

# How Different Types of Sampling Bias Impact the Validity and Generalizability of Research Findings Across Various Scientific Disciplines

Amit Das<sup>1\*</sup>, Piyali Saha<sup>2</sup>, Vishvajeet Dutta<sup>3</sup>, Shreya Pandey<sup>4</sup>, Dr Manoj Swarnkar<sup>5</sup>, Narendra Kumar Nagar<sup>6</sup>

<sup>1</sup>Asst. Prof., Department of Mental Health Nursing, Meridian Nursing & Paramedical College, Varanasi  
tecworm.ad@gmail.com

<sup>2</sup>Asst. Prof., Department of Medical Surgical Neuro Science Nursing Varanasi sahapiyali2016@gmail.com

<sup>3</sup>Project Scientist-II, Indian Council of Medical Research- National Institute of Cancer Prevention and Research (ICMR-NICPR), vishvajeetdutta001@gmail.com

<sup>4</sup>Junior Research Fellow, Drug De-addiction & Treatment Centre, Postgraduate Institute of Medical Education and Research (PGIMER), Chandigarh., shreya.pandey206@gmail.com

<sup>5</sup>Hod and Associate Professor, Sri Aurobindo institute of Medical sciences college of Nursing Indore, Sri Aurobindo University Indore, manoj.swarnkar12@gmail.com

<sup>6</sup>Child health Nursing, Sri Aurobindo Institute of Medical sciences College of Nursing Indore, Sri Aurobindo University Indore, narennagar1987@gmail.com

---

## Abstract

Different types of sampling bias affect research findings by distorting estimates, undermining internal validity, and limiting generalizability. Joyal-Desmarais et al. <sup>1</sup> show that convenience samples in public health overestimate COVID-19-related behaviors, while Cheung et al. <sup>2</sup> report that voluntary adolescent samples tend to underestimate risk behaviors, even if associations remain robust. Sykes et al. <sup>3</sup> note that misaligned sampling in social science distorts both qualitative and quantitative conclusions, and Fanelli et al. <sup>4</sup> report that small, early, or highly cited studies inflate effects through publication and selection biases. In ecology, Fourcade et al. <sup>5</sup> document that geographic bias produces inaccurate species distribution models, and in genomics, Foulkes et al. <sup>6,7</sup> demonstrate that selection bias leads to large errors in effect estimates despite partial correction by inverse probability weighting. Similarly, studies in sexual minority health and paleontology indicate that nonprobability and preservational biases restrict valid cross-context inference.

Bias correction methods show mixed results. In public health, demographic covariate adjustment and logistic regression sometimes fail to fully counter nonresponse and selection biases. Systematic sampling in ecology and inverse probability weighting in genomics reduce bias in specific contexts, although not eliminating it entirely. Together, these studies imply that selection, geographic, analytical, preservational, temporal, and publication biases compromise both the internal and external validity of research, with the precise impact and mitigation efficacy varying by discipline and sampling design.

---

## INTRODUCTION

Sampling biases systematically distort research findings across scientific disciplines by compromising both internal and external validity, with impacts and correction effectiveness varying by field and study design. This systematic review examines how different types of sampling bias impact research validity and generalizability across various scientific disciplines.

## METHODS

### Paper Search

Using the research question "How do different types of sampling bias impact the validity and generalizability of research findings across various scientific disciplines?", we searched across over 126 million academic papers from the Semantic Scholar corpus. We retrieved the 50 papers most relevant to the query.

