Data gaps, tenuity and measurement challenges in achieving sustainable maternal child healthcare ecosystem: an evidence based statistical evaluation

Naseer Uddin⁴ | Majid Zaman⁵ | Sheikh Amir Fayaz⁶

⁴Public Health Analyst and Oral Surgeon, Directorate of Health Services, Department of Health and Medical Education, J&K, India.
⁵Director of IT & SS, University of Kashmir, J&K, India.
⁶Department of Computer Sciences, University of Kashmir, J&K, India.

Abstract Data metrics in public health programs and interventions play a crucial role in reflecting the health and well-being of the population, along with the structural and biosyndemic factors influencing health, public health interventions, and system resources. The management of health data has emerged as a key focus area within the sustainable global agenda, aiming to ensure effective, efficient, timely, and safe quality healthcare through evidence-based clinical practices and high-quality data systems that support quality improvement strategies and interventions to achieve optimal standards of care. Recognized internationally as an integral part of the unfinished health goals and the Sustainable Development Agenda 2030, a high-quality health information system is deemed essential for positively impacting health outcomes. However, emerging economies like India face the significant challenge of developing complex data platforms to track SDG health goals effectively, thereby providing useful insights for policymakers, academics, researchers, and program implementers. This paper aims to contribute to the growing knowledge on the implications of poor data quality and the performance of health information systems. It outlines the availability, completeness, and accuracy of data related to key performance indicators in the maternal and child health domain, evaluates the capacity and readiness of healthcare systems to generate high-quality data metrics, and explores the challenges, determinants, and gaps in achieving a sustainable and effective data management system. The study utilizes a mixed-methods approach and the WHO-DQAT data quality framework, employing objective-subjective assessments, desk analysis, cross-sectional field validation surveys, and integrated mixed-triangulation data tools. By measuring data quality dimensions and identifying potential gaps and determinants, the paper sheds light on deficiencies in the structure, function, and capabilities of the system, particularly in data management processes. It emphasizes the importance of strengthening statistical capacities and creating dedicated data management systems at state and national levels, and urges stakeholders to recognize the significance of data in health information systems for evidence-based policymaking.

Keywords: health data-gaps, data quality, maternal-child health, measurement challenges, good sustainable data practices

1. Introduction

The greatest challenge in the success of any robust health information system is the availability, quality, comprehensiveness, and reliability of data. Poor and inadequate quality data can have numerous negative consequences and significantly impact an organization or program economically and socially. The importance of a high-quality health information system is increasingly recognized globally as a critical aspect of achieving health goals and aligning with the Sustainable Development Agenda 2030. The quality and availability of healthcare data metrics have a profound impact on decision-making throughout the public healthcare continuum. In the context of maternal and child health care, the country's health system faces significant negative consequences and economic and social impacts due to data gaps caused by poor and inadequate quality data on service processes and structure. Despite fifteen years of MDG (Millennium Development Goals) implementation, there is still a persistent challenge of producing only around 50-70% of data, resulting in substantial data gaps and quality issues across various health indicators. This limited and unreliable data hinders the assessment, monitoring, and evaluation of intervention programs. Studies have highlighted the discordance, gaps, and quality issues across multiple health indicators within the policy framework, which impede countries' capacity to measure performance accurately (Murray et al 2000). A significant portion of MDG indicators lack data for more than half of the countries (Muchhala et al 2015) (Poku et al 2015), leading to the reliance on estimated or modelled data for performance assessment (Chen et al 2013) (Dimitri et al 2011) (Hák et al 2016). Poor data quality and quantity have been identified as contributing factors to India’s challenges in achieving MDG indicators, particularly in relation to maternal and child mortality (Sachs et al 2012).
1.1. Where We Stand: Data Health Gaps

In India, numerous policies and programs have been implemented over the years to address the health needs of underprivileged and deprived populations. Implementers, demographers, public health researchers, and academia are continuously striving to assess the availability and quality of health data, as well as to address the significant regional disparities and inequalities that exist in the country. Despite the valuable insights provided by surveys such as the National Family Health Survey (NFHS) and National Sample Survey (NSS), they often fall short in providing more detailed and actionable data to effectively tackle these extreme inequalities and disparities in a targeted manner.

Evidence suggests that India faces significant challenges in terms of data quality and quantity, particularly regarding periodicity and coverage. A study conducted by (Mikkelsen et al 2015), revealed that India’s performance in the Vital Statistics Performance Index (VSPI), a composite index assessing the timely generation of mortality and birth data, is very low (<0.25). This study categorized India among the group of 39 worst-performing countries out of 148 countries. The health of mothers and children is a priority both globally and nationally. Improving their health outcomes remains a persistent socio-economic priority, as evidenced by various social policy intervention programs such as the National Health Mission and other inclusive health policy schemes implemented by the government.

In India, the government and various statistical agencies have long recognized and acknowledged the lack of a coherent and robust systemic data system. The National Health Policy of 2002 and 2017 have emphasized the critical need for systematic and scientific population health statistics. Numerous studies have been conducted on data availability and quality, all indicating the urgent need for India to improve the quality, availability, and timeliness of under-utilized health datasets necessary to track national health goals (Raban et al 2009). These studies have also highlighted the importance of addressing the core issues of data fragmentation and discordance to make the data useful for policymakers, academics, researchers, and frontline workers. Evidence reveals gaps in data management practices and a lack of focus on data recording and reporting by data producers (Pandey et al 2010) (Tripathi et al 2018). The incorporation of evidence-based practices and interventions to strengthen the holistic healthcare ecosystem, particularly in fragile and low-resource regions like the Kashmir valley, has been lacking. This has hindered the achievement of Millennium Development Goals (MDGs) and now Sustainable Development Goals (SDGs).

