FACHHOCHSCHULE
WIENER NEUSTADT
University of Applied Sciences – Austria

Made Invisible? Measuring Gender Visibility Across WHO GHO and World Bank Health Indicators

Master´s Thesis

Submitted by: **Katrin Hainzl**

at **Master´s Programme**
Health Care Informatics

Supervisor: **Prof. (FH) Dr. Anne Busch**

Wiener Neustadt, date 15.06.2026

FACHHOCHSCHULE
WIENER NEUSTADT
Austrian Network for Higher Education

DECLARATION OF INTEGRITY

I hereby affirm that this research work was written solely by me and that I have not previously submitted this work at another educational institution for the purpose of receiving an academic degree. In particular, contributions by other persons and artificial intelligence in this work have been appropriately cited and the data gathered through the methods described have been accurately reproduced.

Ich erkläre Eides statt, dass diese Arbeit ausschließlich von mir selbst verfasst wurde und ich diese Arbeit nicht zuvor an einer anderen Bildungseinrichtung zum Zwecke der Erlangung eines akademischen Grades vorgelegt habe. Insbesondere wurden Beiträge anderer Personen und künstlicher Intelligenz entsprechend kenntlich gemacht sowie die in dieser Arbeit verwendeten Daten entsprechend der dargestellten Verfahren gewonnen und richtig wiedergegeben.

Wiener Neustadt, 15.06.2026 _____

Date Signature

FACHHOCHSCHULE
WIENER NEUSTADT
Austrian Network for Higher Education

Abstract in English:

Women spend 25% more of their lives in poor health than men [1], yet the gender dimensions of this disparity remain structurally invisible in many of the data systems used to monitor global health. While extensive research documents the underrepresentation of women in primary medical research, the extent to which gender data gaps manifest at the level of aggregated international health statistics has received comparatively little systematic attention.

This study examines the structural visibility of women in two major international health databases — the WHO Global Health Observatory (GHO) and the World Bank Open Data platform — through a quantitative, indicator-based secondary data analysis. A set of ten health indicators, matched across both databases using a reproducible fuzzy string matching procedure, was analysed for the period 2000–2023. Three complementary metrics were applied: a binary Gender Visibility Score (GVS), which assesses whether sex-disaggregated data is structurally present; a Coverage ratio, which measures the completeness of available female and male observations across country–year combinations; and a Missingness Asymmetry Index, which tests whether missing data is distributed symmetrically between the two sexes.

The results reveal a fundamental asymmetry between the two databases. The WHO GHO provides sex-disaggregated data for 5 of the 10 indicators, while the World Bank provides disaggregated data for none through its standard aggregate indicator interface. Where the WHO disaggregates, it does so with near-complete coverage of 91.7% to 100%. No systematic asymmetry between female and male missingness was identified. The gender data gap therefore manifests not as differential treatment of the two sexes within disaggregated indicators, but as a binary architectural question of whether sex-disaggregation is structurally embedded at all.

These findings contribute a replicable methodological framework for auditing sex-disaggregation in international health databases and provide empirical grounding for policy recommendations targeting data architecture standards, institutional accountability mechanisms, and technical interoperability across global health data platforms.

Keywords:

gender data gap, sex-disaggregation, global health indicators, WHO Global Health Observatory, World Bank, gender visibility

Abstract in German:

Frauen verbringen 25 % mehr ihrer Lebenszeit in schlechter Gesundheit als Männer [1] – dennoch bleiben die geschlechtsspezifischen Dimensionen dieser Ungleichheit in vielen der Datensysteme, die zur Beobachtung globaler Gesundheit eingesetzt werden, strukturell unsichtbar. Während zahlreiche Studien die Unterrepräsentation von Frauen in der primären medizinischen Forschung dokumentieren, wurde bislang kaum systematisch untersucht, in welchem Ausmaß geschlechtsbezogene Datenlücken auch auf der Ebene aggregierter internationaler Gesundheitsstatistiken auftreten.

Diese Studie untersucht die strukturelle Sichtbarkeit von Frauen in zwei bedeutenden internationalen Gesundheitsdatenbanken — dem WHO Global Health Observatory (GHO) und der World Bank Open Data-Plattform — mittels einer quantitativen, indikatorbasierten Sekundärdatenanalyse. Ein Set von zehn Gesundheitsindikatoren, das mithilfe eines reproduzierbaren Fuzzy-String-Matching-Verfahrens über beide Datenbanken hinweg abgeglichen wurde, wurde für den Zeitraum 2000–2023 analysiert. Es wurden drei komplementäre Metriken angewendet: ein binärer Gender Visibility Score (GVS), der erfasst, ob geschlechtsdifferenzierte Daten strukturell vorhanden sind; eine Coverage-Rate, die die Vollständigkeit verfügbarer weiblicher und männlicher Beobachtungen über Länder-Jahr-Kombinationen misst; sowie ein Missingness Asymmetry Index, der prüft, ob fehlende Daten symmetrisch zwischen den beiden Geschlechtern verteilt sind.

Die Ergebnisse zeigen eine grundlegende Asymmetrie zwischen den beiden Datenbanken. Das WHO GHO stellt für 5 der 10 Indikatoren geschlechtsdifferenzierte Daten bereit, während die World Bank über ihre standardmäßige Aggregatindikatorschnittstelle für keinen der Indikatoren disaggregierte Daten liefert. Wo die WHO disaggregiert, tut sie dies mit nahezu vollständiger Abdeckung von 91,7 % bis 100 %. Es wurde keine systematische Asymmetrie zwischen weiblichen und männlichen Fehlwerten festgestellt. Die geschlechtsspezifische Datenlücke manifestiert sich damit nicht als unterschiedliche Behandlung der beiden Geschlechter innerhalb disaggregierter Indikatoren, sondern als binäre

architektonische Frage, ob eine Geschlechtsdifferenzierung überhaupt strukturell verankert ist.

Die Befunde leisten einen Beitrag zu einem reproduzierbaren methodischen Rahmen für die Überprüfung von Geschlechtsdisaggregation in internationalen Gesundheitsdatenbanken und bieten eine empirische Grundlage für politische Empfehlungen zu Datenarchitekturstandards, institutionellen Rechenschaftsmechanismen und technischer Interoperabilität globaler Gesundheitsdatenplattformen.

Keywords:

Gender Data Gap, Geschlechtsdisaggregation, globale Gesundheitsindikatoren, WHO Global Health Observatory, Weltbank, Gender Visibility

FACHHOCHSCHULE
WIENER NEUSTADT
Austrian Network for Higher Education

Table of Contents

1.	**Introduction**	**1**
1.1	Background and Problem Statement	1
1.2	Research Gap	5
1.3	Objectives and Contribution of the Study	6
2.	**Theoretical foundations**	**8**
2.1	Sex and Gender as Analytical Categories	8
2.2	Gender Bias & Gender Mainstream	9
2.3	Gender Health Data Gap	12
2.4	Conceptual Framework: From Gender Bias to Data Visibility	14
3.	**State of Research**	**16**
3.1	Underrepresentation of Women in Medical Research	16
3.2	Gender Data Gaps Beyond Clinical Trials	17
3.3	Research Gap and Positioning of This Study	18
4.	**Research Questions & Scope**	**20**
4.1	Research Questions and Scope	20
4.1.1	Territorial and institutional scope	21
4.1.2	Time scope	21
4.1.3	Content scope and delimitation	22
4.2	Analytical Framework and Data Quality Dimensions	22
5.	**Methodology**	**23**

5.1	Research Design and Approach	23
5.2	Data Sources: WHO & World Bank	24
5.3	Indicator Selection	25
5.3.1	Theoretical selection criteria	25
5.3.2	Empirical selection criteria	26
5.3.3	Final Indicator Set	28
5.4	Operationalisation: Coverage, GVS, Asymmetry	30
5.4.1	Gender Visibility Score (GVS)	30
5.4.2	Coverage/Completeness	30
5.4.3	Gender Missingness Asymmetry	31
5.5	Data Processing & Harmonisation	32
5.5.1	Data Retrieval	32
5.5.2	Harmonisation	33
5.5.3	Metric Computation	33
5.5.4	Data Quality and Transparency	34
5.6	Reliability, Validity, and Ethical Considerations	35
5.6.1	Reliability	35
5.6.2	Construct Validity	36
5.6.3	Internal and External Validity	37
5.6.4	Ethical Considerations	37
6.	**Preliminary Results**	**38**
6.1	Gender Visibility across Indicators (SRQ1)	38

6.2 Data Coverage for Sex-Disaggregated Indicators 41

6.3 Gender Missingness Asymmetry 44

6.4 Comparison between WHO and World Bank 45

7. Discussion 47

7.1 Interpretation of Gender Visibility Findings 47

7.2 Patterns of Invisibility across Health Domains 49

7.3 Limitations of the Study 52

7.4 Implications for Global Health Data Policy 55

8. Conclusion & Outlook 59

8.1 Summary of Findings 59

8.2 Scientific Contribution 60

8.3 Outlook and Future Research 61

List of Figures 65

List of Tables 66

Bibliography 67

DECLARATION ON THE USE OF ARTIFICIAL INTELLIGENCE 71

Appendix 72

FACHHOCHSCHULE
WIENER NEUSTADT
Austrian Network for Higher Education

List of Abbreviations

Abbreviation	Full Term
AI	Artificial Intelligence
API	Application Programming Interface
CSV	Comma-Separated Values
FAIR	Findable, Accessible, Interoperable, Reusable
FHIR	Fast Healthcare Interoperability Resources
GHO	Global Health Observatory
GVS	Gender Visibility Score
MAI	Missingness Asymmetry Index
SDG	Sustainable Development Goals
SRQ	Sub-Research Question
UHC	Universal Health Coverage
WB	World Bank
WHO	World Health Organization

1. Introduction

This chapter introduces the study and establishes the context within which the research is situated. Section 1.1 outlines the background and problem statement, tracing the structural roots of gender data gaps in global health and their implications for international health monitoring. Section 1.2 identifies the specific research gap this study addresses. Section 1.3 presents the objectives of the study and its intended contribution to the field.

1.1 Background and Problem Statement

According to the McKinsey Health Institute, women spend 25% more of their lives in poor health than men, even though they live longer. This gender health gap results in a loss of 75 million healthy years of life each year. That equates to approximately seven lost healthy days per woman per year worldwide [1]. These inequalities are not only a medical issue but also reflect systemic biases in data collection and health research. Historically, female participants were often excluded from medical studies, and research data was collected from men and then generalised to women and intersex individuals. There are several reasons for the preference for men in research, including concerns about reduced fertility or impaired pregnancy, the bias of predominantly male researchers, and the perception of men as representatives of the human species and thus as the norm [2, pp. 49-59]. It was incorrectly assumed that data collected from female subjects was too variable due to the female cycle and that this could lead to misinterpretations. In fact, however, it has been proven that men exhibit the same or even greater variability than women for some characteristics due to fluctuations in their testosterone levels [3, pp. 137–146]. The discrepancy between women's participation in studies has devastating consequences. For example, of the 10 prescription drugs that were withdrawn from the market in the United States between 1997 and 2001, 8 were more harmful to women than to men [4, p. 2].

Although scientific awareness of this imbalance has increased, numerous studies show that sex and gender continue to be insufficiently considered in data collection and analysis [3, pp. 137–146]. These historical biases in medical research have shaped today's datasets, which are increasingly used in AI systems and global health statistics. If the data is not adjusted, even more distortions can arise, and social injustice can be exacerbated. Even though

numerous 'principles' for human-centred AI have been formulated, there is a lack of mechanisms with which technologies can put these principles into practice. One particular challenge is deciding when it is appropriate for an algorithm to use gender-specific information and in which cases, such as when posting job advertisements, this is not desirable [3, pp. 137–146].

Long-standing inequalities in health data collection and analysis still shape current research practices. As modern digital health systems and AI tools draw increasingly on historical datasets, the underlying gender biases embedded in these data may be reproduced and amplified. The persistence of these structural biases raises an important question: to what extent are women currently visible in international health data? The aim of this study is therefore to examine the extent to which the gender data health gap is represented in two major international health databases — the World Health Organization (WHO) Global Health Observatory (GHO) and the World Bank (WB) Open Data platform — by systematically analysing the represented gender-specific information and the extent to which women are visible in these datasets. By quantifying the visibility of women in published health indicators, it seeks to contribute to the discussion on gender-sensitive data governance and provide evidence of structural disadvantage.

International health databases such as those maintained by the WHO and the World Bank play a central role in shaping global health agendas, as they translate complex health phenomena into standardized indicators used for monitoring, comparison, and policy decision-making.

International health databases are based on standardised indicators that have been prioritised by the global health community in order to clearly present the health situation and health trends at national and global level. These indicators do not provide a complete picture of complex health realities, but serve as simplified representations of multi-layered health phenomena. Through standardisation, complex issues are reduced to a limited number of measurable variables to enable comparability and controllability.

The selection and definition of these indicators explicitly aims to support monitoring processes at national, regional and global levels. Indicators are used not only to describe health conditions, but also to observe trends, evaluate the achievement of policy objectives and guide investment in health systems and data collection structures. They thus form a central basis for political decision-making processes and international reporting.

At the same time, indicators are based on normative decisions. The prioritisation of certain health indicators follows established criteria such as international relevance, scientific robustness, measurability and existing use in national programmes. This determines which aspects of health are considered reportable and comparable and which are given less consideration. Indicators thus determine not only how health is measured, but also what is visible in global health monitoring.

A key feature of the indicators defined in the Global Reference List is the planned disaggregation of data. Standardised metadata is provided for each indicator, including breakdowns by age, gender and other equity stratifications such as geographical location or socio-economic status. These differentiations are an integral part of the indicator definitions and are not designed as optional additional information. Nevertheless, the intended disaggregation at the conceptual level does not automatically guarantee its consistent implementation in international databases. This results in a possible discrepancy between normative requirements and empirical data representation, which is of central importance for the visibility of certain population groups, especially women [5, pp. 1-18].

The 2030 Agenda enshrines gender equality as a prominent, comprehensive and cross-cutting principle of sustainable development, thereby linking it to the assertion that development is only sustainable if its benefits are enjoyed equally by women and men. Monitoring is central to the implementation of this assertion, as it allows progress, gaps and obstacles to implementation to be made visible and accountability to be demanded. At the same time, UN Women shows that gender-responsive monitoring is currently significantly limited by data problems: only a small proportion of gender-relevant indicators can be reliably monitored at the global level; for many others, methods are available, but country coverage is insufficient, and a remaining proportion still requires conceptual and/or methodological development before these indicators can be used. These findings illustrate that political objectives and indicator frameworks alone are not sufficient to make gender-specific developments empirically visible as long as data availability, quality and comparability are not systematically guaranteed [6, pp. 1-12].

The Data2X initiative emphasises that evidence-based policies promoting gender equality depend significantly on the availability and quality of appropriate gender data. The report ‘Mapping Gender Data Gaps’ maps data gaps in developing and emerging countries across

key empowerment domains (health, education, economic opportunities, political participation and human security) and selects relevant gaps based on need, population coverage and political relevance. At the same time, it highlights that many international data sets would in principle allow for gender-disaggregated data, but that this is often not done; therefore, existing data should first be systematically 'mined' before new data collection is started. Empirically, the report shows a clear imbalance in data production: while around 80% of countries regularly provide gender-disaggregated statistics on mortality, labour market participation and education, for example, less than a third of countries regularly provide data on informal employment, entrepreneurship, violence against women and unpaid work, among other things. This uneven distribution is reflected, among other things, in a minimum list of 52 gender indicators, which are divided into tiers according to conceptual clarity, international standards and regularity of production, thus highlighting the areas in which gender monitoring remains structurally limited by data gaps [7, pp. 1-6].

International health databases should not be understood as neutral repositories of existing information. Rather, they are based on specific selection, aggregation and classification decisions that determine which health phenomena are made visible and which remain unnoticed. The inclusion of an indicator in global reference lists signals political relevance, while missing or incompletely recorded variables are structurally removed from the focus of health policy attention. Visibility in international databases is therefore not synonymous with the mere existence of data, but rather the result of institutional, methodological and normative decisions.

Although global health targets and associated indicators are defined internationally, their actual visibility in data products is limited by the availability and timeliness of the underlying measurement data. The Global Monitoring Report on Financial Protection in Health by the WHO and the World Bank shows that reporting on global progress in financial protection in the context of universal health coverage (UHC) is based on a few clearly defined indicators (including SDG indicator 3.8.2 on catastrophic health spending and indicators on impoverishing health expenditure through direct payments). At the same time, it emphasises that comprehensive monitoring of financial protection requires access to household data on both out-of-pocket expenditure and consumption or income, which is typically generated through rare, costly face-to-face surveys. This results in substantial delays in indicator production; the report cites an average time lag of four years, which is why global estimates are only available up to 2017 at the time of writing. These structural delays in data collection

and production – further exacerbated by pandemic-related disruptions and changes in data collection – illustrate that the mere existence of an indicator as a monitoring standard does not equate to its timely, regular and fully visible representation in global statistics [8, pp. 3–10, 37–45].

While much of the existing research on gender bias in healthcare focuses on primary data collection – such as the underrepresentation of women in clinical trials or gender bias in medical research – health policy governance primarily takes effect at the level of aggregated indicators and international comparative data. Global health databases such as those of the World Health Organisation or the World Bank serve as central reference systems for monitoring, priority setting and political decision-making processes. Distortions in these secondary data therefore have not only analytical but also direct normative and political implications. Nevertheless, the extent to which gender-specific data gaps and asymmetries already exist at the level of international health indicators remains largely unexplored.

