of extrapolating beyond the cutoff.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Understanding the specific features of the new context, such as demographic differences, economic conditions, and institutional environments, is crucial. These factors can significantly influence whether the causal relationships identified near the cutoff are applicable to the broader population.

Examples of Problematic Extrapolation / External Validity

The paper discusses the inherent difficulties of generalizing findings from RDDs due to the narrow subpopulation at the cutoff. However, it does not provide specific examples of problematic extrapolations. It highlights the risk of assuming that treatment effects at the cutoff are representative of effects in other subpopulations.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not detailed, the authors emphasize that the proposed methodologies can improve the robustness of extrapolation. For instance, combining RDD with covariate matching has been shown to enhance the credibility of extrapolated treatment effects by validating assumptions of conditional independence.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, using comparative RDD to find a valid comparison group, and applying covariate matching to adjust for contextual differences can help ensure the program's effectiveness. Pilot studies and robust statistical tests for local selection bias can further enhance the reliability of the extrapolated findings.

Wing and Bello-Gomez underscore the importance of methodological rigor in addressing the challenges of external validity in RDDs. By applying advanced techniques such as comparative RDD, covariate matching, TED, and selection bias

tests, researchers can improve the generalizability of their findings and provide more reliable evidence for policy decisions across diverse settings.

2.86 Relabeling Internal and External Validity for Applied Social Scientists (Campbell, 1986)

Donald T. Campbell (1986) addresses the confusion surrounding the concepts of internal and external validity in quasi-experimental research. He proposes a relabeling of these concepts to make them more accessible and relevant to applied social scientists. Specifically, he suggests the terms "local molar causal validity" for internal validity and "the principle of proximal similarity" for external validity.

Main Problems in Extrapolation / External Validity

Campbell identifies a fundamental issue in the distinction between internal and external validity: the misinterpretation and oversimplification of these concepts by researchers. He points out that internal validity often gets more attention, leading to a neglect of external validity in practical applications. This neglect can result in findings that are not applicable to broader contexts due to differences in populations, settings, and implementation conditions.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Campbell suggests adopting a more nuanced approach to understanding and applying validity concepts. For external validity, he introduces the "principle of proximal similarity," which emphasizes the importance of considering how similar the new context is to the original study context. He also recommends purposive sampling to explore the limits of generalizability and to understand the dimensions that affect the applicability of findings to new settings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study emphasizes the need to consider various features of the new context, such as demographic characteristics, cultural norms, and environmental conditions. Campbell argues that these factors significantly influence the generalizability of study findings and should be carefully evaluated when attempting to extrapolate results to new populations.

Examples of Problematic Extrapolation / External Validity

Campbell does not provide specific examples of problematic extrapolation but highlights general issues such as the failure to account for contextual differences

when applying findings from controlled environments to real-world settings. He emphasizes the importance of understanding the nuances of both the original study and the new context to avoid misapplication.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not detailed, Campbell discusses the potential benefits of using the principle of proximal similarity and purposive sampling to enhance the external validity of findings. By systematically exploring and understanding the relevant dimensions of similarity, researchers can improve the likelihood of successful generalization.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context and using purposive sampling to identify key similarities and differences are crucial steps. By focusing on the principle of proximal similarity and understanding the local conditions, policymakers can enhance the effectiveness of the program in the new environment.

Campbell's paper underscores the importance of rethinking and clarifying the concepts of internal and external validity. By adopting a more nuanced approach and considering the principle of proximal similarity, researchers and policymakers can improve the applicability of their findings and make more informed decisions about implementing programs in diverse settings.

2.87 Replication, Experiments, and Knowledge in Public Management Research (Walker et al., 2017)

Richard M. Walker, Oliver James, and Gene A. Brewer (2017) explore the significance of replication in public management research, emphasizing the role of experimental methods. The paper reviews the state of replication practices in the field and provides a framework for conducting replication studies to enhance the reliability and validity of research findings.

Main Problems in Extrapolation / External Validity

The authors identify a significant issue: the lack of replication studies in public management, which hampers the ability to generalize findings across different contexts and populations. They point out that the tendency to prioritize novel findings over replication leads to an overemphasis on positive results, which may not be robust

across varied settings. This situation contributes to a potential misrepresentation of the reliability and applicability of research outcomes.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Walker et al. propose a systematic framework for replication, based on Tsang and Kwan's (1999) classification of replication types. The framework includes:

1. **Checking of Analysis**: Using the same dataset and analysis to verify the original findings.
2. **Reanalysis of Data**: Applying different measurement or analysis techniques to the same dataset.
3. **Exact Replication**: Using the same procedures and analysis with a different sample from the same population.
4. **Conceptual Replication**: Applying different measurement and analysis while using a sample from the same population.
5. **Empirical Generalization**: Applying the same measurement and analysis to a different population.
6. **Generalization and Extension**: Using different measurement and analysis in a different population to broaden the scope of the original findings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study emphasizes the importance of context and population characteristics, such as demographic variables, institutional settings, and cultural norms. Understanding these factors is crucial for assessing the external validity of findings. The authors argue that detailed contextual analysis and appropriate sampling strategies are essential to ensure that research findings are applicable to new settings.

Examples of Problematic Extrapolation / External Validity

While specific problematic examples are not extensively detailed, the paper discusses general issues, such as the difficulty of replicating experimental findings in different contexts. The variability in implementation conditions and population characteristics often leads to discrepancies between original findings and replication results.

Examples of Successful Extrapolation / External Validity

The paper highlights several successful replications that have contributed to validating and extending theoretical models in public management. For instance,

replications of studies on citizen satisfaction and performance information have provided valuable insights into the robustness of these findings across different settings and populations.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context, using the appropriate replication type (e.g., empirical generalization or generalization and extension), and conducting pilot studies can help ensure the program's effectiveness. Systematic replication and validation efforts can enhance the reliability of the findings and their applicability to the new environment.

Walker et al. emphasize the critical role of replication in building robust and generalizable knowledge in public management research. By adopting a systematic framework for replication, researchers can improve the external validity of their findings and contribute to more reliable and applicable evidence for policy and practice.

2.88 Reverse Engineering and Policy Design (Weaver, 2019)

R. Kent Weaver (2019) examines the concept of reverse engineering in the context of policy design, particularly focusing on the application of causal mechanisms in public policy. The paper uses the case of retirement savings policies to explore the potential and limitations of reverse engineering in improving policy outcomes.

Main Problems in Extrapolation / External Validity

Weaver identifies several key issues with extrapolation in policy design. One major problem is the complexity of causal mechanisms in different contexts. The heterogeneity of populations and the variability of environmental factors make it challenging to apply findings from one context to another. Additionally, the dynamic nature of policy environments and the interplay of multiple factors complicate the process of accurately predicting policy outcomes.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Weaver suggests a detailed analysis of first-order and second-order causal mechanisms. First-order mechanisms involve direct effects of policy activators on individual and group behaviors, while second-order mechanisms involve feedback loops that modify these behaviors over time. By understanding these mechanisms, policymakers can better predict the effects of policy changes and adjust

their strategies accordingly. Weaver also emphasizes the importance of considering contextual factors and using a combination of qualitative and quantitative data to inform policy design.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Weaver highlights the significance of demographic variables, economic conditions, and cultural factors in shaping the effectiveness of policies. He argues that policies must be tailored to the specific characteristics of the target population to ensure their success. For example, different populations may respond differently to incentives for retirement savings, necessitating customized approaches for different demographic groups.

Examples of Problematic Extrapolation / External Validity

The paper discusses general issues with extrapolation, such as the difficulty of transferring successful policy models from one country to another due to differing institutional contexts and social norms. Weaver notes that even well-designed policies can fail if they do not account for local conditions and the specific needs of the target population.

Examples of Successful Extrapolation / External Validity

Weaver provides examples of successful policy adaptations, such as the use of automatic enrollment in retirement savings plans, which has been effective in various contexts. He highlights how understanding the underlying causal mechanisms and adapting policies to local conditions can enhance their effectiveness.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context, understanding the causal mechanisms at play, and tailoring the program to fit local conditions are crucial steps. Conducting pilot studies and continuously monitoring and adjusting the program based on feedback can help ensure its success in the new environment.

Weaver emphasizes the importance of a nuanced and context-sensitive approach to policy design. By leveraging the concept of reverse engineering and thoroughly understanding causal mechanisms, policymakers can improve the external validity of their interventions and achieve better outcomes across diverse settings.

2.89 Scaling buhmUp What Works: Experimental Evidence on External Validity in Kenyan Education (Bold et al., 2013)

Tessa Bold, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur (2013) investigate the external validity of scaling up successful educational interventions. They focus on a randomized trial involving contract teachers in Kenyan primary schools, comparing the effectiveness of NGO-led and government-led implementations. The study examines the challenges of replicating positive outcomes from small-scale, NGO-led projects when implemented by the government at a national scale.

Main Problems in Extrapolation / External Validity

The primary issue identified is the variability in implementation between NGOs and government agencies. The study found that while NGO implementation resulted in significant improvements in student test scores, government implementation yielded no measurable effect. This discrepancy highlights the challenge of maintaining intervention fidelity and effectiveness when scaling up from controlled, small-scale pilots to broader, government-led programs.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, the authors suggest several approaches:

1. **Rigorous Monitoring and Accountability**: Enhanced monitoring and accountability mechanisms, as NGOs typically have better oversight and are less susceptible to local capture and corruption.
2. **Context-Specific Adjustments**: Tailoring interventions to the specific political and economic context of the target implementation area, recognizing that government structures and incentives can differ significantly from those of NGOs.
3. **Pilot Programs and Iterative Testing**: Implementing pilot programs and iterative testing to identify potential issues in government implementation before full-scale rollouts.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study highlights several contextual factors affecting the external validity of educational interventions:

- **Institutional Capacity**: Differences in the capacity and efficiency of implementing agencies, with NGOs typically having more flexibility and resources for oversight.
- **Political Economy**: The influence of political dynamics, such as union opposition and bureaucratic constraints, which can affect the fidelity and effectiveness of government-led implementations.
- **Local Context**: Variability in local conditions, including socio-economic factors, educational infrastructure, and community involvement, which can impact the outcomes of scaled-up interventions.

Examples of Problematic Extrapolation / External Validity

The paper provides clear examples of problematic extrapolation. The government's nationwide contract teacher program did not replicate the positive results of the NGO-led pilot. The authors attribute this failure to implementation constraints, such as inadequate monitoring, delays in salary payments, and political resistance from teacher unions.

Examples of Successful Extrapolation / External Validity

Successful extrapolation is evident in the NGO implementation of the contract teacher program, which produced significant improvements in student test scores across diverse settings in Kenya. This success underscores the importance of maintaining intervention fidelity and the ability to adapt to local contexts.

Transferring an Urban Food Security Program

While the article does not directly address transferring an urban food security program, the principles discussed are relevant. Collecting detailed data on the new city's context, ensuring rigorous monitoring and accountability, and conducting pilot studies can help ensure the program's success. Adapting the intervention to fit local political and economic conditions, and addressing potential implementation challenges proactively, can enhance the program's effectiveness.

Bold et al. emphasize the critical role of implementation fidelity and contextual adaptation in scaling up successful interventions. By understanding and addressing the specific challenges of government-led implementations, policymakers can improve the external validity and overall impact of scaled-up programs in diverse settings.

2.90 Species Extrapolation for the 21st Century (Celander et al., 2011)

Malin C. Celander and colleagues (2011) explore the challenges and methodologies for extrapolating toxicological data across different species. The focus is on enhancing the accuracy of species extrapolation in ecological risk assessments by leveraging advances in molecular biology and bioinformatics.

