Literature Review: Data Quality Assessment Methods Of Tuberculosis Recording And Reporting Electronically

Atha Rifqia Pradana¹, Cahya Tri Purnami², Atik Mawarni³

¹,²,³ Department of Biostatistics and Population Studies, Diponegoro University, Semarang
Email: atha.rifqia@gmail.com

Abstract

On 2019, number of tuberculosis (TB) cases in the world reaches 10 million people with deaths reaching 1.7 million people. Data quality of recording and reporting needs to be considered as one of TB control strategies. To optimize TB control, some countries conduct data quality assessments using a method by assessing the dimensions. This study aims to examine several methods of assessing data quality of electronic TB recording and reporting through literature review with simplified approach.

Search of indexed and reputable articles is done on databases of Google Scholar, Scopus, PubMed, Web of Science, CrossRef, PROQUEST. Articles review show that there are 3 methods used to assess data quality, that is Centers for Disease Control and Prevention (CDC) tool (85%), Routine Data Quality Assessment (RDQA) (5%), and independently developed methods (40%). There were differences from 3 methods, specifically CDC was appropriate for assessing data completeness more accurately, RDQA was appropriate for assessing data reported routinely, and independently developed methods were used according to a region conditions. Based on the used of several methods, data quality dimensions most frequently assessed were completeness (85%), followed by dimensions of timeliness (40%) and accuracy (30%). Assessment of TB data quality is still limited, a health institution needs to conduct assessment of data quality regularly so TB control efforts are more optimal.

Keywords: Assessment Method, Data Quality, Electronic Tuberculosis Recording and Reporting.

To cite this article:

INTRODUCTION

Tuberculosis (TB) is an infectious disease caused by infection of bacterium named Mycobacterium tuberculosis which attacks lung organs (National Institute for Health and Care Excellence, 2016). Based on the Global Tuberculosis Report of 2019, there are 10 million sufferers with the number of deaths due to TB reaching 1.7 people in the world (World Health Organization, 2019). One of the strategies that is considered appropriate to control and accelerate the reduction in TB incidence is through improved electronic recording and reporting (Theron et al., 2015; World Health Organization, 2018). Data of electronic recording and reporting has an important role in monitoring trends in the TB epidemic, planning and implementing TB control efforts, and as a basis for implementing other TB control programs, such as National Tuberculosis Program and TB surveillance (Jakubowiak et al., 2001; Yamaguti et al., 2017).

Data related to TB is recorded by health facilities, such as hospitals, primary health care facilities, clinics, and then compiled and reported to sub regional, regional, provincial and national levels (World Health Organization, 2012). Process and analysis of TB data is carried out to produce information that is useful in TB control efforts to be more optimal. Quality information will be generated if the data reported is also has good quality (Fund, 2011; Roomaney et al., 2017; World Health Organization, 2014, 2018). Thus, TB data quality in electronic recording and reporting is a thing that needs attention in TB control (World Health Organization, 2012). Several dimensions of data quality that need to be considered include complete and timely (Batini et al., 2009). TB data quality is important to be considered in order to avoid low data quality because it can result in decisions about TB management that are not appropriate, invalid, and wrong (Vidyanto, 2018; Warwick et al., 2015).

To avoid the low data quality of electronic TB recording and reporting, data quality assessment can be carried out by concerning several data quality dimensions, such as completeness, accuracy, and timeliness (Nturibi, 2010). Data quality assessment of electronic TB recording and reporting can be assessed using several methods. From the previous studies, several methods are used, including Routine Data Quality Assessment (RDQA) tool and Centers for Disease Control and Prevention (CDC) tool (Aqil et al., 2012; Centers for Disease Control (CDC), 2001; MEASURE Evaluation, 2015). In addition, assessment of data quality can also use methods with frameworks and metrics that have been developed independently according to region conditions and TB control goals that desires to be achieved by a region (Chen et al., 2014; Listyorini et al., 2017; Theron et al., 2015).

Although data quality is one of the important topics in public health, a review of the literature regarding methods for assessing data quality of electronic TB recording and reporting is still limited. For this reason, this study was conducted with the aim of reviewing methods of assessing data quality of electronic TB recording and reporting based on literature review from several previous studies.
METHODS

This research was conducted by using literature review with simplified approach to collect and summarize data from previous research. The research was carried out by searching indexed and reputable articles in several databases. Aims the use of Boolean Operators "AND" and "OR" in this study are to combine several keywords, that is "methods", "data quality assessment", "electronic tuberculosis recording and reporting", "tuberculosis surveillance system", "tuberculosis information system". The total number of articles according to the inclusion criteria examined in this study was 20 articles.