1.2 Research Gap

While extensive research documents gender bias in health research and the underrepresentation of women in primary evidence generation—particularly in clinical trials and biomedical studies—the extent to which these biases materialise at the level of aggregated global health statistics remains insufficiently explored. Existing studies largely focus on the production of primary data, whereas international health governance and evidence-based decision-making increasingly rely on secondary data products, such as standardised indicators published by organisations including the World Health Organization and the World Bank.

This represents a critical gap, because global monitoring systems translate complex health phenomena into a limited set of comparable indicators and thereby shape what becomes measurable, visible, and ultimately actionable in policy. Although major indicator frameworks conceptually define sex-disaggregation as an integral component of reporting, it remains unclear to what extent this requirement is consistently implemented in published datasets across indicators, countries, and time.

Moreover, the current evidence base lacks a systematic and reproducible approach to quantify gender visibility at the indicator level. Specifically, there is limited research that

operationalises gender-related data gaps through measurable metrics capturing (1) whether sex-disaggregated values are available, (2) how complete female and male observations are across country–year combinations, and (3) whether missing data occur asymmetrically between genders. As a result, the structural visibility of women in international health databases—and potential gendered patterns of missingness—remain largely unmeasured and therefore under-discussed.

1.3 Objectives and Contribution of the Study

The objective of this study is to assess the structural visibility of women in international health databases at the level of standardised health indicators. Rather than analysing gender differences in health outcomes, the study evaluates how sex is represented in published indicator data and how consistently sex-disaggregated reporting is implemented across databases, indicators, countries, and time.

To achieve this objective, the study applies a quantitative, indicator-based secondary data analysis of harmonised data extracted from the WHO Global Health Observatory and the World Bank data products. The unit of analysis is an aggregated observation defined by Indicator × Country × Year × Gender, enabling a systematic examination of data presence and missingness patterns in sex-disaggregated reporting.

The study makes three main contributions. First, it provides a reproducible operational framework to quantify gender visibility at the indicator level using three complementary metrics: a binary Gender Visibility Score, gender-specific coverage across country–year combinations, and a missingness asymmetry measure to detect potential gendered data gaps. Second, it offers a comparative empirical assessment of how women are represented in two major global health databases, thereby identifying structural differences in reporting practices and data availability. Third, it contributes to the broader discussion on gender-sensitive data governance by making visible the gap between normative indicator requirements (sex-disaggregation as intended in monitoring frameworks) and the empirical reality of published global health statistics. The research questions guiding this study are presented in Chapter 4. At the core of this investigation stands the question of to what extent women are visible in international health databases at the level of standardised health indicators, and how consistently sex-disaggregated reporting is implemented across databases, indicators, countries, and time. Four sub-research questions examine, in turn,

whether sex-disaggregated data exists for a given indicator (gender visibility), how complete this data is across country–year combinations (coverage), whether missing data is distributed asymmetrically between female and male observations (gendered missingness), and how the WHO Global Health Observatory and the World Bank compare in terms of gender data representation.

2. Theoretical foundations

This chapter establishes the conceptual foundations that underpin the study's analytical approach. The theoretical constructs introduced here — sex and gender as analytical categories, gender bias, and the gender health data gap — provide the framework through which the research questions developed in Chapter 4 are grounded in existing scholarship.

2.1 Sex and Gender as Analytical Categories

In order to describe and research the gender health gap, it is first necessary to discuss the term 'gender' and to clearly define and distinguish it from the term 'sex'. Misinterpretations of these terms lead to Data distortion. 'Sex' refers to the biological characteristics that distinguish between male, female and intersex. The term 'gender', on the other hand, refers to socio-cultural attitudes, identities and behaviours that have been structured by society. 'Gender' is multidimensional and overlaps with other social categories such as sex, socio-economic status and age [9, pp. 11-17].

A critical challenge in health data collection is the frequent conflation of sex and gender. As Kaufman et al. [10] emphasize, data intended to capture gender may in practice only reflect sex assigned at birth, and vice versa — leading to misinterpretation of whether observed disparities are biologically or socially driven. This distinction carries direct methodological implications: sex-based differences may point toward biological mechanisms, whereas gender-based differences signal modifiable social determinants that can be addressed through policy interventions. A practical approach to capturing both dimensions is a two-step data collection process, in which sex assigned at birth and self-reported gender identity are recorded separately [10].

The distinction between sex and gender carries direct consequences not only for the design of individual studies, but for the architecture of institutional data systems. In practice, global health databases have predominantly operationalised this dimension through binary sex categories — typically labelled "male" and "female" — derived from administrative records, civil registration systems, or survey instruments that do not systematically distinguish between biological sex and self-reported gender identity. This conflation is not merely a conceptual imprecision; it shapes what is measurable within the data infrastructure and thereby determines the boundaries of any analysis conducted upon it.

The WHO Health Inequality Data Repository, launched in 2023 as one of the most comprehensive gender-equity-oriented data initiatives in global health, illustrates this structural limitation. Despite reporting disaggregated data across more than 1,000 health indicators, the Repository provides only binary sex-disaggregated data and makes no provision for gender identity beyond the male/female categorisation [11]. As Shirzad and Kaufman [11] argue, this design choice renders the repository structurally unable to track health disparities affecting gender-diverse populations — including transgender, non-binary, and gender-nonconforming individuals — whose health outcomes remain largely invisible in aggregated international statistics.

For the present study, this structural reality has a direct methodological implication. Because neither the WHO Global Health Observatory nor the World Bank Open Data platform systematically captures gender identity beyond the binary, the analysis is necessarily limited to assessing the availability of sex-disaggregated data in the male/female sense. The term "sex-disaggregation" is therefore used throughout this thesis to reflect the operationalisation as it actually exists in the databases under examination — not as a normative endorsement of the binary categorisation, but as a precise description of what the available data infrastructure currently supports. This limitation is acknowledged further in Section 5.6.

The binary framing of sex and gender, though predominant in global health databases, fails to capture the full spectrum of gender diversity. Gender-diverse individuals — including transgender, non-binary, and gender-nonconforming persons — face distinct health disparities that remain largely invisible in aggregated international statistics, as most datasets report exclusively binary sex categories [11]. Shirzad and Kaufman [11] further argue that even the WHO's Health Inequality Data Repository, while reporting data for more than 1,000 indicators, provides only sex-disaggregated data and lacks comprehensive gender-disaggregated data — making it structurally unable to track health disparities for gender-diverse populations.

2.2 Gender Bias & Gender Mainstream

The term gender mainstreaming has developed over the years to counteract gender inequalities that are shaped by social, economic and political structures. Instead of focusing purely on the disadvantages faced by women, this approach tackles the core of the problem and addresses both male and female stereotypes that are detrimental to both genders. The

aim is to systematically incorporate gender perspectives, not just discrimination against women [12, pp. 7-15].

Bias generally refers to the tendency to favour one's own group at the expense of others, which promotes social inequality. Gender bias describes the conscious or unconscious favouring or prejudice of one gender over another. This bias is also evident in much medical data. For example, biological models used to represent systems biology mainly apply to men and are underrepresented for other genders. The reason for this is the gender-specific bias in omics data, which describe the metabolic state of a tissue over time [9, pp. 11-17]. These biases in scientific knowledge production are a core reason for the gender health data gap described below.

The institutional response to gender bias in health research and policy has been shaped by a series of landmark commitments at the international level. The 1995 Beijing Platform for Action, adopted at the Fourth World Conference on Women, established gender mainstreaming as an explicit mandate for United Nations agencies and intergovernmental organisations, requiring that gender perspectives be systematically integrated into all policies, programmes, and institutional processes [13]. Within the global health context, this mandate was operationalised most prominently through the World Health Organization's Gender, Equity and Human Rights programme, which since the early 2000s has sought to embed gender analysis into WHO's normative guidance, technical work, and data systems [12].

Despite these formal commitments, the translation of gender mainstreaming principles into the technical infrastructure of health data production has remained incomplete. The WHO's own Gender Responsive Assessment Scale — a tool developed to evaluate the degree to which WHO documents integrate gender perspectives — distinguishes between gender-blind outputs, which ignore sex and gender entirely, gender-sensitive outputs, which acknowledge gender differences without addressing them, and gender-transformative outputs, which actively address the structural determinants of gender inequity. Applied to the domain of data publication, this taxonomy reveals a persistent gap: the majority of standardised health indicators published in international databases remain gender-blind or at best gender-sensitive at the level of data structure, providing either no sex dimension at all or only aggregate totals that obscure gender-differentiated patterns.

This institutionalised incompleteness is not incidental. Bustreo et al. [14] argue that it reflects a structural feature of international health organisations themselves: despite decades of gender mainstreaming commitments, these institutions remain male-dominated in their norms, decision-making processes, and technical outputs — a dynamic that shapes not only strategic priorities but also the data architectures and reporting standards these institutions produce and maintain. Gender mainstreaming in health data systems therefore cannot be reduced to the addition of a sex variable to existing indicators; it requires a fundamental reorientation of what is measured, for whom, and with what granularity. As the empirical findings of this study demonstrate, this reorientation has not yet been consistently achieved in the two most widely used international health data platforms.

Despite three decades of institutional commitment to gender mainstreaming within the United Nations system and global health organisations, the transformative potential of this approach has not been fully realised in practice. Bustreo et al. [14] argue that inadequate leadership, inconsistent support from member states, and a persistent shift toward technocratic rather than transformative solutions continue to limit the impact of gender mainstreaming initiatives. Crucially, international institutions remain structurally male-dominated in their norms, attitudes, and procedures — a dynamic that shapes not only decision-making processes but also the data infrastructures and monitoring systems these institutions produce and maintain [14]. The authors call for a feminist economic approach to gender mainstreaming in global health, one that explicitly addresses power dynamics and structural inequalities rather than treating gender as a technical add-on.

This structural critique is directly relevant to health data systems. Morgan et al. [15] demonstrate that gender-responsive monitoring and evaluation remains an aspirational goal in many institutional contexts: even where sex-disaggregated data collection is nominally required, implementation is inconsistent, and digital data collection tools frequently fail to account for the unequal technology access of women and other underrepresented groups. Gender mainstreaming in health data therefore cannot be reduced to a checkbox exercise of adding a sex variable to existing indicators. It requires a systematic rethinking of what is measured, for whom, and with what granularity — a challenge that, as this study shows, remains unmet in major international health databases.

2.3 Gender Health Data Gap

Di Lego [16, pp. 1-3] provides a more granular taxonomy of the gender health data gap by distinguishing two interrelated problem types. The Type 1 problem refers to missing or incomplete evidence for diseases that disproportionately impact women, arising from the systematic exclusion of female subjects from clinical trials and the underfunding of female-specific conditions. As a result, women experience significantly higher rates of adverse drug reactions — documented across more than 86 medications approved by the US Food and Drug Administration — due to the fact that most drugs were approved based on trials conducted predominantly or exclusively on men [16, pp. 1-3]. The Type 2 problem describes the dominance of male-centred medical evidence as the interpretive default: symptom patterns, diagnostic criteria, and treatment standards derived from male bodies are generalised to the entire population, rendering women's divergent health experiences systematically invisible. Di Lego [16, pp. 1-3] terms this dynamic "healthcare gender bias" and notes that it leads to delays in diagnosis, misattribution of symptoms, and preventable mortality — for example in cardiovascular disease, where women's presentations are less likely to follow guideline-recommended care pathways. Both problem types are structurally interrelated and mutually reinforcing: the absence of female-centred primary data (Type 1) perpetuates the male default in medical interpretation (Type 2), which in turn reduces the perceived necessity of collecting female-specific data. Critically, neither problem is confined to clinical settings. Di Lego [16, pp. 1-3] argues that these biases extend into population-level summary statistics and aggregate health indicators — precisely the domain examined in this study. Gender bias in primary research can propagate through the data value chain into the aggregated international health statistics that form the evidence base for global health policy.

The imperative for sex-disaggregated data is increasingly recognised in the methodological literature on health research. Peters and Woodward [17] argue that disaggregation by sex or gender represents a minimum requirement for health research, as aggregate analyses systematically obscure differences in disease risk, presentation, diagnosis, treatment, and outcomes. They propose a three-phase roadmap for sex- and gender-disaggregated research: first, the exploration of where sex and gender differences exist (and where they do not); second, the explanation of underlying mechanisms; and third, the translation of

findings into policy and clinical practice. Crucially, they note that a statistically significant result in one sex but not the other does not, by itself, constitute evidence of a sex difference — a methodological pitfall particularly relevant when one group is systematically underrepresented in data reporting [17]. This framework directly informs the analytical approach of the present study, in which the GVS, Coverage, and Asymmetry Index serve as instruments for the first phase: identifying where sex-disaggregated data exist and where structural gaps render one or both sexes statistically invisible.

The propagation of gender bias through multiple levels of the health data system is not merely a theoretical concern — it is empirically documented. Hawkes et al. [18] demonstrate, through 18 months of systematic data collection across more than 200 countries, that the absence of sex-disaggregated data is consistent across all levels of health data production simultaneously: it is absent in peer-reviewed publications, in programme data produced by international organisations, and in national vital registration systems. Crucially, this multi-level absence is not uniform: low-income countries show the lowest rates of sex-disaggregated reporting, while even high-income countries fail to consistently report sex-disaggregated data across all points of the care pathway — from testing through hospitalisation to death [18]. This pattern suggests that the problem is not reducible to a failure at any single stage of data collection. Instead, it reflects a structural undervaluation of sex-disaggregation as a reporting norm across the data system as a whole.

This finding has a direct methodological implication for the present study. Because the gap in sex-disaggregated data operates across multiple levels simultaneously, an analysis conducted solely at the level of primary research — examining clinical trial populations, as in the studies reviewed in Section 3.1 — will only capture part of the problem. The gap at the level of international health databases may persist independently of improvements in primary research, if the aggregation and publication decisions of major data platforms do not embed sex as a structural dimension of their indicator architecture. Conversely, sex-disaggregation may be present in aggregate databases for some indicators while remaining absent from the underlying primary evidence base for others. The three metrics developed in this study — the GVS, Coverage, and Asymmetry Index — are designed to detect the endpoint of this chain: the published indicator in institutional databases. They do not identify the origin of the gap within the data value chain, but they provide a systematic, reproducible measure of whether the gap is present at the level where global health policy draws its evidence.

2.4 Conceptual Framework: From Gender Bias to Data Visibility

The theoretical constructs developed in Sections 2.1 through 2.3 are not merely background context for the empirical analysis — they constitute a layered analytical framework that directly motivates the operationalisation of the three study metrics. Each theoretical layer generates a specific analytical question, and each metric is the answer to that question.

The first layer is the conceptual distinction between sex and gender as analytical categories (Section 2.1). Kaufman et al. [10] argue that conflating sex and gender in health research produces misleading results, because biological and socially constructed differences require distinct analytical treatment. For the present study, this distinction has a structural implication: if sex-disaggregated data does not exist in a database, the distinction cannot be made at all at the data level. The Gender Visibility Score operationalises this first-layer concern directly. It does not ask whether a database correctly distinguishes between sex and gender — a conceptually demanding question — but rather the logically prior question: has the database made any structural provision for distinguishing between female and male populations at all? A GVS of 0 means that this most basic analytical distinction is structurally absent from the published data product.

The second layer is the institutional dimension of gender bias and gender mainstreaming (Section 2.2). Bustreo et al. [14] argue that gender mainstreaming in global health institutions is structurally incomplete: even where sex-disaggregation is nominally required, implementation is inconsistent, and the male body continues to function as the implicit default in institutional norms and data systems. Morgan et al. [15] document this gap empirically in monitoring and evaluation systems: sex-disaggregated data collection may be mandated but not enforced, required but not resourced, and present as a formal requirement but absent in practice. The Coverage metric operationalises this second-layer concern: it asks not whether a database has decided to disaggregate, but whether that decision is consistently implemented across the full observation space. An indicator with GVS = 1 but Coverage of 0.3 represents a database where the decision to disaggregate has nominally been made, but where three quarters of the theoretically possible female observations are missing — a gap between institutional commitment and institutional performance.

The third layer is the structural logic of the gender health data gap and the multi-level propagation of bias through the data value chain (Section 2.3). Di Lego's [16, pp. 1-3]

taxonomy of Type 1 and Type 2 gaps predicts that when data gaps exist, they may affect female and male populations differentially: the systematic exclusion of female subjects and the generalisation from male-centred evidence both increase the probability that female observations are more frequently absent than male ones. Hawkes et al. [18] confirm empirically that this differential is present across multiple levels of the data system simultaneously. The Missingness Asymmetry Index operationalises this third-layer prediction: it tests whether missing data is distributed symmetrically between the two sexes, or whether one sex is systematically less represented. A positive asymmetry index — indicating higher female missingness — would confirm the directional prediction derived from the gender health data gap literature.

The conceptual framework thus moves from the foundational (sex/gender distinction → binary visibility question → GVS) through the institutional (mainstreaming commitment → consistency of implementation → Coverage) to the structural (data value chain bias → directional prediction about missingness → Asymmetry Index). This architecture is not merely descriptive: it generates testable expectations against which the empirical findings in Chapter 6 can be evaluated, and provides the interpretive vocabulary through which those findings are discussed in Chapter 7. The three metrics do not stand alone as technical constructs; each is grounded in a specific strand of the theoretical literature and addresses a distinct dimension of the overarching research question.

