NIJAMSS Page: 01-06

NEXUS INTERNATIONAL JOURNAL OF APPLIED SCIENCE, MEDICINE AND SOCIAL SCIENCES (NIJAMSS)

Critical Review
Volume: 01
Issue: 01

Open Access | www.nijamss.org | Vol. 1, Issue 1, May 2026 | Paper No. 001

Article Received: 2nd May 2026 | Published: 31st May 2026 | Pages: 1-06

# BIAS IN EPIDEMIOLOGICAL RESEARCH AND ITS CONSEQUENCES FOR PUBLIC HEALTH DECISIONS: A CRITICAL APPRAISAL OF SELECTION BIAS, INFORMATION BIAS AND CONFOUNDING

**Dr. Daniel A. Otwori, PhD**

[1]School of Public Health, Africa Research University, Zambia

Thematic Area: Epidemiology, Public Health Surveillance and Health Data Science | Manuscript Type: Critical Narrative Review

***Corresponding Author: Dr. Daniel A. Otwori**
**School of Public Health, Africa Research University, Zambia**
**Email: otworidaniel@gmail.com**

## ABSTRACT

Bias is a systematic deviation from the truth that can arise during study design, participant selection, measurement, analysis, interpretation or publication. Because epidemiological findings guide prevention, clinical practice and allocation of public resources, even modest bias can produce large population-level consequences. This critical narrative review examines selection bias, information bias and confounding, with comparative attention to high-income settings and resource-constrained contexts such as Kenya. Evidence was synthesised from foundational epidemiology texts, reporting guidelines, recent methodological scholarship, World Health Organization materials and studies of routine health-information systems. The review shows that bias rarely occurs as an isolated error. Selection mechanisms can alter the distribution of confounders; measurement error can be differential; and analytical adjustment can introduce additional distortion when investigators condition on colliders or use poorly measured covariates. Contemporary causal-inference literature has refined selection bias as the difference between the target-population effect and the effect estimated in the selected sample. In low-resource settings, methodological bias is compounded by structural sources of error, including incomplete diagnosis, under-reporting, weak record linkage, changing denominators and unequal access to care. Kenya has made important progress through digital reporting and surveillance reform, yet fragmented systems, manual delays and uneven data quality remain relevant risks. The paper proposes a bias-control framework covering design, measurement, analysis, reporting and health-system strengthening. It concludes that bias cannot be eliminated completely, but it can be anticipated, diagnosed, quantified and transparently reported through rigorous design, directed acyclic graphs, validation studies, sensitivity analysis, equity-focused reporting and sustained investment in surveillance quality.

**KEYWORDS: Selection bias; information bias; confounding; epidemiology; causal inference; surveillance; Kenya.**

## I. INTRODUCTION

Epidemiological research is central to identifying determinants of disease, estimating intervention effects and informing public-health action. Its value depends on validity: the degree to which a study estimates the quantity it intends to estimate. Bias threatens validity because it produces systematic rather than random error. Increasing sample size narrows confidence intervals but does not correct a systematically distorted estimate. Consequently, a large study can be precise and still be wrong.

Three broad sources of bias dominate introductory and applied epidemiology: selection bias, information bias and confounding. Selection bias arises when the relation between exposure and outcome among those included or retained in a study differs from the relation in the population the study is intended to represent. Information bias arises when exposure, outcome or covariate information is measured or classified inaccurately. Confounding occurs when a common cause of exposure and outcome creates a non-causal association or obscures a causal one. These categories are analytically useful, but in practice they interact across the research process.

Recent methodological work has sharpened the concept of selection bias by distinguishing the target population from the selected analytic sample and by using causal diagrams to identify collider structures and transportability problems. Contemporary guidance also emphasises transparent reporting. The STROBE statement remains the principal reporting framework for cohort, case-control and cross-sectional studies, while the 2025 STROBE-Equity

extension adds explicit attention to avoidable and unjust health differences. These developments are especially relevant where participation, diagnosis and reporting differ by social position.