Duration of Research : May-July 2020
Inclusion Criteria :
- Publication of articles in the period 2010-2020
- Original research
- Articles published in Indonesian and/or English
- Full-text and accessible
- Related to the research topic (assessment methods of TB data quality)
- Research locations are not limited to either developed or developing countries

Stages of Research :
1. Problem formulation
   Observe the background, research objectives and research questions.
2. Data collection
   Aggregation of articles according to keywords is done through the search on Google Scholar, Scopus, PubMed, Web of Science, CrossRef, PROQUEST databases, where articles are published in national and international journals, proceedings, and research reports.
3. Data evaluation
   Identify and eliminate articles according to inclusion criteria.
4. Analysis and interpretation
   Conduct data analysis and interpretation of results according to the selected article.
5. Public presentation
   Presentation the data in narrative text, graphics, charts and other forms.
6. Withdrawal of conclusions
   Clarify and detail the information obtained in accordance with articles review results.
RESULTS

From the search results, data characteristics obtained from 20 articles are shown in Table 1.

Table 1. Data Characteristics of Reviewed Articles Based on Research Locations and Data Quality Assessment Methods

<table>
<thead>
<tr>
<th>NO</th>
<th>Researcher</th>
<th>Country</th>
<th>Recording and Reporting Model</th>
<th>Data Quality Assessment Method</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Kwaghe et al., 2020)</td>
<td>Nigeria</td>
<td>TB surveillance system</td>
<td>CDC tool</td>
<td>Timeliness (100%)</td>
</tr>
<tr>
<td>2</td>
<td>(Zhou et al., 2019)</td>
<td>China</td>
<td>TB information system</td>
<td>CDC tool</td>
<td>Completeness (74%)</td>
</tr>
<tr>
<td>3</td>
<td>(Jamieson et al., 2019)</td>
<td>South Africa</td>
<td>TB information system</td>
<td>Independent method</td>
<td>Completeness (59%)</td>
</tr>
<tr>
<td>4</td>
<td>(Frimpong-Mansoh et al., 2018)</td>
<td>Ghana</td>
<td>TB surveillance system</td>
<td>CDC tool</td>
<td>Completeness (73%) and timeliness (55.6%)</td>
</tr>
<tr>
<td>5</td>
<td>(Ali et al., 2018)</td>
<td>Pakistan</td>
<td>TB information system</td>
<td>Independent method</td>
<td>Completeness (36.5%), timeliness (81%), accuracy (87%), understandability (96%), consistency (99.5%)</td>
</tr>
<tr>
<td>6</td>
<td>(Chaintarli et al., 2018)</td>
<td>Ireland</td>
<td>TB surveillance system</td>
<td>European CDC tool</td>
<td>Completeness (62.5%), timeliness (58.3%), flexibility (70.8%)</td>
</tr>
<tr>
<td>7</td>
<td>(da Silva et al., 2017)</td>
<td>Brazil</td>
<td>TB surveillance system</td>
<td>CDC tool</td>
<td>Completeness (70-100%), consistency (&gt;93.3%), timeliness (24.9-62.8%)</td>
</tr>
<tr>
<td>8</td>
<td>(Mlotshwa et al., 2017)</td>
<td>South Africa</td>
<td>TB surveillance system</td>
<td>U.S. CDC tool</td>
<td>Completeness (95-100%) and reliability (89-99%)</td>
</tr>
<tr>
<td>9</td>
<td>(Yamaguti et al., 2017)</td>
<td>Brazil</td>
<td>TB information system</td>
<td>Independent method</td>
<td>Completeness (19-42%) and reliability (63.5-75.6%)</td>
</tr>
<tr>
<td>10</td>
<td>(Kleinman et al., 2017)</td>
<td>Botswana</td>
<td>TB information system</td>
<td>Independent method</td>
<td>Completeness (46.5-93.4%) and accuracy (41.5-90.2%)</td>
</tr>
<tr>
<td>11</td>
<td>(Majerovich et al., 2017)</td>
<td>Canada</td>
<td>TB information system</td>
<td>European CDC tool</td>
<td>Completeness (nearly 100%), internal validity (&gt;94%) and external validity (20%)</td>
</tr>
<tr>
<td>12</td>
<td>(Kang et al., 2016)</td>
<td>Republic of Korea</td>
<td>TB surveillance system</td>
<td>Independent method</td>
<td>Completeness (90%, 2012; 93%, 2013; 94%, 2014) and timeliness (81.6%, 2012; 80.4%, 2014)</td>
</tr>
<tr>
<td>13</td>
<td>(Sharma et al., 2016)</td>
<td>Kenya</td>
<td>TB surveillance system</td>
<td>Independent method</td>
<td>Completeness (99%)</td>
</tr>
</tbody>
</table>
From Figure 1, it can be seen that TB data quality assessment in mostly done in developed countries (52%) than developing countries (48%). In developed countries, studies are carried out in China, South Africa, Ireland, Canada, Republic of Korea and Taiwan, while studies in developing country are carried out in Nigeria, Pakistan, Brazil, Botswana, Kenya, Haiti, Afghanistan, Cambodia, and Vietnam. Through this study, it was found that there are 2 models of electronic TB recording and reporting that presented in Figure 2, that is TB information system (50%) and TB surveillance system (50%). Utilization of these two information technologies aim to help and surmount deficiencies that are done manually.