3. State of Research

This chapter reviews the existing evidence base relevant to the study's research questions. It situates the present analysis within the broader literature on gender bias in health research and documents the gap that the study addresses.

3.1 Underrepresentation of Women in Medical Research

Numerous studies consistently document that women are underrepresented in medical research, and that this underrepresentation is structural, persistent, and not reducible to women's lack of willingness to participate.

The scale and persistence of the problem across regulatory periods is established by a sequence of studies spanning more than two decades. Geller et al. [19] found that in a sample of 86 randomised controlled trials across nine leading journals, women constituted only 37% of participants, with fewer than 4% of studies acknowledging this imbalance as a limitation. Despite the 1993 NIH Revitalisation Act — which required the inclusion of women in NIH-funded clinical trials proportionate to disease prevalence — a follow-up analysis by the same authors more than two decades later found no meaningful improvement: 15% of non-NIH-funded studies still included fewer than 30% women, and 72% did not include sex in their analyses or explain why [20, pp. 630–635]. Poon et al. [21, pp. 604–616] illustrate the clinical consequences of this pattern with a concrete case: for a gout medication approved in 2009, only 5.3% of trial participants were female, despite women accounting for 30% of gout patients. These findings establish that regulatory intervention alone has been insufficient to close the participation gap.

The structural rather than motivational nature of the gap is confirmed by two complementary lines of evidence. Kannan et al. [22, pp. 703–713] demonstrate through patient portal data that women are not less willing to participate in research than men — the underrepresentation is attributable to biases in trial design and recruitment processes, not to participant behaviour. Feldman et al. [23] extend this conclusion to a quantitative scale: using machine-reading of millions of publications, they calculated sex bias as the difference between the proportion of female trial participants and the proportion of female patients according to global disease prevalence. Across 11 disease categories, women were underrepresented in 7, despite

constituting 49% of study participants overall — most starkly in cardiovascular disease, where women represent 51% of patients but only 37% of research subjects.

More recent systematic analyses confirm that these patterns persist into the 2020s. Daitch et al. [24] found systematic underrepresentation across cardiovascular disease, infectious diseases, and cancer in randomised trials from 2017, and identified a particularly pronounced age effect: older women, who represent a majority of many patient populations, are included at far lower rates than their disease burden warrants, creating a knowledge gap specifically about treatment efficacy and adverse effects in this group. Merone et al. [2, pp. 49-59] examined 2,176 publications and found that the gender gap in medical research continues to manifest not only in participant proportions but in more subtle ways, including the framing of research questions and the absence of sex-specific subgroup analyses. Together, these findings establish that the underrepresentation of women in clinical research is not a historical artefact addressed by prior regulation, but a continuing structural feature of the medical evidence base.

3.2 Gender Data Gaps Beyond Clinical Trials

Beyond clinical trials, gender data gaps manifest at the level of global health monitoring systems. Shirzad and Kaufman [11] examined the WHO's Health Inequality Data Repository — a major global initiative launched in 2023 — and found that despite reporting across more than 1,000 indicators, the repository provides only sex-disaggregated data and makes no provision for gender identity beyond the binary male/female categorisation. The authors argue that this structural limitation renders the repository unable to track health disparities for gender-diverse populations and that, more broadly, treating sex as a proxy for gender obscures the social determinants underlying health inequities.

Gender-responsive monitoring and evaluation frameworks have emerged as a response to these structural limitations. Morgan et al. [15] distinguish between three types of gender data and indicators relevant to health systems: sex- or gender-specific indicators, sex- or gender-disaggregated indicators, and gender power relations and systems indicators. The latter category goes beyond identifying where disparities exist to explain why they occur — capturing structural factors such as inequitable access to resources, gender norms, and decision-making power. Existing global health databases, including those examined in this study, primarily report the first two types at best, leaving the explanatory dimension structurally absent from international health statistics.

The gap in sex-disaggregated data is not evenly distributed across the health data value chain. Feraldi et al. [25] conducted a scoping review of global datasets for eight major health conditions and found sex-disaggregated care cascade data — covering diagnosis, treatment, and disease control — for only three: hypertension, diabetes, and HIV/AIDS. For conditions such as tuberculosis, no global sex-disaggregated cascade data could be identified at all. This finding illustrates that even when aggregate data exist, the sex-specific dimensions of health pathways — from risk exposure through disease prevalence to treatment outcomes — are incompletely captured. Sex differences were found to be significant at multiple points along these pathways: males exhibited higher disease prevalence and death rates across the three analysed conditions, while females demonstrated better progression through the care cascade in a majority of countries [25]. These asymmetries in sex-specific data availability are structurally analogous to the gender visibility gaps examined in this study.

Peters and Woodward [17] further identify two distinct dimensions at which sex differences in risk factors manifest in health data: differences in prevalence (the proportion of each sex exposed to a given risk factor) and differences in the magnitude of the risk factor association (the strength of the link between a risk factor and a disease outcome). Both dimensions can differ systematically between women and men, yet aggregate indicators that do not disaggregate by sex collapse these distinctions into a single population-level figure. This means that not only the presence of a condition but also the mechanisms and severity of its association with underlying risk factors may be misrepresented when sex-disaggregated data are unavailable. For global health databases, this implies that the absence of sex-disaggregation is not merely a reporting gap — it is a structural distortion of the epidemiological information available to policymakers.

3.3 Research Gap and Positioning of This Study

The literature reviewed in this chapter establishes two related but analytically distinct bodies of evidence. The first — documented in Section 3.1 — concerns the underrepresentation of women in the primary production of medical research: clinical trials systematically enrol fewer female subjects than disease prevalence warrants, and this imbalance has persisted across more than three decades of regulatory intervention. The second — reviewed in Section 3.2 — documents gender data gaps at the level of global health monitoring systems: sex-

disaggregated data is absent, structurally inaccessible, or incompletely implemented in major international databases, with direct consequences for the evidence base on which health policy decisions are made.

Between these two bodies of evidence, however, lies a methodological gap that the present study is specifically designed to address. Existing research on gender bias in primary research — the studies reviewed in Section 3.1 — operates at the level of trial populations and answers questions about who is enrolled in research. The studies reviewed in Section 3.2 describe gender data gaps in monitoring systems, but typically do so for specific health conditions [25], specific databases [11], or specific indicator types [15], without a systematic cross-database comparative framework that applies uniform metrics across a matched indicator set.

What is absent from the existing literature is a reproducible, quantitative methodology for assessing gender visibility simultaneously across multiple international databases, using matched indicator pairs and consistent operationalisation. Buvinić et al. [7, pp. 14-20] mapped gender data gaps across multiple domains and found pervasive absences, but their analysis is descriptive rather than metric-based and does not enable comparison of structural disaggregation practices across specific database pairs. Shirzad and Kaufman [11] assess one database in depth. Feraldi et al. [25] examine availability of sex-disaggregated pathway data across conditions, but focus on data existence rather than on the structural architecture of how databases embed or exclude the sex dimension. Morgan et al. [15] provide a framework for gender-responsive monitoring but do not apply it empirically to the indicator architecture of the WHO GHO or the World Bank platform.

This study fills precisely this gap. By operationalising gender visibility through three complementary metrics — the binary GVS, the Coverage ratio, and the Missingness Asymmetry Index — and applying them systematically to matched indicator pairs across the WHO GHO and the World Bank Open Data platform, it provides the first reproducible cross-database comparison of sex-disaggregation practices for a theoretically grounded indicator set. The contribution is both empirical and methodological: empirically, it establishes a documented baseline of gender visibility for two of the most widely used international health data platforms; methodologically, it introduces a framework that can be replicated, extended to additional databases, and applied longitudinally to track changes in sex-disaggregation over time. In doing so, it responds directly to the call by Morgan et al. [15] and Bustreo et al. [14] for systematic, auditable instruments for gender-responsive data

governance — instruments that, prior to this study, did not exist in the form of a standardised, cross-database metric framework.

4. Research Questions & Scope

This chapter formally introduces the research questions guiding the study and defines the analytical scope within which they are investigated.

4.1 Research Questions and Scope

To what extent are women visible in international health databases at the level of standardised health indicators, and how consistently is sex-disaggregated information implemented across databases, indicators, countries and time?

SRQ1 – Gender Visibility

To what extent do international health indicators provide sex-disaggregated data distinguishing between female and male populations?

SRQ2 – Data Coverage

For indicators that include a sex dimension, how complete is the available data coverage for women and men across country–year combinations?

SRQ3 – Gendered Missingness

Are there systematic asymmetries in data completeness between female and male observations, indicating gender-specific data gaps?

SRQ4 – Database Comparison

How does the representation of women differ between the WHO Global Health Observatory and the World Bank Gender Data Portal for comparable indicators?

The scope of this study is defined along three dimensions — institutional, temporal, and content-related — and operationalises the research questions above as follows.

This paper examines gender-related data gaps in global health statistics using standardised health indicators. The unit of analysis is not the individual person, but an aggregated observation at the indicator level, specifically:

Indicator × *Country* × *Year* × *Gender*

Each observation thus describes whether and in what form gender-differentiated data (male/female) are available for a specific health indicator in a given country and year.
The analytical focus is on the availability, completeness and symmetry of gender-disaggregated indicator values, not on the content-related evaluation of the health outcomes themselves. The aim is to systematically record structural patterns of the visibility or invisibility of women in global health data. The OECD Health Statistics were not taken into account in the empirical indicator selection process, as it became apparent that only a very small number of indicators could be consistently mapped with the WHO and World Bank data. In order to ensure the comparability of the analysis and to avoid methodological distortions, the institutional scope of the study was limited to the WHO and World Bank.

4.1.1 Territorial and institutional scope

The analysis is limited to globally available health data from two international databases:

- the World Health Organisation (WHO)
- the World Bank (WB)

The OECD database was systematically excluded during the preliminary methodological analysis, as it became apparent that only a very limited number of indicators are consistently available between the WHO, World Bank and OECD. In order to enable a comparable and methodologically consistent analysis, the scope was therefore deliberately reduced to those databases that show sufficient overlap in indicators and metadata.

4.1.2 Time scope

The analysis covers the period 2000–2023. This time frame was chosen to align with the post-Millennium Development Goals era of global health monitoring, during which international

reporting standards and sex-disaggregation expectations have been increasingly formalised. The period is applied uniformly across all indicators and both databases

4.1.3 Content scope and delimitation

This paper focuses exclusively on:

- the existence of a gender dimension (male/female),
- the coverage of possible country–year combinations,
- and asymmetric missingness between women and men.

The study does not cover:

- causal explanations for data gaps,
- comparisons of health levels in terms of content,
- or evaluations of national health systems.

The study is therefore intended as a methodological and exploratory analysis of the data structure, not as a health economic or epidemiological impact analysis.

4.2 Analytical Framework and Data Quality Dimensions

This analysis is based on the World Health Organisation's Global Reference List of 100 Core Health Indicators as a normative reference framework for the evaluation of international health data. These indicators were developed to create a uniform basis for monitoring key health priorities across countries and over time, and to enable the harmonisation of reporting requirements and data systems. The Global Reference List comprises indicators from four key domains – health status, risk factors, coverage of care and health systems – and is designed to cover the entire spectrum of global health priorities. Standardised metadata is defined for each indicator, including a clear indicator definition, information on numerators and denominators, recommended measurement methods and preferred data sources. An essential component of this metadata is the provision for disaggregation by age, gender and other equity stratifications.

The explicit inclusion of gender as a dimension of disaggregation makes it clear that gender-specific differentiations are understood as a conceptual component of quality-assured global health monitoring. The Global Reference List thus stipulates that a differentiated presentation

of health indicators by gender is part of the intended information base for highlighting health disparities within populations [5, pp. 15-18].

Against this background, the Global Reference List serves as a normative benchmark in this study for analysing the empirical presentation of selected health indicators in international databases. The study addresses the question of the extent to which the conceptually intended disaggregations are actually implemented in the published data representation. Thus, it is not the content of health outcomes that is analysed, but rather the way in which health information is structured and made visible at the indicator level.

5. Methodology

This chapter describes the methodological approach through which the research questions are operationalised. It covers the research design, data sources, indicator selection, and the three metrics — Gender Visibility Score, Coverage, and Missingness Asymmetry — applied to answer SRQ1 through SRQ4.

5.1 Research Design and Approach

This study follows a quantitative, indicator-based secondary data analysis focusing on the structural visibility of women in international health databases. Rather than analysing health outcomes, the study examines how gender is represented, differentiated and documented at the level of standardised health indicators.

The study adopts a secondary data analysis design, which enables the systematic evaluation of data originally collected for other primary purposes [27, pp. 619–626]. This is methodologically appropriate here because the object of investigation is not the health status of populations — for which primary data collection would be required — but the structural properties of how health data are recorded, differentiated, and published in institutional databases. The research question can therefore be answered entirely through the analysis of existing data products, without any need for new data collection.

The analytical unit is a structured observation defined by the combination of Indicator × Country × Year × Gender. This unit allows data presence and missingness to be measured systematically across the full population of theoretically possible observations, for each indicator and each database independently. The approach is quantitative in the sense that all three metrics — the Gender Visibility Score, Coverage, and the Missingness Asymmetry

Index — yield numerical outputs that can be compared across indicators, databases, and time periods.

To ensure methodological transparency and reproducibility, the entire data collection and metric calculation process was implemented in Python. Indicators were retrieved via the WHO GHO API and from bulk exports of the World Bank Open Data platform. All data processing steps — including harmonisation, missing value identification, and metric computation — were automated through dedicated scripts and are traceable to the raw source files. This design minimises the risk of human error in data handling and ensures that results are directly reproducible from the original source data [28, pp. 24634–24642].

5.2 Data Sources: WHO & World Bank

The empirical analysis draws on data from two international health databases: the WHO Global Health Observatory (GHO) and the World Bank Open Data platform. These two sources were selected because they represent the primary institutional repositories for standardised, cross-national health indicators at the global level, and because both are publicly accessible without restriction, ensuring full transparency and reproducibility of the analysis.

The WHO Global Health Observatory is the World Health Organisation's central data repository, providing cross-nationally comparable health statistics compiled from member state reporting, WHO technical programmes, and modelled estimates [29]. It covers a broad range of health domains — including mortality, morbidity, risk factors, and health system performance — and serves as the principal data source for global health monitoring, including the Sustainable Development Goal health targets. The GHO makes its data available via a public REST API in JSON format, which was used in this study to retrieve indicator-level data programmatically.

The World Bank Open Data platform aggregates health-related indicators from multiple international sources, including WHO, UNICEF, and UNAIDS, and provides them as a harmonised, cross-country time-series dataset [30]. Data is accessible both via a public API and as bulk Excel exports. For this study, indicator catalogues were downloaded as bulk exports to serve as the basis for the matching procedure described in Section 5.3.2, while country-level time-series data were retrieved via the API.

Both databases cover the full study period of 2000–2023 for the indicators selected. The time frame was chosen to align with the post-Millennium Development Goals era of global

health monitoring, during which international reporting standards and sex-disaggregation expectations have been increasingly formalised [5, pp. 1-18]. The choice of two databases — rather than a single source — is deliberate: by comparing gender visibility patterns across the GHO and the World Bank platform, the analysis can assess whether structural data gaps are platform-specific or systemic across international health data infrastructure.

It should be noted that the World Bank Open Data platform and the World Bank Gender Data Portal are distinct products. The Gender Data Portal is a thematic interface curated specifically for gender-related indicators and is not used in this study. The analysis draws exclusively on the general World Bank Open Data platform, which contains aggregate health indicators for which the presence or absence of sex-disaggregation can be assessed as an empirical question.

5.3 Indicator Selection

The health indicators analysed in this paper were selected in a multi-stage, systematic selection process that took into account both theoretical and practical data criteria.

5.3.1 Theoretical selection criteria

The selection of indicators for this study was guided by three theoretical principles. First, all indicators were required to be traceable to established international health monitoring frameworks — specifically the WHO Global Reference List of 100 Core Health Indicators and the SDG monitoring framework — as a post-selection validation of their relevance for global health policy. Second, each indicator had to be conceptually applicable to both male and female populations, meaning that sex-disaggregation is not only possible but expected from a monitoring perspective. Indicators that are inherently limited to one biological sex — such as maternal mortality — were explicitly excluded, as they do not allow for a meaningful comparison of gender visibility across both sexes. Third, the selected indicators were required to span multiple health domains, including mortality, morbidity, risk factors, and healthcare outcomes, in order to avoid structural patterns being attributable to a single domain. These three criteria are grounded in established frameworks for indicator-based health research. The requirement for anchoring in international monitoring frameworks follows the logic of the WHO Global Reference List of 100 Core Health Indicators, which defines a minimum set of indicators specifically designed for cross-national comparability and policy relevance [5, pp.

1-18]. The requirement for bilateral applicability — excluding inherently sex-specific indicators — is consistent with the methodological guidance of Peters and Woodward [17], who argue that sex-disaggregation is only analytically meaningful for indicators where both sexes represent a relevant population. The requirement for domain heterogeneity reflects a structural safeguard against domain-specific confounding, as domain clustering in indicator sets can produce results that reflect measurement conventions of a single field rather than systemic patterns [11, pp. 16–17].

