ORIGINAL ARTICLE

# Improving routine health data in Indonesia: Utilising the WHO data quality tool for *Aplikasi Satu Data Kesehatan*

**Guardian Yoki Sanjaya[1], Khairani Fauziah[1], Rio Aditya Pratama[1], Nur Ayu Fitriani[1], Mohammad Yusuf Setiawan[1], Ina Amali Fauziah[1], Lutfan Lazuardi[1], Surahyo Sumarsono[1], Sandra Hakiem Afrizal[1,2], Farida Sibuea[3], Evida Veronika Manullang[3]**

[1]Department of Health Policy and Management, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia, [2]Public Health Master Programs, Faculty of Health Sciences and Technology, Universitas Binawan, Jakarta, Indonesia, [3]Ministry of Health of the Republic of Indonesia

## ABSTRACT

**Introduction: Assessment of data quality in the era of big data is crucial for effective data management and use. However, there are gaps in data quality assessment for routine health data to ensure accountability. Therefore, this research aims to improve the routine health data quality that have been collected and integrated into *Aplikasi Satu Data Kesehatan* (ASDK) as the primary health data system in Indonesia.**

**Materials and Methods: This descriptive study utilises a desk review approach and employs the WHO Data Quality Assurance (DQA) Tool to assess data quality of ASDK. The analysis involves measuring eight health indicators from ASDK and *Survei Status Gizi Indonesia* (SSGI) conducted in 2022. The assessment focuses on various dimensions of data quality, including completeness of variables, consistency over time, consistency between indicators, outliers and external consistency.**

**Results: Current study shows that routine health data in Indonesia performs high-quality data in terms of completeness and internal consistency. The dimension of data completeness demonstrates high levels of variable completeness with most variables achieving 100% of the completeness.**

**Conclusion: Based on the analysis of eight routine health data variables using five dimensions of data quality namely completeness of variables, consistency over time, consistency between indicators, outliers. and external consistency. It shows that completeness and internal consistency of data in ASDK has demonstrated a high data quality.**

**KEYWORDS:**
*Data quality, Completeness, consistency*

## INTRODUCTION

Big data is important to monitor population health status to evaluate population-based health service quality, and to conduct research for innovative solutions which bring great promise for public health.[1] However, poor data quality has resulted in several consequences, including misallocation of health resources, misguided decisions and plans, missed opportunities, incorrect reports and follow-up actions and the indirect impact to the future direction of health plans.[2,3] To avoid the poor data quality, the assessment of data quality of health programs is conducted to review activities during the implementation period of the health program that must be carried out on a routine basis and using a standardised data quality assessment tool.[4–6] WHO has released a data quality tools (DQA) as a method for determining data quality criteria through routine data analysis by using the data quality dimensions such as completeness, consistency and timeliness.[7]

The utilisation of the data quality tool plays a crucial role in monitoring population health status, assessing the quality of population-based health services and supporting evidence-based research for innovative public health solutions.[8–10] Previous research shows that the quality of routine data remains poor in the primary health care facility as the main source of the health data.[11,12] Based on the study, a health management information system is not an effective tool for monitoring health-care performance and as a source of data for planning and decision-making. The similar condition happened in Indonesia where in 2012, the Ministry of Health had issued a data quality assessment tool called *Penilaian Mandiri Kualitas Data Rutin* (PMKDR) (Independent Routine Data Quality Assessment).[13,14] However, several barriers had been raised during the data quality assessment using the PMKDR tool i.e. limited data entered into the system, limited operating system and manual input using excel template which may cause the data inaccuracy.[15]

To streamline the diverse range of data from multiple health information systems in healthcare facilities across Indonesia, the Ministry of Health has introduced a unified system called ASDK, short for Aplikasi Satu Data Kesehatan (ASDK) or Single Health Data Application.[16] ASDK leverages District Health Information System (DHIS2), an open-source platform that enables data analysis, reporting and visualisation of various health program data. The integrated information system ensures that all data entered adhere to a standardised data quality assessment system, specifically employing the dimensions of aggregate data quality, given that reporting primarily relies on aggregate data.[17] Conducting a comprehensive review of data quality dimensions is crucial at

*This article was accepted: 10 March 2024*
*Corresponding Author: Sandra Hakiem Afrizal*
*Email: sandra.hakiem@binawan.ac.id*

all levels, starting from the local health care facilities up to the national level.[16,17] The data collected from health facilities, most are individual data being reported to the higher levels with multiple purposes including reporting, program review, planning and monitoring activities to enhance the program quality as the aggregate data.

The assessment of aggregate data involves three processes: data review by using the aggregate data quality dimension approach, data verification by comparing reported data with field data and monitoring and evaluation systems that identify unqualified data. On the other hand, individual data undergoes assessment during the data collection on the early stage using the dimensions of individual data quality. Thus. the research is aimed to analyse the routine data quality using the WHO DQA which is integrated to the ASDK platform as the main health data resource in Indonesia.

## MATERIALS AND METHODS

A cross-sectional study was carried out from January to February 2023 which involves 34 provinces in Indonesia. Convenient sampling technique was used to select the region from west to east of Indonesia. The study involved reviews of documents from ASDK and *Survei Status Gizi Indonesia* (SSGI) in the year 2022. The source of data was transferred from the monthly summary report in electronic form. Eight variables were selected namely number of births in health facilities, number of antenatal visits, number of maternal deaths, number of neonatal visits, number of stunting cases, number of tuberculosis cases treated, prevalence of stunting from SSGI (variable number 7) and ASDK (variable number 8). The detail of variables is as follow:

All the variables considered are key indicators of the health program and are required to be regularly collected on a monthly basis. The analysis of these variables was conducted using the desk review method. employing the WHO DQA on the ASDK platform. Initially, patient data is recorded in a healthcare information management system or register book and these records are compiled and submitted at the end of the month to the District Health Office as aggregate data. The report is then transferred to an electronic system known as the DHIS2. Additionally, the report for each program is submitted to the ASDK platform for further analysis, such as data quality assessment using the WHO DQA tool.