**Figure 1.** Data Quality Assessment Method  
**Figure 2.** Research Locations
Several kind of TB data quality assessments are carried out using different methods. Methods used to assess data quality of electronic TB recording and reporting are shown in Figure 3. The most frequently used methods to assess data quality are method of Centers for Disease Control and Prevention (CDC) tool (55%), then followed by other methods, that is independently developed method (40%) and Routine Data Quality Assessment (5%). From the 20 articles examined in this study, it can be seen that dimension of data quality most assessed on electronic tuberculosis (TB) recording and reporting is completeness (85%), then followed by other dimensions, that is timeliness (40%) and accuracy (30%). Different dimensions of data quality assessment from the results of the article review are presented in Table 2.

Table 2. Dimensions of Data Quality Assessment

<table>
<thead>
<tr>
<th>Researcher</th>
<th>Assessment Methods</th>
<th>Data Quality Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data Quality Dimensions</td>
<td>Completeness</td>
</tr>
<tr>
<td>(Kwaghe et al., 2020)</td>
<td>CDC tool</td>
<td>-</td>
</tr>
<tr>
<td>(Zhou et al., 2019)</td>
<td>CDC tool</td>
<td>-</td>
</tr>
<tr>
<td>(Jamieson et al., 2019)</td>
<td>Independent method</td>
<td>√</td>
</tr>
<tr>
<td>(Frimpong-Mansoh et al., 2018)</td>
<td>CDC tool</td>
<td>√</td>
</tr>
<tr>
<td>(Ali et al., 2018)</td>
<td>Independent method</td>
<td>√</td>
</tr>
<tr>
<td>(Chaintarli et al., 2018)</td>
<td>European CDC tool</td>
<td>√</td>
</tr>
<tr>
<td>(da Silva et al., 2017)</td>
<td>CDC tool</td>
<td>√</td>
</tr>
<tr>
<td>(Mlotshwa et al., 2017)</td>
<td>U.S. CDC tool</td>
<td>√</td>
</tr>
<tr>
<td>(Yamaguti et al., 2017)</td>
<td>Independent method</td>
<td>√</td>
</tr>
<tr>
<td>(Kleinman et al., 2017)</td>
<td>Independent method</td>
<td>√</td>
</tr>
<tr>
<td>(Majerovich et al., 2017)</td>
<td>European CDC tool</td>
<td>√</td>
</tr>
</tbody>
</table>
Based on Table 2., it can be seen that dimension most frequently assessed is completeness (85%). Data quality assessments regarding completeness were carried out on different TB recording and reporting data. The difference in data assessed regarding the completeness of the data can be seen in Table 3.

Table 3. Data Completeness Assessed in Electronic Tuberculosis Recording and Reporting

<table>
<thead>
<tr>
<th>NO</th>
<th>Researcher</th>
<th>Completeness of Assessed Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Kwaghe et al., 2020)</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>(Zhou et al., 2019)</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>(Jamieson et al., 2019)</td>
<td>1) Individual data (name, sex); 2) Clinical data (treatment)</td>
</tr>
<tr>
<td>4</td>
<td>(Frimpong-Mansoh et al., 2018)</td>
<td>1) Individual data (name, age, sex, registration date, number of TB facility, address, phone number); 2) Clinical data (bacteriology and X-ray examination, treatment)</td>
</tr>
<tr>
<td>5</td>
<td>(Ali et al., 2018)</td>
<td>1) Individual data (name, age, weight, identification code, national ID number, address, phone number); 2) Clinical data (treatment)</td>
</tr>
<tr>
<td>6</td>
<td>(Chaintarli et al., 2018)</td>
<td>1) Individual data (ID, address, sex, date of birth); 2) Clinical data (risk factor, X-ray examination, treatment)</td>
</tr>
<tr>
<td>7</td>
<td>(da Silva et al., 2017)</td>
<td>1) Clinical data (treatment)</td>
</tr>
</tbody>
</table>