Main Problems in Extrapolation / External Validity

The authors identify the primary challenge as the variability in species' responses to toxicants. Traditional approaches often use a few laboratory test species to predict responses for a broad group of environmental species, which can lead to inaccuracies. The difficulty lies in accounting for interspecies differences in physiology, life history traits, and molecular pathways that influence toxicological responses.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, the authors propose a conceptual approach that integrates molecular and computational techniques. Key methods include:

- **Identification of Molecular Pathways**: Understanding the mechanistic basis for toxicological responses at the molecular level.
- **Homology Modeling**: Using computational models to predict interactions of chemicals with enzymes across different species.
- **Adverse Outcome Pathways (AOPs)**: Identifying molecular perturbations that lead to adverse outcomes at the population level.
- **Bioinformatics Tools**: Utilizing databases and computational tools to integrate and analyze molecular data for predicting species-specific responses.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Understanding the specific biological and ecological characteristics of the new context is crucial. Factors such as species' reproductive strategies, tissue-specific gene expression, and metabolic pathways can significantly influence toxicological outcomes. The authors emphasize the importance of detailed molecular and ecological data to improve the accuracy of species extrapolation.

Examples of Problematic Extrapolation / External Validity

The paper discusses general issues with traditional safety factor approaches, which often fail to capture species-specific responses accurately. This can lead to

under- or over-estimation of environmental risks. Specific examples are not detailed, but the authors highlight the need for more sophisticated models to reduce uncertainty.

Examples of Successful Extrapolation / External Validity

Celander et al. present a case study on the effects of steroidogenesis inhibitors (fadrozole and prochloraz) in fish. The study shows similar toxicological responses across three fish species (fathead minnow, medaka, and zebrafish), attributed to similar inhibitor pharmacokinetics and molecular homologies. This demonstrates the potential for successful extrapolation when using detailed molecular data and homology models.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed molecular and ecological data on the new city's context, using bioinformatics tools to predict species-specific responses, and conducting pilot studies can help ensure the program's success. Tailoring interventions based on local biological characteristics and continuously monitoring outcomes are essential steps.

Celander et al. emphasize the importance of integrating molecular biology and computational techniques to enhance the accuracy of species extrapolation. By leveraging advances in bioinformatics and understanding the mechanistic basis of toxicological responses, researchers can improve the external validity of ecological risk assessments and better protect environmental health across diverse species and contexts.

2.91 Target Validity: Bringing Treatment of External Validity in Line with Internal Validity (Lesko et al., 2020)

Catherine R. Lesko and colleagues (2020) discuss the concept of target validity, which encompasses both internal and external validity, in the context of epidemiological research. They emphasize the importance of addressing external bias to ensure that study findings are applicable to target populations.

Main Problems in Extrapolation / External Validity

The authors identify several key issues related to external validity, including the differences in the distribution of effect modifiers between the study sample and the target population, variations in the version of treatments delivered, and the presence

of interference effects where the treatment of one individual affects the outcome of another. These factors can lead to significant biases when attempting to generalize study findings.

Methods for Fixing Problems of Extrapolation / External Validity

Lesko et al. propose several methodologies to address these challenges, focusing on both the design and analysis stages of research:

- **In Design**: They suggest using purposive stratified sampling and pragmatic clinical trials with less restrictive inclusion criteria to improve the representativeness of study samples.
- **In Analysis**: Methods include modeling the probability of the outcome, modeling sample membership, and using doubly robust methods that combine both approaches. These methods help adjust for non-random sampling and enhance the generalizability of findings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study highlights the importance of understanding the specific characteristics of the target population, such as demographic variables and the distribution of effect modifiers. Ensuring that the study sample is representative of the target population on these characteristics is crucial for improving external validity.

Examples of Problematic Extrapolation / External Validity

The authors provide examples where the lack of external validity has led to misleading conclusions. For instance, trials on the effects of medication-assisted therapy for substance use and the effects of antidepressants on suicidal ideation in youth may have yielded different results if conducted in more representative samples.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not extensively detailed, the paper discusses how methodologies like reweighting samples to match target populations have improved the generalizability of findings in various studies.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, using appropriate sampling methods to ensure representativeness, and employing robust statistical adjustments can help ensure the program's effectiveness.

Pilot studies and continuous monitoring based on feedback can further enhance the program's success.

Lesko et al. underscore the importance of integrating considerations of both internal and external validity in epidemiological research. By focusing on target validity and employing robust methodologies, researchers can improve the applicability of their findings to real-world settings, thereby enhancing the relevance and impact of their work on public health decisions.

2.92 The Concept of External Validity (Calder et al., 1982)

Bobby J. Calder, Lynn W. Phillips, and Alice M. Tybout (1982) critically examine the role and importance of external validity in theoretical research. They address the debates surrounding the prioritization of external validity in the context of theory testing and argue against its necessity in every single study.

Main Problems in Extrapolation / External Validity

The authors identify significant issues in the application of external validity, particularly in theoretical research. They argue that the focus on external validity can be misguided due to the complexity and impracticality of accounting for all potential interacting background variables. This complexity often makes it difficult to generalize findings across different measures, persons, settings, and times. The authors point out that theories are generally stated at a universal level, making the specific details of external validity less critical in theory testing.

Methods for Fixing Problems of Extrapolation / External Validity

Calder et al. suggest that instead of striving for external validity in every study, researchers should focus on theory development. They propose that external validity should be addressed through cumulative research and theory refinement rather than in individual studies. The authors advocate for holding background factors constant in single studies to concentrate on internal and construct validity. This approach allows researchers to build robust theories that can later be tested across various contexts.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study emphasizes the importance of understanding specific background factors that might interact with theoretical variables. Factors such as demographic differences, environmental conditions, and socio-economic variables can impact the generalizability of research findings. However, the authors argue that it is practically impossible to account for all these variables in a single study.

Examples of Problematic Extrapolation / External Validity

The authors discuss the theoretical implications of failing to account for interacting background variables, highlighting that earlier studies on attitude change failed to consider the strength of message arguments. This omission led to findings that were later refuted by more comprehensive studies. They argue that attempting to achieve external validity by including all possible background variables can dilute research efforts and reduce the focus on internal and construct validity.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not detailed, the authors emphasize the importance of theory development through cumulative research. They suggest that a robust theory, developed and refined through multiple studies, can achieve external validity over time as it is tested across different contexts.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, focusing on cumulative research, and refining the program through iterative testing in different settings can help ensure its success. Understanding the local variables and gradually adapting the program based on empirical evidence and theoretical development can enhance its effectiveness.

Calder et al. argue that while external validity is important, it should not overshadow the necessity of internal and construct validity in theoretical research. By focusing on theory development and cumulative research, researchers can build robust theories that achieve external validity through extensive testing and refinement across various contexts.

2.93 The External Validity of Experiments (Bracht & Glass, 1968)

Glenn H. Bracht and Gene V. Glass (1968) provide an in-depth analysis of external validity in experimental research, emphasizing the challenges and considerations necessary for generalizing findings across different populations and settings. Their work builds on the foundational concepts introduced by Campbell and Stanley (1963).

Main Problems in Extrapolation / External Validity

The authors identify numerous threats to external validity, which they categorize into two main types: population validity and ecological validity. Population validity concerns the generalizability of findings from the study sample to the target population. Key issues include the differences between experimentally accessible populations and target populations and interactions between personological variables and treatment effects. Ecological validity addresses the conditions under which the results can be generalized, highlighting factors such as multiple-treatment interference, Hawthorne effects, and experimenter effects.

Methods for Fixing Problems of Extrapolation / External Validity

Bracht and Glass propose several strategies to enhance external validity:

- **Explicitly Describing the Independent Variable**: Providing detailed descriptions of the experimental procedures to enable replication and generalization.
- **Sampling Across Conditions**: Ensuring that the experimental design includes a variety of conditions and settings to detect meaningful interactions and enhance generalizability.
- **Conducting Transition Experiments**: Varying aspects of the treatment hypothesized to be independent of the theory to identify which components are truly influential.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study emphasizes the importance of understanding the specific characteristics of both the experimentally accessible population and the target population. Factors such as demographic characteristics, prior experiences, and socio-economic conditions can significantly influence the applicability of experimental findings. The authors argue that detailed knowledge of these characteristics is essential for making informed generalizations.

Examples of Problematic Extrapolation / External Validity

The paper discusses several examples where extrapolation was problematic due to differences in population characteristics or experimental conditions. For instance, the study by Brownell (1966) comparing instructional programs in England and Scotland revealed differing results due to varying levels of familiarity with curricular innovations. Similarly, Kendler and Kendler (1959) found that mediational S-R theory applied differently to fast and slow learners, illustrating the interaction between learning strategies and student characteristics.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not detailed extensively, the authors highlight the importance of considering both population and ecological validity. Studies that carefully account for these factors and use robust methodologies can achieve more reliable and generalizable results.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, understanding the specific characteristics of the target population, and considering potential interactions between program components and local

conditions are crucial steps. Conducting pilot studies and continuously monitoring and adjusting the program based on empirical evidence can help ensure its success.

Bracht and Glass emphasize the complexity and importance of external validity in experimental research. By carefully considering population and ecological factors and employing rigorous methodologies, researchers can enhance the generalizability and practical relevance of their findings, leading to more effective and applicable policy decisions.

2.94 The External Validity of Laboratory Experiments: Qualitative Rather than Quantitative Effects (Kessler & Vesterlund, 2015)

Judd Kessler and Lise Vesterlund (2015) address the debate on the external validity of laboratory experiments in economics. They argue that while laboratory experiments are often criticized for their lack of quantitative external validity, the qualitative insights they provide are generally valid and crucial for understanding economic behaviors.

Main Problems in Extrapolation / External Validity

The primary issue identified is the discrepancy between laboratory settings and real-world environments. Critics argue that the high level of scrutiny, lack of anonymity, specific contexts, stakes involved, and the population used in lab experiments can lead to results that do not generalize to broader, more natural settings. This skepticism is particularly focused on the quantitative aspects of results, such as the exact magnitude of effects observed in the lab versus the field.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Kessler and Vesterlund emphasize the importance of focusing on qualitative rather than quantitative external validity. They argue that laboratory experiments should aim to identify general principles of behavior that can be applied across various settings. The authors suggest that:

- **Modifying Laboratory Designs**: Adjusting laboratory experiments to better mirror real-world conditions when necessary to improve generalizability.
- **Emphasizing Qualitative Results**: Prioritizing the identification of directional effects rather than exact magnitudes, as these are more likely to generalize across different contexts.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study highlights the necessity of understanding the specific features of new contexts and populations that may impact the generalizability of findings. These include the socio-economic background of participants, the cultural and institutional environment, and the specifics of the economic behaviors being studied. The authors argue that these factors should be considered when interpreting the external validity of experimental results.

Examples of Problematic Extrapolation / External Validity

The paper discusses general concerns about the external validity of laboratory experiments but does not provide specific problematic examples. Instead, it focuses on the broader debate, referencing critics like Levitt and List, who have highlighted issues such as the context-specific nature of lab findings and their limited applicability to real-world settings.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not detailed, Kessler and Vesterlund argue that many laboratory findings have been successfully applied in broader contexts by focusing on qualitative insights. They cite the general principles of behavior identified in lab settings, which often hold true in the field, even if the quantitative effects differ.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context and focusing on the qualitative aspects of findings—such as the general principles of behavior regarding food security—can help ensure the program's effectiveness. Pilot studies and adjustments based on initial findings can further enhance the program's success.

Kessler and Vesterlund highlight the value of laboratory experiments in identifying general principles of economic behavior. By emphasizing qualitative external validity and adapting laboratory designs to better reflect real-world conditions, researchers can enhance the generalizability and practical relevance of their findings.

2.95 The Extrapolation Problem and How Population Modeling Can Help (Forbes et al., 2008)

Valery E. Forbes, Peter Calow, and Richard M. Sibly (2008) address the challenges of extrapolating toxicological data from individual-level effects to population-level impacts in ecological risk assessments. They review various extrapolation methods

and advocate for the use of population modeling to improve the accuracy and reliability of these assessments.

Main Problems in Extrapolation / External Validity

The authors identify several key issues with current extrapolation methods, such as the reliance on application factors and species sensitivity distributions (SSDs), which often fail to capture the complexity and variability of ecological systems. The primary problem is the difficulty in translating individual-level toxicological responses to population-level outcomes due to non-linear interactions and the influence of ecological context. This can lead to inaccurate risk assessments, resulting in either over- or underestimation of ecological risks.