Poor data quality and quantity have contributed to India’s inadequate performance in terms of service availability and coverage indicators related to maternal and child health. While health outcomes have shown improvement, there is a growing demand for open and high-quality basic health data on the population, driven by the global transition to the Sustainable Development Goals (SDGs) (Kumar et al 2005) (Bhatta et al 2014) (Kerber et al 2007). The healthcare ecosystem is facing challenges from both consumers and policymakers due to rapid social-epidemiological transitions and public expectations, leading to an imbalance between the supply and demand for high-quality health information and analytics (Kadir et al 2019) (Rashid et al 2012).

Health data has both statistical and technological implications, and the general public is becoming more sophisticated in their consumption of health information. There is a broader focus on public health dimensions such as equity, quality, dignity, efficiency, and productivity. This increasing demand for health data has resulted in a fragmented and complex information system with multiple stakeholders and overlapping vertical health programs. Efforts are being made to manage and harmonize scattered and incomplete databases, but challenges persist in integrating and consolidating the diverse sources of health information.

The management of health data has emerged as a key focus area in the sustainable global agenda, requiring accelerated efforts to meet international and country-specific needs (AbouZahr et al 2005). With India’s commitment to Universal Health Coverage (UHC) and the ongoing social-epidemiological transitions in the country, there will be a growing demand for basic population data for micro-level planning, program implementation, national strategy development, and progress monitoring against national priorities.

The data revolution for sustainable development is currently underway, with an abundance of new data providing greater granularity, aligning with the PARIS21 declaration. For India to come closer to achieving the ambitious SDG health targets, it is crucial to make disaggregated data publicly available. Investing in healthcare quality data is imperative for India, as the success of the SDGs in the next 15 years largely depends on the performance of the health data measuring system, such as the Health Information Management System (HIMS), at the state or regional level (Wahl et al 2018) (Ghassemi et al 2020). India faces a significant challenge in effectively managing its complex data platforms to track SDG health goals, particularly in the context of maternal and child healthcare. It is essential to ensure that the data collected is usable for policymakers, academics, researchers, and frontline workers.

India currently lacks a functional public policy framework to guide the use and sharing practices of health data. These data systems often operate independently without interoperability, leading to silos of information. Despite the government’s policy on the use of open-source software for public data systems, many of these systems, including the Health Management Information System (HMIS), remain expensive, proprietary, and inflexible. The predominant discourse on health information systems has primarily focused on the technology and the potential of standardized health and demographic data for applications in artificial intelligence. While efforts have been made to improve data quality and utilization for informing policy
and programs, there is a need to address the specific concern of utilizing data for health system management and improving the quality of care in a more effective manner. Studies such as those by (Wahl et al 2018) (Ghassemi et al 2020) highlight this topical concern and the potential for further advancements in leveraging data to enhance health system management and care quality.

The measurement of progress is central to a quality health system. Transparency, openness, and uncensored data measures are essential to provide policymakers and researchers with insights into the implementation, functioning, and monitoring of a democratic and accountable health system. These measures help assess the quality and effectiveness of healthcare delivery.

In India, despite a wealth of micro-level knowledge and literature on the role of quality data in maternal and child healthcare, there have been mixed conclusions and limited empirical findings on its holistic impact on maternal and child mortality and morbidity. However, focusing on data quality will be a critical factor in influencing health outcomes in India. The country's experience has highlighted the importance of enhancing data quality in the health system, particularly through the establishment of a functional data quality assurance system. Strengthening health systems and reducing inequities are key strategies employed by the National Health Mission (NHM), with data quality assurance and fair financing playing significant roles. To effectively plan for the institutionalization of quality improvement systems, it is crucial to critically evaluate India's data management system. This evaluation will help identify areas for improvement and guide efforts to enhance the quality, reliability, and usability of health data in India's health system.

1.2. Objectives and originality of study

The healthcare ecosystem suffers from a chronic lack of data-driven evidence-based research, which is crucial in supporting relevant policy interventions and decision-making processes. While data-driven evidence has demonstrated the success of public health intervention programs, ongoing debates surround the effectiveness of management practices in policy and program implementation. Therefore, it becomes essential to understand the interplay between data management strategies and health processes through evaluation and assessments based on evidence-based principles. This study highlights the significance of iterative data analysis and its role in evidence-to-policy approaches, emphasizing the need for continued engagement between researchers and policymakers to accommodate the rapid evolution of health policy planning and implementation.

This research has significant translational implications for estimating the impacts of poor and inadequate quality data on the welfare of beneficiaries. It aims to enhance various processes to reduce maternal and child health (MCH) morbidity and mortality by proposing best practices for data collection and interventions aimed at delivering high-quality health services through the use of reliable data metrics. Given the absence of a robust mechanism to assess the Health Management Information System (HMIS) in India, this study aims to fill this gap by evaluating the quality of HMIS data and assessing the system's capacity to generate accurate and reliable data. The findings from this research will contribute to strengthening the evidence base and informing strategies to improve data quality and enhance the effectiveness of the HMIS in India.

The present research study was conducted to investigate the measurement of data value and the role of quality data in the health information system. It aimed to assess the availability, complexity, and quality of data in terms of its generation, collection, analysis, monitoring, and dissemination across different levels. The study provides an introductory overview of the gaps in health data that are relevant to national health goals. It highlights the challenges in reporting and monitoring progress towards the achievement of the Sustainable Development Goals (SDGs). Furthermore, the study presents potential strategies and good practices that can be employed to effectively utilize data for sustainable health development in line with the 2030 agenda.