The authors monitored the reporting data of selected variables on the ASDK platform throughout January to December 2022, aiming to analyse the dimensions of data quality including completeness, consistency over time, internal consistency between indicators and external consistency. The WHO DQA tool was employed to analyse the five dimensions of data quality, namely completeness, consistency over time, consistency between indicators, outliers and external consistency.[19]

## RESULTS

*Completeness*

Completeness refers to the availability of data, calculated as a percentage of the reports submitted to the system. A total of eight variables from 34 provinces were evaluated using the ASDK platform. However, only one variable was chosen for assessment and this assessment included eight provinces: Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Lampung and DKI Jakarta. The findings reveal that all provinces successfully submitted their annual reports for the year 2022 and no records were missing for the variable representing the number of pregnant women giving birth in a health facility (Figure 1).

### Internal Consistency

*Internal consistency over time*

Internal consistency over time is analysed by comparing the data for the year to be analysed with the average data for the previous three years.[19] In this activity, the primary data was taken from the year 2022 and compared to previous 2 years from 2019 to 2021. The position of the dot on the vertical axis on this chart represents the numerator value for the month selected. The dot on the horizontal axis represents the average value in the same district over the previous 11 months. A difference of more than ± 33% would indicate inconsistency.

Based on the result as seen in Figure 2, it shows that most of the data is located between the lines which means that most of the data have consistency over time.

*Consistency between indicators*

Consistency between indicators is analysed by comparing two related variables.[19] In this study, two variables were being compared such as the number of pregnant women giving birth in a health facility and the number of antenatal visits. Comparison ratio between these two variables in each region is then compared to the national ratio. A difference of more than ± 33% would indicate inconsistency in the observed year. Internal consistency over time can be analysed using the consistency analysis menu with the type of analysis between indicators. Each dot on the scatter plot represents the total values for one district over the last 12 months. A diamond shape represents provinces with values that fall

**Table I: List of variables**

| No | Variables | Metadata | Source |
|---|---|---|---|
| 1 | Number of pregnant women giving birth in a health facility | Data element | ASDK |
| 2 | Number of antenatal visits | Data element | ASDK |
| 3 | Number of maternal deaths | Data element | ASDK |
| 4 | Number of neonatal visits | Data element | ASDK |
| 5 | Number of stunting children | Data element | ASDK |
| 6 | Number of tuberculosis cases treated | Data element | ASDK |
| 7 | Prevalence of stunting | Data element | SSGI 2022 |
| 8 | Prevalence of stunting (ASDK) | Indicator | ASDK |

