Picturing diphtheria outbreak in Indonesia using national annual report data: what are the lessons learned?

T Rinanda\textsuperscript{1,4}, M Mudatsir\textsuperscript{4}, R Raihan\textsuperscript{2} and S Sakdiah\textsuperscript{3}

\textsuperscript{1}Department of Microbiology, Faculty of Medicine, Universitas Syiah Kuala, Banda Aceh, Indonesia
\textsuperscript{2}Department of Pediatric, Universitas Syiah Kuala-Zainoel Abidin Hospital, Banda Aceh, Indonesia
\textsuperscript{3}Department of Family Medicine, Faculty of Medicine, Universitas Syiah Kuala, Banda Aceh, Indonesia

E-mail: tristia.rinanda@unsyiah.ac.id

Abstract. Diphtheria is one of the vaccine preventable diseases. Due to the advances in medical technology and vaccine availability, this disease should no longer be a health problem. Its outbreak which reached the highest point in 2017 has opened our eyes that diphtheria remains a major threat, is highly contagious, and leads to morbidity and mortality. The large number of diphtheria cases also indicates that the protection of immunization in the community has decreased. This paper aims to discuss the usage of the reliable official big data namely the Indonesia Health Profile from year 2004 to 2017 released by the Ministry of Health to describe the magnitude of diphtheria in Indonesia. Providing big open data related to infectious disease is very challenging. The released data should be managed properly so it can be used by the government and related institutions to generate better policies and recommendations in order to end the outbreak and provide adequate protection against diphtheria amongst Indonesian community.

1. Introduction
Nowadays the collection and analysis of big data in the health sector play a very important role in predicting the outcomes of diseases and epidemic, improving therapies and quality of life, preventing the progression of diseases, and formulating effective health policies. \cite{1, 2}. Big data represent the health information implicating the health care system. This information is not only important for the patients but also for the healthcare professionals and providers, government and non-governmental institutions, researchers and pharmaceutical industries. Many countries have invested in the development of a system which enables the efficient usage of big data in the health sector \cite{2}. Malaysia, the United States of America, and the European Union countries have already constructed a system to manage big data on health informatics and the implementation-related policies \cite{1-4}.

The concern about the usage of big data in the health sector is triggered, amongst others, by the ‘personalized medicine’ approach. This approach allows an individual to receive medical treatments based on his/her needs and characteristics through the analysis of medical datasets \cite{1, 4, 5}. The methodology provides not only a better healthcare system but also the best individual care. On the

\textsuperscript{4} To whom any correspondence should be addressed.
other hand, there are many issues to be encountered and anticipated, such as ethical (confidentiality and security of data, users trust, access to data) and technical (reliability and management of data) issues [4,6].

Big data in the health sector are not only limited to medical record data [1]. There are many other data which can be exposed to the public without any confidentiality issues, namely open data. Unlike medical records, open data are accessible to the public without restriction and can be used to formulate solutions concerning health issues [7] by engaging cognitive activities such as methodical reasoning, decision-making, interpreting and problem solving [8].

Indonesia also has big open data managed by the Ministry of Health, namely Indonesia Health Profile. The data are compiled based on routine data and survey data from technical units within the Ministry of Health and other related institutions such as the Central Statistics Agency (BPS), Social Insurance Administration Organization (BPJS), the Ministry of Home Affairs and the National Population and Family Planning Board (BKKBN) [9]. The health profile provides an abundance of information on demography, healthcare facilities, healthcare professionals, health financing, environmental health, and epidemiology of infectious diseases [9].

In this paper we use diphtheria outbreaks in Indonesia as a model in using open big data related to infectious diseases. Diphtheria in 2017 reached its largest outbreak in the history of the disease in Indonesia. Indonesia has become one of the countries with the highest frequency of diphtheria outbreaks in the same year [10]. The resurgent of diphtheria is very alarming since Indonesia once succeeded in eradicating the disease in 1998. The irony is that diphtheria is one of the vaccine preventable diseases, but despite the implementation of a massive vaccination program, Indonesia still has to face the reality of diphtheria outbreaks in 21 provinces in 2017 [11]. Indonesia Health Profile has presented large number of data in describing the Diphtheria outbreak in Indonesia year by year. Those data must be managed wisely in order to provide useful information. This paper will further explore the data about diphtheria which have been displayed in Indonesia Health Profile and summarize some points as lessons learnt. We believe that the lessons learnt will be beneficial in formulating policies and strategies to eradicate diphtheria in Indonesia based on the annual health report data.