2. Materials and Methods

2.1. Methodological Approach

2.1.1. Conceptual Framework Used for Evaluating Data Quality and System Capacity

The conceptual framework of this study was developed based on a mixed theory approach, incorporating objective-subjective assessments within the WHO-DQAT (2008) data quality framework. This framework allows for the verification of data at field sites while also assessing the monitoring and evaluation component of the data management system. The reviewed studies have shown that integrating intuitive, empirical, and theoretical approaches in a quality framework leads to greater validity, credibility, rigor, and scope. To ensure efficient practice, this research implements the criteria from the (Price and Shanks 2004) framework, which prompts researchers to examine the objectivity of the dataset and covers a broad range of research and data management processes. The study combines methods from the Data–HIS Assessment Framework of WHO-DQA 2020, European Statistical System-DaTQAM 2020, Health Metrics Evaluation Network 2006 (WHO-HMN), and ME-PRISM Framework 2011, aligning with the conceptual mixed theories proposed by (Creswell et al 2003) to create a design topology. Figure 1 illustrates the criteria, measurements, and definitions of the quality framework employed in this study.
The evaluation strategy employed in this study involved a multiphase approach, including objective desk analysis, cross-sectional field validation surveys, and an integrated mixed-triangulation data tool. This approach utilized a range of complementary methods and interpretative synthesis, drawing on multiple sources of actionable evidence. By incorporating various methods such as those outlined by (Ako-Arrey et al 2016) (Eugene et al 1996) (Creswell et al 2011) (Campbell et al 1959), the study aimed to address technical and data gaps that may arise when relying on a single method. This multiphase approach also helped to compensate for any methodological weaknesses, resulting in rigorous and impactful information and conclusions. The use of multiple methods added credibility and judgment to the analysis of outcome evaluations.

2.1.2. Health Management Information System under National Health Mission

The World Health Organization (WHO) recognized the Health Management Information System (HMIS) as a critical component of health systems in 2007. In India, the National Health Mission (NHM) has placed great emphasis on the establishment and utilization of an effective HMIS. Since its inception in 2005, the HMIS has been the primary data tool for supervising, managing, evaluating, and implementing NHM programs and other inclusive health initiatives (Prinja et al 2013) (Kumar et al 2011). An effective HMIS not only monitors the performance and quality of health services but also provides a reliable evidence base for decision-making (Pandey et al 2010). The use of routine Management Information Systems (MIS) is not limited to policymakers and program managers, as researchers also rely on it to address critical questions about the effectiveness and efficiency of health programs in India. However, the HMIS has been consistently reported to have gaps in terms of data completeness, over-reporting, irregularities in report generation, data duplication, and inconsistencies at all levels of healthcare delivery. Despite its vital role in routine decision-making, there is currently no robust mechanism to assess the HMIS in India (Verma et al 2007). This study aims to fill this gap by evaluating the quality of HMIS data and assessing the system’s capacity to generate this data in the Union Territory of Jammu and Kashmir.

2.1.3. Health Management Information System under National Health Mission

Defining data quality and its dimensions is not without challenges. There is no universally accepted definition of ‘data quality’ as it is a contextual concept. The specific qualities that constitute good data quality depend on the intended use of the data and are defined by the user within that context (Klein et al 1999). The need for “defects free data” that is fit for the
intended task highlights the importance of data quality in ensuring accurate and reliable information for decision-making and analysis.

Data quality is recognized as a multidimensional concept in public health and other sectors (Chen et al 2013) (Wang et al 1996) (Karr et al 2006) (Hardee et al 2008). There is no universally accepted definition of data quality, as it is a contextual concept, with the qualities of good data being defined by the user based on the intended use of the data. Modern definitions of data quality go beyond the traditional characteristic of accuracy and encompass a wider range of attributes (Klein et al 1999). Quality data is considered to be data that is fit for use within the information chain process and from the user’s perspective. The concept of quality data aligns with the consumer-focused view that data should be fit for use, as widely adopted in the literature (Canada et al 2009) (Strong et al 1997). High-quality data is crucial for knowledge management and enables organizations to maximize the benefits derived from their resources (Wang et al 1996) (Brown et al 2002) (Kerr et al 2004) (Gray et al 2005). Figure 2 presents the ideal dimensions of data quality.

In the application of data quality, it is not always necessary to consider all dimensions of data quality. Instead, the most essential quality dimensions can be combined for specific applications or purposes, as suggested by (Jesiļevska et al 2017). Various authors have studied the dimensions of data quality, and the dimensions proposed by (Wang and Strong et al 1996) are widely accepted as quality dimensions of data. It is widely recognized in the literature that poor data quality hinders effective decision-making at both the organizational and country levels and becomes a significant cost factor due to the time and resources spent on error detection and correction (Wang et al 1996). Unidentified and uncorrected data can have negative socio-economic impacts, emphasizing the importance of ensuring data accuracy and reliability. Decision-making based on inaccurate and unreliable data can lead to detrimental consequences.

2.1.4. Use of Data Quality Audit Tool for Assessment of Data Dimensions and System capacity

The Data Quality Audit (DQA) tool is a measurement instrument developed by multilateral global organizations, such as the World Health Organization (WHO) in 2003. Its purpose is to provide a standardized approach for assessing, monitoring, and improving data quality. The DQA tool enables programs and projects to assess the quality of their data and enhance their data management and dissemination systems (Jesiļevska et al 2017). It evaluates the quality of reported data for key indicators at selected facilities and simultaneously assesses the capacity of the data management system to produce, collect, and report high-quality data. The tool generates findings, recommendations, and actions to strengthen the data management system of the implementing program (Ballou et al 1985) (Goodchild et al 1993). During field visits, the DQA tool was employed to assess the quality of data and the functional capacity of the data management system, particularly within the monitoring and evaluation units of the National Health Mission.

The Data Quality Audit (DQA) tool assesses the minimum system at three levels: service points (data producers), intermediate data aggregators, and M&E units (data acquirers). The tool is based on the components of data quality dimensions, including completeness, accuracy, reliability, timeliness, and availability, aligned with the objectives of the implementing program. It consists of two protocols: the System Assessment protocol and the Data Verification protocol, which are administered at all three levels of assessment.