**Table II: Data quality analysis using WHO data quality tools in ASDK**

| Province | 1 | | | 2 | | | 3 | | | 4 | | | 5 | | | 6 | | | 1 and 2 (Internal consistency) | 7 and 8 (External consistency) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Dimensions | | | | | | | | | | | | | | | | | | | |
| | 1 | 2 | 4 | 1 | 2 | 4 | 1 | 2 | 4 | 1 | 2 | 4 | 1 | 2 | 4 | 1 | 2 | 4 | 3 | 5 |
| Aceh | 100 | -0.29 | 33.33 | 100 | -0.95 | 33.33 | 58.3 | -0.34 | 0.00 | 100 | -0.17 | 33.33 | 100 | 1.61 | N/A | 100 | -0.12 | 8.33 | -0.05 | 0.73 |
| North Sumatra | 100 | 0.03 | 25.00 | 100 | 0.61 | 25.00 | 58.3 | -0.36 | 0.00 | 100 | 0.37 | 41.67 | 100 | -0.14 | 16.67 | 100 | 0.00 | N/A | -0.02 | 0.74 |
| West Sumatra | 100 | 0.05 | 33.33 | 100 | 0.70 | 33.33 | 50 | 0.17 | 0.00 | 100 | 0.40 | 25.00 | 100 | -0.17 | 16.67 | 100 | -0.03 | N/A | 0.00 | 0.53 |
| Riau | 100 | 0.01 | 25.00 | 100 | 0.64 | 25.00 | 58.3 | 0.12 | 0.00 | 100 | 0.26 | 25.00 | 100 | 0.25 | 16.67 | 100 | -0.11 | N/A | -0.09 | 0.79 |
| Jambi | 100 | -0.03 | 41.67 | 100 | 0.66 | 41.67 | 58.3 | 0.71 | 0.00 | 100 | 0.39 | 41.67 | 100 | 0.02 | 16.67 | 100 | -0.10 | N/A | -0.08 | 0.76 |
| South Sumatra | 100 | 0.13 | 25.00 | 100 | 0.92 | 25.00 | 58.3 | 0.07 | 0.00 | 100 | 0.52 | 41.67 | 100 | 0.49 | N/A | 100 | -0.14 | N/A | -0.03 | 0.88 |
| Bengkulu | 100 | 0.03 | N/A | 100 | 0.67 | N/A | 50 | 0.68 | 0.00 | 100 | 0.39 | N/A | 100 | -0.53 | 16.67 | 100 | -0.02 | N/A | -0.05 | 0.78 |
| Lampung | 100 | 0.09 | 33.33 | 100 | 0.75 | 33.33 | 50 | -0.04 | 0.00 | 100 | 0.53 | 33.33 | 100 | -0.07 | 16.67 | 100 | -0.09 | N/A | -0.03 | 0.65 |
| Bangka Belitung | 100 | -0.01 | N/A | 100 | 0.57 | N/A | 58.3 | 0.72 | 0.00 | 100 | 0.38 | N/A | 100 | -0.49 | 8.33 | 100 | 0.07 | N/A | 0.02 | 0.66 |
| Riau islands | 100 | 0.07 | N/A | 100 | 0.74 | 8.33 | 50 | -0.24 | 0.00 | 100 | 0.55 | 0.00 | 100 | 0.37 | 16.67 | 100 | -0.14 | N/A | -0.11 | 0.51 |
| DKI Jakarta | 100 | -0.13 | 8.33 | 100 | 0.40 | 8.33 | 58.3 | -0.33 | 0.00 | 100 | 0.30 | N/A | 100 | 0.71 | N/A | 0 | 0.11 | N/A | -0.05 | 0.78 |
| West Java | 100 | 0.06 | 33.33 | 100 | 0.65 | 25.00 | 58.3 | 0.23 | 8.33 | 100 | 0.44 | 33.33 | 100 | 0.18 | 16.67 | 100 | 0.02 | N/A | -0.02 | 0.59 |
| Central Java | 100 | 0.02 | 8.33 | 100 | 0.60 | 8.33 | 50 | -0.41 | 0.00 | 100 | 0.58 | 8.33 | 100 | 5.00 | 8.33 | 100 | 0.07 | 8.33 | 0.01 | 0.43 |
| In Yogyakarta | 100 | -0.01 | N/A | 100 | 0.63 | 8.33 | 50 | 0.14 | 8.33 | 100 | 0.30 | N/A | 100 | 0.28 | 8.33 | 100 | 0.02 | N/A | 0.00 | 0.36 |
| East Java | 100 | -0.12 | N/A | 100 | 0.20 | 8.33 | 58.3 | -0.25 | 8.33 | 100 | 0.19 | 8.33 | 100 | 0.27 | 16.67 | 100 | -0.01 | 8.33 | 0.18 | 0.56 |
| Banten | 100 | -0.02 | 8.33 | 100 | 0.47 | 8.33 | 58.3 | 0.18 | 16.67 | 100 | 0.50 | 8.33 | 100 | 0.27 | 16.67 | 100 | -0.12 | 8.33 | -0.05 | 0.73 |
| Bali | 100 | -0.14 | 25.00 | 100 | 0.39 | 33.33 | 50 | 0.11 | 0.00 | 100 | 0.22 | 25.00 | 100 | 0.31 | N/A | 100 | -0.06 | N/A | 0.00 | 0.51 |
| West Nusa Tenggara | 100 | -0.08 | 33.33 | 100 | 0.33 | 33.33 | 50 | 0.27 | 0.00 | 100 | 0.26 | 33.33 | 100 | 1.85 | 16.67 | 100 | -0.07 | N/A | 0.01 | 0.40 |
| East Nusa Tenggara | 100 | 0.01 | 8.33 | 100 | 0.09 | 8.33 | 50 | 0.48 | 8.33 | 100 | 0.38 | 8.33 | 100 | -0.79 | 8.33 | 100 | -0.03 | N/A | 0.08 | 0.28 |
| West Kalimantan | 100 | 0.03 | 8.33 | 100 | 0.66 | 8.33 | 50 | 0.27 | 0.00 | 100 | 0.56 | 25.00 | 100 | 0.12 | 16.67 | 100 | -0.08 | N/A | -0.13 | 0.45 |
| Central Kalimantan | 100 | 0.00 | 33.33 | 100 | 0.62 | 41.67 | 50 | 0.22 | 8.33 | 100 | 0.45 | 41.67 | 100 | 0.22 | N/A | 100 | 0.13 | N/A | -0.18 | 0.51 |
| South Kalimantan | 100 | -0.07 | 8.33 | 100 | 0.51 | 8.33 | 58.3 | 0.16 | 0.00 | 100 | 0.26 | 33.33 | 100 | 0.02 | 8.33 | 100 | -0.05 | 8.33 | -0.01 | 0.55 |
| East Kalimantan | 100 | -0.06 | 8.33 | 100 | 0.51 | 8.33 | 50 | -0.26 | 0.00 | 100 | 0.66 | 16.67 | 100 | 0.84 | 16.67 | 100 | -0.02 | 8.33 | 0.03 | 0.35 |
| North Kalimantan | 100 | 0.07 | 25.00 | 100 | 0.71 | 16.67 | 50 | 0.80 | 0.00 | 100 | 0.49 | 16.67 | 100 | 0.78 | 16.67 | 100 | -0.15 | 8.33 | 0.01 | 0.31 |
| North Sulawesi | 100 | -0.17 | N/A | 100 | 0.47 | N/A | 50 | -0.53 | 8.33 | 100 | 0.30 | 41.67 | 100 | 0.62 | 8.33 | 100 | -0.06 | N/A | 0.00 | 0.81 |
| Central Sulawesi | 100 | 0.08 | 8.33 | 100 | 0.61 | 8.33 | 58.3 | 0.05 | 0.00 | 100 | 0.44 | 33.33 | 100 | 0.22 | 16.67 | 100 | -0.11 | N/A | 0.04 | 0.44 |
| South Sulawesi | 100 | 0.11 | N/A | 100 | 0.79 | 41.67 | 58.3 | 0.34 | 0.00 | 100 | -0.87 | N/A | 100 | -0.22 | 16.67 | 100 | -0.14 | N/A | 0.01 | 0.55 |
| Southeast Sulawesi | 100 | 0.02 | 41.67 | 100 | 0.62 | N/A | 58.3 | 0.33 | 0.00 | 100 | 0.31 | 41.67 | 100 | 1.08 | 16.67 | 100 | -0.20 | N/A | 0.07 | 0.57 |
| Gorontalo | 100 | 0.47 | 8.33 | 100 | 1.29 | N/A | 50 | 0.07 | 0.00 | 100 | 0.97 | 8.33 | 100 | 1.94 | 8.33 | 100 | -0.20 | N/A | 0.14 | 0.56 |
| West Sulawesi | 100 | 0.23 | 16.67 | 100 | 0.90 | 25.00 | 58.3 | 0.50 | 0.00 | 100 | 0.77 | N/A | 100 | 0.07 | 16.67 | 100 | -0.13 | 8.33 | 0.09 | 0.23 |
| Maluku | 100 | -0.17 | 41.67 | 100 | 0.27 | 41.67 | 58.3 | -0.29 | 8.33 | 100 | 0.12 | 33.33 | 100 | 1.03 | N/A | 100 | -0.09 | N/A | -0.11 | 0.64 |
| North Maluku | 100 | 0.03 | 33.33 | 100 | 0.58 | 33.33 | 58.3 | 0.60 | 0.00 | 100 | 0.49 | 41.67 | 100 | 0.80 | 8.33 | 100 | -0.06 | N/A | -0.01 | 0.58 |
| Papua | 100 | 0.02 | N/A | 100 | 0.70 | N/A | 50 | 1.34 | 16.67 | 100 | 0.70 | N/A | 100 | 1.09 | 8.33 | 100 | -0.08 | N/A | 0.19 | 0.39 |
| Papua Barat | 100 | 0.85 | N/A | 100 | 0.67 | N/A | 50 | -0.77 | 0.00 | 100 | 2.37 | N/A | 0 | -1.00 | 0.00 | 100 | | N/A | -0.19 | N/A |