### A. Problem Statement

Despite widespread recognition of bias, many epidemiological studies continue to report adjusted estimates without adequately explaining the assumptions that make those estimates valid. Variables may be selected solely by statistical significance, loss to follow-up may be treated as harmless, self-reported exposures may be assumed accurate and missing records may be ignored. These practices can generate results that appear methodologically sophisticated while remaining causally ambiguous.

The challenge is intensified in low- and middle-income countries. When access to diagnosis and care is unequal, facility-based datasets can systematically exclude people who are poor, geographically isolated or socially marginalised. Routine surveillance may also be affected by delayed reporting, fragmented systems, changing case definitions and incomplete private-sector participation. A recent review of Kenya's malaria surveillance highlighted fragmented data systems and manual reporting delays as persistent constraints. Thus, methodological bias and health-system weakness can reinforce each other.

### B. Aim, Objectives and Research Questions

The aim of this paper is to critically appraise the principal forms of bias in epidemiological research and to propose an applied framework for reducing their impact on evidence and public-health decision-making.

i. Clarify the conceptual and causal foundations of selection bias, information bias and confounding.

ii. Examine how bias arises across study design, recruitment, measurement, follow-up, analysis and reporting.

iii. Compare the manifestation of bias in high-income and resource-constrained settings, with specific reference to Kenya.

iv. Assess the consequences of biased evidence for public-health priorities, intervention estimates and equity.

v. Propose methodological and system-level strategies for preventing, diagnosing and quantifying bias.

The central question is: How do selection bias, information bias and confounding distort epidemiological estimates, and which design, analytical, reporting and health-system strategies are most effective for limiting their impact?

## II. MATERIALS, METHODS AND TECHNICAL PROCEDURES

### A. Review Design

A critical narrative review design was used to integrate epidemiological theory, recent causal-inference scholarship, reporting guidance and evidence on surveillance-system quality. The purpose was conceptual and applied synthesis rather than statistical pooling.

### B. Sources and Search Concepts

Sources included standard epidemiology texts, peer-reviewed methodological papers, STROBE resources, World Health Organization materials and literature on health-information systems in sub-Saharan Africa and Kenya. Search concepts included selection bias, information bias, misclassification, confounding, collider bias, missing data, loss to follow-up, transportability, surveillance data quality, STROBE and Kenya DHIS2.

### C. Eligibility and Appraisal

Sources were included when they provided a clear definition, mechanism, empirical example, diagnostic method or mitigation strategy. Priority was given to peer-reviewed and official sources with explicit methods. Evidence was appraised for conceptual clarity, relevance, transparency and applicability to public-health settings.

### D. Analytical Framework

The synthesis classified bias by the stage at which it arises: target-population definition, participant selection, measurement, follow-up, causal adjustment, analysis and dissemination. Each bias was examined in relation to direction, likely magnitude, detectability, preventability and consequences for internal validity, external validity and equity.

### E. Limitations

This paper is not a systematic review and does not estimate the prevalence of biased studies. Some recent Kenyan data-quality evidence is based on programme-specific or pre-publication studies and should be interpreted cautiously. The paper therefore emphasises principles and practical safeguards rather than definitive national estimates.

## III. CONCEPTUAL AND TECHNICAL FOUNDATIONS

### A. Selection Bias

Selection bias occurs when inclusion in, retention in or analysis of a study depends on variables related to exposure and outcome, causing the estimated association in the selected sample to differ from the target-population association. Contemporary causal literature distinguishes selection bias created by conditioning on a collider from selection processes that also affect generalisability or transportability. Examples include inappropriate control selection in case-control studies, differential loss to follow-up, healthy-worker effects, volunteer bias, non-response and restricting analysis to people who survive long enough to be enrolled.

### B. Information Bias

Information bias arises from systematic error in measuring exposure, outcome or covariates. Misclassification may be differential when accuracy differs by another study variable, or non-differential when error is unrelated to that variable. Recall bias, interviewer bias, diagnostic suspicion bias and surveillance artefacts are common forms. The familiar claim that non-differential misclassification always biases toward the null is unsafe; the direction depends on the structure of the error, the number of categories and whether exposure or outcome is misclassified.