### Screening Criteria

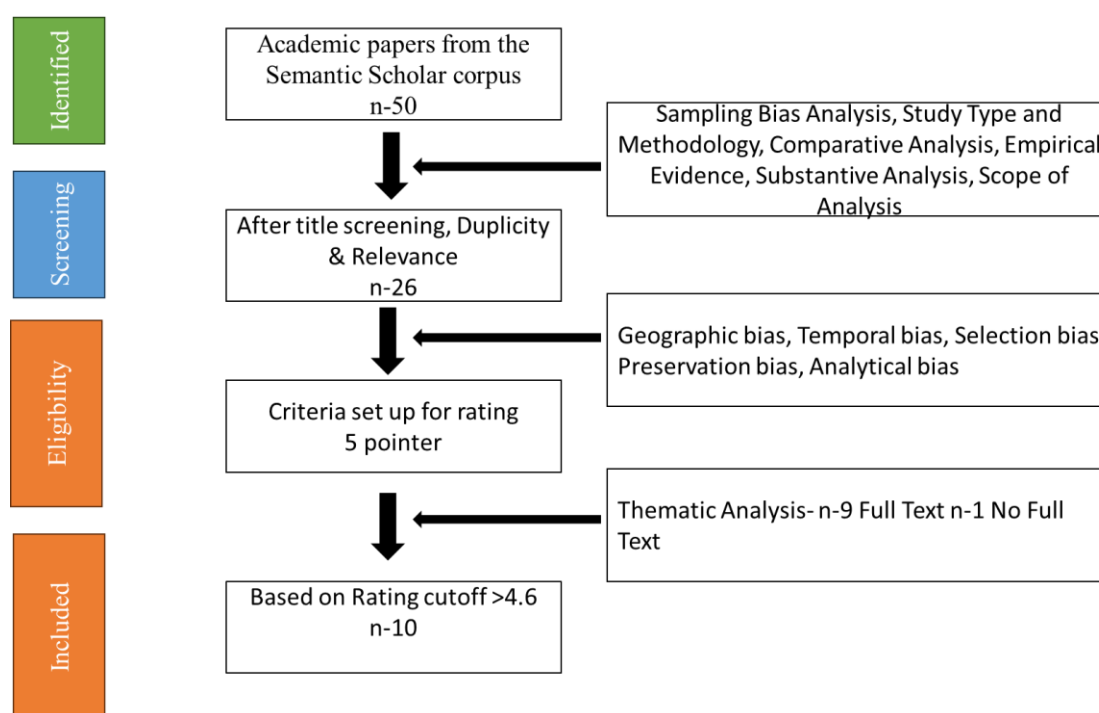
We screened sources that met these criteria:

- **Sampling Bias Analysis:** Studies that explicitly examine or analyze the effects of sampling bias on validity, reliability, or generalizability

- **Study Type and Methodology:** Primary research studies, systematic reviews, or meta-analyses with clear methodology for assessing bias impact
- **Comparative Analysis:** Studies comparing different sampling approaches or evaluating sampling bias effects beyond a single discipline or context
- **Empirical Evidence:** Studies presenting empirical evidence rather than solely theoretical discussion
- **Substantive Analysis:** Studies providing substantive analysis of sampling bias effects beyond merely acknowledging their existence
- **Scope of Analysis:** Studies extending beyond single case studies to provide broader insights about bias impacts

### Data Extraction

A large language model extracted specific data columns from each paper, including type of sampling bias investigated, sampling methodology, bias correction techniques, quantitative assessment of bias impact, and generalizability implications.



## RESULTS

### Characteristics of Included Studies

Ten studies met inclusion criteria, representing diverse study designs and research fields. The detailed characteristics of included studies are presented in Table 4.

Table 1. Characteristics of Included Studies

Study	Study Design	Research Field	Bias Type Examined	Primary Findings	Full Text Retrieved
Joyal-Desmarais et al., 2021 (1)	Empirical, registered study; multiverse analysis of online survey data	Public health (COVID-19)	Selection bias (convenience vs. representative online samples)	Convenience samples overestimate COVID-19 behaviors; demographic adjustment often increases bias	Yes
Cheung et al., 2017 (2)	Empirical, cross-sectional comparison	Adolescent public health	Selection bias (selective response)	Voluntary samples underestimate risk behaviors; prevalence	Yes