*The red result value indicates that the value exceeds the threshold on the analysis of data quality dimensions.

Variables:
1. Number of mothers giving birth at health facilities
2. Number of antenatal visits
3. Number of maternal deaths
4. Number of neonatal visits
5. Number of stunting children
6. Number of tuberculosis cases treated
7. The prevalence of stunting from SSGI
8. The prevalence of stunting data from ASDK

Quality dimensions:
1. Variable completeness
2. Consistency over time
3. Consistency between indicators
4. Outlier
5. External consistency

| Unit | Data | Jan 22 | Feb 22 | Mar 22 | Apr 22 | May 22 | Jun 22 | Jul 22 | Aug 22 | Sep 22 | Oct 22 | Nov 22 | Dec 22 | Weight / Missing | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 - ACEH | dqa_num_Jumlah Ibu melahirkan di faskes | 6708.0 | 12157.0 | 18068.0 | 24327.0 | 25245.0 | 26144.0 | 26115.0 | 26128.0 | 26156.0 | 26151.0 | 26124.0 | 32945.0 | 0 | |
| 12 - SUMATERA UTARA | dqa_num_Jumlah Ibu melahirkan di faskes | 16574.0 | 32195.0 | 51846.0 | 71982.0 | 92523.0 | 111628.0 | 111628.0 | 111628.0 | 111628.0 | 111628.0 | 111628.0 | 133681.0 | 0 | |
| 13 - SUMATERA BARAT | dqa_num_Jumlah Ibu melahirkan di faskes | 6617.0 | 12946.0 | 19113.0 | 25261.0 | 31973.0 | 41019.0 | 41019.0 | 41019.0 | 41019.0 | 41019.0 | 41019.0 | 44859.0 | 0 | |
| 14 - RIAU | dqa_num_Jumlah Ibu melahirkan di faskes | 8536.0 | 15861.0 | 23764.0 | 31800.0 | 39243.0 | 41786.0 | 41786.0 | 42810.0 | 43089.0 | 43748.0 | 43748.0 | 48910.0 | 0 | |
| 15 - JAMBI | dqa_num_Jumlah Ibu melahirkan di faskes | 4596.0 | 9178.0 | 12664.0 | 17115.0 | 22562.0 | 26605.0 | 26712.0 | 26803.0 | 26885.0 | 27023.0 | 27171.0 | 27328.0 | 0 | |
| 16 - SUMATERA SELATAN | dqa_num_Jumlah Ibu melahirkan di faskes | 11512.0 | 22671.0 | 33721.0 | 45060.0 | 55351.0 | 68559.0 | 68733.0 | 69054.0 | 69227.0 | 69451.0 | 69684.0 | 77203.0 | 0 | |
| 18 - LAMPUNG | dqa_num_Jumlah Ibu melahirkan di faskes | 10611.0 | 21224.0 | 32535.0 | 44002.0 | 56013.0 | 67404.0 | 67404.0 | 67404.0 | 67404.0 | 67404.0 | 67404.0 | 72465.0 | 0 | |
| 31 - DKI JAKARTA | dqa_num_Jumlah Ibu melahirkan di faskes | 25077.0 | 30072.0 | 44417.0 | 53628.0 | 62387.0 | 68361.0 | 68361.0 | 68361.0 | 68361.0 | 68361.0 | 68361.0 | 68361.0 | 0 | |