2. The lessons learned
Indonesia Health Profile has been published by the Ministry of Health as annual report data. Besides diphtheria, there are other infectious diseases reported such as measles, pertussis, tetanus, tuberculosis, leprosy, Human Immunodeficiency Virus (HIV) infection, hepatitis B, Dengue Hemorrhagic Fever, filariasis, polio, malaria, rabies and leptospirosis. Diphtheria, pertussis, tetanus, polio, hepatitis B and measles are vaccine-preventable diseases and highlighted on the report due to their role as indicators of vaccine efficacy.

Diphtheria has been reported annually since 2004, but some epidemiological parameters are not reported consistently every year. In table 1, we had summarized the parameters which were published on Indonesia Health Profile from 2004 to 2017 [12-25].

Table 1 shows that there were some parameters which were not consistently provided, such as outbreak frequency, Case Fatality Rate (CFR) and proportion of vaccinated persons amongst diphtheria cases. Meanwhile, the number of cases or incidence has been annually reported. In 2008, information related to outbreak frequency was not recorded when there were actually 77 cases reported nationwide [26]. The absence of outbreak frequency data from 2008 to 2017 was quite disconcerting because in terms of diphtheria, one case of suspected diphtheria is already defined as an outbreak [27]. Information related to outbreak frequency and CFR was not found in the Indonesia Health Profile year 2009, whereas both parameters were reported in a journal managed by Ministry of Health [28]. Although it was not officially reported in the Health Profile year 2010, CFR actually could be determined by comparing the number of deaths related to diphtheria per total number of cases multiplied by 100 percent [29].
The number of diphtheria cases (incidence) was provided every year from 2004 to 2017. These data were classified into age-specific groups except for the report in 2008 where the data only mentioned the total number of cases [17]. The data reported from 2004 to 2006 were grouped as the following: under 1 year old, 1 to 4 years old, 5 to 14 years old, 15 to 44 years old, and 45 years old and above [12,13,14]. For the next consecutive years, the data were reported using different age classifications: under 1 year old, 1 to 4 years old, 5 to 9 years old, 10 to 14 years old, 15 to 18 years old, and 40 years old and above [15,17-25]. The information about the incidence based on age-specific groups is very important to define the magnitude of risk to certain ages related to the history of immunization against diphtheria [30]. The different classifications of ages will result in difficulties in summarizing the data. It is very important to maintain a consistent pattern of data reporting [31]. Delivering the big data for infectious diseases also requires the improved standardization of disease reports [32].

One of the major challenges in using and analyzing big data is ensuring their consistency and reliability [33]. A previous study showed that there was some discrepancy between the Dengue surveillance data provided by the Indonesian and Brazilian Ministries of Health and the online data on Dengue virus infection provided by World Health Organization (DengueNet). This problem was mainly related to the large variety of surveillance methodology and lack of proper documentation resulting in unreliable and inefficient use of data [34]. Consistent reporting of public health data over time plays an important role in providing conclusive assessments of the status and trends in healthcare system and leading to the forefront of policy discussions and decisions [35].

Big data in the health sector must be used optimally to improve the health and well-being of the community, otherwise they are only a large collection of meaningless information [7]. In order to avoid misinterpretation, the data must be collected, analyzed and interpreted in a proper manner [33]. Upon reviewing the Indonesian Health Profile, we find that there are some data which require further elaboration. The proportion of vaccinated persons amongst all diphtheria cases has been reported regularly, except for 2004-2006 and 2009 data (Table 2) [12-25]. The data must be interpreted correctly and properly, especially when they are presented as open data accessible to anyone with or without proper background knowledge in epidemiology. Table 2 indicates that there are a large number of vaccinated people who still suffer from diphtheria. This condition leads people to begin

| Year of Publication | Outbreak Frequency | Number of cases | Number of deaths | Case Fatality Rate | Proportion of vaccinated persons amongst cases |
|---------------------|--------------------|----------------|------------------|-------------------|-----------------------------------------------|
| 2004                | Available          | Available      | Available        | Available         | ND                                            |
| 2005                | Available          | Available      | Available        | Available         | ND                                            |
| 2006                | Available          | Available      | Available        | Available         | ND                                            |
| 2007                | Available          | Available      | Available        | Available         | ND                                            |
| 2008                | ND                 | Available      | ND               | ND                | Available                                     |
| 2009                | ND                 | Available      | ND               | ND                | Available                                     |
| 2010                | ND                 | Available      | ND               | ND                | Available                                     |
| 2011                | ND                 | Available      | Available        | Available         | Available                                     |
| 2012                | ND                 | Available      | Available        | Available         | Available                                     |
| 2013                | ND                 | Available      | Available        | Available         | Available                                     |
| 2014                | ND                 | Available      | Available        | Available         | Available                                     |
| 2015                | ND                 | Available      | Available        | Available         | Available                                     |
| 2016                | ND                 | Available      | Available        | Available         | Available                                     |
| 2017                | ND                 | Available      | Available        | Available         | Available                                     |