Beyond these criteria, the selection was guided by a deliberate analytical logic: the indicator set was designed to produce variation in observable gender visibility patterns. Indicators were therefore selected to span a spectrum — from those where gender disparities are scientifically well-documented and sex-disaggregation is particularly important (Group A), to those where sex-disaggregation is conceptually expected but its empirical implementation remains an open empirical question (Group B). This spectrum-based design allows the analysis to identify not only whether gender data gaps exist, but where in the indicator landscape they tend to cluster. This analytical spectrum is informed by the existing literature on sex-disaggregated health data. Indicators in Group A correspond to health domains for which gender disparities are empirically established and for which the absence of sex-disaggregated reporting has been explicitly flagged as a data gap [11], [17]. Indicators in Group B represent domains where disaggregation is normatively expected — as evidenced by their inclusion in the WHO Global Reference List and SDG monitoring frameworks [5, pp. 1-18] — but where actual disaggregation practice in international databases has not been systematically evaluated.

5.3.2 Empirical selection criteria

At the empirical level, the indicator selection followed a systematic, reproducible matching procedure. As a first step, all available indicators were downloaded in full from both databases: the complete WHO indicator list was retrieved via the WHO Global Health Observatory API in JSON format, and the World Bank indicator catalogue was downloaded as a bulk Excel export from the World Bank Open Data platform. This resulted in a dataset of all indicators available in each database at the time of data collection, without any pre-filtering by topic or relevance. This exhaustive retrieval approach ensures that the candidate

pool is not limited by prior topic assumptions, in line with recommendations for systematic database comparison studies [28, pp. 24634-24642].

In a second step, a Python-based fuzzy string matching algorithm was applied to identify conceptually corresponding indicator pairs across the two databases. For each World Bank indicator, the algorithm calculated a similarity score against all WHO indicators using the token_set_ratio metric of the RapidFuzz library [31]. This metric decomposes both strings into individual tokens, sorts them alphabetically, and computes the Levenshtein similarity on the resulting canonical form — making it robust against differences in word order, which is a common source of mismatch between indicator names across institutional databases. All pairs with a similarity score of 60 or above were retained as candidate matches, resulting in 222 candidate pairs. The threshold of 60 was determined through iterative pilot testing on a random sample of 50 indicator pairs, at which it produced the optimal balance between recall (retaining conceptually relevant pairs) and precision (excluding clearly unrelated matches). This approach follows established practice in automated text-matching for bibliometric and health data linkage tasks, where threshold selection is typically empirically validated rather than theoretically prescribed [28, pp. 24634-24642].

From this candidate list, the final indicator set was selected based on three empirical criteria. First, an indicator had to be identifiable in both databases with a sufficiently high match score to ensure conceptual alignment. Second, the matched WHO and World Bank indicator names had to refer to the same underlying health construct — pairs where the algorithm had matched conceptually unrelated indicators (e.g. suicide mortality matched against malaria mortality) were manually identified and excluded. Third, indicators whose World Bank code explicitly referenced a single sex (e.g. containing .FE. or .MA. in the code) were excluded, as these represent pre-disaggregated variables rather than aggregate indicators for which the presence or absence of sex-disaggregation can be assessed.

It should be noted that for five of the ten selected indicators, the World Bank Open Data platform does provide sex-specific data — but not through disaggregation of the aggregate indicator. Instead, the platform offers separate, pre-disaggregated indicator codes for female and male populations as distinct time series. These are: life expectancy at birth (SP.DYN.LE00.FE.IN / SP.DYN.LE00.MA.IN), suicide mortality rate (SH.STA.SUIC.FE.P5 / SH.STA.SUIC.MA.P5), tobacco use prevalence (SH.PRV.SMOK.FE / SH.PRV.SMOK.MA), under-5 mortality rate (SH.DYN.MORT.FE / SH.DYN.MORT.MA), and alcohol consumption per capita (SH.ALC.PCAP.FE.LI / SH.ALC.PCAP.MA.LI). These codes were excluded from the

analysis by design, as they represent pre-disaggregated variables rather than aggregate indicators for which sex-disaggregation can be assessed as a structural property. The existence of these codes is however documented and discussed as a finding in Section 6.4.

Together, these three criteria ensure that the final indicator set is methodologically coherent and analytically interpretable. The combination of algorithmic pre-selection with manual validation reflects best practice in systematic data linkage procedures, where automated methods are used for efficiency but human review remains necessary for conceptual accuracy [27, pp. 619–626].

Match scores for all retained indicators are reported in the final indicator table (Section 5.3.3) as a transparency measure. Where the WHO match is imperfect or refers to a broader concept, this is noted as a methodological limitation and discussed further in Section 5.6.

5.3.3 Final Indicator Set

Based on the theoretical and empirical criteria described above, a final set of 10 health indicators was defined. The indicators are grouped into two analytical categories that reflect their expected position on the gender visibility spectrum.

Group A comprises indicators for which gender disparities are scientifically well-documented in the literature. This grouping builds on the empirical record established in the literature: Feraldi et al. [25] document systematic sex disaggregation gaps for HIV and diabetes specifically, while Shirzad and Kaufman [11] identify life expectancy, suicide, and HIV as indicators for which sex-disaggregated reporting is particularly critical for health equity monitoring. These studies confirm that the indicators assigned to Group A represent domains where the absence of disaggregation would constitute a documented analytical blind spot. For these indicators, sex-disaggregated reporting is considered particularly important from a public health perspective, and the absence of disaggregation would represent a notable gap relative to existing knowledge.

Group B comprises indicators for which sex-disaggregation is conceptually expected — as both sexes are affected and monitoring frameworks foresee disaggregation — but where the empirical implementation in international databases remains an open question to be answered by this analysis. These indicators form the core of the empirical investigation.

The heterogeneity of the indicator set in terms of health domains, measurement units, and data sources is acknowledged as a structural feature of this analysis. Comparability across indicators is not defined at the level of health content, but at the level of data structure: the

unit of analysis is Indicator × Country × Year × Gender, and the GVS, Coverage, and Asymmetry metrics are applied uniformly across all indicators regardless of their substantive domain.

Nr	Indicator	WB Code	WHO Code	Match Score	Group
1	Life expectancy at birth	SP.DYN.LE00.IN	WHOSIS_000001	90.1	A
2	HIV prevalence (ages 15–49)	SH.DYN.AIDS.ZS	MDG_0000000029	100.0	A
3	Suicide mortality rate	SH.STA.SUIC.P5	SDGSUICIDE	81.3	A
4	Diabetes prevalence	SH.STA.DIAB.ZS	NCD_DIABETES_PREVALENCE_AGESTD	100.0	B
5	Tuberculosis incidence	SH.TBS.INCD	MDG_0000000020	83.8	B
6	Tuberculosis treatment success rate	SH.TBS.CURE.ZS	TB_c_new_snep_tsr	85.7	B
7	Tobacco use prevalence	SH.PRV.SMOK	M_Est_cig_curr_std	91.5	B
8	Under-5 mortality rate	SH.DYN.MORT	MDG_0000000007	82.4	B
9	Neonatal mortality rate	SH.DYN.NMRT	WHOSIS_000003	97.9	B
10	Alcohol consumption per capita	SH.ALC.PCAP.LI	SA_0000001760	79.6	B

Table 1 | Final indicator set with database codes, match scores, and analytical grouping. Source: WHO Global Health Observatory and World Bank Open Data. Own calculation.

It should be noted that for indicators 2 and 4, the automated matching procedure initially returned a generic population code (RS_1845). The correct disease-specific WHO codes were identified through manual validation and are reported in Table 1. This reflects a limitation of the automated matching procedure and does not affect the World Bank side of the analysis. This limitation is further discussed in Section 5.6.

5.4 Operationalisation: Coverage, GVS, Asymmetry

Three complementary metrics are used for the quantitative analysis of gender-related data gaps. These metrics capture different dimensions of visibility and data availability and are applied sequentially.

5.4.1 Gender Visibility Score (GVS)

The Gender Visibility Score measures whether a health indicator is reported in a sex-disaggregated manner in the accessible raw data of a given database. It captures the most fundamental dimension of gender visibility: whether the data infrastructure makes a distinction between female and male observations at all.

The GVS is defined as a binary variable at the level of each indicator–database combination. A score of 1 is assigned if an explicit sex dimension is present in the data and at least one non-missing numerical value exists for both female and male categories. A score of 0 is assigned in all other cases — including where no sex dimension exists in the data structure, where only aggregate totals are reported, or where sex categories are present structurally but contain no numerical values. The binary construction is deliberately conservative: partial disaggregation, such as the presence of sex categories without populated values, does not qualify as gender visibility.

Indicators with a GVS of 0 are not analysed further with respect to Coverage or Asymmetry, as no sex-disaggregated observations exist on which these metrics could be computed. The GVS thus functions as a gate-keeping criterion that determines which indicators enter the subsequent stages of analysis.

5.4.2 Coverage/Completeness

For all indicators with GVS = 1, Coverage measures the completeness of sex-disaggregated data across the theoretically possible observation space. The possible observation space is defined as the number of distinct countries in the database multiplied by the number of years in the study period (2000–2023, corresponding to 24 years). Coverage is calculated separately for female and male observations:

$$Coverage_{female} = \frac{Number\ of\ female\ values\ available}{Number\ of\ possible\ country - year\ combinations}$$

$$Coverage_{male} = \frac{Number\ of\ male\ values\ available}{Number\ of\ possible\ country - year\ combinations}$$

The result is a proportion between 0 and 1, where 1 indicates complete data availability and 0 indicates total absence. As an example: if an indicator covers 75 countries over 24 years, the possible observation count is 1,800. If 1,200 female values are available, $Coverage_{female}$= 1,200 / 1,800 = 0.67, indicating that 67% of theoretically possible female observations are present. Coverage is reported separately per sex to allow direct comparison of data completeness between female and male observations for the same indicator.

5.4.3 Gender Missingness Asymmetry

The Missingness Asymmetry Index captures whether missing data is distributed symmetrically between female and male observations, or whether one sex is systematically less represented than the other. It is calculated only for indicators with GVS = 1, since it requires sex-disaggregated observations to be present.
Missingness for each sex is defined as the complement of Coverage:

$$Missingness_{female} = 1 - Coverage_{female}$$

$$Missingness_{male} = 1 - Coverage_{male}$$

The Asymmetry Index is then calculated as:

$$AsymetryIndex = MissingRate_{female} - MissingRate_{male}$$

A positive value indicates that female observations are missing at a higher rate than male observations; a negative value indicates the reverse; a value of zero indicates symmetric data availability. The metric is intentionally descriptive: it identifies the presence and direction of asymmetric missingness without making assumptions about its causes. A finding of asymmetry does not by itself imply intentional exclusion — it may reflect data collection practices, reporting conventions, or structural features of the underlying data sources — but it does indicate that the two sexes are not equally represented in the available data.

The three metrics are applied in strict sequence: GVS first, then Coverage, then Asymmetry. This ensures that Coverage and Asymmetry are only computed where the data structure supports meaningful calculation, and prevents misinterpretation arising from structurally incomplete indicators.

5.5 Data Processing & Harmonisation

The data collection and processing pipeline was implemented entirely in Python and consists of three sequential scripts: collect_data.py for data retrieval, calculate_metrics.py for metric computation, and calculate_missingness.py for the derivation of the missingness analysis. All scripts operate on a shared file structure, with raw indicator data stored in the indicator_data/ subdirectory and processed outputs written to the results/ subdirectory. This structure ensures full traceability from raw source data to reported results.

5.5.1 Data Retrieval

WHO data was retrieved programmatically via the GHO REST API. For each indicator, a GET request was sent to the endpoint https://ghoapi.azureedge.net/api/{indicator_code}, with a time filter applied to restrict results to the period 2000–2023. The API returns data in JSON format, with each record containing a spatial dimension (country code), a time dimension (year), a sex dimension, and a numeric value. Pagination was handled automatically by following @odata.nextLink references until the full dataset was retrieved. World Bank data was retrieved via the World Bank API using the endpoint https://api.worldbank.org/v2/country/all/indicator/{indicator_code}, with results

paginated and collected in full. For each indicator, the combined WHO and World Bank data was saved as a separate CSV file, resulting in ten indicator-level datasets.

5.5.2 Harmonisation

To enable consistent analysis across both databases, several harmonisation steps were applied. First, only observations with valid ISO 3166-1 alpha-3 country codes (three uppercase letters, no digits) were retained, excluding regional aggregates and non-country entries that both APIs include by default. Second, the sex dimension was standardised to a common vocabulary: WHO sex codes (FMLE, MLE, BTSX, SEX_FMLE, SEX_MLE, SEX_BTSX) were mapped to the unified labels FEMALE, MALE, and TOTAL. World Bank data does not include a sex dimension in the standard aggregate indicator endpoint; all World Bank observations were therefore assigned the value TOTAL, reflecting the absence of disaggregation rather than an explicit sex category. Third, all datasets were filtered to the study period 2000–2023 and restricted to observations at the national level.

5.5.3 Metric Computation

The GVS, Coverage, and Asymmetry Index were computed for each indicator–database combination by calculate_metrics.py. For the GVS, the script checks whether both FEMALE and MALE labels are present in the sex column and whether at least one non-missing numerical value exists for each. Coverage was computed as the ratio of non-missing observations to theoretically possible observations, where the denominator is defined as the number of distinct country codes multiplied by the number of years in the study period (24 years). The Asymmetry Index was derived as missingness_female − missingness_male, where missingness is defined as 1 − Coverage. Results were written to metrics_summary.csv (one row per indicator–database combination) and metrics_detail.csv (the full harmonised observation-level dataset). A separate script, calculate_missingness.py, produced a focused output file metrics_missingness.csv containing missingness and coverage values in both decimal and percentage form, restricted to WHO indicators with GVS = 1.

5.5.4 Data Quality and Transparency

Several data quality issues were identified and documented during the collection process. For two indicators — HIV prevalence and diabetes prevalence — the automated matching procedure returned the generic WHO population code RS_1845 rather than a disease-specific code. For these indicators, data was retrieved by indicator name via the GHO API rather than by code, and the conceptual validity of the match was confirmed manually. Match scores for all indicators are reported in Table 1 (Section 5.3.3) as a transparency measure, and cases where the WHO code refers to a broader concept than the World Bank indicator are noted as methodological limitations and discussed further in Section 5.6. All scripts used for data collection and processing are documented and available for inspection as part of the supplementary materials of this thesis.

The transparency and reproducibility standards applied in this study are further grounded in the FAIR data principles — a framework for research data management that has gained broad institutional recognition since its publication by Wilkinson et al. [32]. FAIR requires that research data and the processes that generate them be Findable, Accessible, Interoperable, and Reusable. While originally developed for scientific datasets, the FAIR framework provides a productive analytical lens for evaluating the structural properties of international health databases — and in particular for understanding the nature of the gender visibility gap identified in this study.

Applied to the databases examined here, the FAIR criteria reveal a distinction that is central to interpreting the World Bank findings. The sex-specific data available through the World Bank's .FE. and .MA. indicator codes is technically present within the platform and is therefore not absent in the Accessible sense — it can be retrieved by a user who knows the specific code variants. However, it fails the Findable criterion: a user accessing the standard aggregate indicator receives no sex dimension and no indication that disaggregated variants exist. Sex-disaggregated data is present in the platform but structurally unfindable through the standard data interface. This distinction between data accessibility and data findability is precisely the kind of structural quality gap that the FAIR framework is designed to surface, and it provides a principled methodological vocabulary for the finding documented in Section 6.4 and discussed further in Section 7.2.

The Interoperability criterion is equally relevant. For sex-disaggregation to be actionable in cross-database analyses — such as the comparison conducted in this study — it must be

embedded as a retrievable dimension within the aggregate indicator rather than externalised into separate variable codes with non-standardised naming conventions. The World Bank's architectural approach, which uses platform-specific .FE./.MA. suffixes rather than a shared metadata standard, limits interoperability with other databases that embed sex as an internal dimension of the indicator structure, such as the WHO GHO. This interoperability gap means that automated or systematic cross-database comparisons of sex-disaggregated data cannot be conducted without platform-specific adaptations — a structural barrier to the kind of reproducible, scalable gender data monitoring that the present study advocates.

5.6 Reliability, Validity, and Ethical Considerations

This section addresses the quality dimensions and ethical grounding of the study. Section 5.6.1 discusses the reliability of the analytical pipeline and the conditions under which results are reproducible. Section 5.6.2 examines construct validity, assessing the degree to which the metrics operationalise the underlying theoretical concepts. Section 5.6.3 addresses internal and external validity, including the generalisability of findings beyond the indicators and databases analysed. Section 5.6.4 outlines the ethical considerations relevant to this secondary data analysis.

5.6.1 Reliability

The reliability of this study rests on the reproducibility of the data collection and metric computation pipeline. Because all processing steps are implemented in Python scripts that operate deterministically on fixed source data, the results can be reproduced in full by re-running the scripts against the same API endpoints and bulk export files. No manual data entry was performed at any stage; all values reported in Chapter 6 are directly traceable to the output files metrics_summary.csv, metrics_detail.csv, and metrics_missingness.csv, which are themselves generated automatically from the raw indicator CSVs [28, pp. 24634-24642].

A potential source of unreliability is temporal: both the WHO GHO API and the World Bank Open Data platform are living databases that are updated and revised over time. The data underlying this analysis was collected at a fixed point in time (May 2026), and results may differ if the same indicators are retrieved at a later date due to retrospective corrections,

methodological revisions, or newly added observations. This limitation is inherent to all secondary analyses of dynamic institutional databases and does not affect the internal consistency of the results as reported.