**Fig. 1:** Missing data analysis.

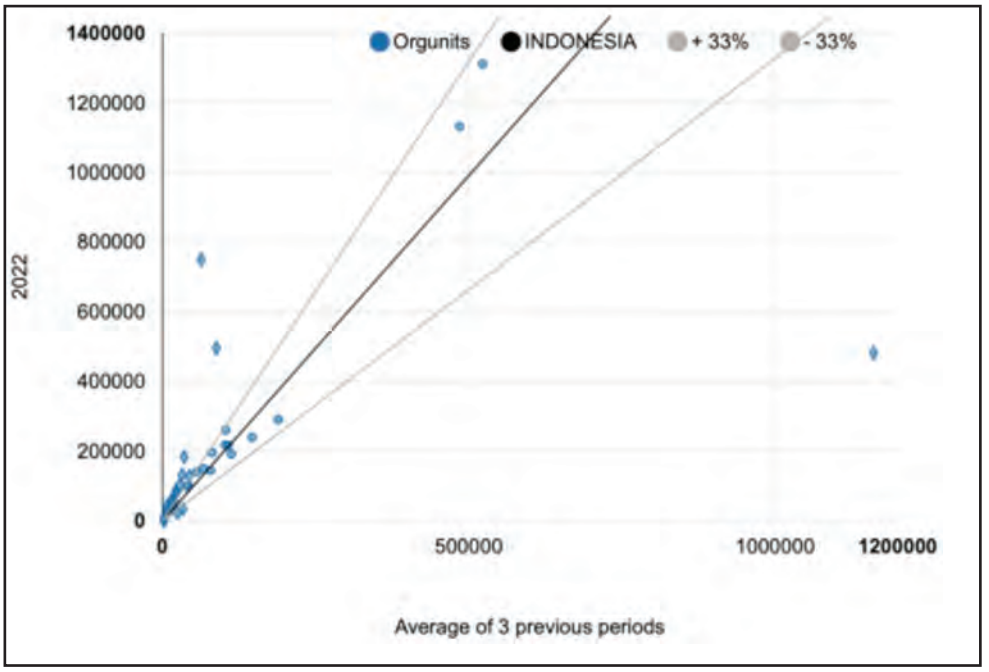


**Fig. 2:** Internal consistency analysis over time.

outside of the grey threshold lines. This result shows there was no diamond shape outside the lines which means a good internal consistency (Figure 3) and this variable is increasing over time.

*Outlier*

In this study, outliers are determined automatically by the WHO data quality tool that was identified by comparing monthly values to the mean of values for the year for the same unit. Data is classified as an 'extreme' outlier if it is more than three standard deviations and 'moderate' outliers for those between two and three standard deviations. The shaded values are indicative of data quality problems. This study shows that several data from 12 provinces in the earlier 2022 (January to April) was classified as outliers by the system (Figure 4). Based on DQA tool, grey shading is for moderate outliers while pink shading is for extreme outliers.[19]