The System Assessment protocol, conducted in two phases (offsite desk review and on-site assessment), identifies potential determinants and challenges of data quality within the system. It assesses service delivery units and M&E units across
five functional areas, highlighting strengths and weaknesses in each area of data management and reporting. It also evaluates data capturing events and data service structure and processes.

The Data Verification protocol focuses on the accurate and timely collection and reporting of data by service delivery units and data aggregation units. This is done through the verification of selected data elements for key performance indicators related to MCH service delivery. The data verification process includes in-depth verification at service delivery units (health facilities) and follow-up verification at intermediate data aggregation units and higher-level M&E units.

2.1.5. Selection of Key Performance MCH indicators for Data Quality Assessment

The selection of indicators for evaluation during the Data Quality Assessment is a crucial decision that depends on the objectives and evaluation needs. In this study, five core indicators representing a snapshot of MCH (Maternal and Child Health) service delivery were selected. These indicators were chosen based on their alignment with the objectives of the National Health Mission (NHM) program theory and routine monitoring and evaluation system. Additionally, they were also included in the national health family survey (NFHS-5), which allowed for comparability and cross-validation of the data (cross-cutting indicators). Furthermore, these indicators were relevant to the Sustainable Development Goal 3 (SDG 3) and Universal Health Coverage (UHC) goals, highlighting their significance in assessing health outcomes and access to healthcare services.

Table 1 MCH Core Indicators (Data Elements) used in Evaluation of Data Quality.

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of Pregnant women who had at least 4 ANC check-ups completed</td>
</tr>
<tr>
<td>2</td>
<td>Number of Pregnant women who have been given 180 IFA tablets (during pregnancy)</td>
</tr>
<tr>
<td>3</td>
<td>Number of Pregnant women who had Hb &lt; 11g/dl</td>
</tr>
<tr>
<td>4</td>
<td>Number of deliveries conducted in Health Facility</td>
</tr>
<tr>
<td>5</td>
<td>Number of Home deliveries conducted by SBA</td>
</tr>
<tr>
<td>6</td>
<td>Number of infant/children (0-23months) who were given BCG vaccine</td>
</tr>
</tbody>
</table>

2.1.5. Data Sets and Time periods used

The data available in the public domain from the NFHS-5 survey and routine data from HMIS of NHM was taken into the account.

The time period for assessing data quality using the DQA tool aligned with the data that was submitted or reported to higher levels during the previous month when the health facility or service unit was visited. This allowed for a current assessment of the data quality at the facility level.

For the desk review and external cross-validation, the reporting period considered was from April 2019 to March 2020, which corresponds to the annual HMIS report. This timeframe was chosen to facilitate a comparison with the National Health and Family Survey (NFHS-5), which conducted its fieldwork between July 2019 and January 2020. By aligning the reporting periods, it allowed for a comprehensive evaluation of the data quality and facilitated cross-validation between the HMIS report and the NFHS-5 data.

2.2. Study Design

The study was conducted in two phases. Phase I involved a desk review and interpretative deductive analysis using secondary data. This phase focused on reviewing relevant literature to understand the phenomenon of Health Information Systems (HIS) and its interaction with the environment. It aimed to identify crucial issues and gaps that affected the HIS. Detailed secondary data analysis was performed using the HMIS and NFHS datasets to draw conclusions about data validation and its parameters.

Phase II consisted of a cross-sectional assessment through a field validation survey using primary mixed data. This phase included two modules. Module I involved the assessment of facility services and implementation through observations and questionnaires. Module II focused on data quality and system assessment using the Data Quality Audit (DQA) Tool. The purpose of this phase was to assess the functionality of the health information system, explore the dimensions of data quality for MCH indicators, and identify any gaps, potential determinants, or risks associated with data quality.

By conducting both desk review and field validation survey, the study aimed to gain a comprehensive understanding of the health information system, assess the quality of MCH data, and uncover key factors influencing data quality.
Table 2 Sources of Data.

<table>
<thead>
<tr>
<th>Data Sets</th>
<th>Source and period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secondary Data</td>
<td></td>
</tr>
<tr>
<td>National Family Health Survey (NFHS-5) (Population based Survey) (Data set V)</td>
<td>NFHS-5 (2019-20)</td>
</tr>
<tr>
<td>Submitted HMIS monthly reports (Routine Data) (Data set III)</td>
<td>NHM-J&amp;K (2019-20)</td>
</tr>
<tr>
<td>Primary Data</td>
<td></td>
</tr>
<tr>
<td>Data obtained from interviewed beneficiaries (evaluative coverage assessment)- level I discordance.</td>
<td>Field survey (2019-20)</td>
</tr>
<tr>
<td>Data obtained through checklist and observational assessment of Health service delivery points (assessment for readiness and capacity)</td>
<td>Field survey (2019-20)</td>
</tr>
<tr>
<td>Data obtained through observation and interview of data management unit personnel (assessment of data management system)</td>
<td>Field survey (2019-20)</td>
</tr>
<tr>
<td>(Data set II)</td>
<td></td>
</tr>
<tr>
<td>Data got through DQA Tool assessment</td>
<td>Field survey (2019-20)</td>
</tr>
</tbody>
</table>

2.3.1. Phase II Field Survey Setting and Procedure

The cross-sectional field validation survey was conducted between April 2019 and March 2020 in six diverse districts of the Kashmir valley, with a specific focus on the high-priority aspirational hilly districts of Kupwara and Baramulla. The study aimed to evaluate the Health Management Information System (HMIS) under the National Health Programme for Mother and Child in Jammu and Kashmir. The units of analysis included health facilities such as Sub-Centers, Primary Health Centers, Community Health Centers, and District Hospitals (as data producers), as well as Monitoring and Evaluation (M&E) units at the block and district levels under the State Health Mission (as data acquirers and implementers). The end-users or beneficiaries of the data were mothers and children.