5.6.2 Construct Validity

Construct validity refers to the degree to which the operationalised metrics actually measure the theoretical concept they are intended to capture [28, pp. 24634-24642]. The central construct in this study is gender visibility — understood as the extent to which international health databases structurally differentiate between female and male populations at the indicator level.

The GVS operationalises this construct as a binary gate-keeping criterion: an indicator is considered gender-visible if and only if non-missing numerical values exist for both female and male observations. This is a conservative operationalisation that deliberately sets a low threshold — it does not require high coverage or symmetrical missingness, only the structural presence of disaggregated data. This conservatism is intentional: the GVS is designed to capture the most fundamental dimension of visibility, not its quality or completeness, which are addressed by the Coverage and Asymmetry metrics respectively.

One construct validity limitation concerns the WHO code mismatch for indicators 2 (HIV prevalence) and 4 (diabetes prevalence), for which the automated matching procedure returned the generic population code RS_1845. For these indicators, WHO data was retrieved by indicator name rather than code, and the conceptual alignment was confirmed through manual inspection of the returned data structure. The risk of retrieving conceptually incorrect data for these two indicators is considered low, but cannot be entirely excluded.

A further construct validity consideration concerns the scope of "gender visibility" as operationalised here. The metrics capture sex-disaggregation in the binary male/female sense — they do not assess whether databases provide data for non-binary gender categories, nor whether reported sex categories accurately reflect gender identity in the populations concerned. This limitation is consistent with the current state of international health data infrastructure, which does not yet systematically support gender identity beyond the binary categorisation [11], but it means that the findings should be interpreted as measuring binary sex disaggregation rather than gender visibility in its fuller sense.

5.6.3 Internal and External Validity

The internal validity of the analysis is supported by the consistent application of identical metrics and processing procedures across all ten indicators and both databases. Because the analytical unit (Indicator × Country × Year × Gender) and the computation logic are defined uniformly, cross-indicator and cross-database comparisons are methodologically valid within the scope of this study.

External validity — the generalisability of findings beyond the ten indicators and two databases examined — is limited by design. The indicator set was deliberately selected to represent a range of health domains and expected gender visibility patterns, but it is not a random sample of all possible health indicators. The findings regarding the WHO GHO and World Bank Open Data platforms are specific to these databases as accessed via their public APIs and bulk exports; they cannot be automatically generalised to other international health databases, institutional data repositories, or national-level data systems. The cross-database comparison is further limited by the fact that WHO and World Bank indicators, while matched by name similarity, may differ in their precise definitions, reference populations, or data sources — a structural heterogeneity that is acknowledged as a feature of the study design rather than a flaw, since the goal is to assess visibility as it actually presents itself in the publicly accessible data products [27, pp. 619–626].

5.6.4 Ethical Considerations

This study does not involve the collection, processing, or analysis of personal data. All data used are publicly available aggregate statistics published by international organisations at the country level. No individual-level data was accessed at any point, and no data subjects are identifiable from the datasets used. The study therefore does not fall within the scope of data protection regulations applicable to personal data, and no ethics committee approval was required.

The use of artificial intelligence tools in the preparation of this thesis is declared in the AI usage statement. ChatGPT and Claude were used for idea development, literature identification, and the structuring of data collection scripts. All analytical decisions, code implementation, and written content were reviewed and verified by the author. No AI-generated text was adopted without critical review and adaptation.

6. Preliminary Results

This chapter presents the empirical findings of the study, addressing SRQ1 through SRQ4 sequentially. All results are based on data collected for the period 2000–2023 - a timeframe chosen to align with the post-MDG era of formalised global health monitoring, as discussed in Section 5.2 - and computed using the metrics defined in Chapter 5.

6.1 Gender Visibility across Indicators (SRQ1)

This section addresses SRQ1 by examining the extent to which the 10 selected health indicators provide sex-disaggregated data in the WHO Global Health Observatory and the World Bank Open Data platform.

The Gender Visibility Score (GVS) was calculated for each indicator and each database separately. A GVS of 1 was assigned where both female and male disaggregated values were present with at least one non-missing numerical observation; a GVS of 0 was assigned in all other cases, as defined in Section 5.4.1. Table 2 presents the results across all 10 indicators.

Indicator	Group	WHO GVS	World Bank GVS
Life expectancy at birth	A	1	0
HIV prevalence (ages 15–49)	A	0	0
Suicide mortality rate	A	1	0
Diabetes prevalence	B	1	0
Tuberculosis incidence	B	0	0
TB treatment success rate	B	0	0
Tobacco use prevalence	B	1	0
Under-5 mortality rate	B	1	0
Neonatal mortality rate	B	0	0

Alcohol consumption per capita	B	0	0

Table 2 | Gender Visibility Score by indicator and database. Source: WHO Global Health Observatory and World Bank Open Data. Own calculation.

This binary operationalisation is consistent with approaches used in comparable assessments of sex-disaggregated data availability in global health databases, where the presence or absence of disaggregation is treated as the foundational criterion before completeness is assessed [11], [15].

The results reveal a fundamental asymmetry between the two databases. WHO provides sex-disaggregated data for 5 of the 10 indicators (GVS = 1), while the World Bank provides no sex-disaggregated data for any of the 10 indicators (GVS = 0 across the board). This finding was verified both through the automated data collection pipeline and through manual inspection of the raw data exports from both platforms.

Among the five indicators with GVS = 0 in the WHO database, three distinct patterns can be observed. For HIV prevalence, tuberculosis incidence, and TB treatment success rate, the GHO API returns no sex dimension at all — these indicators are reported exclusively as aggregate values with no disaggregation structure present. For neonatal mortality rate, a sex dimension exists structurally but contains only "Both sexes" values, with no female or male breakdown available. For alcohol consumption per capita, female and male categories are present in the data structure but contain no numerical values, suggesting the disaggregation is planned but not populated. These three patterns are documented as distinct structural characteristics of the WHO database and are discussed further in Section 7.

A separate structural observation concerns the World Bank platform. While the GVS is 0 for all 10 indicators in the World Bank dataset as accessed via the aggregate indicator codes, the platform does provide sex-specific data for five of the ten indicators through a parallel set of pre-disaggregated indicator codes. Table 3 documents these codes alongside the aggregate codes used in this analysis.

Indicator	Aggregate WB Code (used)	Female Code	Male Code
Life expectancy at birth	SP.DYN.LE00.IN (total, years)	SP.DYN.LE00.FE.IN (female, years)	SP.DYN.LE00.MA.IN (male, years)
Suicide mortality rate	SH.STA.SUIC.P5 (per 100,000 population)	SH.STA.SUIC.FE.P5 (per 100,000 female population)	SH.STA.SUIC.MA.P5 (per 100,000 male population)
Tobacco use prevalence	SH.PRV.SMOK (% of adults)	SH.PRV.SMOK.FE (% of female adults)	SH.PRV.SMOK.MA (% of male adults)
Under-5 mortality rate	SH.DYN.MORT (per 1,000 live births)	SH.DYN.MORT.FE (per 1,000 live births, female)	SH.DYN.MORT.MA (per 1,000 live births, male)
Alcohol consumption per capita	SH.ALC.PCAP.LI (litres, 15+)	SH.ALC.PCAP.FE.LI (litres, female 15+)	SH.ALC.PCAP.MA.LI (litres, male 15+)

Table 3 | World Bank aggregate codes used in this analysis and corresponding pre-disaggregated sex-specific codes. Source: World Bank Open Data. Own compilation.

These pre-disaggregated codes cannot be treated as equivalent to sex-disaggregated reporting within a single aggregate indicator for three reasons. First, the reference populations differ: SH.STA.SUIC.P5 measures the suicide rate per 100,000 of the total population, while SH.STA.SUIC.FE.P5 measures it per 100,000 female population — these are different denominators yielding non-comparable rates. Second, the female and male codes are separate, independent time series with no structural link to the aggregate indicator; they are not decompositions of the aggregate value. Third, and most importantly for the research question of this study, this architectural choice reflects a fundamentally different approach to sex-disaggregation: rather than embedding disaggregation as a dimension within a single indicator, the World Bank externalises it into separate variables — making sex-disaggregated data structurally inaccessible through the standard aggregate indicator

interface. This represents a distinct type of gender visibility gap: the data exists within the platform, but is not accessible as a disaggregated property of the standard indicator. This finding is interpreted further in Section 7.

6.2 Data Coverage for Sex-Disaggregated Indicators

This section addresses SRQ2 by analysing the completeness of sex-disaggregated data for those indicators where disaggregation is available. As established in Section 6.1, sex-disaggregated data is only present in the WHO database, and only for 5 of the 10 indicators. Coverage is therefore calculated exclusively for these five indicators. For all indicators with GVS = 0 — in both databases — coverage of sex-disaggregated data is by definition zero and is not reported separately.

Coverage was calculated as the ratio of available non-missing observations to the total number of theoretically possible observations, defined as the number of distinct countries multiplied by the number of years in the observation period (2000–2023, 24 years). Coverage is reported separately for female and male disaggregated values.

Indicator	Countries	Coverage Female	Coverage Male
Life expectancy at birth	190	91.7%	91.7%
Suicide mortality rate	190	91.7%	91.7%
Diabetes prevalence	204	95.8%	95.8%
Tobacco use prevalence	190	2.4%	2.2%
Under-5 mortality rate	205	100.0%	100.0%

Table 4 | Data coverage for sex-disaggregated WHO indicators, 2000–2023. Source: WHO Global Health Observatory. Own calculation.

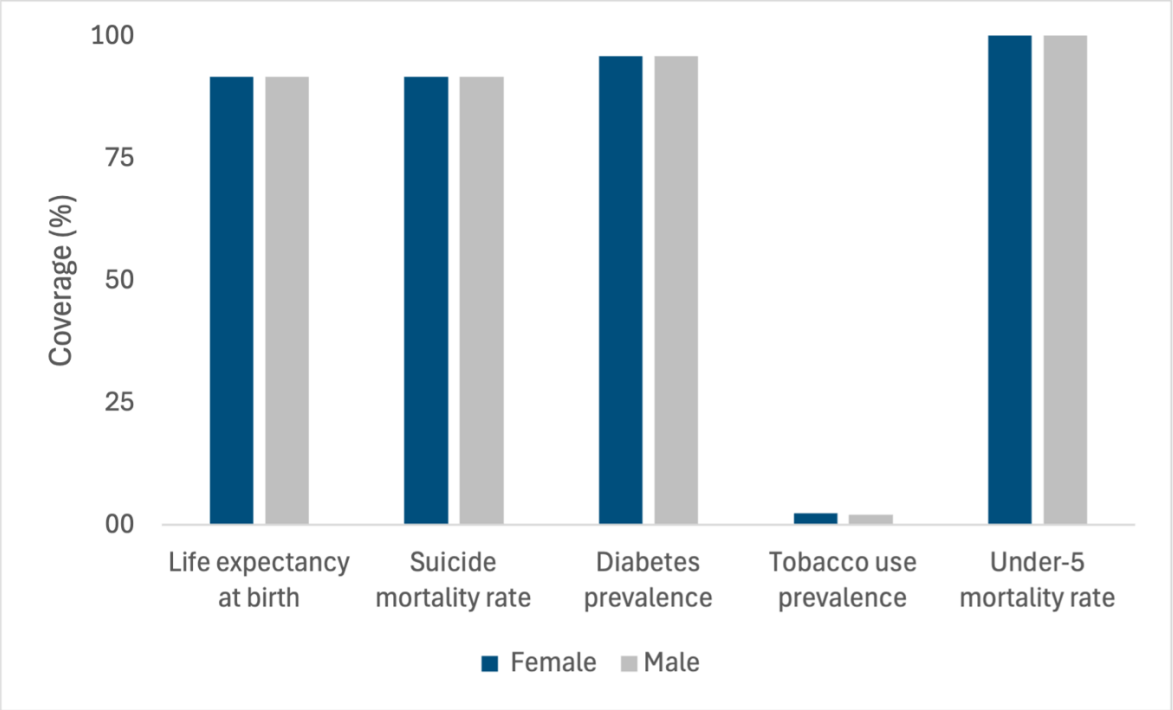

Figure 1 | Data coverage for sex-disaggregated WHO indicators, female and male observations, 2000–2023. Coverage is defined as the ratio of available non-missing observations to theoretically possible country-year combinations. The low coverage for tobacco use prevalence reflects the survey-based nature of data collection for this indicator. Source: WHO Global Health Observatory. Own calculation.

The results show a clear split within the five disaggregated indicators. Four indicators — life expectancy at birth, suicide mortality rate, diabetes prevalence, and under-5 mortality rate — achieve high coverage ranging from 91.7% to 100.0%, indicating near-complete data availability across countries and the observation period. Under-5 mortality rate achieves full coverage for all 205 countries across all 24 years.

Tobacco use prevalence is a notable exception, with coverage of only 2.4% for female and 2.2% for male values. This low coverage reflects the survey-based nature of tobacco data collection: the WHO does not estimate tobacco use annually for all countries, but reports values only for years in which national surveys were conducted. As a result, most country-year combinations have no data, and the effective observation window is substantially narrower than the 2000–2023 period. This structural characteristic of the indicator must be taken into account when interpreting the coverage result — it does not indicate a gender data gap in the conventional sense, but rather a general data sparsity inherent to survey-based reporting.

The structural nature of the low coverage value for tobacco use prevalence warrants a more detailed methodological examination. WHO tobacco data is derived primarily from nationally representative surveys — most notably the Global Adult Tobacco Survey (GATS), the STEPwise Approach to NCD Risk Factor Surveillance (STEPS), and equivalent national health surveys conducted by member states in collaboration with WHO and the Bloomberg Initiative to Reduce Tobacco Use [33]. These surveys are not conducted annually: the typical cycle between nationally representative tobacco surveys ranges from five to ten years, depending on country capacity and funding availability. As a result, the effective observation density for tobacco use prevalence is structurally limited to approximately three to four data points per country across the 2000–2023 study period, rather than the 24 annual observations that define the denominator in the coverage calculation.

This has a concrete arithmetic consequence. For the 190 countries included in the tobacco use indicator, the theoretically possible observation space is 190 × 24 = 4,560 country-year combinations. Under a realistic survey cycle of one observation per five to seven years, the empirically plausible observation count is closer to 190 × 3–4 = 570–760 observations — meaning that even a perfectly complete survey-based dataset would achieve a coverage rate of only 12–17% under the denominator used in this study. The observed coverage of 2.4% for female and 2.2% for male values therefore reflects not a failure to report available data, but the structural mismatch between the annual-observation denominator and the survey-cycle-based data supply.

This distinction is methodologically important for interpreting the gender missingness findings in Section 6.3. Unlike the four indicators with high coverage — where near-zero asymmetry between female and male observations reflects a genuine structural symmetry in data availability — the near-zero asymmetry for tobacco use prevalence reflects the near-universal absence of both female and male observations in non-survey years. The tobacco indicator therefore cannot contribute meaningful evidence to SRQ3, and its inclusion in the asymmetry analysis is retained only for completeness. Future applications of the GVS framework to survey-based indicators should consider adopting a survey-adjusted denominator — defined as the number of countries multiplied by the expected number of survey cycles within the observation period — to avoid structurally misleading coverage values of this kind.

6.3 Gender Missingness Asymmetry

This section addresses SRQ3 by examining whether missing data is distributed symmetrically between female and male observations for indicators where sex-disaggregation is available. The Asymmetry Index was calculated as the difference between female missingness and male missingness, where missingness is defined as 1 minus coverage. A positive value indicates higher data gaps for female observations; a negative value indicates higher data gaps for male observations; a value of zero indicates perfect symmetry.

Indicator	Coverage Female	Coverage Male	Missingness Female	Missingness Male	Asymmetry Index
Life expectancy at birth	91.7%	91.7%	8.3%	8.3%	0.0000
Suicide mortality rate	91.7%	91.7%	8.3%	8.3%	0.0000
Diabetes prevalence	95.8%	95.8%	4.2%	4.2%	0.0000
Tobacco use prevalence	2.4%	2.2%	97.6%	97.8%	−0.0013
Under-5 mortality rate	100.0%	100.0%	0.0%	0.0%	0.0000

Table 5 | Gender missingness asymmetry for sex-disaggregated WHO indicators, 2000–2023. Source: WHO Global Health Observatory. Own calculation.

The results show no meaningful asymmetry between female and male data availability for any of the five indicators. Four indicators return an Asymmetry Index of exactly 0.0000, indicating that wherever data exists for one sex, it exists equally for the other. Tobacco use prevalence shows a marginally negative index of −0.0013, suggesting a very slight tendency toward higher male coverage — however, given the overall sparsity of tobacco data

(coverage below 3% for both sexes), this difference is negligible and should not be interpreted as a systematic pattern.

These findings suggest that within the WHO Global Health Observatory, gender data gaps do not manifest as asymmetric missingness between sexes. Where WHO reports sex-disaggregated data, it does so consistently for both female and male populations. The gender data gap identified in this study therefore lies not in differential treatment of the two sexes within disaggregated indicators, but in the more fundamental question of whether disaggregation is provided at all — as documented in Section 6.1.

6.4 Comparison between WHO and World Bank

This section addresses SRQ4 by comparing the gender visibility of health data between the WHO Global Health Observatory and the World Bank Open Data platform across the 10 selected indicators.