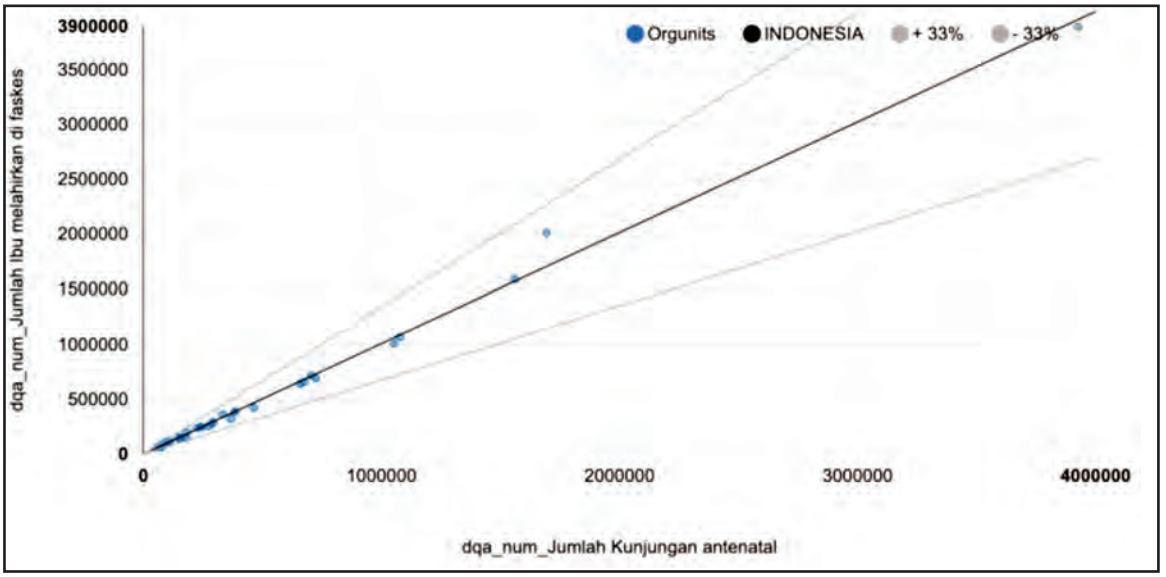


**Fig. 3:** Internal consistency analysis between indicators.

| Unit | Data | Jan 22 | Feb 22 | Mar 22 | Apr 22 | May 22 | Jun 22 | Jul 22 | Aug 22 | Sep 22 | Oct 22 | Nov 22 | Dec 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 - ACEH | dqa_num_Jumlah Ibu melahirkan di faskes | 6708.0 | 12157.0 | 18068.0 | 24327.0 | 25245.0 | 26144.0 | 26115.0 | 26128.0 | 26156.0 | 26151.0 | 26124.0 | 32945.0 |
| 12 - SUMATERA UTARA | dqa_num_Jumlah Ibu melahirkan di faskes | 16574.0 | 32195.0 | 51846.0 | 71982.0 | 92523.0 | 111628.0 | 111628.0 | 111628.0 | 111628.0 | 111628.0 | 111628.0 | 133681.0 |
| 13 - SUMATERA BARAT | dqa_num_Jumlah Ibu melahirkan di faskes | 6617.0 | 12946.0 | 19113.0 | 25261.0 | 31973.0 | 41019.0 | 41019.0 | 41019.0 | 41019.0 | 41019.0 | 41019.0 | 44859.0 |
| 14 - RIAU | dqa_num_Jumlah Ibu melahirkan di faskes | 8536.0 | 15861.0 | 23764.0 | 31800.0 | 39243.0 | 41786.0 | 41786.0 | 42810.0 | 43089.0 | 43748.0 | 43748.0 | 48910.0 |
| 15 - JAMBI | dqa_num_Jumlah Ibu melahirkan di faskes | 4596.0 | 9178.0 | 12664.0 | 17115.0 | 22562.0 | 26605.0 | 26712.0 | 26803.0 | 26885.0 | 27023.0 | 27171.0 | 27328.0 |
| 16 - SUMATERA SELATAN | dqa_num_Jumlah Ibu melahirkan di faskes | 11512.0 | 22671.0 | 33721.0 | 45060.0 | 55351.0 | 68559.0 | 68733.0 | 69054.0 | 69227.0 | 69451.0 | 69684.0 | 77203.0 |
| 18 - LAMPUNG | dqa_num_Jumlah Ibu melahirkan di faskes | 10611.0 | 21224.0 | 32535.0 | 44002.0 | 56013.0 | 67404.0 | 67404.0 | 67404.0 | 67404.0 | 67404.0 | 67404.0 | 72465.0 |
| 31 - DKI JAKARTA | dqa_num_Jumlah Ibu melahirkan di faskes | 25077.0 | 30072.0 | 44417.0 | 53628.0 | 62387.0 | 68361.0 | 68361.0 | 68361.0 | 68361.0 | 68361.0 | 68361.0 | 68361.0 |
| 32 - JAWA BARAT | dqa_num_Jumlah Ibu melahirkan di faskes | 76834.0 | 129392.0 | 198496.0 | 263541.0 | 338285.0 | 400335.0 | 403994.0 | 407749.0 | 411572.0 | 411572.0 | 419050.0 | 434068.0 |
| 33 - JAWA TENGAH | dqa_num_Jumlah Ibu melahirkan di faskes | 35005.0 | 66106.0 | 104465.0 | 134205.0 | 142171.0 | 158301.0 | 158301.0 | 158301.0 | 158301.0 | 158301.0 | 158301.0 | 158301.0 |
| 36 - BANTEN | dqa_num_Jumlah Ibu melahirkan di faskes | 20040.0 | 37090.0 | 53820.0 | 71037.0 | 91938.0 | 104721.0 | 104721.0 | 104721.0 | 104721.0 | 104721.0 | 104721.0 | 104721.0 |
| 51 - BALI | dqa_num_Jumlah Ibu melahirkan di faskes | 5010.0 | 9524.0 | 14813.0 | 16570.0 | 20027.0 | 23734.0 | 23734.0 | 23734.0 | 23734.0 | 23734.0 | 23734.0 | 38906.0 |

**Fig. 4:** Outlier analysis.

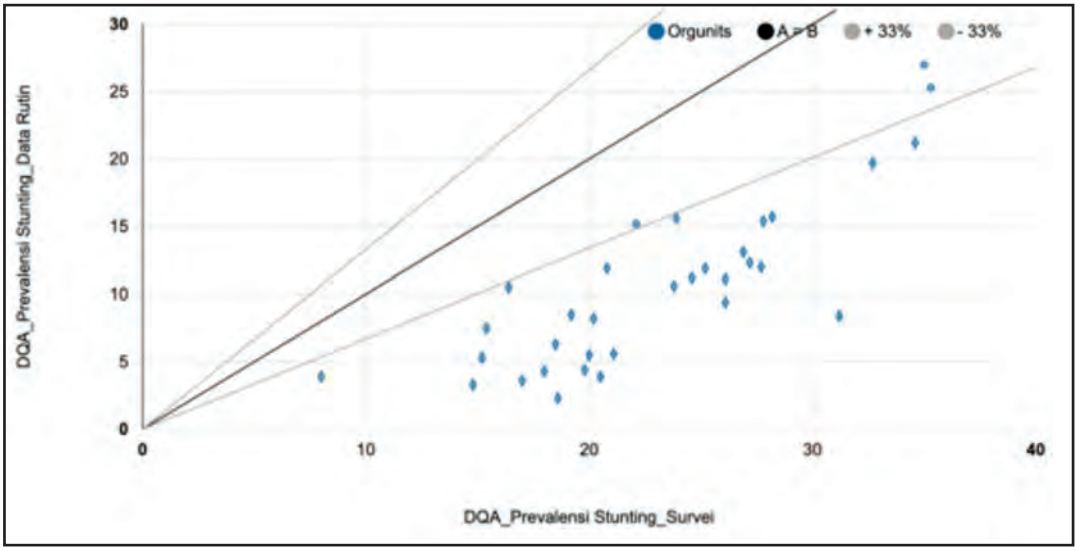


**Fig. 5:** External consistency analysis.

*External consistency*

External consistency is analysed by comparing the routine data in ASDK with other survey data sources. The two data sources being compared must have the same variables. In this study, we used the prevalence of stunting variables from two data sources, ASDK and SSGI in the year 2022. The ratio of the routine value to the survey value is then calculated. If the result is more than 0.33, it is said to be inconsistent. The output of the outlier analysis is a scatterplot with the national-level ratio of data source 1 compared to data source 2 depicted by the dark grey line. Subnational unit values that fall above or below the thresholds are potential data quality problems (Figure 5).

The study shows that most of the values fall below the thresholds for quality which represent differences from the national-level ratio that are greater than standard and are therefore potential data quality problems. In general, data quality analysis of the ASDK platform for the year 2022 is represented Table II.