Primary data, both quantitative and qualitative, was collected from 215 MCH service providing health institutions and 45 program M&E units. Interviews were conducted with 505 beneficiaries, including mothers with children under 2 years, who had availed services at these health facilities. Additionally, interviews were conducted with 90 data managerial staff personnel from the M&E units. All data elements of the Data Quality Audit (DQA) tool were captured in Excel format during the survey. Prior to the field validation survey, a pilot study was conducted in the Ganderbal district to address any potential obstacles or issues. The study included observational surveys and pilot interview sessions to ensure the smooth implementation of the main survey.

2.3.2. Data capturing elements and instruments— Data Collection

The final questionnaire and schedules used in the study were developed based on considerations of convergent and discriminant validity, incorporating feedback and insights from the pilot study. They were specifically tailored to address unresolved system design issues related to the Health Management Information System (HMIS). The Data Quality Audit (DQA) tool was utilized alongside three different types of schedules: M&E Unit, Health Facility, and End-user Woman. These schedules were administered based on predefined criteria to ensure comprehensive coverage of relevant perspectives.

The questionnaires aimed to capture a 360-degree view of stakeholders’ perspectives on data management system issues and implementation challenges. They were designed to gather information about data activities, planning, and working processes, including the structure, function, and capacity of the system. The questionnaires also sought to identify gaps in knowledge and skills related to data management. Throughout the study, all elements, indicators, and variables followed the
definitions and frameworks used by the World Health Organization (WHO) and Measure Evaluation (ME) frameworks to ensure consistency and comparability of data across different contexts.

3. Findings

The findings of this study were derived from a combination of interpretative secondary data analysis of HMIS-NFHS datasets and primary quantitative-qualitative data collected through objective and subjective assessments of health and M&E units. The analysis focused on assessing the internal data validation and external consistency of data pertaining to maternal and child health (MCH) indicators related to service availability and delivery.

Descriptive and inferential data analysis was conducted on the primary data collected, and the results were compared with the findings from both Phase I (desk review) and Phase II (cross-sectional field validation survey) of the study. The analysis aimed to evaluate the internal and external consistency of the data related to MCH indicators.

To measure the dimensions of data quality, a predefined Data Quality Audit (DQA) tool was utilized, which included outcome and dashboard analysis features. The qualitative data obtained from interviews and other qualitative assessments were processed manually, organized thematically, and categorized based on the role of the respondents.

The findings derived from the analysis of both phases of the study provided valuable insights and information regarding the quality and consistency of data in relation to MCH indicators.

3.1. Interpretative secondary data analysis

3.1.1. Challenges and Tenuity in Sources and Availability of Data on Health

The health data in Jammu and Kashmir is derived from multiple sources, including census, civil registration system, sample registration system, sample surveys, and official statistics. However, these data sources suffer from various limitations such as under enumeration, misreporting, coverage errors, and content errors. The data quality and completeness are hindered by irregular tabular categories and changing definitions of indicators over different census periods. The sample registration system also faces challenges with incomplete birth and death registration and limited availability of data at lower administrative levels. The civil registration system is characterized by gross underreporting and incompleteness, leading to unreliable data on vital events. The health management information system (HMIS) data captured in public facilities is facility-based and may not be representative of the general population (Mari et al 1984). These issues have resulted in divergent and fragmented population demographic figures, causing difficulties in interpreting vital statistics like infant mortality rate (IMR) and birth sex ratio.

3.1.2. Gaps in Data Gathering and Capturing Practices

During the descriptive review and analysis of the field assessment data, several challenges and practices were identified that affected the quality of output data. These challenges included variations in standardized data collection practices, limited knowledge about data elements and their significance, reporting of inflated data to higher facilities, and a narrow focus on a limited range of indicators. These factors resulted in incomplete and inaccurate information being recorded. Additionally, there was a lack of evidence-based and technology-enabled practices at all levels of data management implementation, and no mechanism was in place to verify the collected data for correctness and completeness. Overall, there was a lack of a robust data quality assurance system in place.

3.1.3. Insufficient Private Sector Data and Lack of Segregated Data

The data collection and reporting systems in Jammu and Kashmir face challenges in capturing information from the private healthcare sector, which holds a significant share in healthcare services utilization. The private sector contributes to about 51% of the health-seeking behaviour in J&K, including OPD services, hospitalizations, and maternal and childcare services. However, the current HMIS and surveys like NFHS primarily focus on public facilities, resulting in limited data availability and lack of disaggregation for specific population groups, particularly those residing in remote and mountainous regions like Kupwara, Bandipora, and Rajouri. This poses a limitation in assessing the reach and impact of programs on vulnerable and marginalized populations, especially those working in the informal sectors. It is crucial to have indicators disaggregated by background characteristics to ensure that programs are reaching the poor and deprived sections of the population. Extensive reforms are needed in the HMIS to accommodate data from the private healthcare sector and to improve data coverage and disaggregation for a more comprehensive understanding of health services utilization and outcomes.

3.1.4. Gaps in Linkages, Dissemination and Use of Data

The data management system in Jammu and Kashmir faces several challenges. There is limited coverage and a lack of coordination among the various agencies responsible for data capturing and reporting. This hinders effective data sharing, processing, and compilation. Moreover, there has been no dissemination or analysis of data on levels and trends of maternal
and child health (MCH) mortality indicators in the region. The absence of a data resource center further exacerbates these issues, as there is no central entity to coordinate and collaborate with the national data resource center for decentralized data collection and dissemination. The lack of a centralized mechanism for linking data across different sources results in data overlapping, duplication of efforts, and consistency issues. Addressing these challenges and establishing a robust data management system with proper coordination and data stewardship is crucial for improving the overall data quality and strengthening the long-term data management system in Jammu and Kashmir.