ND: No Data
questioning the protection of the vaccines. High proportion appears to reflect the low protection provided by the vaccination against diphtheria. This misinterpretation can lead to a false conclusion which may hamper the implementation of vaccination program amongst the community. The Indonesian Pediatric Society (Ikatan Dokter Anak Indonesia) has stated that the data misinterpretation related to the proportion of vaccinated persons amongst all cases has been used by anti-vaccines group for their propaganda [36].

Table 2. Proportion of vaccinated persons amongst all diphtheria cases from 2004 to 2017 provided by Indonesia Health Profile from 2004 to 2017.

| Year of Publication | Proportion of vaccinated persons amongst all cases (%) |
|---------------------|-------------------------------------------------------|
| 2004                | ND                                                    |
| 2005                | ND                                                    |
| 2006                | ND                                                    |
| 2007                | 17.50                                                 |
| 2008                | 38.00                                                 |
| 2009                | ND                                                    |
| 2010                | 56.88                                                 |
| 2011                | 54.96                                                 |
| 2012                | 47.42                                                 |
| 2013                | 52.19                                                 |
| 2014                | 62.88                                                 |
| 2015                | 63.49                                                 |
| 2016                | 50.80                                                 |
| 2017                | 18.70                                                 |

ND = No Data

The effectiveness of the vaccines in providing necessary protection cannot be simply extrapolated using the proportion of vaccinated persons over the total cases. Vaccines can have effects at the level of individual and population. At the population level, the vaccinated persons “share” the immunity with the surrounding community (herd immunity). This condition depends on the coverage of vaccination and the interaction between unvaccinated and vaccinated groups [37]. Determining the protection of diphtheria vaccines only by considering the proportion is not enough. Therefore, the data must be followed by proper explanations in order to avoid misinterpretation by the reader.

3. Conclusion
In this paper we used big open data namely Indonesia Health Profile to picture the diphtheria in Indonesia from 2004 to 2017. From these annual reports, we highlighted some challenging issues in the form of inconsistencies of data. These issues will complicate data generation and summarization and eventually lead to inappropriate use of data. Some data (especially open data) can be potentially misinterpreted. Therefore, it is important to maintain consistency and reliability of the data during the whole stage. In addition, proper explanation of this type of data is crucial to avoid distorted conclusion by the reader. In order to be used to formulate better health strategies and policies, the collection, analysis and reporting process of big data must fulfil the proper requirements.