### C. Confounding

Confounding is a causal-mixing problem rather than a simple statistical association. A confounder is typically a pre-exposure common cause of exposure and outcome that is not on the causal pathway. Control strategies include randomisation, restriction, matching, stratification, standardisation, regression, inverse-probability weighting and g-methods. Adjustment should be guided by subject-matter knowledge and causal diagrams rather than automated variable selection. Controlling for mediators, colliders or consequences of exposure can create rather than remove bias.

**Table 1. Principal forms of epidemiological bias and their likely consequences**

| Bias | Typical mechanism | Illustrative example | Main validity threat | Preferred control |
|---|---|---|---|---|
| Selection bias | Selection or retention depends on exposure, outcome or their causes | Differential loss to follow-up; hospital-control bias | Distorted association and limited transportability | Population-based sampling, high retention, weighting, sensitivity analysis |
| Information bias | Systematic error in measuring exposure, outcome or covariates | Recall bias; diagnostic suspicion; miscoding | Misclassification and incorrect effect size | Validated tools, blinding, standardisation, repeat measurement |
| Confounding | Common cause influences both exposure and outcome | Socioeconomic status in lifestyle-disease studies | Spurious or masked causal association | Randomisation, design restriction, DAG-guided adjustment, weighting |
| Collider bias | Analysis conditions on a common effect of exposure and outcome | Restriction to hospital patients or tested individuals | Creates a non-causal association | Avoid conditioning on colliders; model selection process |
| Reporting bias | Results or outcomes are selectively reported | Publication of significant outcomes only | Distorted evidence base and meta-analysis | Prospective registration, protocols, complete reporting |

## IV. RESULTS AND APPLIED FINDINGS

### A. Bias Is a Process, Not a Single Event

The reviewed evidence indicates that bias is best understood as a process running from the research question to dissemination. A poorly defined target population can produce selection problems before recruitment begins. Differential participation can alter the distribution of confounders. Measurement error can then weaken or reverse associations, and inappropriate adjustment can amplify the distortion. Finally, selective outcome reporting can present only the most favourable version of the findings. Bias assessment should therefore be built into every stage rather than added as a paragraph at the end of a study.

### B. The Importance of Target Populations and Selection Mechanisms

Modern selection-bias theory requires investigators to state the population for which the causal effect is intended. The selected study sample may estimate a different effect if inclusion depends on effect modifiers or on variables related to exposure and outcome. This distinction helps reconcile traditional definitions based on non-representativeness with causal definitions based on conditioning. It also clarifies why representativeness alone is not sufficient: a non-representative sample can still yield an internally valid effect estimate under appropriate assumptions, while a superficially representative sample can be biased by selective missingness.

### C. Measurement Error and Diagnostic Inequality

Measurement quality is rarely uniform across participants. In high-income settings, self-report, wearable devices, electronic records and algorithmic classifications can still produce systematic error. In resource-constrained settings, diagnostic inequality adds a structural dimension: people with better access to facilities are more likely to be tested, diagnosed and recorded. Consequently, observed disease patterns can partly reflect service distribution rather than biological risk. This is especially important when facility data are used to compare counties, informal settlements or rural populations.

### D. Confounding and the Limits of Statistical Adjustment

The historical contrast between observational hormone-therapy findings and later randomised evidence remains a useful illustration of healthy-user confounding. However, the broader lesson is not that observational studies are inherently unreliable. Rather, causal claims require explicit assumptions, adequate measurement of common causes and methods appropriate to time-varying exposures and confounders. Residual confounding can persist after regression when covariates are measured poorly, categorised crudely or omitted.