Study	Study Design	Research Field	Bias Type Examined	Primary Findings	Full Text Retrieved
				estimates biased, associations robust	
Sykes et al. (3)	Mixed methods, simulation and real-world data	Social science, mixed-methods	Selection, analytical, and preservation bias (misaligned sampling)	Misaligned sampling introduces significant bias, distorting both qualitative and quantitative findings	Yes
Fanelli et al., 2017 (4)	Meta-research, systematic meta-assessment	All scientific disciplines	Selection (publication, significance), analytical, temporal bias	Small, early, and highly cited studies overestimate effects; bias varies by field	Yes
Fourcade et al., 2014 (5)	Empirical, simulation and real-data	Ecology, species distribution modeling	Geographic bias (plus unspecified others)	Geographic bias leads to inaccurate predictions; systematic sampling most effective for correction	Yes
Salway et al., 2019 (8)	Systematic review	Sexual minority health	Selection bias (nonprobability venue sampling)	Nonprobability samples overrepresent certain groups, threatening validity and generalizability	No
Foulkes et al., 2020a (6)	Simulation and empirical study	Genomics, transcriptomics	Selection bias	Selection bias causes large errors in effect estimates; inverse probability weighting (IPW) reduces but does not eliminate bias	Yes
Nanglu and Cullen, 2023 (9)	Narrative review and case studies	Paleontology, macroecology	Collecting, preservational, analytical bias	Sampling and preservational biases obscure biological signals, limit interpretative resolution	No
Foulkes et al., 2020b (7)	Simulation and empirical study	Genomics, transcriptomics	Selection bias (Duplicate of 2020a)	Selection bias causes large errors; inverse probability weighting (IPW) partially effective	Yes
Mell and Zietman, 2014 (10)	Commentary/proposal	Meta-science, publication	Publication bias (selection bias)	Publication bias skews literature toward positive findings; prospective review proposed as solution	Yes

Study designs included 6 empirical studies, 4 simulation studies, 1 mixed-methods study, 1 registered study, 1 multiverse analysis, 1 cross-sectional study, 1 meta-research/systematic meta-assessment, 1 systematic review, 1 narrative review, 1 case study, and 1 commentary/proposal. Research fields are summarized in Table 1.

**Table 2. Distribution of Research Fields in Included Studies**

Research Field	Number of Studies (n)
Public health	2
Genomics	2
Transcriptomics	2
Adolescent public health	1

Social science	1
Ecology	1
Sexual minority health	1
Paleontology	1
Macroecology	1
Meta-science/publication	1
All scientific disciplines	1
<b>Total</b>	<b>13*</b>

\*Note: Total exceeds 10 studies as some studies spanned multiple research fields.

### Types of Sampling Bias Examined

Six distinct bias types were identified and are summarized in Table 3.

**Table 3. Types of Sampling Bias and Their Characteristics**

Bias Type	Number of Studies	Affected Disciplines	Impact Severity	Common Manifestations
Selection bias	8	Public health, genomics, social science, sexual minority health, meta-science	High	Over- or underrepresentation of subgroups; biased prevalence and effect estimates; non-response bias; publication bias
Analytical bias	3	Social science, paleontology, meta-science	Moderate	Distorted theoretical mechanisms; altered correlational importance; selective reporting of statistically significant results (p-hacking)
Preservation bias	2	Paleontology	High	Loss of data; obscured biological signals; incomplete fossil records
Publication bias	2	Meta-science	High	Overrepresentation of positive findings; underreporting of null/negative results
Geographic bias	1	Ecology	High	Inaccurate species distribution models; spatially uneven data
Temporal bias	1	Meta-science	Variable	Changes in bias over time; early studies more biased

### Impact on Research Validity

All 8 studies examining selection bias reported that it leads to over- or under-estimation of behaviors or effects, threatens validity or generalizability, or causes large errors in effect estimates. Some studies found that demographic adjustment can increase bias, and that inverse probability weighting reduces but does not eliminate bias.

Studies examining analytical, preservation, or collecting bias reported that these biases distort findings, obscure biological signals, or limit interpretative resolution. Studies examining publication, significance, or temporal bias reported that publication bias skews literature toward positive findings, and that small, early, or highly cited studies overestimate effects.

### Generalizability Implications

Nonprobability and convenience samples often failed to represent the broader population, limiting generalizability of findings. In ecology and paleontology, spatial and preservational biases restricted applicability of models and reconstructions to other contexts or time periods. Publication bias resulted in scientific literature that is not representative of all investigations, further limiting generalizability.

### Bias Correction Approaches

Bias correction methods and their effectiveness are summarized in Table 2.