Methods for Fixing Problems of Extrapolation / External Validity

Forbes et al. propose the use of population models as a solution to these challenges. They categorize population models into three broad classes:

1. **Demographic Models**: These models focus on recruitment and survival, structured by size or age classes, and can include spatial structures.
2. **Energy Budget Models**: These models describe individual responses in terms of energy intake and output, relating to growth rates and reproductive performance.
3. **Individual-Based Models (IBMs)**: These models treat each individual within a population as distinct, with population dynamics emerging from the interactions of individual responses.

By integrating these models, researchers can link individual-level effects to population dynamics, incorporating greater ecological complexity and providing more accurate predictions of population-level impacts.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study emphasizes the importance of understanding the specific ecological context and population characteristics when extrapolating findings. Factors such as demographic variables, life-history traits, and ecological interactions (e.g., competition and trophic relationships) can significantly influence the applicability of toxicological data to new populations and environments.

Examples of Problematic Extrapolation / External Validity

The paper highlights several limitations of current methods, such as the arbitrary nature of application factors and the empirical basis of SSDs, which often do not

account for the complex interactions and variability within ecosystems. These limitations can result in significant discrepancies between predicted and actual ecological impacts.

Examples of Successful Extrapolation / External Validity

Forbes et al. discuss the potential of population models to improve the accuracy of ecological risk assessments. While specific successful examples are not detailed, they highlight the ability of these models to integrate individual-level responses and predict population dynamics more reliably, thus enhancing the external validity of extrapolations.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, using appropriate population models to simulate the program's impact on different demographic groups, and considering ecological interactions are crucial steps. Conducting pilot studies and continuously monitoring outcomes can further enhance the program's success.

Forbes et al. advocate for the integration of population modeling in ecological risk assessments to address the limitations of current extrapolation methods. By incorporating greater ecological complexity and understanding the specific context of target populations, researchers can improve the accuracy and reliability of their assessments, leading to better-informed environmental and policy decisions.

2.96 The Extrapolation Problem: How Can We Learn from the Experience of Others? (Bardach, 2004)

Eugene Bardach (2004) addresses the challenge of extrapolating successful practices from one context to another in policy analysis. He examines how policy analysts can learn from the experiences of others while acknowledging the complexity and potential pitfalls of this process.

Main Problems in Extrapolation / External Validity

Bardach identifies the central issue in extrapolation as the difficulty in accurately transferring successful practices from one context to another. This includes understanding which elements of a practice are essential for its success and which can be adapted. He highlights the risk of misinterpreting the context and mechanisms

that drive the effectiveness of a practice, leading to failures when these practices are applied in different settings.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Bardach proposes a systematic approach to extrapolation, focusing on identifying and understanding the basic mechanisms that make a practice successful. He emphasizes the importance of:

- **Mechanism Identification**: Recognizing the causal mechanisms that underlie the success of a practice.
- **Contextual Adaptation**: Adapting practices to the specific conditions of the new context, rather than attempting strict replication.
- **Conceptual Framework**: Using a robust conceptual framework to guide the adaptation process and ensure that the core mechanisms are preserved.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Bardach underscores the importance of considering various features of the new context, such as institutional, political, economic, and cultural factors. These features can significantly impact the success of a transplanted practice. Understanding these local conditions helps in tailoring the practice to fit the new environment effectively.

Examples of Problematic Extrapolation / External Validity

The paper provides examples of common mistakes in extrapolation, such as underestimating the need for public education campaigns or failing to account for local political dynamics. These errors often stem from a lack of understanding of the essential mechanisms and the specific conditions required for a practice's success.

Examples of Successful Extrapolation / External Validity

While the paper does not detail specific successful examples, Bardach emphasizes that successful extrapolation involves thorough understanding and careful adaptation. He discusses the concept of "smart practices" that are both effective and adaptable, providing general guidelines for identifying and implementing these practices in new contexts.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Collecting detailed data on the new city's context, understanding the key mechanisms that drive the success of the program, and adapting it to fit local conditions are crucial steps. Conducting

pilot studies and continuously monitoring and adjusting the program based on empirical evidence can help ensure its success.

Bardach's paper highlights the complexity of extrapolation in policy analysis and the importance of a systematic, mechanism-focused approach. By identifying the core elements that drive the success of a practice and adapting them to fit new contexts, policymakers can enhance the likelihood of successful implementation and avoid common pitfalls.

2.97 The Generalizability Puzzle: Rigorous Impact Evaluations and Their Relevance (Bates & Glennerster, 2017)

Mary Ann Bates and Rachel Glennerster (2017) explore the challenges of generalizing findings from impact evaluations to new contexts. They discuss how policymakers can effectively use both local and global evidence to inform policy decisions, emphasizing the importance of understanding the underlying mechanisms behind successful interventions.

Main Problems in Extrapolation / External Validity

The authors identify the central issue as the generalizability puzzle: whether the results of a specific program can be expected to hold in different contexts. Key problems include differences in local conditions, such as socio-economic factors, cultural norms, and institutional structures, which can affect the applicability of findings from one setting to another. The authors also highlight the difficulty in deciding whether to prioritize local data over more rigorous but contextually different global evidence.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Bates and Glennerster propose a framework that integrates local and global evidence:

1. **Mechanism Focus**: Emphasizing the understanding of causal mechanisms behind successful interventions, which can help determine whether similar mechanisms are likely to operate in a new context.
2. **Contextual Similarity**: Assessing whether the local conditions in the new context match those of the original study setting, including factors such as demographic similarities and local infrastructure.
3. **Local Testing**: Conducting localized pilot studies or randomized evaluations to test the intervention under new conditions before scaling up.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study underscores the importance of understanding specific features of the new context that might affect the success of an intervention. These features include local economic conditions, social norms, and the capacity of local institutions to implement and sustain the intervention. Detailed contextual analysis is crucial to adapting the intervention effectively.

Examples of Problematic Extrapolation / External Validity

The authors provide examples of common mistakes in extrapolation, such as the assumption that a program successful in one country will work similarly in another without considering local differences. For instance, the failure of a Kenyan program to be effective in Rwanda due to differences in local perceptions of HIV risk illustrates the importance of understanding local conditions.

Examples of Successful Extrapolation / External Validity

Successful extrapolation is discussed through the example of a tutoring program implemented in Chicago, which drew on evidence from similar programs in India and Kenya. By focusing on the underlying mechanisms—such as the need for tailored, small-group instruction for students who had fallen behind—the program was adapted successfully to a new urban context.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are relevant. Collecting detailed data on the new city's context, understanding the mechanisms behind successful food security interventions, and adapting the program to fit local conditions are crucial steps. Conducting pilot studies and monitoring outcomes to refine the intervention can further ensure its success.

Bates and Glennerster highlight the importance of a nuanced approach to generalization, focusing on mechanisms and contextual similarities. By integrating local and global evidence and conducting careful contextual analysis, policymakers can improve the likelihood of successful adaptation and implementation of programs in new settings.

2.98 The Importance of External Validity (Steckler & McLeroy, 2008)

Allan Steckler and Kenneth R. McLeroy (2008) emphasize the critical importance of external validity in public health research. They argue that the focus on internal validity has often overshadowed the need for generalizability, which is crucial for translating research findings into effective public health practice.

Main Problems in Extrapolation / External Validity

The authors identify the primary issue as the historical emphasis on internal validity, often at the expense of external validity. This focus results in highly controlled studies that may not generalize well to real-world settings. They highlight that this imbalance has contributed to a significant lag in translating research into practical public health interventions, with effective interventions often failing to reach or impact diverse populations.

Methods for Fixing Problems of Extrapolation / External Validity

Steckler and McLeroy propose several methods to improve external validity:

1. **Efficacy and Effectiveness Trials**: They advocate for a research progression from efficacy trials, which test interventions in controlled environments, to effectiveness trials, which evaluate interventions in more real-world settings.
2. **Detailed Reporting**: Emphasizing the need for detailed reporting on external validity factors, including participant recruitment, selection procedures, and implementation consistency.
3. **Adoption of Models and Guidelines**: Utilizing frameworks like CONSORT, TREND, and RE-AIM to systematically enhance the reporting and evaluation of external validity in research publications.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study highlights the importance of various contextual features that can impact the generalizability of research findings. These include demographic variables, socio-economic conditions, cultural norms, and local organizational structures. Understanding these features is essential for adapting interventions to new settings and populations effectively.

Examples of Problematic Extrapolation / External Validity

The paper discusses the issue of a substantial lag between discovery and delivery of effective interventions, particularly in cancer prevention and control. This lag is attributed to the failure to consider external validity adequately, resulting in

interventions that are not widely adopted or effective in different populations and settings.

Examples of Successful Extrapolation / External Validity

Specific successful examples are not extensively detailed, but the authors discuss the potential benefits of adopting models like RE-AIM to improve the generalizability of research findings. By focusing on real-world applicability and systematic reporting, researchers can enhance the likelihood of successful extrapolation.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, ensuring that interventions are tested in real-world settings through effectiveness trials, and using systematic reporting frameworks can help ensure the program's success. Conducting pilot studies and continuous monitoring based on empirical evidence can further enhance the program's applicability and effectiveness.

Steckler and McLeroy emphasize the need to balance internal and external validity in public health research. By focusing on the generalizability of findings and adopting systematic approaches to enhance external validity, researchers can improve the translation of research into practice, ultimately leading to more effective public health interventions.

2.99 The Logic of Generalization From Systematic Reviews and Meta-Analyses of Impact Evaluations (Littell, 2024)

Julia H. Littell (2024) delves into the challenges and methodologies for generalizing findings from systematic reviews and meta-analyses (SRMAs) of impact evaluations. The article emphasizes the need for robust approaches to assess the generalizability of intervention effects across diverse policy and practice contexts.

Main Problems in Extrapolation / External Validity

Littell identifies several key issues in generalizing findings from SRMAs:

- **Heterogeneity of Effects**: Variations in treatment effects across different studies, populations, and settings complicate the generalization process.
- **Sampling and Representativeness**: The reliance on non-probability samples in most studies included in SRMAs limits the ability to generalize findings to broader populations.

- **Measurement and Reporting**: Inconsistencies and gaps in the descriptive data provided by primary studies hinder the ability to assess external validity comprehensively.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Littell proposes a pragmatic approach that includes:

- **Proximal Similarity**: Using the principle of proximal similarity to assess how closely the contexts of the original studies match the target contexts.
- **Heterogeneous Irrelevancies**: Demonstrating that findings hold across variations in conceptually irrelevant factors.
- **Discriminant Validity**: Ensuring that the observed effects are due to the intervention itself and not confounded by other variables.
- **Interpolation and Extrapolation**: Identifying the range of contexts over which the findings are likely to hold, and specifying potential moderators of effects.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study underscores the importance of considering various contextual features, including demographic characteristics, socio-economic conditions, and institutional settings, which can significantly impact the applicability of intervention findings. Detailed knowledge of these attributes helps in tailoring interventions to new contexts.

Examples of Problematic Extrapolation / External Validity

Littell discusses the issue of relying on simplified rubrics, such as the "two-studies rule" and single point estimates, for generalization. These approaches often fail to account for the heterogeneity of effects and the complexity of social interventions, leading to unreliable generalizations.

Examples of Successful Extrapolation / External Validity

The article does not provide specific successful examples but emphasizes the potential of SRMAs to test generalizability claims and explore heterogeneity. By systematically synthesizing data from multiple studies, SRMAs can reveal broader trends and identify conditions under which interventions are more or less effective.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, using systematic reviews to understand potential variations in

effectiveness, and conducting pilot studies can help ensure the program's success. Adapting the intervention based on empirical evidence and continuously monitoring outcomes can further enhance its applicability and effectiveness.

Littell's paper highlights the complexity of generalizing findings from SRMAs and the need for a nuanced, systematic approach to assessing external validity. By focusing on the principles of proximal similarity, heterogeneous irrelevancies, discriminant validity, and interpolation and extrapolation, researchers can improve the generalizability and practical relevance of their findings in diverse contexts.