*N.B.: √ (Assessed); - (Not Assessed)*
DISCUSSION

Research Locations and Electronic Tuberculosis Recording Reporting Models

Assessment of tuberculosis (TB) data quality is still limited by several countries because it requires collaboration between several related sectors and supporting variables (Xiao et al., 2017). Some developed countries carry out data quality assessments with the aim that results of data quality assessments can be used to strengthen TB control interventions such as National Tuberculosis Program (NTP) and TB surveillance (Qader et al., 2013; Salyer et al., 2015; Sharma et al., 2016). Results of data quality assessment can also be used as a reference for the government in improving data quality so it helps in making decisions about TB control more quickly and accurately (Rockwood et al., 2016). World Health Organization states that there are 2 models of electronic TB recording and reporting, that is through TB information system and TB surveillance.
system (World Health Organization, 2012). Utilization of both information technologies are expected to overcome deficiencies that found in manual recording and reporting, therefore it can supporting to record and report of TB data (Ali et al., 2018; Manangan et al., 2012). This condition is in line with study of Kwaghe, et al. (2020) in Nigeria that utilization of TB surveillance system can detect TB cases timely and diagnose accurately (Kwaghe et al., 2020). Another study by Majerovich, et al. (2017) said that data reported on TB information system was 100% useful in eliminate TB program (Majerovich et al., 2017). There are variations of data kind that reported in both of systems, including individual data (demography data such as name, gender, address) and clinical data (such as risk factor, sputum result, X-ray, pre-treatment smears, and treatment) (Jamieson et al., 2019). However, there are still found several problems in the implementation of TB information system or TB surveillance system. This is in line with study by Heijden, et al. (2019) that there are still found low data quality that reported on TB information system in South Africa, such as completeness (30%), reliability (22%), and errors in matching data (26%), where this conditions causing laboratory results can not be connected directly to TB treatment facilities (Heijden et al., 2019; Rafli et al., 2017).

**Methods Used in Assessing Data Quality of Electronic Tuberculosis Recording and Reporting**

a. **Centers for Disease Control and Prevention Tool**

The results of articles review indicate that as many as 55% studies assess data quality using method of CDC tool. Method of Centers for Disease Control and Prevention (CDC) tool developed by CDC is appropriate for assessing data completeness reported on TB information system more accurately (Centers for Disease Control (CDC), 2001). This is in line with study by Pillaye, et al. (2003) which said that assessment of data quality using CDC tool in United Kingdom uses several data sources and capture-recapture (CR) method that can combine more than 2 data sources so complete data can be assessed more accurately, where results of data completeness are 7-27% (Pillaye & Clarke, 2003). Another study by Pratt, et al. (2020) in United States also stated that CDC tool can assess TB data completeness more accurately in 3 data types needed in TB control, that is data of diagnosis examination results, data of risk factors and data of treatment (Pratt et al., 2020). Development of CDC tool is in accordance to data quality assessment standards so it can be used by the public health surveillance institution, which one is TB surveillance. Development of CDC tool method is in accordance with data quality assessment standards so it can be used by all institutions that correlate in surveillance. Therefore, this method is appropriate for assessing reported of TB data completeness more accurately. Method of CDC tool in its entirety can assess 3 data quality dimensions, that is completeness, validity, and timeliness. Result of TB data quality assessment using CDC tool can support implementation of TB data analysis and TB surveillance more effective and efficient.

b. **Routine Data Quality Assessment**

Routine Data Quality Assessment (RDQA) has been arranged by World Health Organization to assess data quality in tuberculosis (TB) control that reported routinely
(MEASURE Evaluation, 2015). Result of articles review showed that there was 1 study by Salyer, et al. (2015) that use RDQA to assess routine data completeness (40-90%) concerning to AFB (Acid-Fast Bacilli) test result, HIV (human Immunodeficiency) test result, contact tracing, and treatment result reported on TB surveillance system (Salyer et al., 2015). TB treatment requires routine data as biomarker in monitoring treatment of patient TB for follow-up bacteriological findings (Perrin et al., 2007). This is in line with study by Rockwood, et al. (2016) that stated routine data such as AFB test results affect the success of TB treatment, as well as to predict recurrence that may happen to TB patients (Rockwood et al., 2016). RDQA method used in its entirety can assess TB data quality dimensions regarding to completeness, accuracy, reliability, timeliness, confidentiality, precision, and integrity. Based on previous studies, it can be concluded that RDQA method can use to assess TB data quality of recording and reporting electronically that are reported routinely so that is expected to improve TB control efforts.