References
[1] Gisele R K 2016 J. Healthc. Commun. 1 1–3
[2] Fatt Q K and Ramadas A 2018 J. Healthc. Commun. 3 1–4
[3] Vayena E, Dzenowagis J, Brownstein J S and Sheikh A 2018 Bull. World Health Organ. 96 66–8
[4] Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B and Hart G, et al 2016 Front Public Heal. 4 1-6
[5] Directorate-General for Health and Consumers 2014 The Use of Big Data in Public Health Policy and Research Background information document Available from: http://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20141118_co07b_en.pdf
[6] Salas-Vega S, Haimann A and Mossialos E 2015 Heal. Syst. Reform. 1 285–300
[7] D’Agostino M, Samuel N O, Sarol M J, de Cosio F G, Marti M and Luo T, et al 2018 Rev. Panam. Salud. Pública 42 1-6
[8] Sedig K and Ola O 2014 The Challenge of Big Data in Public Health: An Opportunity for Visual Analytics. Online J Public Health Inform 5 1–21
[9] Kementerian Kesehatan Republik Indonesia 2014 Profil Kesehatan Indonesia Available from: http://www.pusdatin.kemkes.go.id/folder/view/01/structure-publikasi-pusdatin-profil-kesehatan.html
[10] Kapoor K 2017 Indonesia fights “extraordinary” diphtheria outbreak that has killed 32 Available from: https://www.reuters.com/article/us-indonesia-health-diphtheria/indonesia-fights-extraordinary-diphtheria-outbreak-that-has-killed-32-idUSKBN1E117C
[11] Harapan H, Anwar S, Dimiati H, Hayati Z and Mudatsir M 2018 Clin. Epidemiol. Glob. Heal. 1–2. Available from: http://linkinghub.elsevier.com/retrieve/pii/S2213398418300265
[12] Kementerian Kesehatan Republik Indonesia 2005 Profil Kesehatan Indonesia Tahun 2004 Available from: http://www.depkes.go.id
[13] Kementerian Kesehatan Republik Indonesia 2006 Profil Kesehatan Indonesia Tahun 2006 Available from: http://www.depkes.go.id
[14] Kementerian Kesehatan Republik Indonesia 2007 Profil Kesehatan Indonesia Tahun 2007 Available from: http://www.depkes.go.id
[15] Kementerian Kesehatan Republik Indonesia 2008 Profil Kesehatan Indonesia Tahun 2008 Available from: http://www.depkes.go.id
[16] Kementerian Kesehatan Republik Indonesia 2009 Profil Kesehatan Indonesia Tahun 2008 Available from: http://www.depkes.go.id
[17] Kementerian Kesehatan Republik Indonesia 2010 Profil Kesehatan Indonesia Tahun 2009 Available from: http://www.depkes.go.id
[18] Kementerian Kesehatan Republik Indonesia 2011 Profil Kesehatan Indonesia Tahun 2010 Available from: http://www.depkes.go.id
[19] Kementerian Kesehatan Republik Indonesia 2012 Profil Kesehatan Indonesia Tahun 2011 Available from: http://www.depkes.go.id
[20] Kementerian Kesehatan Republik Indonesia 2013 Profil Kesehatan Indonesia Tahun 2012 Available from: http://www.depkes.go.id
[21] Kementerian Kesehatan Republik Indonesia 2014 Profil Kesehatan Indonesia Tahun 2013 Available from: http://www.depkes.go.id
[22] Kementerian Kesehatan Republik Indonesia 2015 Profil Kesehatan Indonesia Tahun 2014 Available from: http://www.depkes.go.id
[23] Kementerian Kesehatan Republik Indonesia 2016 Profil Kesehatan Indonesia Tahun 2015 Available from: http://www.depkes.go.id
[24] Kementerian Kesehatan Republik Indonesia 2017 Profil Kesehatan Indonesia Tahun 2016 Available from: http://www.depkes.go.id
[25] Kementerian Kesehatan Republik Indonesia 2018 Profil Kesehatan Indonesia Tahun 2017 Available from: http://www.depkes.go.id
[26] Kompas 2012 Difteri Masih Menguancam Available from: https://nasional.kompas.com/read/2012/04/19/02244023/difteri.masih.menguancam
[27] Direktorat Surveilans dan Karantina Kesehatan 2017 Pedoman Pencegahan dan Pengendalian Difteri Available from: http://sehatnegeriku.kemkes.go.id/wp-content/uploads/2018/01/buku-
pedoman-pencegahan-dan-penanggulangan-difteri.pdf

[28] Sudjatmiko 2012 *Buletin Jendela Data dan Informasi Kesehatan* 1 2–6

[29] Zeegers M P, Bours M J L and Freeman M D 2016 *Methods Used in Forensic Epidemiologic Analysis: Forensic Epidemiology: Principles and Practice* (Academic Press) 71–110

[30] Clarke K E N 2017 *Review of the epidemiology of diphtheria 2000-2016* Available from: http://www.who.int/immunization/sage/meetings/2017/april/1_Final_report_Clarke_april3.pdf

[31] World Health Organization 2007 *Reporting of health statistics from service data: Country best practices* Available from: https://www.who.int/healthinfo/Report/20Workshop/20Annual/20Health/20Statistics/20Reporting_final(14.11.07).pdf

[32] Bansal S, Chowell G, Simonsen L, Vespignani A and Viboud C 2016 *J. Infect. Dis.* 214 375–9

[33] Lee E C, Asher J M, Goldlust S, Kraemer J D, Lawson A B and Bansal S 2016 *J. Infect. Dis.* 214 409–13

[34] Ruberto I, Marques E, Burke D S, and Van Panhuis W G 2015 *PLoS Negl. Trop. Dis.* 9 1–10

[35] Claxton G, Cox C, Gonzales S, Kamal R and Levitt L 2015 *Measuring the quality of healthcare in the U.S.* Available from: https://www.healthsystemtracker.org/brief/measuring-the-quality-of-healthcare-in-the-u-s/

[36] Ikatan Dokter Anak Indonesia 2014 *Persepsi yang salah tentang Imunisasi* Available from: http://www.idai.or.id/artikel/klinik/imunisasi/persepsi-yang-salah-tentang-imunisasi-bagian-2

[37] Lhariya C 2016 *J. Fam. Med. nad. Prim. Care* 5 7–15