2.100 The Necessity of Construct and External Validity for Generalized Causal Claims (Esterling et al., 2023)

Kevin M. Esterling, David Brady, and Eric Schwitzgebel (2023) explore the critical roles of construct and external validity in making generalized causal claims. They argue that the Credibility Revolution's emphasis on internal validity has led to a neglect of these essential components, which are necessary for robust and meaningful causal generalizations.

Main Problems in Extrapolation / External Validity

The authors identify the primary problem as the overemphasis on internal validity, often at the expense of construct and external validity. This focus can lead to causal claims that, while internally valid, fail to generalize beyond the specific conditions of the original study. Mislabeling causes or outcomes (construct validity) and misunderstanding the conditions under which the cause has an effect (external validity) are central issues that undermine the generalizability of findings.

Methods for Fixing Problems of Extrapolation / External Validity

Esterling et al. propose a framework called causal specification to address these challenges. This framework integrates assumptions about internal, construct, and external validity into a single causal expression. Key methods include:

- **Clarifying Constructs**: Ensuring that the cause and outcome are correctly labeled and conceptualized, avoiding misinterpretation of what is being measured and its implications.
- **Contextual Analysis**: Identifying and explicitly stating the conditions under which the causal relationship holds, allowing for more accurate generalizations.
- **Theoretical and Qualitative Integration**: Using qualitative research and theoretical insights to support and validate causal claims, thus ensuring that

statistical findings are grounded in a broader understanding of the causal process.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study emphasizes the necessity of understanding specific features of the new context, such as demographic characteristics, socio-economic conditions, and institutional settings. These features can significantly impact whether the causal relationship observed in the original study will hold in the new context. Detailed contextual analysis and the integration of local knowledge are crucial for accurate extrapolation.

Examples of Problematic Extrapolation / External Validity

The authors discuss examples where reliance solely on internal validity has led to problematic generalizations. For instance, findings from randomized controlled trials (RCTs) in one context may not generalize to other settings due to differences in local conditions that were not accounted for. This highlights the limitations of purely statistical approaches that do not consider the broader context.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not extensively detailed, the authors emphasize the potential of their causal specification framework to improve generalizability. By integrating construct and external validity into causal claims, researchers can make more robust and applicable generalizations. The framework helps ensure that findings are relevant and valid across different settings and populations.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, ensuring the program's constructs are valid, and explicitly stating the conditions under which the program works are crucial steps. Conducting pilot studies and continuously monitoring and adjusting the program based on empirical evidence can further enhance its success.

Esterling et al. argue that for causal claims to be genuinely credible and generalizable, researchers must balance internal, construct, and external validity. By adopting their causal specification framework, integrating theoretical and qualitative insights, and conducting thorough contextual analyses, researchers can improve the robustness

and applicability of their findings, leading to more effective policy and practice interventions.

2.101 The Role of External Validity in Theoretical Research (Lynch, 1983)

John G. Lynch, Jr. (1983) examines the debate on the relevance of external validity in theoretical research, specifically within consumer research. He critiques the position that external validity is less important than internal validity for theory testing and argues for a more balanced approach.

Main Problems in Extrapolation / External Validity

Lynch identifies the central issue as the tendency to dismiss external validity in favor of internal validity, following the argument by Calder et al. (1982). This dismissal can lead to theories that are statistically robust in controlled settings but fail to generalize to broader, real-world contexts. The main problem is the interaction between treatment effects and background variables, which can mask significant findings when not properly accounted for.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Lynch proposes several methods:

- **Deliberate Sampling for Heterogeneity**: This involves sampling across a variety of conditions to identify interactions between treatment effects and background variables. This method aims to ensure that the results are not overly dependent on specific, controlled conditions.
- **Selective Approach**: This method involves systematically varying specific background factors that are hypothesized to interact with the treatment. By doing so, researchers can identify boundary conditions and understand the limits of their theoretical claims.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Lynch emphasizes the importance of understanding and explicitly testing the interaction between theoretical variables and background factors. Features such as demographic differences, socio-economic conditions, and cultural contexts can significantly affect the applicability of findings. By identifying and testing these interactions, researchers can improve the external validity of their theories.

Examples of Problematic Extrapolation / External Validity

The paper discusses general issues with ignoring external validity, such as the risk of developing theories that do not hold up in different contexts. Lynch points to memory research, where holding the meaningfulness of stimuli constant (e.g., using nonsense syllables) inhibited understanding of real-world memory processes, as an example of how ignoring external variables can hinder theoretical progress.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not detailed, Lynch highlights the potential benefits of adopting his proposed methods. By using deliberate sampling for heterogeneity and the selective approach, researchers can gain a better understanding of how their theories hold up across different conditions and improve their generalizability.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, using deliberate sampling to test the program under various conditions, and identifying key background factors that may interact with the intervention can help ensure its success. Pilot studies and continuous monitoring based on these principles can further enhance the program's applicability and effectiveness.

Lynch's paper underscores the importance of balancing internal and external validity in theoretical research. By adopting methods that account for the interaction between theoretical variables and background factors, researchers can develop more robust and generalizable theories, leading to better-informed policy and practice interventions.

2.102 The Science of Using Science: Towards an Understanding of the Threats to Scaling Experiments (Al-Ubaydli et al., 2019)

Omar Al-Ubaydli, John A. List, and Dana Suskind (2019) explore the challenges and threats associated with scaling experimental findings to broader populations. They provide a theoretical framework to understand these issues and suggest methods to address them.

Main Problems in Extrapolation / External Validity

The authors identify the primary problem as the "scale-up problem," which revolves around three key threats to scalability:

1. **Statistical Inference**: Determining when evidence becomes actionable.

2. **Population Properties**: The representativeness of the experimental population.
3. **Situation Properties**: The representativeness of the experimental situation.

They argue that failing to understand and address these areas can lead to significant vulnerabilities when scaling up experiments, potentially wasting resources and diminishing public trust in scientific methods.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, the authors propose:

- **Rigorous Statistical Methods**: Implementing more precise statistical summaries and frequent replication of studies to address inference problems.
- **Representative Sampling**: Ensuring that the sample population closely mirrors the broader population to which the results will be generalized.
- **Contextual Adaptation**: Identifying and adapting to the key situational features that affect scalability, including fidelity to the core components of the intervention and understanding the mechanisms behind its effects.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study emphasizes the importance of understanding specific features of the new context and target population that can impact scalability. These include demographic characteristics, socio-economic conditions, and the operational environment. Ensuring that these factors are considered can significantly improve the chances of successful extrapolation.

Examples of Problematic Extrapolation / External Validity

The authors discuss several examples where interventions that were successful in initial studies failed when scaled up. One notable example is the Collaborative Strategic Reading (CSR) intervention, which showed promising results in initial trials but had no discernible effect when replicated in broader settings. This illustrates the importance of broad replication before wide implementation.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not extensively detailed, the authors highlight the importance of addressing the identified threats to scalability. By using their proposed methods, researchers can improve the generalizability and robustness of their findings, leading to more effective policy implementations.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, ensuring that the core components of the program are maintained, and adapting the intervention to local conditions are crucial steps. Conducting pilot studies and continuously monitoring outcomes can further enhance the program's success.

Al-Ubaydli et al. emphasize the complexity of scaling experimental findings and the need for a comprehensive understanding of the factors affecting scalability. By addressing statistical inference, population properties, and situational features, researchers and policymakers can improve the likelihood of successful extrapolation and implementation of interventions at scale.

2.103 The Sequential Scale-Up of an Evidence-Based Intervention: A Case Study (Thomas et al., 2018)

Jaime Thomas, Thomas D. Cook, Alice Klein, Prentice Starkey, and Lydia DeFlorio (2018) analyze the process of scaling up the Pre-K Mathematics program from controlled studies to a statewide implementation. They explore the challenges and outcomes of this expansion, providing insights into the factors that influence the effectiveness of interventions at larger scales.

Main Problems in Extrapolation / External Validity

The authors identify several issues with scaling up interventions, primarily the challenge of maintaining program effectiveness in larger, more heterogeneous settings. Key problems include variations in program implementation, changes in the counterfactual conditions, and differences in the populations and settings involved. As the scale increases, the consistency of program delivery and fidelity to the original design can diminish, potentially leading to reduced effectiveness.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Thomas et al. propose a sequential approach to scaling up interventions, which involves:

1. **Pilot Studies**: Testing the feasibility of the intervention in highly controlled conditions.
2. **Efficacy Research**: Assessing the intervention under less controlled conditions but still involving the developers.

3. **Effectiveness Studies**: Testing the intervention in real-world conditions with less developer involvement.
4. **Scale-Up Studies**: Implementing the intervention on a larger scale to evaluate its effectiveness in broader, more diverse populations.

They also recommend using frameworks to explain why effect sizes might change as scale increases, and comparing causal estimates from different stages of research to understand these dynamics.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study emphasizes the importance of considering the number and heterogeneity of settings and study participants, quality of program content and delivery, changes in the counterfactual condition, variations in outcome measurement quality, and quality of evaluation design and execution. These factors can significantly impact the success of scaling up an intervention.

Examples of Problematic Extrapolation / External Validity

The paper discusses how increased sample heterogeneity and changes in implementation fidelity can lead to reduced effect sizes. For instance, in the case of Pre-K Mathematics, larger and more diverse samples presented challenges in maintaining the quality and consistency of program delivery, which affected the program's effectiveness.

Examples of Successful Extrapolation / External Validity

The study provides a nuanced view of how scaling up can be successful by carefully managing the factors mentioned above. The Pre-K Mathematics program remained effective at the state level, though the effect sizes were smaller than in earlier, smaller-scale studies. This success was attributed to the systematic approach to scaling up and the continuous adaptation of the program based on empirical evidence.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, maintaining the core components of the program, and adapting the intervention to fit local conditions are crucial steps. Conducting pilot studies and continuously monitoring and adjusting the program based on empirical evidence can further enhance its success.

Thomas et al. highlight the complexity of scaling up interventions and the importance of a sequential, evidence-based approach. By understanding and addressing the factors that affect program effectiveness at larger scales, policymakers and researchers can improve the likelihood of successful implementation and broader adoption of evidence-based interventions.

2.104 The Theory-Driven Approach to Validity (Chen & Rossi, 1987)

Huey-Tsyh Chen and Peter H. Rossi (1987) propose a theory-driven approach to validity in program evaluation, arguing that current approaches tend to prioritize one type of validity over others, leading to incomplete and potentially misleading results.

Main Problems in Extrapolation / External Validity

The authors identify the primary problem as the trade-off between different types of validity, particularly the dominance of internal validity in experimental and quasi-experimental designs. This focus often comes at the expense of external validity, resulting in findings that are not generalizable to other contexts. They argue that methods emphasizing internal validity, such as randomization, can undermine external validity by creating artificial conditions that do not reflect real-world scenarios.

Methods for Fixing Problems of Extrapolation / External Validity

Chen and Rossi propose a balanced, theory-driven approach to validity, which involves:

1. **Formulating a Theory**: Developing a detailed model or theory to guide the evaluation, identifying potential threats to validity across all types.
2. **Model Specification**: Including extraneous variables and interactions in the model to account for their influence on the outcomes, rather than relying solely on randomization.
3. **Mixed Methods**: Combining randomization with model specification and qualitative methods to enhance both internal and external validity.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The theory-driven approach emphasizes the importance of understanding the specific context and population where the intervention will be applied. Factors such as demographic characteristics, socio-economic conditions, and the operational environment must be considered to ensure that the findings are applicable to

the new setting. This involves detailed modeling of the relationships between variables and careful consideration of how these relationships might change in different contexts.