c. Independently Developed Methods

Based on articles review conducted, as many as 40% studies assess tuberculosis (TB) data quality using independently developed methods. A health institution can assess data quality using methods with frameworks and metrics adopted from existing data quality assessment methods, where the methods is adjusted to the conditions and TB control policies of each region (Chen et al., 2014). This is in line with Huang, et al. (2014) study which assesses data quality using independently developed methods so it can be known that there are delays in reporting sputum results of TB patient and low of internal data consistency (Huang et al., 2014). Assessment of health data quality can regard to 3 crucial dimensions in assessment, that is completeness, timeliness and accuracy, to help the decision making quickly and avoid duplication of TB reporting data (Alipour & Ahmadi, 2016; Zulkiffli et al., 2019). Therefore, through assessment of data quality using frameworks and metrics that are developed independently, it can be known the problems of TB data quality of recording and reporting electronically in a region. Implementation of data quality assessment using independently developed methods is also adjusted to users that involved in TB monitoring so epidemiological analysis and TB control policies are taken according to the conditions of each region.

Dominant Dimensions in Data Quality Assessment

Data has an important role in decision making. The results of articles review shown that as many as 85% studies assess data completeness of electronic tuberculosis (TB) recording and reporting. This research is in line with Alipour, et al. study in 2017, where completeness is one of the dimensions that becomes the main standard in assessing health data quality (Alipour & Ahmadi, 2016). According to Centers for Disease Control and Prevention (CDC), standard completeness of TB data reporting variable is 95% (Centers for Disease Control (CDC), 2001). Several completeness data that need to be considered, that is risk factor data, treatment data, and sputum results data. Results from Majerovich, et al. (2017) study said that completeness of TB risk
factor data in Canada is low (10.9%), whereas risk factor data is useful for identifying population groups at high risk of TB (Majerovich et al., 2017). Complete treatment data regarding drug surveillance and DR-TB (Drug-resistant Tuberculosis) need to be considered to monitor TB trends in a region (Chaintarli et al., 2018). This is in line with Dunbar, et al. (2011) study that the low completeness of treatment data (65%) triggers a new source of ongoing TB transmission (Dunbar et al., 2011). In addition, data completeness of culture examination for TB patients need to be considered because it becomes the gold standard for determining TB diagnoses (Rockwood et al., 2016). This is in line with study of Podewils, et al. (2015), that the low of complete cultural examination data (10.9%) has potential for the continuation of TB spread in the community (Podewils et al., 2015). Result of Straetemans, et al. (2020) study stated that completeness of reporting and detection of TB cases have important role in decreasing TB incidence (Straetemans et al., 2020). In another side, data completeness problem of TB notification not only found in developing countries, but also in developed countries that have well-organized TB control systems, such as United Kingdom (data completeness is 27%) (Pillaye & Clarke, 2003). The low completeness of TB reporting data causes data quality of recording and reporting needs to be considered, especially regarding TB notification completeness.

CONCLUSIONS AND RECOMMENDATIONS

From the research conducted, it can be concluded that there are 3 methods used in the data quality assessment of electronic TB recording and reporting. Centers for Disease Control and Prevention (CDC) tool is method most frequently used in study, then followed by independently methods and Routine Data Quality Assessment (RDQA). There are some differences found from utilization of 3 methods in assessing TB data quality. CDC tool method is appropriate to be used to assess data completeness more accurately, RDQA is appropriate to be used to assess routine data reported, while independently developed methods can be used according to user needs and TB control policies in each region. From the used of 3 data quality assessment methods, dimension of data quality most frequently assessed is completeness, then followed by other dimensions that is timeliness and accuracy.

Data quality assessment needs to be conducted regularly by a health institution to know the data quality that has a role in optimizing the TB control. Data quality assessment can be carried out using the Centers for Disease Control and Prevention (CDC) tool developed by CDC as a center for control and prevention of communicable diseases at a global level Development of CDC tool is in accordance with general assessment standards so it can be used by all regions.

REFERENCES


www.ijhes.com


National Institute for Health and Care Excellence. (2016). *Tuberculosis: Prevention, diagnosis,
management and service organisation (Issue January).


Straetemans, M., Bakker, M. I., Alba, S., Mergenthaler, C., Rood, E., Andersen, P. H., Schimmel, H., Simunovic, A., Svetina, P., Carvalho, C., Lyytikäinen, O., Abubakar, I., Harris, R. J.,