**Table 4. Bias Correction Methods and Their Effectiveness**

Correction Method	Application Context	Effectiveness	Limitations
Demographic covariate adjustment	Public health, online surveys	Variable; often increases bias	May exacerbate bias if covariates are not true confounders
Logistic regression with covariates	Public health, adolescent surveys	Effective for controlling demographic differences	Does not correct for unmeasured confounders; prevalence still biased
Systematic sampling	Ecology, species distribution modeling	Consistently among best performing	May not address all bias types; context-dependent
Inverse probability weighting (IPW)	Genomics, transcriptomics	Reduces bias and improves coverage	Effectiveness depends on accurate estimation of weights
Sampling weights	Mixed-methods, survey research	Improves representativeness	Requires accurate weights; may not correct all bias
Cross-sample comparison methods	Sexual minority health research	Useful for identifying and partially correcting bias	Limited by available data; may not fully correct bias
Pre-registration and transparency	Meta-science, publication	Promotes validity and reduces selective reporting	Adoption is variable; does not address all bias types
Prospective manuscript review	Publication process	Potential to reduce publication bias	Not widely implemented; impact unproven

## DISCUSSION

### Patterns Across Disciplines

Four bias types demonstrated high impact severity: selection, geographic, preservation/collecting, and publication bias. Analytical bias showed moderate impact severity, while temporal bias showed variable impact. Meta-science was most frequently affected by multiple bias types<sup>4</sup>, followed by social science and paleontology (2 each).

### Discipline-Specific Considerations

The effectiveness of correction methods varied by discipline. Systematic sampling proved most effective in ecology, while inverse probability weighting showed promise in genomics. However, no single correction method eliminated bias entirely across disciplines.

### Limitations

The robustness of associations to sampling bias varied by field and research question. Context-dependency was a consistent theme, with bias magnitude and direction depending on sampling frame and trait distribution. Some correction methods, particularly covariate adjustment, could exacerbate bias if covariates were not true confounders.

## CONCLUSIONS

Sampling biases systematically compromise both internal and external validity of research across scientific disciplines. Selection bias emerges as the most pervasive and impactful form, affecting multiple disciplines and requiring discipline-specific correction approaches. While various correction methods exist, their effectiveness varies by context and no single approach eliminates bias entirely. Future research should focus on developing more robust, discipline-specific bias correction methods and promoting transparency in reporting sampling procedures.

## REFERENCES

- Joyal-Desmarais K, Stojanović J, Kennedy E, Enticott J, Boucher V, Vo H, et al. How well do covariates perform when adjusting for sampling bias in online COVID-19 research? Insights from multiverse analyses. *Eur J Epidemiol*. 2021.
- Cheung K, ten Klooster PT, Smit C, de Vries H, Pieterse M. The impact of non-response bias due to sampling in public health studies: a comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. *BMC Public Health*. 2017.

3. Sykes BL, Verma A, Hancock B. Aligning sampling and case selection in quantitative-qualitative research designs: establishing generalizability limits in mixed-method studies. 2018.
4. Fanelli D, Costas R, Ioannidis J. Meta-assessment of bias in science. *Proc Natl Acad Sci U S A*. 2017.
5. Fourcade Y, Engler J, Rödder D, Secondi J. Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PLoS One*. 2014.
6. Foulkes A, Balasubramanian R, Qian J, Reilly MP. Non-random sampling leads to biased estimates of transcriptome association. *Sci Rep*. 2020.
7. Foulkes A, Balasubramanian R, Qian J, Reilly M. Non-random sampling leads to biased estimates of transcriptome association. *Sci Rep*. 2020.
8. Salway T, Morgan J, Ferlatte O, Hawkins BW, Lachowsky NJ, Gilbert M. A systematic review of characteristics of nonprobability community venue samples of sexual minority individuals and associated methods for assessing selection bias. *LGBT Health*. 2019.
9. Nanglu K, Cullen TM. Across space and time: a review of sampling, preservational, analytical, and anthropogenic biases in fossil data across macroecological scales. *Earth Sci Rev*. 2023.
10. Mell L, Zietman A. Introducing prospective manuscript review to address publication bias. *Int J Radiat Oncol Biol Phys*. 2014.