Examples of Problematic Extrapolation / External Validity

The authors critique traditional experimental designs for their limited applicability to broader contexts. They reference the case of cross-sectional surveys, which often fail to account for self-selection biases and other confounding factors, leading to misleading conclusions about causal relationships.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not detailed, Chen and Rossi highlight the potential of their theory-driven approach to improve generalizability. By incorporating a comprehensive understanding of the underlying causal mechanisms and contextual factors, researchers can make more accurate and applicable generalizations.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Developing a detailed theoretical model of the program, including potential contextual variables and interactions, can help ensure its success in a new city. Conducting pilot studies, using both quantitative and qualitative methods, and continuously refining the program based on empirical evidence can further enhance its applicability and effectiveness.

Chen and Rossi argue for a balanced approach to validity that integrates multiple types of validity through a theory-driven framework. By focusing on the underlying causal mechanisms and the specific context of the intervention, researchers can improve the robustness and generalizability of their findings, leading to more effective policy and practice implementations.

2.105 Theorizing How Interventions Work in Evaluation: Process-Tracing Methods and Theorizing Process Theories of Change (Camacho Garland & Beach, 2023)

Gabriela Camacho Garland and Derek Beach (2023) present a framework for using process-tracing methods to theorize process theories of change (pToC) in evaluation. This approach aims to provide a more detailed understanding of how interventions work by unpacking the causal linkages and interactions that produce outcomes.

Main Problems in Extrapolation / External Validity

The authors identify a key issue in evaluation as the failure to unpack the sequence of actions and interactions between program actors and stakeholders. Traditional evaluation methods, such as Realist Evaluation (RE) and Contribution Analysis (CA), often treat the process as a series of static, one-off activities without detailing how these activities interact to produce outcomes. This oversight can lead to challenges in generalizing findings to different contexts.

Methods for Fixing Problems of Extrapolation / External Validity

Camacho Garland and Beach propose a three-step process to develop more granular pToCs:

1. **Defining the Intervention and Potential Contribution**: Clearly articulate the intervention and the outcomes it aims to achieve.
2. **Theorizing Potential Contribution Pathways**: Identify and theorize the different pathways through which the intervention could lead to the desired outcomes, considering potential challenges and barriers.
3. **Unpacking the Process**: Break down the process into key episodes of interaction, detailing the actions and causal principles that link them.

This method allows for a detailed examination of how interventions work, providing a robust framework for understanding and testing the causal linkages in different contexts.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study emphasizes the importance of understanding the specific contextual factors that may influence the success of an intervention. These include demographic characteristics, socio-economic conditions, cultural factors, and institutional settings. By identifying and incorporating these factors into the pToC, evaluators can better assess the potential for successful extrapolation to new contexts.

Examples of Problematic Extrapolation / External Validity

The paper highlights issues with traditional theories of change that fail to unpack interactions between actors, leading to static models that do not account for the dynamic nature of real-world interventions. This can result in evaluations that do not provide a clear understanding of how and why an intervention succeeded or failed, limiting their applicability to other settings.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not extensively detailed, the authors illustrate the use of their framework with a hypothetical example involving vaccination programs. By unpacking the process of overcoming vaccine hesitancy through community engagement, tailored communication, and interpersonal dialogue, they demonstrate how detailed pToCs can provide actionable insights for scaling interventions to new contexts.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are relevant. Collecting detailed data on the new city's context, defining clear pathways for how the program can achieve its goals, and unpacking the interactions between program actors and stakeholders can help ensure its success. Conducting pilot studies and continuously refining the program based on empirical evidence can further enhance its applicability and effectiveness.

Camacho Garland and Beach's framework highlights the importance of detailed, theory-driven evaluations that unpack the causal processes of interventions. By focusing on the interactions and mechanisms that drive outcomes, their approach provides a robust tool for understanding and improving the external validity of evaluations, making them more applicable across different contexts and populations.

2.106 Theory and External Validity (Lynch, 1999)

John G. Lynch, Jr. (1999) discusses the importance of external validity in theoretical research, particularly within the context of consumer behavior studies. He critiques common methods aimed at enhancing external validity and offers a theoretical framework to better understand and assess it.

Main Problems in Extrapolation / External Validity

Lynch identifies several key problems with current approaches to external validity:

- **Overemphasis on Method**: Many researchers focus on methodological procedures like probability sampling and realistic settings, which Lynch argues do not necessarily enhance external validity.
- **Interaction with Background Factors**: There is often a lack of understanding about how the focal variables in a theory interact with various background factors, which can lead to misleading conclusions about generalizability.

Methods for Fixing Problems of Extrapolation / External Validity

Lynch proposes a shift from focusing on method to focusing on theory. Key methods include:

- **Specifying Moderator Variables**: Developing theories that explicitly include moderator variables and boundary conditions.
- **Testing Interactions**: Designing studies to test for interactions between the focal variables and background factors.
- **Deliberate Sampling for Heterogeneity**: Ensuring that studies include diverse settings and populations to identify robust findings and potential interactions.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study highlights the importance of understanding specific contextual features that can impact the success of generalizing findings. These include demographic differences, cultural factors, and institutional settings. By specifying and testing these variables, researchers can better assess whether their findings will hold in different contexts.

Examples of Problematic Extrapolation / External Validity

Lynch critiques the belief that findings from realistic field settings are inherently more generalizable than those from laboratory settings. He argues that both settings can suffer from limited external validity if the interaction with background factors is not well understood. An example provided is the use of probability sampling, which often fails to account for the true complexity of real-world behaviors.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not detailed extensively, Lynch emphasizes that theories incorporating well-specified moderator variables and boundary conditions are more likely to produce generalizable findings. The approach allows researchers to understand how and why certain results occur, making it easier to apply these findings to new contexts.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Developing a detailed theoretical model that includes potential moderator variables and boundary conditions, conducting pilot studies to test these interactions, and continuously refining the program based on empirical evidence can help ensure its success in a new city.

Lynch's paper advocates for a theory-driven approach to external validity, emphasizing the need to understand and test the interactions between focal variables and background factors. By focusing on theory rather than method, researchers can improve the robustness and generalizability of their findings, leading to more effective and applicable interventions.

2.107 Theory-Testing, Generalization, and the Problem of External Validity (Lucas, 2003)

Jeffrey W. Lucas (2003) examines the issue of external validity in social science research, particularly in the context of theory-testing. He argues that external validity is often misunderstood and misapplied, especially in experimental research, and emphasizes the importance of theoretical considerations in assessing external validity.

Main Problems in Extrapolation / External Validity

Lucas identifies the primary problem as the misconception that external validity is solely a methodological issue. He argues that external validity is primarily a theoretical concern and should be assessed based on the interplay between theory and method. The main problems include:

- **Overemphasis on Sampling**: The belief that probability sampling alone can ensure external validity is flawed because it does not account for the theoretical constructs being tested.
- **Misinterpretation of Artificiality**: Experimental settings are often criticized for their artificiality, but Lucas argues that this criticism overlooks the advantage of controlling theoretically relevant variables.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Lucas proposes several methods:

- **Theoretical Scope Conditions**: Clearly defining the scope conditions of the theory being tested, which specify the circumstances under which the theory is expected to hold.
- **Replication and Reproducibility**: Conducting replication studies in diverse settings to increase confidence in the theory's applicability across different contexts.
- **Construct Validity**: Ensuring that empirical measures accurately reflect the theoretical constructs, which is essential for meaningful generalization.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Lucas highlights the importance of understanding the theoretical scope conditions that define the applicability of research findings. These include demographic characteristics, socio-economic conditions, and cultural factors. By specifying and testing these conditions, researchers can better assess the potential for successful generalization to new contexts.

Examples of Problematic Extrapolation / External Validity

The paper discusses general issues with assuming that findings from highly controlled experimental settings can be easily generalized to natural settings without considering theoretical scope conditions. This can lead to misleading conclusions about the applicability of the findings.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not extensively detailed, Lucas emphasizes that findings that are supported by diverse tests across different settings enhance the external validity of the theory. Successful generalization occurs when the scope conditions of the theory are well-specified and consistently supported.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Developing a detailed theoretical model with clear scope conditions, conducting replication studies in different urban settings, and ensuring that empirical measures accurately reflect theoretical constructs can help ensure the program's success. Continuous monitoring and refinement based on empirical evidence can further enhance its applicability and effectiveness.

Lucas argues that external validity should be treated primarily as a theoretical issue rather than purely a methodological one. By focusing on the interplay between theory and method, clearly defining scope conditions, and ensuring construct validity, researchers can improve the generalizability and robustness of their findings, leading to more effective policy and practice interventions.

2.108 Theory, External Validity, and Experimental Inference: Some Conjectures (Martel Garcia & Wantchekon, 2010)

Fernando Martel Garcia and Leonard Wantchekon (2010) explore the relationship between theory and external validity in experimental research, particularly in the context of policy evaluations. They propose theoretical approaches to improve the generalizability of experimental findings.

Main Problems in Extrapolation / External Validity

The authors identify a significant problem in experimental research: the trade-off between internal and external validity. While experiments are strong in identifying causal relationships (internal validity), they often struggle with generalizability (external validity). They point out that variations in background conditions, covariates, and settings can limit the external validity of experimental findings.

Methods for Fixing Problems of Extrapolation / External Validity

Martel Garcia and Wantchekon suggest two main approaches to address these issues:

1. **Robustness Approach**: This involves performing replication studies across different settings, treatments, outcome measures, and units to test the consistency of findings. While this method can reduce uncertainty, it is often impractical and costly.
2. **Analytical Approach**: This approach relies on theoretically motivated replications. It involves developing a series of experiments that explicitly test various aspects of a theory, incorporating mediator and moderator variables that may affect the generalizability of findings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The study emphasizes the importance of understanding and theorizing about the mediator and moderator variables that might influence the outcomes in new contexts. Key features include socio-economic conditions, cultural factors, and specific attributes of the intervention that need to be considered and tested to ensure successful extrapolation.

Examples of Problematic Extrapolation / External Validity

The authors discuss the case of Opportunity NYC, a conditional cash transfer (CCT) program inspired by Mexico's Progresa. Despite similarities, significant differences in context and target populations between New York City and rural Mexico posed

challenges to generalizing the success of Progresa to Opportunity NYC. Critics argued that the socio-economic conditions and policy needs were sufficiently different to question the applicability of the Mexican program to New York.

Examples of Successful Extrapolation / External Validity

While the paper does not provide detailed examples of successful extrapolation, it highlights the potential of the analytical approach. By connecting individual experiments with a coherent theoretical framework, researchers can better predict and understand the applicability of findings across different contexts.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are relevant. Developing a detailed theoretical model that includes potential mediator and moderator variables, conducting pilot studies, and continuously refining the program based on empirical evidence can help ensure its success in a new city. This approach can reduce uncertainty and enhance the program's applicability and effectiveness.

Martel Garcia and Wantchekon argue that theory-driven approaches can significantly improve the external validity of experimental research. By focusing on theoretically motivated replications and understanding the role of mediator and moderator variables, researchers can enhance the generalizability and robustness of their findings, leading to more effective policy and practice interventions.

2.109 To Scale or Not to Scale: The Principles of Dose Extrapolation (Sharma & McNeill, 2009)

Vijay Sharma and John H. McNeill (2009) discuss the principles and challenges of dose extrapolation across species, emphasizing the need for accurate methods to ensure the safety and efficacy of drug dosages in different species, including humans.

Main Problems in Extrapolation / External Validity

The authors identify several issues with dose extrapolation, primarily:

- **Species Differences**: Variations in drug metabolism, protein binding, and physiological processes between species can lead to incorrect dose extrapolation.
- **Allometric Scaling Limitations**: Simple allometric scaling based on body weight often fails to account for these differences, leading to potential toxicity or inefficacy.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Sharma and McNeill propose several methods:

- **Allometric Scaling with Adjustments**: Adjusting for species-specific pharmacokinetic parameters like clearance and volume of distribution.
- **Correction for Protein Binding**: Accounting for differences in drug-protein binding across species.
- **Physiologically Based Pharmacokinetic (PBPK) Modeling**: Using detailed physiological data to create models that predict how drugs behave in different species.
- **In Vitro-In Vivo Extrapolation (IVIVE)**: Combining in vitro data with in vivo studies to improve predictions of human pharmacokinetics.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key factors include:

- **Metabolic Rate and Body Size**: Larger animals generally have slower metabolic rates, affecting drug absorption, distribution, metabolism, and excretion.
- **Species-Specific Differences**: Differences in liver enzyme activity, renal excretion, and biliary excretion can impact drug clearance and necessitate adjustments in dosing.