The comparison is based on the GVS values reported in Table 2 (Section 6.1), which documents results for all 10 indicators across both databases.

The comparison reveals a stark difference between the two databases. WHO achieves a GVS of 1 for 5 out of 10 indicators, while the World Bank achieves a GVS of 1 for none. For the 5 indicators where WHO provides sex-disaggregated data, the World Bank provides only aggregate totals. There is no indicator in the set for which the World Bank matches or exceeds the WHO in terms of gender visibility.

It is notable that the 5 indicators with GVS = 1 in WHO include both Group A indicators — life expectancy at birth and suicide mortality rate — as well as three Group B indicators — diabetes prevalence, tobacco use prevalence, and under-5 mortality rate. The remaining Group A indicator, HIV prevalence, receives a GVS of 0 in both databases despite being one of the most extensively monitored global health indicators and one where gender disparities are particularly well-documented in the literature.

The World Bank result warrants particular attention. The absence of sex-disaggregated data across all 10 indicators does not reflect a limitation of the indicators themselves, as several of these indicators are available in disaggregated form in other sources including the WHO database. Rather, it reflects a structural characteristic of the World Bank Open Data platform as accessed via its bulk download and API interface: the platform provides aggregate

indicator values without a sex dimension, regardless of whether disaggregated data exists elsewhere.

The architectural difference between the two databases is particularly revealing in this regard. The WHO GHO embeds sex-disaggregation as a dimension within the indicator itself: a single indicator code returns female, male, and total observations in a unified dataset, and the presence or absence of disaggregation is directly assessable. The World Bank, by contrast, externalises sex-disaggregation into separate indicator codes for five of the ten indicators examined. This means that sex-specific data is not absent from the World Bank platform — it exists, but only through a parallel set of pre-disaggregated variables that are architecturally distinct from the aggregate indicator. A user accessing the standard World Bank indicator would receive only aggregate data with no sex dimension, while a user aware of the .FE./.MA. code variants could access sex-specific values separately. This design places the burden of awareness on the user and makes sex-disaggregation structurally invisible at the level of standard data access.

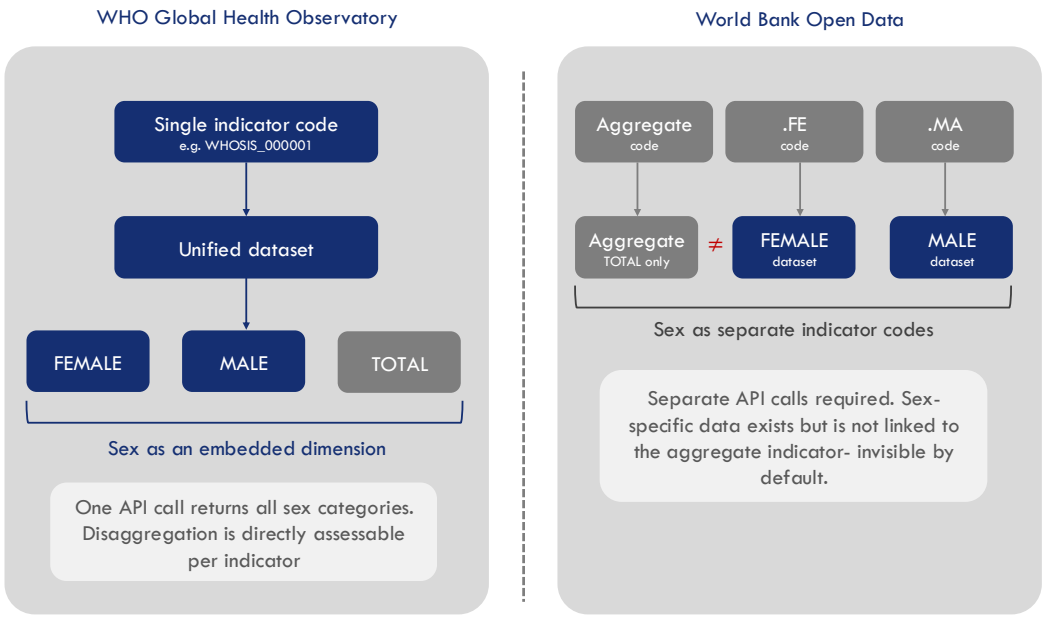

Figure 2 | Disaggregation architectures of the WHO Global Health Observatory and World Bank Open Data platform. The WHO embeds sex as a dimension within a single aggregate indicator; the World Bank provides sex-specific data through separate pre-disaggregated indicator codes that are architecturally distinct from the aggregate indicator. Source: Own illustration.

7. Discussion

This chapter interprets the empirical findings reported in Chapter 6 in light of the theoretical framework and the four sub-research questions. Section 7.1 discusses the gender visibility findings in relation to existing literature. Section 7.2 examines patterns of invisibility across health domains and databases. Section 7.3 addresses the limitations of the study. Section 7.4 derives implications for global health data policy.

7.1 Interpretation of Gender Visibility Findings

The central finding of this study is a fundamental asymmetry between the WHO Global Health Observatory and the World Bank Open Data platform in their provision of sex-disaggregated health data. The WHO provides disaggregated data for 5 of the 10 selected indicators; the World Bank provides disaggregated data for none. This finding is robust: it was verified both through the automated data collection pipeline and through manual inspection of the raw data exports, and it holds across all ten indicators regardless of health domain or analytical group.

This result is consistent with and extends the findings of Shirzad and Kaufman [11], who document structural gaps in sex-disaggregated reporting in the WHO Health Inequality Data Repository, and with Peters and Woodward [17], who argue that sex-disaggregation in global health monitoring remains normatively expected but empirically inconsistent. The present study adds a comparative dimension: by examining two major international databases simultaneously, it demonstrates that the gap is not uniform across platforms but takes platform-specific forms.

The platform-specific nature of the gap is a finding that requires theoretical contextualisation. Di Lego's [16, pp. 1-3] taxonomy, as introduced in Chapter 2, distinguishes between two types of gender health data gaps: those arising from the absence of data collection, and those arising from failures in data reporting and aggregation [16, pp. 1-3]. The present findings align predominantly with the second type. Both the WHO GHO and the World Bank collect or compile data on all ten indicators examined — the issue is not that the underlying information does not exist, but that the institutional data publication process does not systematically expose sex as a retrievable dimension of the aggregate indicator. This distinction is consequential: it implies that closing the gap does not necessarily require new

primary data collection, but rather a structural reorientation of how existing data is packaged and made accessible at the institutional level.

The theoretical framework further situates this finding within the broader concept of gender mainstreaming in global health [12, pp. 7-15], [14]. Bustreo et al. [14] argue that gender mainstreaming cannot be limited to policy commitments at the strategic level but must extend to the technical infrastructure through which health data is produced and disseminated. The GVS findings operationalise this argument empirically: a database that collects health data without structurally embedding sex as a disaggregation dimension fails the gender mainstreaming standard at the most fundamental technical level, regardless of the normative commitments documented in the surrounding policy frameworks. The WHO Global Reference List [5, pp. 1-18] and the SDG monitoring framework both formally require sex-disaggregated reporting — yet this requirement is not consistently implemented in the GHO's own standard indicator interface, and is entirely absent from the World Bank's aggregate indicator layer.

Kaufman et al. [10] argue that the conflation of sex and gender in health research data is itself a structural problem: when databases do not systematically distinguish between biological sex and gender identity, the data infrastructure reproduces rather than corrects existing epistemic blind spots. While the present study focuses on the more basic question of whether sex disaggregation exists at all, the findings are consistent with Kaufman et al.'s broader diagnosis. Databases that provide only aggregate totals — without any sex dimension — make it impossible to apply even the most rudimentary sex-based analysis, let alone the more nuanced gender-sensitive approaches that the literature advocates.

The finding that HIV prevalence receives a GVS of 0 in both databases is particularly noteworthy. HIV is among the most extensively monitored indicators in global health, and gender disparities in HIV prevalence and care access are among the best-documented in the literature [11], [25]. The absence of sex-disaggregated data in the aggregate indicator codes of both the WHO GHO and the World Bank — despite the existence of extensive HIV surveillance infrastructure — suggests that the problem is not one of data unavailability per se, but of how data is structured and published at the level of standardised indicator databases. This distinction between data existence and data accessibility is central to the conceptual contribution of this study.

The symmetry findings in Section 6.3 add a further dimension to this picture. Where the WHO does provide sex-disaggregated data, it does so with equal coverage for both sexes —

Asymmetry Indices of 0.0000 for four of the five disaggregated indicators. This means that the gender data gap identified in this study is not a within-indicator problem of differential treatment of female and male observations, but a structural problem at the level of whether disaggregation is provided at all. The gap operates upstream of coverage: it is a binary, architectural feature of the database rather than a quantitative bias in data completeness.

This upstream location of the gender data gap has important implications for how the problem should be understood and addressed. Peters and Woodward [17] propose a roadmap for sex- and gender-disaggregated health research that distinguishes between data generation, data analysis, and data reporting as three distinct intervention points. The present findings suggest that the primary failure in the databases examined is at the data reporting stage: the data may exist in underlying surveillance systems, but is not surfaced as a disaggregated dimension in the standardised indicator products that most users access. This framing shifts the policy conversation from data collection — which often receives the most attention — to data architecture and publication standards, which are less visible but equally consequential for the accessibility of gender-disaggregated information.

The high coverage values for the five WHO indicators with GVS = 1 — ranging from 91.7% to 100% — reinforce this point. Where the WHO has decided to structure an indicator with a sex dimension, it does so comprehensively and consistently, achieving near-complete geographic and temporal coverage. The problem is therefore not one of capacity or effort once the decision to disaggregate has been made, but one of whether that decision is made in the first place. This selectivity in the application of sex-disaggregation — present for some indicators, absent for others, with no apparent systematic rationale based on health domain or data availability — is itself a finding that warrants attention in future research and institutional review.

7.2 Patterns of Invisibility across Health Domains

The distribution of GVS values across the ten indicators reveals patterns that are both domain-specific and cross-cutting. Within the WHO database, sex-disaggregated data is available for life expectancy, suicide mortality, diabetes prevalence, tobacco use, and under-5 mortality — but not for HIV prevalence, tuberculosis incidence, TB treatment success rate, neonatal mortality, and alcohol consumption. This is not a simple Group A versus Group B

divide: three of the five Group B indicators achieve GVS = 1 in the WHO, while one Group A indicator (HIV) achieves GVS = 0 in both databases.

This pattern challenges a simple narrative in which gender data gaps are most severe for indicators where gender disparities are least recognised. HIV prevalence is precisely the kind of indicator for which the literature most strongly advocates sex-disaggregated reporting: Feraldi et al. [25] and Shirzad and Kaufman [11] both identify HIV as a domain where gender-differentiated patterns of exposure, access to testing, and treatment uptake are well-documented and policy-relevant. Yet it is precisely this indicator that fails to provide a sex dimension in either database's aggregate interface. The absence of disaggregation for HIV is therefore not a reflection of normative ambiguity about whether disaggregation is needed — it is a structural failure in the translation of normative consensus into data architecture.

Conversely, the presence of GVS = 1 for three Group B indicators — diabetes prevalence, tobacco use, and under-5 mortality — suggests that sex-disaggregation in the WHO GHO is not strictly determined by the strength of the documented gender disparity. Diabetes and tobacco use are indicators for which the evidence on sex-differentiated patterns is growing but less consolidated than for HIV or suicide [25]. Their inclusion in the WHO's disaggregated reporting may reflect broader institutional decisions about data quality and model-based estimation capacity rather than a systematic prioritisation based on gender relevance. This interpretation is consistent with the WHO's practice of providing modelled estimates — which inherently require sex-specific input data — for certain indicators such as diabetes prevalence and under-5 mortality, while relying on member state reporting for others.

The three qualitatively distinct patterns of non-disaggregation identified in the WHO data — complete absence of a sex dimension, structural presence without populated values, and sex categories present but empty — map onto Di Lego's [16, pp. 1-3] taxonomy of gender health data gaps in important ways. The complete absence of a sex dimension for HIV, TB incidence, and TB treatment success corresponds to what Di Lego terms a Type 1 gap: data that is collected but not differentiated by sex at the point of aggregation. The structural presence without populated values for alcohol consumption suggests a Type 2 gap: disaggregation is anticipated in the data architecture but not yet operationalised, pointing to a transitional state in reporting practice.

The neonatal mortality indicator presents a particularly instructive case. The GHO data structure contains a sex dimension for this indicator — labelled "Both sexes" — but provides no female or male breakdown. This is not a case of missing data in the conventional sense:

the sex field is populated, but only with a single aggregate category. This pattern suggests that the data infrastructure was designed with sex-disaggregation in mind, but that the actual disaggregated values were either never collected, never submitted by member states, or never published at the aggregate level. Di Lego [16, pp. 1-3] identifies this type of gap — where the structural scaffolding for disaggregation exists but is not populated — as particularly problematic from a monitoring perspective, because it creates the appearance of a sex-aware data system without delivering its functional benefit.

The alcohol consumption indicator presents the inverse of the tobacco use situation. For tobacco, GVS = 1 but coverage is below 3% — disaggregation exists but data is sparse. For alcohol, female and male categories exist in the data structure but contain no numerical values — GVS = 0 despite structural readiness. Together, these two cases illustrate the full spectrum of the relationship between disaggregation architecture and disaggregation practice: from structurally absent, to structurally present but unpopulated, to populated but sparse, to comprehensively populated. This spectrum — empirically demonstrated across the ten indicators of this study — provides a more granular understanding of gender data gaps than the binary present/absent framing that dominates much of the existing literature.

The tuberculosis indicators deserve particular attention. TB incidence and TB treatment success rate both receive GVS = 0 in both databases. Feraldi et al. [25] demonstrate that sex disaggregation along the full TB care cascade — from exposure to treatment — is essential for identifying differential access patterns, yet the aggregate indicators accessible through standard database interfaces provide no sex dimension. This finding corroborates Feraldi et al.'s argument at the level of global indicator infrastructure: the structural absence of disaggregation in the two most commonly accessed international TB indicators forecloses the comparative gender analysis that the literature identifies as necessary.

The absence of sex-disaggregation for both tuberculosis indicators — incidence and treatment success rate — across both databases is compounded by a finding from the data collection process. For TB incidence, the standard WHO indicator code matched by the algorithm (MDG_0000000020) returned data without a sex dimension; a corrected API query using the disease-specific code TB_e_inc_100k confirmed that sex-disaggregated TB incidence data does exist within the GHO system but is not accessible through the standard aggregate indicator code. This illustrates a practical dimension of the gender visibility gap: even where sex-disaggregated data technically exists within a database, it may not be accessible through the standard indicator code that a non-specialist user would employ. The

gap between data existence and data findability is a structural feature of complex institutional databases that has received insufficient attention in the gender data literature — and one that the FAIR Findability criterion, as applied to these databases in Section 5.5.4, is specifically designed to surface. The World Bank's architectural approach — externalising sex-disaggregation into separate .FE./.MA. indicator codes rather than embedding it as a dimension within the aggregate indicator — represents a distinct type of invisibility. For five of the ten indicators, sex-specific data is technically accessible within the platform, but only to users who know to look for it and know the specific code variants. From the perspective of a standard data user or automated analysis pipeline, the World Bank's aggregate indicators appear to have no sex dimension at all. This design choice has direct consequences for gender-responsive monitoring: Morgan et al. [15] argue that gender-disaggregated data must be structurally embedded in monitoring systems to be actionable — data that requires specialist knowledge to locate is functionally invisible for routine policy use.

7.3 Limitations of the Study

Several methodological limitations must be acknowledged. First, the indicator set of ten indicators, while deliberately designed to span multiple health domains and expected visibility patterns, is not a random or exhaustive sample of all indicators available in the two databases. The findings cannot be generalised to the full catalogue of WHO GHO or World Bank indicators; they are specific to the ten indicators selected and should be interpreted as an exploratory, structured assessment rather than a comprehensive audit.

Second, the fuzzy string matching procedure used to identify indicator pairs introduces a degree of uncertainty. While all matches were validated manually and conceptually implausible pairs were excluded, the matching is based on indicator names rather than on formal metadata or conceptual definitions. Indicators with similar names but different measurement approaches — such as the tobacco use prevalence indicator, where the WHO uses a survey-based approach and the World Bank uses modelled estimates — may have been retained as matches despite methodological differences that affect comparability.

Third, the GVS is a binary, conservative measure of gender visibility. It does not capture the degree of disaggregation, the quality of the underlying data, or the temporal trends in disaggregation practice. An indicator that provides disaggregated data for only a single year receives the same GVS as one that provides it consistently across the entire 2000–2023

period — as illustrated by tobacco use prevalence, where GVS = 1 but coverage is below 3%. Future research could extend this framework by incorporating temporal dimensions of disaggregation availability.

Fourth, the study is limited to the binary male/female sex categorisation as implemented in both databases. This reflects the current state of international health data infrastructure rather than a normative choice: as Shirzad and Kaufman [11] document, neither the WHO GHO nor the World Bank systematically captures gender identity beyond the binary, meaning that the visibility of non-binary and gender-diverse populations cannot be assessed through these databases at present.