### E. Bias in Kenya and Comparable Settings

Kenya has expanded routine digital reporting, laboratory networks and public-health surveillance. Nonetheless, data quality varies across programmes and locations. Fragmented systems, manual reporting steps, incomplete interoperability, missing private-sector records and limited local analytic capacity can introduce selection and information bias. A 2025 review of Kenya's malaria-surveillance information systems identified fragmented data systems and reporting delays, while a 2025 multi-county study of DHIS2 use reported continuing concerns about completeness, consistency and practical use. These findings support a systems approach in which methodological training is paired with infrastructure, governance and workforce investment.

**Table 2. Bias risks across the epidemiological research cycle**

| Research stage | Common risk | Diagnostic question | Recommended safeguard |
|---|---|---|---|
| Question and target population | Vague estimand or target group | Whose effect is being estimated, over what time and under what intervention? | Pre-specify estimand, target population and causal contrast |
| Sampling and recruitment | Differential participation | Does inclusion depend on exposure, outcome or their causes? | Probability sampling, recruitment logs, non-response analysis |
| Measurement | Differential accuracy | Are exposure and outcome measured equally across groups? | Blinding, standardised tools, validation subsamples |
| Follow-up | Informative attrition | Are reasons for loss related to exposure and outcome? | Retention plan, inverse-probability weighting, bounds |
| Analysis | Over-adjustment or collider control | Does the adjustment set follow a defensible causal model? | DAG-guided adjustment and sensitivity analysis |
| Reporting | Selective outcomes or models | Were all planned analyses and adverse findings reported? | Registration, protocol publication and reporting checklist |

## V. DISCUSSION AND PUBLIC-HEALTH IMPLICATIONS

### A. Internal Validity, External Validity and Equity

Internal validity concerns whether the estimated association is unbiased for the study population. External validity concerns whether the estimate can be transported to another population. Equity adds a further requirement: investigators should examine whether selection, measurement and missingness differ across socially disadvantaged groups. The STROBE-Equity extension is important because a study can be statistically rigorous overall while still concealing avoidable inequalities in participation, exposure measurement or outcome ascertainment.

### B. Consequences for Policy and Resource Allocation

Biased evidence can exaggerate benefits, obscure harms, misidentify high-risk groups and direct resources away from communities with poor data visibility. Under-counting disease in underserved populations may be interpreted as low need, producing a self-reinforcing cycle of low investment and continued under-detection. Conversely, differential screening can make better-served areas appear to have higher disease risk. Policy interpretation must therefore separate true occurrence from surveillance intensity.

### C. A Practical Bias-Control Framework

**Table 3. Integrated framework for preventing, diagnosing and reporting bias**

| Domain | Minimum practice | Advanced practice | Evidence to report |
|---|---|---|---|
| Design | Clear target population, eligibility criteria and sampling frame | Target-trial emulation and explicit estimand | Flow diagram and recruitment fractions |
| Causal structure | List plausible confounders and mediators | Directed acyclic graph and transport diagram | Rationale for each adjustment variable |
| Measurement | Standardised instruments and staff training | Validation study, repeat measures and error models | Reliability, sensitivity and specificity |
| Missing data | Describe extent and patterns | Multiple imputation, weighting and sensitivity analysis | Assumptions and comparison of complete versus incomplete cases |
| Selection | Track non-response and attrition | Inverse-probability selection weighting and bias bounds | Reasons for exclusion and loss |
| Reporting | Use STROBE checklist | Use STROBE-Equity and share code/protocol | All planned outcomes, models and limitations |
| Health system | Routine data-quality review | Interoperable systems, unique identifiers and independent audits | Completeness, timeliness, concordance and subgroup gaps |

### D. Recommendations

i. **1.** Define the target population, causal estimand and study time scale before selecting data or adjustment variables.

ii. **2.** Use directed acyclic graphs to distinguish confounders, mediators, colliders and selection variables.

iii. **3.** Adopt probability-based or clearly justified sampling and document participation at every stage.

iv. **4.** Use validated measurement instruments, blinded assessment where feasible and validation subsamples for key variables.

v. **5.** Plan for retention, document reasons for loss to follow-up and conduct quantitative sensitivity analyses for informative attrition.

vi. **6.** Report missingness, misclassification and residual confounding transparently rather than treating adjustment as proof that bias has been removed.

vii. **7. Apply STROBE and STROBE-Equity reporting guidance to observational studies.**

viii. **8.** Strengthen Kenya's interoperability, private-sector reporting, laboratory linkage, community surveillance and subnational data-quality audits.

ix. **9.** Build causal-inference and data-quality capacity among researchers, surveillance officers, ethics committees and journal reviewers.

x. **10.** Encourage prospective protocols, code sharing, negative-result publication and independent replication.