Examples of Problematic Extrapolation / External Validity

The paper highlights several cases where improper dose extrapolation led to adverse outcomes:

- **Tusko the Elephant**: A dose of LSD that was effective in cats caused fatal status epilepticus in an elephant due to differences in pharmacokinetics and sensitivity.
- **Resveratrol Media Misinterpretation**: Incorrectly scaled doses from mice to humans led to unrealistic expectations and skepticism about its efficacy.

Examples of Successful Extrapolation / External Validity

Sharma and McNeill emphasize that successful extrapolation requires thorough understanding and adjustment for species-specific factors. They note that antibiotics, which are less metabolized and primarily excreted renally, are more amenable to allometric scaling.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Understanding the specific context of the new city, including socio-economic and demographic factors, and adjusting the program based on local conditions can enhance its success. Pilot studies and continuous monitoring are crucial for making necessary adjustments and ensuring the program's efficacy.

Sharma and McNeill argue for a nuanced approach to dose extrapolation that accounts for species-specific differences and physiological parameters. By using advanced modeling techniques and combining in vitro and in vivo data, researchers can improve the accuracy and safety of drug dosing across different species, ensuring better generalizability and applicability of their findings.

2.110 Toward Causality and Improving External Validity (Bühlmann, 2020)

Peter Bühlmann (2020) discusses the challenges and methodologies for improving external validity in causal inference, with a focus on observational data and genetic studies. He emphasizes the importance of robust causal structures for achieving reliable and generalizable findings.

Main Problems in Extrapolation / External Validity

Bühlmann identifies several key issues with external validity:

- **Directionality and Causality**: Establishing whether a variable is a cause or an effect is often complicated by hidden confounding variables. Observational data alone cannot provide answers to directionality without additional assumptions or data from experimental designs.
- **Replicability and Stability**: Achieving results that are stable and replicable across different subpopulations and conditions is challenging. Many findings fail to generalize due to variations in population and environmental factors.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Bühlmann suggests several approaches:

- **Robust Causal Inference Methods**: Utilizing methods that target causal relations, which inherently improve external validity due to their stability and invariance across different conditions.

- **Perturbation Data**: Using data from various perturbations of the system, such as randomized control trials (RCTs) or non-randomized specific interventions, to enhance causal inference. Perturbation data help identify invariances and robust causal structures that generalize better.
- **Finite-Sample Guarantees**: Implementing statistical techniques that provide guarantees on the false discovery rate, ensuring that findings are not merely artifacts of specific samples.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Bühlmann highlights the need for understanding specific contextual features that can influence external validity. These include demographic differences, genetic variability, and environmental factors. By incorporating these features into causal models, researchers can improve the generalizability of their findings.

Examples of Problematic Extrapolation / External Validity

The article discusses general issues related to the extrapolation of findings from genetic studies, such as the complexity of inferring causality from single-nucleotide polymorphisms (SNPs) to phenotypes without considering hidden confounders. These challenges are exemplified by the difficulty in achieving stable and replicable findings across different populations.

Examples of Successful Extrapolation / External Validity

Bühlmann refers to the methodological advances presented by Bates et al. (2020), which use observational data to infer causal relations in genetic studies. Their approach demonstrates how robust causal inference methods can lead to improved external validity by providing stability and invariance in the findings.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed data on the new city's context, using robust causal inference methods, and incorporating perturbation data to test the program under various conditions can help ensure its success. Conducting pilot studies and continuously monitoring outcomes can further enhance the program’s applicability and effectiveness.

Bühlmann emphasizes the importance of robust causal structures for improving external validity. By using advanced causal inference methods and incorporating data from various perturbations, researchers can achieve more stable and generalizable findings, leading to better-informed policy and practice decisions.

2.111 Towards Greater Understanding of Implementation During Systematic Reviews of Complex Healthcare Interventions: The Framework for Implementation Transferability Applicability Reporting (FITAR) (Baxter et al., 2019)

Susan Baxter, Maxine Johnson, Duncan Chambers, Anthea Sutton, Elizabeth Goyder, and Andrew Booth (2019) present the Framework for Implementation Transferability Applicability Reporting (FITAR) to enhance the understanding and reporting of applicability and transferability in systematic reviews of complex healthcare interventions.

Main Problems in Extrapolation / External Validity

The authors identify several key issues:

- **Contextual Variability**: Differences in local contexts, such as patient demographics, organizational structures, and financial systems, can affect the implementation and outcomes of interventions.
- **Heterogeneity in Reporting**: Inconsistent and insufficient reporting of contextual data in primary studies limits the ability to assess the applicability and transferability of findings.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, the authors propose the FITAR framework, which includes 44 items to guide the extraction and reporting of contextual data. Key methods include:

- **Comprehensive Data Extraction**: Using the framework to identify and report detailed contextual information during systematic reviews.
- **Stakeholder Engagement**: Incorporating input from stakeholders to ensure the relevance and comprehensiveness of the framework.
- **Iterative Development**: Refining the framework through multiple iterations based on feedback and application in systematic reviews.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The framework emphasizes several contextual features:

- **Patient Demographics**: Characteristics such as age, socio-economic status, and health needs.

- **Organizational Structures**: Size, type, and historical relationships between organizations.
- **Financial and Commissioning Processes**: Funding sources, budget arrangements, and resource availability.
- **Systems Leadership**: Leadership roles, project champions, and stakeholder support.
- **Service Characteristics**: Location, alignment with other initiatives, and existing standards of care.
- **Workforce Features**: Staff motivation, employment conditions, and training requirements.
- **Intervention Complexity**: Breadth, longevity, and integration level of initiatives.

Examples of Problematic Extrapolation / External Validity

The paper highlights several examples where lack of contextual data hindered the applicability of findings:

- **Variation in Patient Populations**: Differences in patient demographics and conditions affected the outcomes and applicability of interventions.
- **Organizational Differences**: Variations in organizational size and type influenced the success of interventions, complicating the transferability of findings.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not extensively detailed, the authors illustrate the potential of the FITAR framework to enhance the reporting and analysis of contextual factors. By systematically collecting and analyzing contextual data, researchers can improve the generalizability and applicability of their findings.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Collecting detailed contextual data on the new city's demographics, organizational structures, financial systems, and leadership can help ensure the program's success. Using the FITAR framework to systematically analyze and report this data can enhance the program's applicability and effectiveness.

Baxter et al. argue for a comprehensive approach to understanding and reporting contextual factors in systematic reviews. By using the FITAR framework, researchers can improve the generalizability and applicability of their findings, leading to more effective policy and practice implementations.

2.112 Transferability of Lessons From Program Evaluations: Iron Laws, Hiding Hands and the Evidence Ecosystem (Ling, 2024)

Tom Ling (2024) discusses the complexities and methodologies involved in transferring lessons from program evaluations across different contexts. He explores how different forms of evidence and argumentation can support or hinder the generalizability of findings, highlighting the importance of context and implementation.

Main Problems in Extrapolation / External Validity

Ling identifies key issues with external validity, emphasizing the challenges in generalizing findings due to:

- **Causal Density**: The complex and multiple interacting causes in social programs make it difficult to replicate outcomes in different settings.
- **Program Design and Implementation**: Variations in how programs are designed and implemented across different contexts affect the transferability of results.

Methods for Fixing Problems of Extrapolation / External Validity

To address these challenges, Ling proposes:

- **Mixed Methods Approach**: Combining quantitative and qualitative evidence to provide a richer understanding of how programs work and under what conditions they succeed.
- **Contextual Analysis**: Detailed descriptions of program contexts and implementation processes to better understand the conditions under which programs are effective.
- **Iterative Learning**: Viewing the transfer of lessons as part of an ongoing process of learning and adaptation within the scientific and policy communities.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Ling emphasizes the importance of understanding several contextual features:

- **Causal Pathways**: The complexity and variability of causal pathways in different social settings.
- **Implementation Capacity**: The capacity of organizations to implement programs effectively, including staff skills and organizational structures.

- **Prior Knowledge and Evidence**: The existing body of evidence and how it shapes the understanding and expectations of program outcomes.

Examples of Problematic Extrapolation / External Validity

The paper discusses the example of Multi-Systemic Therapy (MST), which showed mixed results when transferred from the USA to Europe due to differences in implementation contexts and local professional practices. This highlights the risks of assuming that successful programs in one setting will automatically succeed in another without adaptation.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not detailed, Ling suggests that successful extrapolation requires a nuanced understanding of the interplay between program design, implementation, and local contexts. Programs that account for these factors in their evaluation and adaptation processes are more likely to succeed in new settings.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. Conducting thorough contextual analysis, understanding local implementation capacity, and engaging in iterative learning can help ensure the program's success. Pilot studies and continuous monitoring can further enhance the applicability and effectiveness of the program in a new city.

Ling's framework underscores the importance of a comprehensive, context-sensitive approach to transferring lessons from program evaluations. By integrating mixed methods, focusing on detailed contextual analysis, and embracing iterative learning, researchers and policymakers can improve the generalizability and effectiveness of their interventions across different settings.

2.113 Using Case Studies to Explore the External Validity of 'Complex' Development Interventions (Woolcock, 2013)

Michael Woolcock (2013) delves into the application of case studies to evaluate the external validity of complex development interventions. He argues that the unique contexts and intricate nature of these interventions necessitate a more sophisticated approach than what traditional experimental methods can offer.

Main Problems in Extrapolation / External Validity

Woolcock points out that the main challenge in extrapolating findings from development interventions lies in their inherent complexity and the specific nature of the contexts in which they are implemented. The interplay of numerous variables and the distinct characteristics of local settings complicate the process of generalizing results. Furthermore, variations in how interventions are implemented across different contexts significantly impact outcomes, making it hard to predict success in new environments.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Woolcock advocates for the use of analytic case studies. These studies can reveal the conditions under which interventions succeed or fail by identifying key contextual factors and mechanisms that influence outcomes. He proposes developing a structured framework for external validity, which includes considerations of causal density, implementation capability, and reasoned expectations. This framework helps in understanding the complexity of interventions and predicting their success in varied contexts. Emphasizing continuous learning and adaptation, rather than relying on one-time evaluations, is crucial. This approach allows for ongoing refinement of interventions based on new insights from different settings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Woolcock underscores the importance of understanding specific contextual features that can influence the success of interventions. These include the extent to which multiple variables and their interactions impact outcomes (causal density), the capacity of local organizations and staff to effectively implement interventions (implementation capability), and having realistic expectations about the timeframes and trajectories of change (reasoned expectations).

Examples of Problematic Extrapolation / External Validity

The paper provides examples of interventions that succeeded in one context but failed in another due to differences in local conditions. For instance, education programs involving contract teachers were effective in some regions but not in others, primarily due to variations in local implementation. Similarly, a business registration program in Brazil worked well in urban areas but failed in rural areas because of local distrust of government initiatives.

Examples of Successful Extrapolation / External Validity

Although specific successful examples are not extensively detailed, Woolcock suggests that successful extrapolation involves a thorough understanding of local contexts and careful adaptation. Case studies that identify key factors and mechanisms can guide this adaptation process, making interventions more likely to succeed in new settings.

Transferring an Urban Food Security Program

While the article does not directly address transferring an urban food security program, the principles discussed are applicable. Conducting detailed case studies to understand the new city's context, assessing the capability of local organizations to implement the program, and setting realistic expectations based on local conditions can help ensure the program's success. Continuous monitoring and iterative adaptation will be crucial for refining the program and enhancing its effectiveness.

Woolcock argues for a nuanced, context-sensitive approach to evaluating and generalizing complex development interventions. By using analytic case studies and developing a structured framework for external validity, researchers and policymakers can improve the generalizability and robustness of their findings, leading to more effective and adaptable interventions.