## DISCUSSION

ASDK is an integrated data platform which collects information from several *electronic registers such as Komunikasi Data platform* or *Komdat elektronik-Pencatatan dan Pelaporan Gizi Berbasis Masyarakat* or ePPGBM, *Sistem Informasi Tuberkulosis* or SITB and eKohort.[16] The system integration is a potential solution for improving data quality and data collection in developing countries.[20–22] However, some literature conclude that barriers and challenges in low resource settings might impact the quality of the data.[23,24] These studies suggest that the data integration must be followed by a routine data quality assessment. Thus, our study was conducted to assess the data quality using WHO DQA tool which integrates to the ASDK platform using the data quality dimensions such as completeness, Consistency over time, consistency between indicators, outlier and external consistency.

The recent study shows that most variables have improved the completeness of data such as number of mothers giving birth at health facilities. number of antenatal visits. number of neonatal visits. number of stunting in under five children. and number of treated tuberculosis cases. There is only one variable namely the number of maternal deaths which have insufficient data in all provinces. Many instances of incompleteness in this variable seem to arise due to lack of integration system or lack of infrastructure to record relevant data. Thus. the data for this variable is only available from January to May 2022.

The current research shows that the data have good consistency over time in several variables such as the number of mothers giving birth at health facilities. number of stunting children. and number of treated tuberculosis cases. There are no significant data fluctuations in the year 2022 compared to previous years. However. the other variables such as the number of antenatal and neonatal visits have poor consistency over time. These results indicate that the data quality is different in several periods which might be caused by several barriers during the data collection or data report such as the significant decrease number of visits amongst pregnant women to health facilities during pandemic[25] or other technical errors and barriers such as lack of infrastructure. and limited health care resources which brings impact to the technology adoption.[26]

In this study. we conducted a comparison of data between related indicators. specifically examining two variables: the number of pregnant women giving birth in a health facility and the number of antenatal visits. The findings reveal that there were no instances of a diamond shape outside the lines. indicating a satisfactory consistency between related indicators and an overall upward trend in the variables over time. According to the results. most provinces demonstrate good consistency between related indicators (Table II). Despite these positive findings confirmation through other data sources, known as external consistency. is essential. Many indicators in this study exhibit data classified as outliers. However, some areas do not present outliers in their data. The presence of outliers may result from various issues, such as errors in the data entry process in the information system used. The DQA tool can offer insights into the extreme values, explaining where and why a value deviates significantly from others around it by drilling down the data and displaying values by districts within the province.

External consistency was analysed by comparing the consistency of routine data from ASDK with data from other data sources. In this study, we used a variable of prevalence of stunting from ASDK and compared with the prevalence of stunting from SSGI 2022. The results show that the data from the two sources are inconsistent. There are significant differences between routine data and survey data in most provinces in Indonesia. Data consistency between routine data and survey data was only found in three provinces such as East Nusa Tenggara, North Kalimantan and West Sulawesi. This difference might occur due to several barriers including: 1) competency of data collectors, 2) differences in the selection of the respondents that were collected, 3) different data collectors where ASDK was taken by governance health workers while SSGI data was taken by surveyor, 4) data collection periods where routine data are collected in consistently every month while the survey data collection period is carried out only once a year.

This study provides an account of the data quality analysis using the WHO DQA tool, encompassing five data quality dimensions: completeness, consistency over time, consistency between related indicators, outliers and external consistency. The results reveal that the majority of provinces in Indonesia have adopted integrated data collection using the ASDK platform and consistently submit reports. Leading to high percentages of data completeness. However, limitation of this study is the absence of guidelines recommending the minimum acceptable percentage of data incompleteness in routine health data collection. Additionally, the study was unable to directly verify issues related to data quality dimensions at the data sources as it did not scrutinise the data collection process at the health facility level. Instead it relied on secondary data available in the ASDK, collected through an integration mechanism. Regarding data completeness, it is also influenced by the data integration

process from the information system of each variable's data source. It is possible that a health facility or health office has entered and sent data to a higher level but due to delayed data integration, the information is not yet accessible in the ASDK application. Nonetheless, further follow-up is necessary to confirm with the data source by conducting field assessments.

## CONCLUSION

The current study shows that routine data quality in Indonesia performs high-quality data in terms of completeness and internal consistency. The dimension of data completeness demonstrates high levels of variable completeness with most variables achieving 100% of the completeness. Regarding the dimension of internal consistency over time, all variables except variables related to maternal and neonatal health show good consistency. Therefore, it indicates the need for improvement in these areas. Furthermore, the analysis reveals that the majority of provinces demonstrate good consistency between indicators due to the limited number of interrelated variables in *Aplikasi Satu Data Kesehatan* (ASDK). In terms of external consistency, the analysis shows inconsistent results between the prevalence of stunting from ASDK and prevalence of stunting from the SSGI 2022 survey. It might be caused by the data collection process in the field.

Based on the aforementioned results, the utilisation of WHO Data Quality Tools (DQA) on ASDK proves to be a selected tool for assessing the quality of routine data in Indonesia. While there are areas that require improvement such as to provide further training and guidance of the health information managers regarding the assessment methods. Furthermore, monitoring of the data quality on a routine basis using ASDK should be carried out at all levels through the desk review at the local health office. The limitation of this study is that this research solely relies on the secondary data recorded in the ASDK system and does not examine the data collection process from the lowest level nor the data transfer from lower levels to the central system.

## CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest related to the research.

## ACKNOWLEDGEMENT

We are grateful to all informants from national and sub-national level who took part in the study and the World Health Organisation Indonesia. for providing research funding.