Fifth, the data was collected at a fixed point in time (May 2026). Both databases are dynamic and updated continuously; findings may differ if the analysis is replicated at a later date. This is particularly relevant for indicators where the sex dimension appears structurally present but unpopulated — such as alcohol consumption per capita in the WHO — where population of the empty categories may be a matter of ongoing data work rather than a permanent structural feature.

A sixth limitation concerns the conceptual scope of the GVS itself. The present study operationalises gender visibility as the structural presence of sex-disaggregated data in standardised indicator databases. This is a necessary but not sufficient condition for gender-responsive health monitoring. Morgan et al. [15] distinguish between three levels of gender data: sex-disaggregated indicators, gender-specific indicators, and indicators that capture gender power relations and systems. The GVS and associated metrics address only the first level — the existence of a sex dimension in aggregate health statistics. They do not assess whether the indicators themselves are designed to capture gender-specific phenomena, nor whether the databases provide contextual data on the social and structural determinants of gendered health outcomes. Future research extending this framework to the second and third levels of Morgan et al.'s taxonomy would provide a more comprehensive picture of gender visibility in global health data infrastructure.

A seventh limitation relates to the comparability of matched indicator pairs across databases. While the fuzzy string matching procedure and subsequent manual validation were designed to ensure conceptual alignment, some matched pairs may differ in their underlying measurement methodology, reference population, or data source. The life expectancy at birth indicator, for example, uses modelled estimates in both databases but may draw on different methodological assumptions. The tobacco use prevalence indicator is survey-based in the

WHO but uses different survey instruments and reference years across countries. These methodological differences between matched pairs mean that the Coverage and Asymmetry metrics, while internally consistent within each database, may not be directly comparable across databases for specific indicators. This limitation is acknowledged as an inherent feature of cross-database comparative analysis and does not affect the GVS findings, which are computed independently for each database.

An eighth limitation concerns the linguistic and terminological basis of the indicator matching procedure. The fuzzy string matching algorithm applied in this study operates on indicator names as they appear in the English-language catalogues of the WHO GHO and the World Bank Open Data platform. This means that the matching is sensitive to the specific terminology used in each platform's naming conventions rather than to the underlying conceptual definitions or formal classification systems that the indicators instantiate.

This creates two related risks. First, conceptually identical indicators that are named differently across platforms may fail to reach the similarity threshold and be excluded from the candidate pool. An indicator labelled "Prevalence of current tobacco use" in one database and "Current tobacco smoking prevalence" in another will score high on fuzzy similarity; but an indicator labelled "Tobacco use among adults" against "Adult smoking rates" may fall below the threshold of 60 despite referring to the same construct. Because the candidate pool was assembled through exhaustive retrieval followed by algorithmic filtering rather than manual pre-selection, any pairs missed at the matching stage cannot be identified post hoc. The final indicator set of ten may therefore underrepresent health domains where naming conventions diverge substantially between the two platforms.

Second, the matching procedure does not draw on formal semantic resources such as the International Classification of Diseases (ICD-11), SNOMED CT, or the SDG indicator metadata framework maintained by the United Nations Statistics Division [34]. These classification systems assign stable, language-independent identifiers to health concepts that would enable more robust cross-database matching independent of naming variation. A future application of this framework using formal ontology-based matching — mapping indicators to shared concept identifiers rather than name strings — would substantially reduce the risk of missed or mismatched pairs and increase the reproducibility and completeness of the indicator selection procedure. The present study acknowledges this as a structural limitation of the name-based approach and recommends ontology-based matching as a methodological improvement for subsequent research in this area.

7.4 Implications for Global Health Data Policy

The asymmetries in sex-disaggregated data availability documented in this study have concrete consequences for gender-responsive health policy. As Morgan et al. [15] argue, sex- or gender-disaggregated data constitute a necessary entry point for gender-responsive monitoring and evaluation — but only as a first step. Identifying where disparities exist does not, by itself, explain why they arise. The structural absence of sex-disaggregated data in global indicator systems thus not only limits epidemiological analysis but also forecloses the possibility of designing targeted, evidence-based interventions. Feraldi et al. [25] demonstrate this concretely: without sex-disaggregated care cascade data for tuberculosis, it is structurally impossible to identify which population groups are left behind at which point along the pathway from exposure to treatment.

This structural foreclosure is not merely a technical inconvenience — it has measurable consequences for the scope and quality of evidence available to policymakers. The 2030 Agenda for Sustainable Development explicitly commits to leaving no one behind and to disaggregating all SDG monitoring data by sex, age, and other relevant characteristics [6, pp. 1-12]. The findings of this study document the distance between this commitment and its implementation in the two most widely used international health data platforms. For the five indicators with GVS = 0 in both databases, no sex-disaggregated analysis is possible using standard institutional data products — meaning that the SDG commitment to disaggregated monitoring is structurally unenforceable for these indicators at the global level, regardless of national data collection practices.

The McKinsey Health Institute [1] estimates the economic cost of the women's health gap at up to one trillion dollars annually in lost productivity and healthcare costs. While this figure encompasses a broader set of issues than data disaggregation, it illustrates the scale of the problem that adequate gender-disaggregated monitoring is intended to address. Data infrastructure gaps of the kind documented in this study are one mechanism through which this broader women's health gap is perpetuated: without disaggregated data, the gender-differentiated burden of disease remains invisible to the planning systems that allocate resources and design interventions.

The findings of this study point to three specific areas where policy attention is warranted. First, the absence of sex-disaggregated data for HIV prevalence in both databases — despite HIV being one of the most extensively monitored global health indicators — suggests

that the problem is not one of technical capacity but of institutional prioritisation. The normative frameworks already exist: both the WHO Global Reference List [5, pp. 1-18] and the SDG monitoring framework explicitly foresee sex-disaggregation for this indicator. The gap between normative expectation and empirical implementation identified here supports Bustreo et al.'s [14] argument that gender mainstreaming in global health requires not only policy commitments but systematic enforcement mechanisms at the level of data publication standards.

Second, the World Bank's architectural approach to sex-disaggregation — externalising it into separate indicator codes rather than embedding it within aggregate indicators — has implications for the accessibility of gender data in routine monitoring contexts. The burden-of-awareness problem identified in Section 7.2 is directly relevant to the data users most likely to rely on standard indicator interfaces: national health ministries, WHO country offices, and international NGOs using bulk data exports for programme monitoring. A policy implication is that international data platforms should be encouraged to adopt unified disaggregation architectures in which sex is a retrievable dimension of the aggregate indicator rather than a separate product.

Third, the 2023 launch of the WHO Health Inequality Data Repository — documented by Shirzad and Kaufman [11] — represents a significant institutional development that the findings of this study contextualise rather than render obsolete. The Repository is a separate product from the GHO analysed here, and its existence does not resolve the structural gaps in the standard GHO indicator interface that this study documents. However, it does signal an institutional recognition of the problem, and the structural gaps identified in this study provide a baseline against which the Repository's contribution to improved gender data visibility can be assessed in future research.

A fourth area warranting policy attention concerns the absence of technical interoperability standards for sex-disaggregated health data across international platforms. The findings of this study reveal that the WHO GHO and the World Bank Open Data platform not only differ in their gender visibility outcomes, but in the fundamental architectural approach through which sex-disaggregation is — or is not — embedded in the indicator structure. These architectural differences are not merely a matter of institutional preference; they have direct consequences for the usability, findability, and interoperability of gender data in routine monitoring practice, as documented in Sections 6.4 and 7.2.

The absence of a shared technical standard for sex-disaggregation in health indicator publication means that cross-database comparisons of the kind conducted in this study require platform-specific adaptations at every stage — from data retrieval to harmonisation to metric computation. A user or analytical system designed to extract sex-disaggregated data from the WHO GHO via its standard indicator interface will retrieve no sex dimension when querying the equivalent World Bank indicator, not because the data does not exist, but because the two platforms embed the sex dimension through structurally incompatible approaches. This incompatibility constitutes a barrier to the scalable, automated gender data monitoring that increasingly characterises evidence synthesis in global health [15].

The FAIR data principles — specifically the Interoperability and Findability criteria — provide a principled framework for addressing this gap [32]. Applied to the domain of sex-disaggregated health indicators, the Interoperability criterion requires that sex be embedded as a retrievable dimension within the aggregate indicator in a standardised, machine-readable form, rather than externalised into platform-specific variable codes. The Findability criterion requires that sex-disaggregated data be discoverable through the standard indicator interface without requiring specialist knowledge of platform-specific code conventions. Neither criterion is currently met by the World Bank Open Data platform for the indicators examined in this study, and the WHO GHO meets both criteria only partially — providing sex-disaggregation for 5 of 10 indicators, with no transparent rationale for why the remaining 5 are excluded.

A concrete policy step in this direction would be the development of a shared metadata standard for sex-disaggregated health indicator publication, adopted jointly by the WHO, the World Bank, and other major international health data producers. Analogous standards already exist in adjacent domains: the HL7 Fast Healthcare Interoperability Resources (FHIR) standard defines a common data model for clinical health information exchange that has been adopted across national health systems and electronic health record platforms [35]. A comparable standard applied at the level of international health indicator publication — specifying sex as a mandatory disaggregation dimension in indicator metadata, with a standardised vocabulary for sex categories and a uniform API response structure — would resolve the architectural incompatibility documented in this study and make sex-disaggregated data structurally accessible across platforms without platform-specific adaptations. Bustreo et al. [14] argue that gender mainstreaming in global health requires institutional accountability structures that extend beyond policy declarations to the technical

outputs of data systems; the development of such a standard would represent precisely this kind of structural accountability mechanism at the level of data infrastructure.

Beyond these three specific policy areas, the findings of this study carry a broader implication for the governance of global health data infrastructure. The gap between normative expectation and empirical implementation documented here — across two of the most authoritative international health databases, for indicators that are formally required to be sex-disaggregated — suggests that existing accountability mechanisms for data publication standards are insufficient. Bustreo et al. [14] argue that gender mainstreaming in global health requires institutional accountability structures that extend beyond policy declarations to the technical outputs of data systems. The present study provides empirical grounding for this argument: normative frameworks alone have not produced consistent sex-disaggregation in the standard indicator products of the WHO GHO or the World Bank, even for indicators where the need is well-established and the data is known to exist.

A concrete step in this direction would be the development of systematic, periodic assessments of sex-disaggregation coverage in major international health databases — analogous to the data quality reviews already conducted by the WHO for other dimensions of data completeness [26, pp. 15-22]. Such assessments would make the gender visibility gap visible as a trackable, reportable dimension of database quality, creating institutional incentives for improvement. The methodological framework developed in this study — the GVS, Coverage, and Asymmetry Index applied to matched indicator pairs across databases — provides a replicable template for such assessments. Applied to a larger indicator set and extended over time, it could serve as the basis for a gender data visibility index for international health databases.

8. Conclusion & Outlook

8.1 Summary of Findings

This study set out to examine the extent to which women are structurally visible in international health databases at the level of standardised health indicators, and how consistently sex-disaggregated data is implemented across databases, indicators, countries, and time. The analysis was conducted across ten matched health indicators drawn from the WHO Global Health Observatory and the World Bank Open Data platform, covering the period 2000–2023. Three metrics — the Gender Visibility Score, Coverage, and the Missingness Asymmetry Index — were applied to a unified analytical unit of Indicator × Country × Year × Gender. The findings provide clear and empirically grounded answers to all four sub-research questions.

SRQ1 — Gender Visibility: Sex-disaggregated data is available in the WHO GHO for 5 of the 10 selected indicators, and for none of the 10 indicators in the World Bank Open Data platform as accessed via aggregate indicator codes. This represents a fundamental asymmetry between the two databases. The absence of sex-disaggregation is not uniformly distributed across health domains: it affects HIV prevalence — one of the most extensively monitored global health indicators and one for which gender disparities are among the best-documented in the literature — while being absent for indicators such as diabetes prevalence and under-5 mortality, where the documentation of gender disparities is less consolidated.

SRQ2 — Data Coverage: Where the WHO GHO provides sex-disaggregated data, it does so with near-complete geographic and temporal coverage. Four of the five disaggregated indicators achieve coverage rates between 91.7% and 100.0% across all available countries and years. Tobacco use prevalence is a structural exception, with coverage below 3% due to the survey-based nature of data collection — a characteristic of the indicator rather than a gender-specific gap. This finding demonstrates that the WHO has both the capacity and the infrastructure to provide comprehensive sex-disaggregated data when the institutional decision to do so has been made.

SRQ3 — Gendered Missingness: No systematic asymmetry between female and male data availability was identified for any of the five WHO indicators with GVS = 1. Four indicators return an Asymmetry Index of exactly 0.0000; tobacco use prevalence returns a negligible value of −0.0013. The gender data gap identified in this study therefore does not manifest as differential treatment of female and male observations within disaggregated indicators, but as a binary, upstream architectural question of whether sex-disaggregation is structurally embedded in the indicator at all.

SRQ4 — Database Comparison: The comparison between the WHO GHO and the World Bank Open Data platform reveals not only a quantitative difference in gender visibility, but a qualitative difference in disaggregation architecture. The WHO embeds sex as a dimension within the aggregate indicator, making disaggregation directly assessable. The World Bank externalises sex-disaggregation into separate pre-disaggregated indicator codes for five of the ten indicators, making sex-specific data technically present within the platform but structurally inaccessible through the standard aggregate interface. This architectural difference has direct consequences for the accessibility of gender data in routine monitoring practice.

8.2 Scientific Contribution

This study makes three contributions to the existing literature on gender data gaps in global health.

First, it introduces and operationalises a replicable methodological framework — comprising the Gender Visibility Score, Coverage, and Missingness Asymmetry Index — for the systematic, quantitative assessment of sex-disaggregation in international health databases. Unlike qualitative audits or narrative reviews, this framework produces numerical outputs that are directly comparable across indicators, databases, and time periods, and that can be replicated by other researchers using the publicly documented scripts and source data.

Second, it provides the first systematic cross-database comparison of sex-disaggregation availability across matched indicator pairs in the WHO GHO and the World Bank Open Data platform. Prior studies, including Shirzad and Kaufman [11] and Feraldi et al. [25], have documented gender data gaps within specific databases or indicator domains. The present study extends this work by comparing two of the most widely used international health data

platforms simultaneously, demonstrating that the gap takes platform-specific architectural forms rather than being a uniform absence of data.

Third, it introduces a conceptual distinction — between data existence, data accessibility, and data findability — that refines the prevailing binary framing of gender data gaps as present or absent. The empirical finding that sex-disaggregated data may exist within a database but be inaccessible through standard indicator codes, or present in the data architecture but unpopulated, or populated but structurally separated from the aggregate indicator, provides a more granular taxonomy of gender visibility gaps that has direct implications for both research methodology and institutional data governance.

8.3 Outlook and Future Research

Several directions for future research emerge from the findings and limitations of this study.

The most immediate extension would be the application of the same methodological framework to a larger and more systematically selected indicator set. The ten indicators examined here were chosen to span multiple health domains and expected visibility patterns, but they do not constitute a representative sample of either database's full catalogue. A comprehensive application of the GVS framework across all indicators available in the WHO GHO and the World Bank platform would provide a full audit of sex-disaggregation coverage and enable the identification of domain-specific patterns that remain speculative in this study.

A second direction concerns the temporal dimension of gender visibility. The GVS as operationalised here is a binary measure applied to the full 2000–2023 period. Future research could extend the framework to track changes in sex-disaggregation availability over time — examining, for example, whether the introduction of the SDG monitoring framework in 2015 or the WHO Health Inequality Data Repository in 2023 [11] produced measurable shifts in disaggregation practice within the standard indicator databases. Such longitudinal analysis would transform the current cross-sectional baseline into a trend assessment with direct policy relevance.

A third direction concerns the extension of the framework beyond the binary male/female sex categorisation. As Shirzad and Kaufman [11] document, existing international health databases do not systematically capture gender identity beyond the binary, meaning that the visibility of non-binary and gender-diverse populations cannot currently be assessed

through institutional data infrastructure. Future research could explore how emerging national data systems that have begun to incorporate non-binary gender categories might be evaluated using adapted versions of the GVS framework.

Finally, the methodological framework developed here could be extended to additional international health databases beyond the WHO GHO and the World Bank — including UNICEF's Multiple Indicator Cluster Surveys, the Global Burden of Disease database, and regional databases maintained by the European Centre for Disease Prevention and Control or the Pan American Health Organization. A comparative assessment across a broader set of institutional data products would allow the platform-specific architectural differences identified in this study to be situated within a wider landscape of global health data governance, and would strengthen the empirical basis for the policy recommendations derived in Section 7.4.

A fifth direction for future research concerns the temporal dimension of gender visibility and its relationship to international policy milestones. The present study provides a cross-sectional baseline of sex-disaggregation availability for the period 2000–2023 as a whole, treating the observation window as a single analytical unit. This design was appropriate for the study's primary objective — establishing whether sex-disaggregated data exists — but it is not equipped to detect whether disaggregation practices have changed over time in response to institutional interventions. Two policy milestones in particular generate testable hypotheses about temporal change that a longitudinal extension of this framework could evaluate.

The first is the adoption of the 2030 Agenda for Sustainable Development in 2015, which formally committed all UN member states to disaggregating SDG monitoring data by sex, age, and other equity dimensions [6, pp. 1-6]. If this commitment produced a measurable shift in sex-disaggregation practice within the WHO GHO or the World Bank platform, it should be detectable as an increase in GVS = 1 assignments or in Coverage rates for the post-2015 observation years relative to the pre-2015 period. The second milestone is the launch of the WHO Health Inequality Data Repository in 2023 [11], which represents the most recent major institutional initiative explicitly designed to improve gender data availability in global health monitoring. A longitudinal analysis that tracks GVS and Coverage values year by year — rather than aggregating across the full 2000–2023 window — would allow both milestones to be evaluated as natural experiments, testing whether normative commitments and institutional initiatives translate into measurable improvements in sex-disaggregation at the indicator level.