2.114 Validity in Evaluation Research: A Critical Assessment of Current Issues (Chen, 1988)

Huey-Tsyh Chen (1988) critically assesses the major methods of ensuring validity in evaluation research. He argues that the dominance of randomized experiments may not be as powerful as advocates claim, especially given their limitations in practical application and emphasis on internal validity over other types of validity.

Main Problems in Extrapolation / External Validity

Chen identifies the rigid focus on internal validity in randomized experiments as a primary problem, which often comes at the expense of external validity. He points out that the strict control and manipulation required in these experiments can lead to artificial results that are not easily generalizable to real-world settings. Moreover, the reactivity of human subjects and the complex, uncontrolled environments in social research present significant challenges to achieving reliable external validity.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Chen proposes a theory-driven approach that integrates various types of validity. This approach emphasizes incorporating formal models or theory into program evaluation to clearly identify and address potential threats to validity. By considering internal, external, construct, and statistical conclusion validity simultaneously, this method aims to provide a more balanced and comprehensive evaluation framework. Chen also highlights advancements in quasi-experimental designs, such as ARIMA models and selection bias modeling, which have narrowed the gap between randomized experiments and quasi-experiments in terms of internal validity while offering better applicability to real-world settings.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The effectiveness of extrapolation largely depends on understanding the unique characteristics of the new context and target population. These features include demographic variables, socio-economic conditions, organizational structures, and environmental factors that might interact with the treatment. Chen emphasizes the need for detailed contextual analysis and flexibility in adapting interventions to different settings.

Examples of Problematic Extrapolation / External Validity

Chen provides examples of the limitations of applying randomized experimental methods to social programs. One case involves the New Jersey-Pennsylvania Income Maintenance Experiment, where legislative changes and high attrition rates among control groups undermined the reliability of the experiment's findings. Such examples illustrate how external factors can significantly impact the validity of extrapolated results.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not extensively detailed, Chen underscores the potential of quasi-experimental designs and theory-driven approaches in improving external validity. By integrating contextual factors and employing sophisticated modeling techniques, these methods can offer more reliable and generalizable findings.

Transferring an Urban Food Security Program

Although the article does not directly address the transfer of an urban food security program, the principles discussed are relevant. Using a theory-driven approach to incorporate formal models and contextual analysis can help assess the program's applicability in a new city. By understanding the local conditions, demographics, and

socio-economic factors, policymakers can adapt the program to ensure its success. Pilot studies and continuous monitoring would be essential for refining the intervention and enhancing its effectiveness in the new setting.

Chen advocates for a balanced approach to evaluation research that considers multiple types of validity. By integrating theoretical frameworks and contextual analysis, researchers can improve the generalizability and robustness of their findings, leading to more effective and applicable interventions in diverse settings.

2.115 Weighing People Rather Than Food: A Framework for Examining External Validity (Loyka et al., 2020)

Caitlin M. Loyka, John Ruscio, Andrew B. Edelblum, Lindsey Hatch, Brittany Wetreich, and Amanda Zabel (2020) present a framework aimed at enhancing the external validity of research in applied psychological science. Their approach emphasizes the importance of considering populations, settings, outcomes, and timeframes to ensure the generalizability of research findings.

Main Problems in Extrapolation / External Validity

Loyka et al. highlight the trade-off between internal and external validity in research, particularly noting that experimental designs often prioritize internal validity at the expense of external validity. This leads to results that, while methodologically sound, may not generalize well to real-world settings. They point out that the lack of systematic attention to external validity can result in interventions that are effective in controlled environments but fail in practical applications.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, the authors propose a framework grounded in four domains: populations, settings, outcomes, and timeframes. This framework encourages researchers to:

- Clearly define the intended population and ensure the sample is representative.
- Conduct studies in settings that reflect the real-world environment where the intervention will be applied.
- Measure outcomes that are relevant and meaningful in the real world, rather than relying on convenient proxies.
- Consider the appropriate duration for studies to capture long-term effects and sustainability of interventions.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

The framework emphasizes understanding the specific characteristics of the new context and target population. Key factors include demographic variables, the typical settings where interventions will be implemented, the specific outcomes of interest, and the timeframe over which these outcomes will be measured. By tailoring interventions to these factors, researchers can improve the likelihood of successful extrapolation.

Examples of Problematic Extrapolation / External Validity

The authors discuss the limitations of research on mindless eating interventions, which often fail to generalize due to sampling from non-representative populations, conducting studies in artificial settings, and measuring short-term or proxy outcomes. For example, studies that measure food intake in a lab setting do not necessarily translate to weight loss in everyday life.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not extensively detailed, the framework itself is designed to improve the likelihood of successful extrapolation by addressing the key domains systematically. The use of this framework in research design can lead to findings that are more robust and generalizable.

Transferring an Urban Food Security Program

Although the article does not directly address the transfer of an urban food security program, the principles discussed can be applied. Ensuring that the program is tested with representative populations, in real-world urban settings, with meaningful outcomes, and over an adequate timeframe can enhance its success when transferred to a new city. Continuous monitoring and adaptation based on local conditions are crucial.

Loyka et al. advocate for a comprehensive approach to research that balances internal and external validity. By focusing on the generalizability of findings through a structured framework, researchers can design studies that not only demonstrate efficacy in controlled settings but also succeed in practical applications, thereby improving the overall impact of their interventions.

2.116 What Works at Scale? A Framework to Scale Up Workforce Development Programs (Ruder, 2019)

Alexander Ruder (2019) examines the challenges and methodologies for scaling up workforce development programs. He highlights the importance of translating small-scale successes into large-scale implementations that meet regional workforce needs.

Main Problems in Extrapolation / External Validity

Ruder identifies several key issues affecting external validity in workforce development programs. The main problem is that most evaluations are conducted on small, nonrandom samples in specific urban areas, making it difficult to generalize the findings to larger populations or different contexts. Context dependence, randomization bias, and piloting bias are common challenges that limit the scalability of these programs.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Ruder proposes a framework involving two key methods: mechanism mapping and sensitivity analysis. Mechanism mapping helps policymakers systematically identify ways an evidence-based program might fail in a new context by checking the assumptions in the program's theory of change. Sensitivity analysis, commonly used in cost-benefit analysis, assesses how changes in key assumptions affect expected program impacts. By applying these methods, policymakers can better understand and mitigate potential risks when scaling up programs.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features of the new context include demographic and socio-economic characteristics, local labor market conditions, and the capacity of local organizations to implement the program. Understanding these factors is crucial for adapting the program to the new setting and ensuring its success.

Examples of Problematic Extrapolation / External Validity

Ruder discusses various examples where workforce development programs faced challenges in scaling up. For instance, the Quantum Opportunity Program Demonstration showed that site selection bias and local implementation capacity significantly impacted the program's success when transferred to different community-based organizations. Similarly, the Workforce Investment Act Gold

Standard Evaluation highlighted how economic conditions could influence program effectiveness.

Examples of Successful Extrapolation / External Validity

While specific successful examples are not extensively detailed, Ruder illustrates the potential of using mechanism mapping and sensitivity analysis to enhance program scalability. Programs that undergo thorough contextual analysis and adaptation are more likely to succeed in new settings.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed can be applied. Using mechanism mapping to identify potential pitfalls and sensitivity analysis to test assumptions about program outcomes can help ensure the program's success in a new city. Continuous monitoring and adaptation based on local conditions will be essential for refining the program and achieving desired outcomes.

Ruder emphasizes the importance of a systematic approach to scaling up workforce development programs. By employing mechanism mapping and sensitivity analysis, policymakers can better navigate the complexities of translating small-scale successes into large-scale implementations, ultimately leading to more effective and sustainable workforce development initiatives.

2.117 What's (Successful) Extrapolation? (Khosrowi, 2022)

Donal Khosrowi (2022) delves into the complexities and nuances of extrapolating causal effects, particularly in the context of Evidence-Based Policy, development economics, and microeconometrics. He critiques existing methodologies and proposes a more comprehensive framework to understand and achieve successful extrapolation.

Main Problems in Extrapolation / External Validity

Khosrowi identifies the fundamental problem of assuming that causal effects observed in one population (study population) will hold true in another (target population). He discusses how differences in populations, settings, and conditions can lead to significant variations in the effectiveness of interventions. The key issues include the extrapolator's circle, where the resources needed to support extrapolation assumptions might render the original evidence redundant.

Methods for Fixing Problems of Extrapolation / External Validity

To address these issues, Khosrowi proposes a refined framework that includes:

- **Mechanism Mapping**: Identifying and understanding the causal mechanisms and how they function differently across contexts.
- **Sensitivity Analysis**: Testing how changes in key assumptions affect expected outcomes, which helps in understanding the robustness of extrapolations.
- **Comparative Process Tracing**: Using qualitative evidence to draw parallels between different contexts, enhancing the understanding of causal mechanisms.

Features of the New Context and/or Target Population Affecting Extrapolation / External Validity

Key features include demographic and socio-economic characteristics, differences in causal mechanisms, and the presence of moderating variables that can alter the impact of interventions. Understanding these factors is crucial for making accurate extrapolations.

Examples of Problematic Extrapolation / External Validity

The paper highlights the extrapolator's circle, where overly demanding empirical requirements for supporting extrapolation can undermine the relevance of initial causal evidence. For instance, using detailed causal graphs that require extensive knowledge of mechanisms might make the initial experimental data almost irrelevant.

Examples of Successful Extrapolation / External Validity

Khosrowi suggests that successful extrapolation involves balancing the need for additional evidence with maintaining the relevance of initial causal findings. He emphasizes that strategies like mechanism mapping and sensitivity analysis can provide a more nuanced approach to extrapolation, though specific successful examples are not detailed extensively.

Transferring an Urban Food Security Program

Although the article does not directly address transferring an urban food security program, the principles discussed are applicable. By employing mechanism mapping to understand how food security interventions work in different urban settings, and using sensitivity analysis to test key assumptions, policymakers can better adapt and implement these programs in new cities. Continuous monitoring and iterative adaptation are essential for refining the interventions and ensuring their success in varied contexts.

Khosrowi's framework underscores the importance of a comprehensive and nuanced approach to extrapolation. By integrating mechanism mapping, sensitivity analysis, and comparative process tracing, researchers and policymakers can enhance the generalizability and robustness of their findings, leading to more effective and adaptable interventions across different settings.

3. Final Evaluation

The study involved the analysis of a total of 117 papers. Each paper was rated on a scale from 1 to 5 based on its relevance to the topic under investigation. The complete list of papers and their respective ratings is presented in the following table.