## REFERENCES

1. Thorpe JH, Gray EA. Big data and public health: Navigating privacy laws to maximize potential. Public Health Rep 2015; 130(2): 171-5.
2. Puttkammer N, Baseman JG, Devine EB, Valles JS, Hyppolite N, Garilus F. et al. An assessment of data quality in a multi-site electronic medical record system in Haiti. Int J Med Inform 2016; 86: 104-16.
3. Kabakama S, Ngallaba S, Musto R, Montesanti S, Konje E, Kishamawe C. Assessment of four common underfive children's illnesses routine health management information system data for decision making at Ilemela Municipal Council, Northwest Tanzania: a case series analysis. Int J Med Inform 2016; 93: 85-91.
4. Sowe A, Gariboldi MI. An assessment of the quality of vaccination data produced through smart paper technology in The Gambia. Vaccine 2020; 38(42): 6618-26.
5. Álvarez Sánchez R, Beristain Iraola A, Epelde Unanue G, Carlin P. TAQIH, a tool for tabular data quality assessment and improvement in the context of health data. Comput Methods Programs Biomed 2019; 181: 104824.
6. Chen H, Yu P, Hailey D, Cui T. Application of a four-dimensional framework to evaluate the quality of the HIV/AIDS data collection process in China. Int J Med Inform 2021; 145: 104306. Available from: https://doi.org/10.1016/j.ijmedinf.2020.104306
7. World Health Organization. Improving data quality: a guide for developing countries. 2003; 74.
8. Jalloh MF, Namageyo-Funa A, Gleason B, Wallace AS, Friedman M, Sesay T, et al. Assessment of VaxTrac electronic immunization registry in an urban district in Sierra Leone: implications for data quality. defaulter tracking, and policy. Vaccine 2020; 38(39): 6103-11.
9. Wilkins K, Nsubuga P, Mendlein J, Mercer D, Pappaioanou M. The data for decision making project: assessment of surveillance systems in developing countries to improve access to public health information. Public Health 2008; 122(9): 914-22.
10. Medhanyie AA, Spigt M, Yebyo H, Little A, Tadesse K, Dinant GJ, et al. Quality of routine health data collected by health workers using smartphone at primary health care in Ethiopia. Int J Med Inform 2017; 101: 9-14.
11. Rumisha SF, Lyimo EP, Mremi IR, Tungu PK, Mwingira VS, Mbata D, et al. Data quality of the routine health management information system at the primary healthcare facility and district levels in Tanzania. BMC Med Inform Decis Mak 2020; 20(1): 1-22.
12. Ahanhanzo YG, Ouendo EM, Kpozèhouen A, Levêque A, Makoutodé M, Dramaix-Wilmet M. Data quality assessment in the routine health information system: An application of the lot quality assurance sampling in Benin. Health Policy Plan 2015; 30(7): 837-43.
13. Indonesia Ministry of Health. Modul Penilaian Mandiri Kualitas Data Rutin (PMKDR). Jakarta; 2013.
14. Pusdatin. Kebijakan Sistem Informasi Kesehatan Nasional [Internet]. 2013 [cited 2018 Feb 8]. Available from: https://staff.blog.ui.ac.id/r-suti/files/2016/09/SIK_nasional_kebijakan.pdf
15. Lazuardi L, Siahaan RGMS, Firdaus MZ. Evaluasi ddigitalisasi Pelayanan dan Sistem Informasi Kesehatan Ibu dan Anak. 2021.
16. Indonesia Ministry of Health. Peraturan Menteri Kesehatan Tentang Penyelenggaraan Satu Data Bidang Kesehatan Melalui Sistem Informasi. Permenkes No 18 Tahun 2022 Indonesia; 2022.
17. Calciolari S, Buccoliero L. Information integration in health care organizations. Health Care Manage Rev 2010; 35(3): 266-75.
18. Lodge W, Menon G, Kuchukhidze S, Jumbam DT, Maongezi S, Alidina S. et al. Assessing completeness of patient medical records of surgical and obstetric patients in Northern Tanzania. Glob Health Action 2020; 13(1). Available from: https://doi.org/10.1080/16549716.2020.1765526
19. World Health Organization. Data Quality Assurance (DQA). 2021. Available from: https://www.who.int/data/data-collection-tools/health-service-data/data-quality-assurance-dqa
20. Michelsen K, Brand H, Achterberg P, Wilkinson J. Promoting better integration of health information systems: best practices and challenges. Health Evidence Network Synthesis Report. Available from: http://www.euro.who.int/__data/assets/pdf_file/0003/270813/Promoting-better-integration-of-HIS-best-practices-and-challenges.pdf

21. Jardim SVB. The electronic health record and its contribution to healthcare information systems interoperability. Procedia Technol 2013; 9: 940-8.
22. Katapally TR. The SMART framework: Integration of citizen science. community-based participatory research. and systems science for population health science in the digital age. JMIR mHealth uHealth 2019; 7(8): 1-12.
23. Afrizal SH, Handayani PW, Hidayanto AN, Eryando T, Budiharsana M, Martha E. Barriers and challenges to Primary Health Care Information System (PHCIS) adoption from health management perspective: a qualitative study. Informatics Med Unlocked 2019; 17: 100198.
24. Heart T, Ben-Assuli O, Shabtai I. A review of PHR, EMR and EHR integration: a more personalized healthcare and public health policy. Heal Policy Technol 2017; 6(1): 20-5.
25. de Guzman GS, Banal-Silao MJB. Antenatal care utilization during the COVID-19 pandemic: an online cross-sectional survey among Filipino women. BMC Pregnancy Childbirth 2022; 22(1): 1-12.
26. Afrizal SH, Hidayanto AN, Handayani PW, Siregar KN, Besral. Budiharsana M, et al. Survey-based data describing readiness to adopt an electronic pregnancy registration-monitoring system amongst health workers. Data Br 2020; 32: 106192.