### E. Future Research

Future studies should quantify the direction and magnitude of bias rather than merely listing it as a limitation. Kenya-specific priorities include validation of routine diagnoses against laboratory or registry data, evaluation of differential missingness by county and social group, assessment of private-sector under-reporting, and research on algorithmic bias in digital health systems. Simulation studies could also identify when commonly used analytical corrections worsen rather than reduce bias.

## VI. CONCLUSION

Bias is not a minor technical defect; it is a central determinant of whether epidemiological evidence supports or undermines public health. Selection bias, information bias and confounding can occur separately, but they frequently interact. Their effects cannot be corrected by sample size or statistical adjustment alone.

The strongest protection is an integrated approach: explicit causal questions, defensible sampling, accurate measurement, appropriate adjustment, sensitivity analysis, transparent reporting and strong surveillance systems. For Kenya and comparable settings, methodological rigour must be accompanied by investment in diagnostic access, interoperable data systems, reporting completeness and local analytic capacity. Bias cannot be eliminated entirely, but making its mechanisms visible and testing the assumptions behind estimates can substantially improve the credibility, equity and usefulness of epidemiological research.

## VII. DECLARATIONS

### Ethics Approval and Consent to Participate

This review used published literature and publicly available institutional materials and did not involve direct participation by human subjects.

**Funding**

The author received no specific funding for this work.

**Conflict of Interest**

The author declares no conflict of interest.

**Data Availability**

All evidence discussed in this paper is available from the cited sources.

**Author Contribution**

Dr. Daniel A. Otwori conceptualised the paper, conducted the evidence synthesis, developed the analytical framework and prepared the manuscript.

**AI and Digital Tool Disclosure**

Digital tools were used for language refinement, document formatting and reference organisation. The author reviewed the manuscript and accepts responsibility for its accuracy, interpretation and originality.

## VIII. REFERENCES

Bonita, R., Beaglehole, R., & Kjellström, T. (2006). Basic epidemiology (2nd ed.). World Health Organization.

Dewidar, O., et al. (2025). Improving the reporting on health equity in observational studies: STROBE-Equity extension. BMJ, 390, e083882. https://doi.org/10.1136/bmj-2024-083882

Gordis, L. (2014). Epidemiology (5th ed.). Elsevier Saunders.

Griffin, T., et al. (2025). Strengthening health information systems and inherent surveillance: A review of malaria surveillance in Kenya. BMC Public Health.

Lu, H., et al. (2024). The evolution of selection bias in the recent epidemiologic literature. Epidemiology, 35. https://pmc.ncbi.nlm.nih.gov/articles/PMC11879605/

Pearce, N. (2024). Selection bias and other miscellaneous biases. In Encyclopaedia of Epidemiology. NCBI Bookshelf.

Rothman, K. J., Greenland, S., & Lash, T. L. (2021). Modern epidemiology (4th ed.). Wolters Kluwer.

Sylla, B., et al. (2024). Improving health information systems data quality in sub-Saharan Africa: A review. Studies in Health Technology and Informatics.

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement. Annals of Internal Medicine, 147(8), 573–577. https://doi.org/10.7326/0003-4819-147-8-200710160-00010

World Health Organization. (2024). World health statistics 2024: Monitoring health for the SDGs. WHO.

World Health Organization. (2025). Strengthening public health surveillance and data quality in low- and middle-income countries. WHO.

World Health Organization. (2026). Digital public health surveillance architecture and roadmap: Kenya. WHO and Kenya National Public Health Institute.

## IX. COPYRIGHT AND LICENSE