Paper	Evaluation
A Conceptual Framework for External Validity (Averitt et al., 2021)	4
A Design-Based Approach to Improve External Validity in Welfare Policy Evaluations (Tipton and Peck, 2017)	5
A Focus on External Validity (Glasgow et al., 2007)	4
A General Algorithm for Deciding Transportability of Experimental Results (Bareinboim and Pearl, 2013)	5
A Literature Review on the Representativeness of Randomized Controlled Trial Samples and Implications for the External Validity of Trial Results (Kennedy-Martin et al., 2015)	3
A New Approach to Argument by Analogy: Extrapolation and Chain Graphs (Steel, 2010)	4
A Note on Campbell's Distinction Between Internal and External Validity (Hammersley, 1991)	3
A Protocol for the Extrapolation of 'Best' Practices: How to Draw Lessons from One Experience to Improve Public Management in Another Situation (Ongaro, 2010)	5
A Review of Generalizability and Transportability (Degtiar and Rose, 2023)	5
A Review of The Project With Special Emphasis on The Monitoring and Information System (Shekar, 1991)	4
A Simple Approximation for Evaluating External Validity Bias (Andrews and Oster, 2019)	4

A Typology of Useful Evidence: Approaches to Increase the Practical Value of Intervention Research (Hasson et al., 2020)	4
Across the Boundaries: Extrapolation in Biology and Social Science (Steel, 2007)	4
Against External Validity (Reiss, 2019)	3
Assessing External Validity (Bo and Galiani, 2021)	5
Assessing Methods for Generalizing Experimental Impact Estimates to Target Populations (Kern et al., 2016)	5
Assessing the Applicability of Public Health Interventions (Burchett et al., 2018)	4
Beyond 'Context Matters': Context and External Validity in Impact Evaluation (Williams, 2020)	4
Beyond External Validity (Calder et al., 1983)	3
Cambridge Handbook of Experimental Political Science (Druckman et al., 2011)	4
Can Understanding Mechanisms Solve the Problem of Extrapolating from Study to Target Populations (Howick et al., 2013)	4
Causal Interaction and External Validity: Obstacles to the Policy Relevance of Randomized Evaluations (Muller, 2015)	4
Causality is Good for Practice: Policy Design and Reverse Engineering (Busetti, 2023)	5
Checklist for the Qualitative Evaluation of Clinical Studies with Particular Focus on External Validity and Model Validity (Bornhöft et al., 2006)	4
Checklists for External Validity: A Systematic Review (Dyrvig et al., 2014)	4
Conceptual Tools for Assessing Experiments: Some Well-Entrenched Confusions Regarding the Internal/External Validity Distinction (Jiménez-Buedo, 2011)	3
Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix (Pritchett and Sandefur, 2014)	5
Cross-Sample Comparisons and External Validity (Krupnikov and Levine, 2014)	4
Designing Multi-Actor Implementation: A Mechanism-Based Approach (Busetti and Dente, 2018)	5
Designing Randomized Controlled Trials with External Validity in Mind (Chassang and Kapon, 2022)	5

Elements of External Validity: Framework, Design, and Analysis	5
Establishing the Internal and External Validity of Experimental Results	4
Estimates of External Validity Bias When Impact Evaluations Select Sites Nonrandomly (Bell et al., 2016)	4
Evidence-Based Policy: A Practical Guide to Doing It Better (Cartwright and Hardie, 2012)	5
Expanding the Framework of Internal and External Validity in Quantitative Research (Onwuegbuzie, 2000)	4
Experimental and Quasi-Experimental Designs for Generalized Causal Inference (Shadish et al., 2002)	5
Experimental Localism and External Validity (Guala, 2003)	4
Experimental Practices and Objectivity in the Social Sciences (Jiménez-Buedo and Russo, 2021)	4
Experimentation in the 21st Century: The Importance of External Validity (Winer, 1999)	5
External Validity (Findley et al., 2021)	5
External Validity and Evaluation Research: A Codification of Problems (Bernstein et al., 1975)	4
External Validity and Meta-Analysis (Slough and Tyson, 2023)	5
External Validity and Model Validity: A Conceptual Approach (Khorsan and Crawford, 2014)	4
External Validity and Policy Adaptation: From Impact Evaluation to Policy Design (Williams, 2020)	5
External Validity and the Research Process: A Commentary (McGrath and Brinberg, 1983)	4
External Validity and Translation from Research to Implementation (Prohaska and Etkin, 2010)	5
External Validity in IS Survey Research (King and He, 2005)	4
External Validity in Policy Evaluations That Choose Sites Purposively (Olsen et al., 2013)	4
External Validity is More Than Skin Deep: Some Answers to Criticisms of Laboratory Experiments (Berkowitz and Donnerstein, 1982)	3
External Validity of a Framed Field Experiment (Lusk et al., 2006)	4
External Validity of Randomised Controlled Trials: To Whom Do the Results of This Trial Apply? (Rothwell, 2005)	5

External Validity, Generalisability, Applicability, and Directness: A Brief Primer (Murad et al., 2018)	4
External Validity, Generalizability, and Knowledge Utilization (Ferguson, 2004)	4
External Validity: From Do-Calculus to Transportability (Pearl and Bareinboim, 2022)	4
External Validity: Is There Still a Problem? (Marcellesi, 2015)	4
External Validity: The Neglected Dimension in Evidence Ranking (Persaud and Mamdani, 2006)	4
External Validity: The Next Step for Systematic Reviews (Avellar et al., 2017)	5
External Validity: We Need to Do More (Glasgow et al., 2006)	5
ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework (Angrist and Fernandez-Val, 2010)	4
Extrapolation of Causal Effects: Hopes, Assumptions, and Challenges (Khosrowi, 2019)	4
Factors That Can Affect the External Validity of Randomised Controlled Trials (Rothwell, 2006)	4
Fidelity and Adaptation of Programs: Does Adaptation Mean a Loss of Fidelity? (Nolt & Leviton, 2023)	3
From Local to Global: External Validity in a Fertility Experiment (Dehejia et al., 2021)	4
Generalization in the Tropics – Development Policy, Randomized Controlled Trials, and External Validity (Peters et al., 2018)	4
Generalizing about Public Health Interventions: A Mixed-Methods Approach to External Validity (Leviton, 2017)	4
Generalizing Causal Knowledge in the Policy Sciences: External Validity as a Task of Both Multi-Attribute Representation and Multi-Attribute Extrapolation (Cook, 2014)	5
Generalizing Treatment Effect Estimates From Sample to Population: A Case Study in the Difficulties of Finding Sufficient Data (Stuart & Rhodes, 2017)	4
How Do We Know When Research From One Setting Can Be Useful in Another? A Review of External Validity, Applicability and Transferability Frameworks (Burchett et al., 2011)	5
How Much Can We Generalize From Impact Evaluations? (Vivalt, 2020)	5

How to Assess the External Validity of Therapeutic Trials: A Conceptual Approach (Dekkers et al., 2010)	4
Interaction of Theory and Practice to Assess External Validity (Leviton & Trujillo, 2017)	4
Internal and External Validity of the Comparative Interrupted Time-Series Design: A Meta-Analysis (Coopersmith et al., 2022)	3
Internal, External, and Ecological Validity in Research Design, Conduct, and Evaluation (Andrade, 2018)	4
Introduction to Special Issue: External Validity and Policy (Westbrook, 2017)	4
Is it Possible to Overcome Issues of External Validity in Preclinical Animal Research? Why Most Animal Models Are Bound to Fail (Pound & Ritskes-Hoitinga, 2018)	2
Learning from Experiments When Context Matters (Pritchett & Sandefur, 2015)	5
Learning from Second-Hand Experience: Methodology for Extrapolation-Oriented Case Research (Barzelay, 2007)	4
On the External Validity of Laboratory Tax Compliance Experiments (Alm et al., 2015)	3
Evidence, External Validity, and Explanatory Relevance (Cartwright, 2011)	5
Policy Evaluation, Randomized Controlled Trials, and External Validity—A Systematic Review (Peters et al., 2016)	4
Predicting the Efficacy of Future Training Programs Using Past Experiences at Other Locations (Hotz et al., 2005)	4
Problems with Using Mechanisms to Solve the Problem of Extrapolation (Howick et al., 2013)	3
Randomised Trials for Policy: A Review of the External Validity of Treatment Effects (Muller, 2014)	4
Regression Discontinuity and Beyond: Options for Studying External Validity in an Internally Valid Design (Wing & Bello-Gomez, 2018)	4
Relabeling Internal and External Validity for Applied Social Scientists (Campbell, 1986)	3
Replication, Experiments, and Knowledge in Public Management Research (Walker et al., 2017)	3
Reverse Engineering and Policy Design (Weaver, 2019)	4

Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education (Bold et al., 2013)	5
Species Extrapolation for the 21st Century (Celander et al., 2011)	2
Target Validity: Bringing Treatment of External Validity in Line with Internal Validity (Lesko et al., 2020)	4
The Concept of External Validity (Calder et al., 1982)	3
The External Validity of Experiments (Bracht & Glass, 1968)	4
The External Validity of Laboratory Experiments: Qualitative Rather Than Quantitative Effects (Kessler & Vesterlund, 2015)	3
The Extrapolation Problem and How Population Modeling Can Help (Forbes et al., 2008)	4
The Extrapolation Problem: How Can We Learn from the Experience of Others? (Bardach, 2004)	4
The Generalizability Puzzle: Rigorous Impact Evaluations and Their Relevance (Bates & Glennerster, 2017)	5
The Importance of External Validity (Steckler & McLeroy, 2008)	4
The Logic of Generalization From Systematic Reviews and Meta-Analyses of Impact Evaluations (Littell, 2024)	4
The Necessity of Construct and External Validity for Generalized Causal Claims (Esterling et al., 2023)	5
The Role of External Validity in Theoretical Research (Lynch, 1983)	3
The Science of Using Science: Towards an Understanding of the Threats to Scaling Experiments (Al-Ubaydli et al., 2019)	4
The Sequential Scale-Up of an Evidence-Based Intervention: A Case Study (Thomas et al., 2018)	4
The Theory-Driven Approach to Validity (Chen & Rossi, 1987)	3
Theorizing How Interventions Work in Evaluation: Process-Tracing Methods and Theorizing Process Theories of Change (Camacho Garland & Beach, 2023)	4
Theory and External Validity (Lynch, 1999)	3
Theory-Testing, Generalization, and the Problem of External Validity (Lucas, 2003)	4
Theory, External Validity, and Experimental Inference: Some Conjectures (Martel Garcia & Wantchekon, 2010)	4
To Scale or Not to Scale: The Principles of Dose Extrapolation (Sharma & McNeill, 2009)	3
Toward Causality and Improving External Validity (Bühlmann, 2020)	5

Towards Greater Understanding of Implementation During Systematic Reviews of Complex Healthcare Interventions: The Framework for Implementation Transferability Applicability Reporting (FITAR) (Baxter et al., 2019)	4
Transferability of Lessons from Program Evaluations: Iron Laws, Hiding Hands and the Evidence Ecosystem (Ling, 2024)	4
Using Case Studies to Explore the External Validity of 'Complex' Development Interventions (Woolcock, 2013)	3
Validity in Evaluation Research: A Critical Assessment of Current Issues (Chen, 1988)	4
Weighing People Rather Than Food: A Framework for Examining External Validity (Loyka et al., 2020)	3
What Works at Scale? A Framework to Scale Up Workforce Development Programs (Ruder, 2019)	4
What's (Successful) Extrapolation? (Khosrowi, 2022)	5

This list was instrumental in identifying the papers that were truly relevant to this project and that subsequently informed the development of Deliverable 3.3 on the literature review. The table below lists all the papers that received a score of 5.

Paper rated 5
A Design-Based Approach to Improve External Validity in Welfare Policy Evaluations (Tipton and Peck, 2017)
A General Algorithm for Deciding Transportability of Experimental Results (Bareinboim and Pearl, 2013)
A Protocol for the Extrapolation of 'Best' Practices: How to Draw Lessons from One Experience to Improve Public Management in Another Situation (Ongaro, 2010)
A Review of Generalizability and Transportability (Degtiar and Rose, 2023)
Assessing External Validity (Bo and Galiani, 2021)
Assessing Methods for Generalizing Experimental Impact Estimates to Target Populations (Kern et al., 2016)
Causality is Good for Practice: Policy Design and Reverse Engineering (Busetti, 2023)

Context Matters for Size: Why External Validity Claims and Development Practice Don’t Mix (Pritchett and Sandefur, 2014)
Designing Multi-Actor Implementation: A Mechanism-Based Approach (Busetti and Dente, 2018)
Designing Randomized Controlled Trials with External Validity in Mind (Chassang and Kapon, 2022)
Elements of External Validity: Framework, Design, and Analysis
Evidence-Based Policy: A Practical Guide to Doing It Better (Cartwright and Hardie, 2012)
Experimental and Quasi-Experimental Designs for Generalized Causal Inference (Shadish et al., 2002)
Experimentation in the 21st Century: The Importance of External Validity (Winer, 1999)
External Validity (Findley et al., 2021)
External Validity and Meta-Analysis (Slough and Tyson, 2023)
External Validity and Policy Adaptation: From Impact Evaluation to Policy Design (Williams, 2020)
External Validity and Translation from Research to Implementation (Prohaska and Etkin, 2010)
External Validity of Randomised Controlled Trials: To Whom Do the Results of This Trial Apply? (Rothwell, 2005)
External Validity: The Next Step for Systematic Reviews (Avellar et al., 2017)