3.2. DQA tool Assessment Analysis

3.2.1. Availability, Completeness, Accuracy, and Timeliness of data

The analysis of the assessment conducted using the DQA tool in individual districts showed varying levels of completeness, timeliness, and availability of data. The completeness of data ranged from 60% in Kupwara district to 84% in Ganderbal district. Timeliness of reports ranged from 89% in Bandipora to 100% in Srinagar. The availability of data ranged from 96% in Bandipora to 100% in all other districts. The accuracy of uploaded data from service delivery points for most MCH indicators was found to be between 85% to 100%. However, during facility data validation, significant differences were observed in the recording and reporting of data in lower-level health facilities across all surveyed districts. Overall, the uploaded data from service points showed a 90% accuracy and completeness, 98% availability, and 94% timeliness. These findings highlight the need for improving data recording and reporting practices in lower-level health facilities to ensure the reliability and consistency of the data.

![Figure 3](image-url) Overall average percentage of dimensions of data quality in surveyed districts.

![Figure 4](image-url) Internal Validity of HMIS Data from surveyed Health Facilities (HF).

3.2.2. Consistency of HMIS Data-Internal Data Validity

Despite its crucial role in informing routine decisions, the Health Management Information System (HMIS) faces numerous challenges such as irregularities in report generation, data duplication, inconsistencies, and a tendency to over-report outcomes. To assess the internal validity of HMIS data, we conducted a validation analysis using a set of logical checks and validity rules (Husain et al 2012). This analysis aimed to identify outliers, missing data points, and duplicate records within the HMIS dataset. The findings of the internal validation are presented in Figure 4, providing insights into the quality and integrity of the HMIS data.

3.3. Discordance Analysis

3.3.1. Gaps Due to Insufficient Data Capturing--Coverage Gap Analysis (Level 1 Discordance)

To identify data gaps in captured and recorded MCH data elements, the monthly reported data of HMIS from surveyed health facilities and data generated from the field coverage survey underwent appropriate statistical tests. These tests revealed significant gaps in capturing to recording for indicators such as PW given 180 IFA tablets (p=0.003), PW with Hb less than 11g/dl (p=0.05), and infants given BCG vaccine (p=0.000).

3.3.2. Recording to Reporting Data Variability (Level 2 Discordance)
The assessment of facility records involved the verification of source documents and various registers, comparing them with the submitted HMIS data at the portal. This analysis uncovered significant differences between the recorded and reported values of five MCH data elements. Monthly reports sent by the health facilities exhibited overreporting of data elements where higher coverage levels were desired, while underreporting was observed for those with lower coverage requirements. It was also evident that ANMs and service providers aimed to showcase better outcomes in MCH service indicators for their catchment areas, potentially to receive positive recognition. Overreporting and underreporting of MCH elements were prevalent, with maximum variation (48%) observed in home deliveries conducted by SBA, followed by pregnant women with Hb levels less than 11g/dl (25%) and pregnant women who had taken 180 IFA tablets (23%). The only indicator showing minimal variation was the proportion of infants taking BCG vaccine, with just 1% difference.

Further analyses, including linear regression and Bland-Altman analysis, were performed on the two data sets of HMIS (reported) and survey (validated). These analyses revealed a weak to moderate level of agreement in MCH indicators, indicating a significant level of discordance. Numeric differences were calculated for each indicator, and the significance of these differences was computed. Detailed information regarding this comparison and the level II data discordance, specifically recoding to reporting gaps, can be found in Table 3.

### Table 3 Level of Variation in Recording and Reporting of Data (DQA / HMIS ratio comparison) Level 2 data discordance (Recording to Reporting difference).

<table>
<thead>
<tr>
<th>S.No</th>
<th>MCH Indicator data elements</th>
<th>Level of Variation (%)</th>
<th>Mean Difference between Reported and Validated Data</th>
<th>95% Confidence Interval of the Difference</th>
<th>P value</th>
<th>significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of pregnant women who had at least 4 ANC check-ups completed</td>
<td>Over reporting by 16%</td>
<td>1.49</td>
<td>-.53</td>
<td>3.52</td>
<td>.147</td>
</tr>
<tr>
<td>2</td>
<td>Number of pregnant women who have been given/ consumed 180 IFA tablets (during pregnancy)</td>
<td>Over reporting by 23%</td>
<td>2.83</td>
<td>1.94</td>
<td>3.73</td>
<td>.000</td>
</tr>
<tr>
<td>3</td>
<td>Number of pregnant women who had Hb &lt; 11g/dl</td>
<td>Underreporting by 25%</td>
<td>-3.60</td>
<td>-5.38</td>
<td>-1.82</td>
<td>.000</td>
</tr>
<tr>
<td>4</td>
<td>Number of home deliveries conducted by SBA</td>
<td>Underreporting by 48%</td>
<td>-.37</td>
<td>-.47</td>
<td>-.27</td>
<td>.000</td>
</tr>
<tr>
<td>5</td>
<td>Number of infant/ children (0-23months) who were given BCG vaccine</td>
<td>Over reporting by 1%</td>
<td>15.06</td>
<td>-11.17</td>
<td>41.29</td>
<td>.259</td>
</tr>
</tbody>
</table>

### 3.3.3. External Consistency with NFHS- 5 – External Data Discordance

The external discordance analysis revealed weak to moderate linear associations between the HMIS and NFHS-5 data sets, as indicated by the Bland Altman analysis. The values were closer to zero, indicating the variability and level of discordance between the two data sets. The mean differences in MCH indicators between the two data sets were significant for 3 out of 5 indicators, suggesting the presence of data gaps in MCH coverage. The external validation of HMIS data with NFHS-5 survey and validation survey data for MCH data elements showed wide divergence between the comparison data sets. It is challenging to determine which data set is overestimating or underestimating, but the differences observed between the data sets are concerning. These discrepancies across various data sources reflect serious errors in data entry, aggregation, compilation, and validation. These errors can be attributed to systemic issues such as the existence of multiple registers, misinterpretation of data, data duplication, lack of written guidelines and procedures, and staff shortages. Table 4 provides further details of the external discordance analysis.