Such a longitudinal framework would also allow the identification of regression patterns — cases where sex-disaggregation was available in earlier years but subsequently discontinued, for example due to changes in reporting methodology, database restructuring, or shifts in institutional priorities. The present study's binary GVS and aggregate Coverage metrics are insensitive to such patterns by design. A year-by-year extension would transform the current static audit into a dynamic monitoring instrument, directly supporting the kind of periodic, trackable gender data quality assessment that Section 7.4 identifies as a priority for global health data

Beyond these methodological extensions, the findings of this study point to a broader research agenda at the intersection of gender studies, global health governance, and data science. The question of who is visible in global health data — and who is rendered invisible by the structural choices embedded in institutional data infrastructure — is ultimately a question about whose health is considered worth monitoring, and at what level of granularity. The present study addresses this question at the most basic level: the binary sex dimension. Future research could examine how the patterns documented here relate to other dimensions of invisibility in global health data — including age-disaggregation, geographic disaggregation below the national level, and the representation of marginalised population groups whose health outcomes are systematically absent from international indicator systems. The methodological framework developed here, extended to these dimensions, could contribute to a more comprehensive understanding of structural invisibility in global health data as a field of inquiry in its own right.

The urgency of this agenda is underscored by the scale of the problem. The McKinsey Health Institute [1] estimates the economic cost of the women's health gap at up to one trillion dollars annually — a figure that reflects not only differential health outcomes between women and men, but the cumulative effect of decades of under-investment in gender-disaggregated research, monitoring, and policy. Closing this gap requires, as a necessary first step, that the data infrastructure through which global health is monitored makes women structurally visible. The findings of this study document how far the two most authoritative international health data platforms still are from that standard — and provide a replicable, evidence-based starting point for tracking progress toward it.

The central argument of this thesis can be stated simply: you cannot monitor what you cannot see. As long as sex-disaggregated data remains absent from the standard indicator interfaces of major international health databases, the gender dimension of global health will

remain structurally invisible — not because the data does not exist, but because the systems designed to make it accessible have not yet been built to do so consistently. This study is one step toward making that gap visible and measurable.

FACHHOCHSCHULE
WIENER NEUSTADT
Austrian Network for Higher Education

List of Figures

Figure 1| Data coverage for sex-disaggregated WHO indicators, female and male observations, 2000–2023. Coverage is defined as the ratio of available non-missing observations to theoretically possible country-year combinations. The low coverage for tobacco use prevalence reflects the survey-based nature of data collection for this indicator. Source: WHO Global Health Observatory. Own calculation. 42

Figure 2 | Disaggregation architectures of the WHO Global Health Observatory and World Bank Open Data platform. The WHO embeds sex as a dimension within a single aggregate indicator; the World Bank provides sex-specific data through separate pre-disaggregated indicator codes that are architecturally distinct from the aggregate indicator. Source: Own illustration. 46

FACHHOCHSCHULE
WIENER NEUSTADT
Austrian Network for Higher Education

List of Tables

Table 1| Final indicator set with database codes, match scores, and analytical grouping. Source: WHO Global Health Observatory and World Bank Open Data. Own calculation. 29

Table 2 | Gender Visibility Score by indicator and database. Source: WHO Global Health Observatory and World Bank Open Data. Own calculation.....39

Table 3 | World Bank aggregate codes used in this analysis and corresponding pre-disaggregated sex-specific codes. Source: World Bank Open Data. Own compilation.....40

Table 4 | Data coverage for sex-disaggregated WHO indicators, 2000–2023. Source: WHO Global Health Observatory. Own calculation.....41

Table 5 | Gender missingness asymmetry for sex-disaggregated WHO indicators, 2000–2023. Source: WHO Global Health Observatory. Own calculation.....44

Bibliography

[1] K. Ellingrud, L. Pérez, A. Petersen, and V. Sartori, Closing the women's health gap: A $1 trillion opportunity to improve lives and economies, McKinsey Health Institute, 2024. [Online]. Available: https://www.mckinsey.com/mhi/our-insights/

[2] L. Merone, K. Tsey, D. Russell, and C. Nagle, "Sex Inequalities in Medical Research: A Systematic Scoping Review of the Literature," Women's Health Reports, vol. 3, no. 1, pp. 49–59, 2022. doi: 10.1089/whr.2021.0083.

[3] C. Tannenbaum, R. P. Ellis, F. Eyssel, J. Zou, and L. Schiebinger, "Sex and gender analysis improves science and engineering," Nature, vol. 575, no. 7781, pp. 137–146, 2019. doi: 10.1038/s41586-019-1657-6.

[4] S. Heidari, T. F. Babor, P. De Castro, S. Tort, and M. Curno, "Sex and Gender Equity in Research: Rationale for the SAGER guidelines and recommended use," Research Integrity and Peer Review, vol. 1, no. 1, p. 2, 2016. doi: 10.1186/s41073-016-0007-6.

[5] World Health Organization, Global Reference List of 100 Core Health Indicators (2018), Geneva: WHO, 2018, pp. 1–18.

[6] UN Women, Turning Promises into Action: Gender Equality in the 2030 Agenda, New York, 2018, pp. 1–17, 28–35.

[7] M. Buvinić et al., Mapping Gender Data Gaps, Data2X, 2014, pp. 1–6, 14–20.

[8] World Health Organization & World Bank, Global Monitoring Report on Universal Health Coverage, Geneva, 2021, pp. 3–10, 37–45.

[9] European Commission, Gendered Innovations 2: How Inclusive Analysis Contributes to Research and Innovation. Publications Office of the European Union, 2020.

[10] M. R. Kaufman, E. L. Eschlimann, and T. S. Karver, "Differentiating sex and gender in health research to achieve gender equity," Bulletin of the World Health Organization, vol. 101, pp. 666–671, 2023. doi: 10.2471/BLT.22.289310.

[11] M. Shirzad and M. R. Kaufman, "Missing piece: the imperative of gender-disaggregated data in unraveling health inequity," BMJ Global Health, vol. 10, e018266, 2025. doi: 10.1136/bmjgh-2024-018266.

[12] World Health Organization, Gender mainstreaming for health managers: a practical approach, Geneva: WHO, 2011, pp. 7-15.

[13] United Nations, Beijing Declaration and Platform for Action, adopted at the Fourth World Conference on Women, UN Women, New York, 1995. [Online]. Available: https://www.un.org/womenwatch/daw/beijing/pdf/BDPfA%20E.pdf. Accessed: 01-Nov-2025

[14] F. Bustreo, A. G. Ponchia, C. Rocco, and R. Hinton, "Strengthening the transformative potential of gender mainstreaming in global health," EClinicalMedicine, vol. 35, 100858, 2021. doi: 10.1016/j.eclinm.2021.100858.

[15] R. Morgan et al., "Gender-responsive monitoring and evaluation for health systems," Health Policy and Planning, vol. 39, no. 9, pp. 1000–1005, 2024. doi: 10.1093/heapol/czae073.

[16] V. D. Lego, “Uncovering the gender health data gap,” Cadernos de Saúde Pública, vol. 39, no. 7, e00065423, pp. 1-3, 2023. doi: 10.1590/0102-311xen065423.

[17] S. A. E. Peters and M. Woodward, "A roadmap for sex- and gender-disaggregated health research," BMC Medicine, vol. 21, no. 354, 2023. doi: 10.1186/s12916-023-03060-w.

[18] S. Hawkes et al., "Sex-disaggregated data matters: tracking the impact of COVID-19 on the health of women and men," Economia Politica, vol. 39, pp. 55–73, 2022. doi: 10.1007/s40888-021-00254-4.

[19] S. E. Geller, A. Koch, B. Pellettieri, and M. Carnes, "Inclusion, Analysis, and Reporting of Sex and Race/Ethnicity in Clinical Trials: Have We Made Progress?" Journal of Women's Health, vol. 20, no. 3, pp. 315–320, 2011. doi: 10.1089/jwh.2010.2469.

[20] S. E. Geller et al., "The More Things Change, the More They Stay the Same: A Study to Evaluate Compliance With Inclusion and Assessment of Women and Minorities in Randomized Controlled Trials," Academic Medicine, vol. 93, no. 4, pp. 630–635, 2018. doi: 10.1097/ACM.0000000000002027.

[21] R. Poon et al., "Participation of Women and Sex Analyses in Late-Phase Clinical Trials of New Molecular Entity Drugs and Biologics Approved by the FDA in 2007–2009," Journal of Women's Health, vol. 22, no. 7, pp. 604–616, 2013. doi: 10.1089/jwh.2012.3753.

[22] V. Kannan et al., "Count me in: Using a patient portal to minimize implicit bias in clinical research recruitment," Journal of the American Medical Informatics Association, vol. 26, no. 8–9, pp. 703–713, 2019. doi: 10.1093/jamia/ocz038.

[23] S. Feldman, W. Ammar, K. Lo, E. Trepman, M. Van Zuylen, and O. Etzioni, "Quantifying Sex Bias in Clinical Studies at Scale With Automated Data Extraction", JAMA Netw. Open, vol. 2, no. 7, p. e196700, July 2019, doi: 10.1001/jamanetworkopen.2019.6700.

[24] V. Daitch et al., "Underrepresentation of women in randomized controlled trials: a systematic review and meta-analysis", Trials, vol. 23, no. 1, p. 1038, Dec. 2022, doi: 10.1186/s13063-022-07004-2.

[25] A. Feraldi, V. Zarulli, K. Buse, S. Hawkes, and A. Y. Chang, "Sex-disaggregated data along the gendered health pathways: A review and analysis of global data on hypertension, diabetes, HIV, and AIDS," PLOS Medicine, vol. 22, no. 5, e1004592, 2025. doi: 10.1371/journal.pmed.1004592.

[26] World Health Organization, Data Quality Review Framework, Geneva: WHO, 2014, pp. 15–22.

[27] M. P. Johnston, "Secondary Data Analysis: A Method of Which the Time Has Come". *Qualitative and Quantitative Methods in Libraries* (QQML), vol. 3, no. 3, pp. 619-626, 2014. doi:10.29297/QQML2014.3.3.619-626.

[28] C. Cichy and S. Rass, "An Overview of Data Quality Frameworks", IEEE Access, vol. 7, pp. 24634–24648, 2019, doi: 10.1109/ACCESS.2019.2899751.

[29] World Health Organization, "Global Health Observatory (GHO) Data Repository," WHO, Geneva, 2023. [Online]. Available: https://www.who.int/data/gho. Accessed: 01-Nov-2025.

[30] World Bank, "Gender Data Portal," World Bank, Washington, DC, 2023. [Online]. Available: https://datatopics.worldbank.org/gender/. Accessed: 01-Nov-2025.

[31] M. Bachmann, "RapidFuzz: A fast string matching library for Python," Journal of Open Source Software, vol. 6, no. 68, p. 3383, 2021. doi: 10.21105/joss.03383.

[32] M. D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," Scientific Data, vol. 3, no. 160018, 2016. doi: 10.1038/sdata.2016.18.

[33] World Health Organization, Global Adult Tobacco Survey (GATS), WHO, Geneva. [Online]. Available: https://www.who.int/teams/noncommunicable-diseases/surveillance/systems-tools/global-adult-tobacco-survey. Accessed: 11-Jun-2026.

[34] United Nations Statistics Division, SDG Indicators Metadata Repository, United Nations, New York. [Online]. Available: https://unstats.un.org/sdgs/metadata/. Accessed: 11-Jun-2026.

[35] HL7 International, HL7 FHIR Release 4: Fast Healthcare Interoperability Resources, HL7 International, Ann Arbor, MI, 2019. [Online]. Available: https://hl7.org/fhir/R4/. Accessed: 11-Jun-2026.

FACHHOCHSCHULE
WIENER NEUSTADT
Austrian Network for Higher Education

DECLARATION ON THE USE OF ARTIFICIAL INTELLIGENCE

In the course of preparing this thesis, AI-based tools (ChatGPT by OpenAI and Claude by Anthropic) were used in a supporting capacity for the following purposes:
The exploration and structuring of research ideas in the conceptual phase.
The identification of potentially relevant literature as a starting point for subsequent manual verification by the author.
The development and structural planning of data collection scripts, which were subsequently executed and critically reviewed by the author.
All analytical decisions, interpretations, and written content are the author's own work. No AI-generated text was adopted without substantial revision and critical evaluation. The use of AI tools is fully consistent with the Declaration of Integrity submitted with this thesis.

Wiener Neustadt, 15.06.2026 _____
Date / Signature

FACHHOCHSCHULE
WIENER NEUSTADT
Austrian Network for Higher Education

Appendix

Appendix 1| Fuzzy Matching Results: Final WHO–World Bank Indicator Pairs..... 1
Appendix 2 | Gender Visibility Score and Coverage Summary by Indicator and Database 3
Appendix 3| Missingness Asymmetry Index — WHO Indicators with Sex-Disaggregated Data 3

#	WHO Code	WHO Indicator Name	WB Code	WB Indicator Name	Score
1	WHOSIS_000001	Life expectancy at birth (years)	SP.DYN.LE00.IN	Life expectancy at birth, total (years)	90.1
2	MDG_0000000029	Prevalence of HIV among adults aged 15 to 49 (%)	SH.DYN.AIDS.ZS	Prevalence of HIV, total (% of population ages 15-49)	100.0
3	SDGSUICIDE	Crude suicide rates (per 100 000 population)	SH.STA.SUIC.P5	Suicide mortality rate (per 100,000 population)	81.3
4	NCD_DIABETES_PREVALENCE_AGESTD	Prevalence of diabetes, age-standardized	SH.STA.DIAB.ZS	Diabetes prevalence (% of population ages 20 to 79)	100.0
5	MDG_0000000020	Incidence of tuberculosis (per 100 000 population per year)	SH.TBS.INCD	Incidence of tuberculosis (per 100,000 people)	83.8
6	TB_c_new_snep_tsr	Treatment success rate for new pulmonary smear-negative and extrapulmonary TB cases	SH.TBS.CURE.ZS	Tuberculosis treatment success rate (% of new cases)	85.7
7	M_Est_cig_curr_std	Estimate of current cigarette smoking prevalence (%) (age-standardized)	SH.PRV.SMOK	Prevalence of current tobacco use (% of adults)	91.5
8	MDG_0000000007	Under-five mortality rate (probability of dying by age 5 per 1000 live births)	SH.DYN.MORT	Mortality rate, under-5 (per 1,000 live births)	82.4
9	WHOSIS_000003	Neonatal mortality rate (per 1000 live births)	SH.DYN.NMRT	Mortality rate, neonatal (per 1,000 live births)	97.9
10	SA_0000001400	Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol), by beverage type	SH.ALC.PCAP.LI	Total alcohol consumption per capita (liters of pure alcohol, projected estimates, 15+ years of age)	79.6

Appendix 1 | Fuzzy Matching Results: Final WHO–World Bank Indicator Pairs

#	Indicator	Source	GVS	Cov. Female (%)	Cov. Male (%)	Asymmetry Index
1	Life expectancy at birth	WHO	1	91.7	91.7	0.000
1	Life expectancy at birth	World Bank	0	—	—	—
2	HIV prevalence (ages 15–49)	WHO	0	—	—	—
2	HIV prevalence (ages 15–49)	World Bank	0	—	—	—
3	Suicide mortality rate	WHO	1	91.7	91.7	0.000
3	Suicide mortality rate	World Bank	0	—	—	—
4	Diabetes prevalence	WHO	1	95.8	95.8	0.000
4	Diabetes prevalence	World Bank	0	—	—	—
5	Tuberculosis incidence	WHO	0	—	—	—
5	Tuberculosis incidence	World Bank	0	—	—	—
6	TB treatment success rate	WHO	0	—	—	—
6	TB treatment success rate	World Bank	0	—	—	—
7	Tobacco use prevalence	WHO	1	2.4	2.2	−0.001
7	Tobacco use prevalence	World Bank	0	—	—	—
8	Under-5 mortality rate	WHO	1	100.0	100.0	0.000
8	Under-5 mortality rate	World Bank	0	—	—	—
9	Neonatal mortality rate	WHO	0	—	—	—

9	Neonatal mortality rate	World Bank	0	—	—	—
10	Alcohol consumption per capita	WHO	0	—	—	—
10	Alcohol consumption per capita	World Bank	0	—	—	—

Appendix 2 | Gender Visibility Score and Coverage Summary by Indicator and Database

#	Indicator	Cov. Female (%)	Cov. Male (%)	Missingness Female (%)	Missingness Male (%)	Asymmetry Index
1	Life expectancy at birth	91.7	91.7	8.3	8.3	0.000
3	Suicide mortality rate	91.7	91.7	8.3	8.3	0.000
4	Diabetes prevalence	95.8	95.8	4.2	4.2	0.000
7	Tobacco use prevalence	2.4	2.2	97.6	97.8	−0.001
8	Under-5 mortality rate	100.0	100.0	0.0	0.0	0.000

Appendix 3| Missingness Asymmetry Index — WHO Indicators with Sex-Disaggregated Data