### 3.4. Assessment for Capacity and Readiness of Data Management System to Produce Quality Data

A mixed-method approach was employed to validate and understand the processes of the data management system in NHM and HMIS. The analysis included the use of the DQA tool and qualitative data to generate explicit themes for explanation and understanding of local contexts and information reporting systems. The assessment of districts using the DQA tool revealed similar findings across all districts, indicating deficiencies in the M&E structure, function, and capabilities, with significant gaps in data management processes. The spider chart in Figure 5 illustrates the dimensions of the data management system based on five parameters. The findings highlight significant gaps in data management processes, particularly in lower-level health facilities that are directly involved in data capturing, generation, and reporting.
Table 4 Level of external data discordance between HMIS reported coverage and NFHS-5 survey-based coverage.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Mean Health Facility based Record coverage (%)</th>
<th>Mean NFHS-5 based Coverage (%)</th>
<th>Mean difference</th>
<th>95% CI of Mean difference</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Pregnant women who had at least 4 ANC check-ups completed</td>
<td>76.18</td>
<td>82.1</td>
<td>-5.98</td>
<td>-22.85 - 10.89</td>
<td>0.40</td>
</tr>
<tr>
<td>Number of Pregnant women who have been given/ consumed 180 IFA tablets</td>
<td>76.13</td>
<td>14.46</td>
<td>61.66</td>
<td>47.35 - 5.98</td>
<td>0.00*</td>
</tr>
<tr>
<td>Number of Pregnant women who had Hb &lt; 11g/dl</td>
<td>74.31</td>
<td>52.13</td>
<td>22.18</td>
<td>11.86 - 2.50</td>
<td>0.003*</td>
</tr>
<tr>
<td>Number of Home deliveries conducted by SBA</td>
<td>37.56</td>
<td>2.41</td>
<td>35.31</td>
<td>73.08 - 2.40</td>
<td>0.061</td>
</tr>
<tr>
<td>Number of infant/ children (0-23 months) who were given BCG vaccine</td>
<td>116.7</td>
<td>94.2</td>
<td>22.48</td>
<td>37.26 - 3.90</td>
<td>0.011*</td>
</tr>
</tbody>
</table>

Figure 5 Overall Average Values of Five Parameters of Data System.

The HMIS of NHM generates a substantial amount of data, with thousands of data elements being generated monthly. However, only a small percentage of these elements are actually used to generate indicators, while a significant portion remains inactive or returns blank values. This suggests potential gaps in the readiness of health facilities to deliver certain services. Data capturing, collation, cleaning, and uploading processes are not being adequately performed, partly due to a lack of understanding among service providers about the importance and relevance of high-quality data. Insufficient technical staff and supervision contribute to these challenges. Structural and functional capacities for data handling are also lacking, with a majority of workers lacking knowledge and familiarity with the data elements being reported. Additionally, the study found that there are significant gaps in utilizing and sharing data evidence, and there is a lack of clear mechanisms for data transfer and keying. Only a small percentage of Data Management Units (DMUs) have comprehensive statistics on population parameters and denominators, and a minority have the capacity and readiness to produce quality data aligned with national health goals and SDGs (Figure 6).
4. Discussion

There are evidences of data gaps and tenuity with respect to MCH across regions of Jammu & Kashmir. The health statistics for Jammu and Kashmir presents a mixed-bag scenario in terms of the quality of data with little standardization across Jammu and Kashmir, making the data being reported of little practical use. Also, much of the data generated is not easily disseminated with no data flow and no use for monitoring the performance and accountability to the population.

Overall HMIS data showed good completeness, availability, timeliness, and accuracy as per recommended guidelines given by WHO-HM framework of data quality. However, the record to report ratio was depicting high variability of over and underreporting with significant levels of I and II data discordance in most of MCH indicators. The least inflated or most stable indicator was BCG vaccination mainly due to better immunization coverages at place of delivery. Most of the findings of this study were consistent with other studies undertaken in India especially the study undertaken in Haryana by (Sharma et al 2016) and study conducted by (Khandade et al 2013) in Bihar. One of the studies on health information conducted in Kerala in 2013 by (Harikumar et al 2012) found that HMIS of state had very low accuracy of 37% and completeness of 29% which is very much opposite to this study, the reason being NHM was in middle of its implementation in 2013. Our results support those of previous studies which suggest strategies and data practices to improve HMIS data quality within MCH domain (Bhojani et al 2010).

The capacity of the health data system in Jammu and Kashmir needs to be improved especially with reference to data implementation practices both in terms of structures, processes and capacities. Appropriate attention is needed on genuineness of HMIS data reporting and correct recording of user details. J&K achievements with health data remains limited, out of three main issues that concern a health data system—timeliness, completeness and correctness of data, the state has been able to achieve the first two up to some extent but has been lagging in achieving quality health data for which correctness and consistency is must. As far as J&K is concerned the challenges in capturing and collecting data are geographical in nature, different in rural and urban regions of J&K.

In this study, we examined the data quality dimensions through mixed assessment methods based on our proposed conceptual framework. We tried to use framework which included assessment of data quality dimensions as well as processes involved in health information system along with incorporation of triangulation methods for data quality assessment. So far, no such study has been done in UT of Jammu and Kashmir with such large datasets which covers nearly 10% subset of health facilities in UT. Simultaneously, it combined an established systemic approach of data quality assessment at different levels of functionality and evaluative coverage survey to find out level of agreement and data discordance at both levels of recording and reporting to get a broader picture of data ecosystem architecture.

5. Implications and Conclusion

Data metrics are crucial and integral to public health practice and interventions. The last decade has witnessed the boom of digital technology in every field, broadening the scope of its application from both producers’ and users’ perspectives. This has drawn the attention of governments and policy-making bodies to the importance of data quality and its assessment methods. Our study systematically reviewed the current status of data quality assessment for public health, including the dimensions of data quality and the methods used in data collection, collation, reporting, and sharing. Based on our proposed
conceptual framework, we identified gaps in the evaluation of data quality methods. The study highlights the need for the development of comprehensive data quality assessment methods that incorporate perspectives from both users and producers. It is essential to measure end users’ perceptions of data quality to enhance our understanding of data quality issues. Access to high-quality, open, and transparent data is crucial for creating a sustainable and accountable system and measuring success in improving overall health and well-being. The study supports the recommendations made in the National Health Policy 2017 and Universal Health Coverage 2030. Figure 7 presents recommendations and key practices for an ideal Health data platform to address the challenges and limitations identified in the study.

Key Suggestions for policy makers

- Tweaking and tuning existing surveys considering national health goals towards attainment of SDG agenda.
- Incorporating demand based, standardized universal health indicators into health data system.
- Reforming the National-State Health Management Information System (HMIS).
- Incorporating value measurement, triangulation Strategy, and appropriate techniques and mixed approach in estimation of health data metrics.
- Investing in health data System for data revolution.
- Data policy framework that enables interoperability and actionability of information.
- Engaging private sector to collect data so that comprehensive picture of health delivery is got.

Key Features for ideal health platform towards achieving a sustainable MCH goals

A long-term and gradual investment in HMIS is crucial to establish an efficient and effective data system with the capacity and readiness to produce high-quality data. This can be achieved through an institutional capacity-building approach that involves collaboration and shared responsibility among health and statistical stakeholders. It is essential to focus on grassroots levels and redesign processes to identify and address data errors. A systematic and structured approach is needed to streamline data collection, reporting, and real-time updating, taking advantage of technological advancements and capabilities. By implementing these measures, we can eliminate errors, improve data quality, and enhance the overall effectiveness of the health information management system.

One of the key messages of this study is the importance of strengthening both state and national statistical capacities, as well as creating dedicated data management systems. By doing so, we can maximize the value and utility of data from all available sources, whether at the regional, national, or international level. The study emphasizes the need to recognize the significance of data in health information systems and to establish foundations for effective and evidence-based policymaking. It highlights the actions required to improve the health and well-being of women and children, including accelerating data momentum, adopting innovative thinking, conducting cutting-edge research, and embracing a new paradigm for the mother-child ecosystem to achieve sustainable development goals. The paper also emphasizes the need for investment in health data collection, analysis, and research to inform government policies and strategies aimed at addressing the various challenges facing India’s healthcare performance.

In India, there have been ongoing efforts by implementers, demographers, public health researchers, and academia to assess the availability and quality of health data, particularly for underprivileged and deprived populations. These efforts aim to understand the significant regional disparities and inequalities that exist in the country. Recognizing the importance of high-quality data, initiatives such as the National Data Quality Forum (NDQF) have been established. The NDQF, formed in partnership with the Indian Council of Medical Research (ICMR) and the Population Council, focuses on identifying gaps in data collection and compilation, as well as understanding the reasons for data discordance. Some of the major barriers to achieving quality data identified by the NDQF include age-reporting errors, underreporting due to subjective question interpretation, and inadequate data to generate reliable estimates on mortality. These efforts and initiatives highlight the commitment to improving the quality and reliability of health data in India.

In 2020, the Government of India approved the “Health Data Management Policy” as part of the framework for the National Digital Health Mission (NDHM). The NDHM aims to digitize the healthcare system in India, leveraging health data and emerging technologies. The success of the NDHM relies on the adoption and active participation of diverse stakeholders, including states, public and private healthcare providers, policy makers, program managers, and citizens. One of the key components of the NDHM is the establishment of health data registries, which have the potential to provide valuable data for...
decision-making, health system planning, and improvements. These registries can play a crucial role in informing policies and interventions that aim to enhance healthcare delivery and outcomes in the country.

The healthcare sector in India has a significant opportunity to leverage the vast amount of data metrics available across the full spectrum of healthcare services. While currently, the healthcare sector contributes less than 12 percent of the data generated in India, it is expected to grow to around 40 percent by 2025, according to the Digital India report. The increasing availability of data presents an opportunity for healthcare implementers and providers to adopt a holistic approach and incorporate best practices in data management and utilization. The recent Covid-19 crisis, although challenging, has also provided opportunities for the use of data-driven evidence in developing low-cost, scalable, and quick solutions. It has also accelerated the expansion of existing data and health information systems at a global level, highlighting the importance of leveraging data for informed decision-making and addressing healthcare challenges.

The results of this study do not align with the "technocratic presumptions" emphasized by (Reddy et al 2015) regarding the achievement of national or international goals, as highlighted by organizations like NITI Ayog or WHO. Instead, the findings of this study align with the concerns raised by (Fukuda-Parr and Greenstein, 2013) regarding the broader policy and programmatic implications of goal setting and its unintended consequences. The study underscores the importance of critically considering what is being measured, how it is being measured, and when it is being measured, highlighting the need for a thoughtful and reflective approach to data collection and measurement in order to inform effective policies and programs.

Ethical Considerations
Not Applicable.

Funding
This paper is a part of wider research study funded by National Council of Applied Economic Research (NCAER) New Delhi.

Conflict of Interest
The authors declare no conflicts of interest.

References
Canada S (2009) "Health indicators": Canadian Institute for Health Education.
Creswell JW, Plano Clark V, Gutmann ML, Hanson WEATyCT (2003) An expanded typology for classifying mixed methods research into designs.


