GLOBAL BIOSECURITY

Feroza Binti Sulaiman, NK Semara Yanti, Dyah Ayu Shinta Lesmanawati, MJ Trent, CR MacIntyre & AA Chughtai. Language-specific gaps in identifying early epidemic signals – a case study of the Malay language. *Global Biosecurity*, 2019; 1(3).

RESEARCH ARTICLES

Language-specific gaps in identifying early epidemic signals – a case study of the Malay language

Feroza Binti Sulaiman[1], NK Semara Yanti[2], Dyah Ayu Shinta Lesmanawati[3], Mallory Trent[4], Chandini Raina MacIntyre[4], Abrar Ahmad Chughtai[5]

[1]Ministry of Health, Malaysia
[2]Udayana One Health Collaborating Center, Indonesia
[3]Universitas Gadjah Mada, Indonesia
[4]Biosecurity Program, Kirby Institute, University of New South Wales, Australia
[5]School of Public Health and Community Medicine, University of New South Wales, Australia

Abstract

Background: Internet-based surveillance systems have become invaluable tools to complement traditional surveillance in the early detection of infectious disease outbreaks. Search results limited to only English-language sources may lead to missed opportunities in global outbreak monitoring efforts. The Malay language has 290 million native speakers across the Southeast Asian region (Indonesia, Malaysia, Brunei, Singapore and the southern parts of Thailand and the Philippines), which is an important region for outbreaks. The aim of this research is to determine the completeness of reports of Malay language outbreak detection by commonly used open source surveillance platforms.

Methods: We searched ProMED-mail, HealthMap, and Google News to investigate outbreak events in Malaysia and Indonesia between 1st August 2016 and 31st August 2017. We also cross-checked published epidemic reports in Medline to determine the proportion of outbreaks that are published.

Results: A total of 371 entries from ProMED-mail and HealthMap were included. A Google News search query using Malay language keywords resulted in 453 news outbreak reports, 98 of which were missed by either or both ProMED-mail and HealthMap. During the study time frame, only 40 published epidemic reports were found in MEDLINE, showing the importance of informal data sources to identify outbreaks. Some diseases such as mumps, dengue and avian influenza had different local terminology in Indonesia and Malaysia.

Conclusions: Search results limited to only English-language sources may lead to reduced outbreak detection. The inclusion of keywords in the Malay language improves epidemic intelligence in the region, but needs to be nuanced for local terminology, which may differ between countries. Informal Internet-based surveillance can improve compliance with International Health Regulations (IHR) for low income countries, because it is less resource-intensive than formal surveillance. However, there is a need to include local language search terms to improve outbreak detection, especially in regions at high risk of emerging infectious diseases.

Keywords: Malay language, outbreak investigation, epidemic, language, surveillance, detection, timeliness and accuracy, Internet-based surveillance system

Introduction

The timeliness and accuracy of public health surveillance data are critical factors for implementing control measures to minimise the mortality and morbidity caused by an infectious disease outbreak[1]. Traditional disease surveillance uses formal source reporting systems that are limited by the lengthy process of collecting and validating data from health systems and laboratories[1,2]. The time spent in achieving data validation and accuracy is often an opportunity cost of early epidemiologic assessments[3,4]. This time-lag may impede public health efforts to rapidly mitigate and respond to impending outbreaks, such as those seen in post-disaster Haiti's cholera outbreak in 2010 and West Africa's Ebola virus outbreak in 2014[4,5]. Severe diseases that have potential political or economic consequences may also suffer from bias and less transparent formal information source reporting[6].

The World Health Organization's (WHO) 2005 International Health Regulations (IHR) is a legal framework which mandates governments to have fully functional surveillance capacities and report potential public health events of concern to the WHO within 24 hours[6]. This puts pressure on governments to rapidly detect, assess and report outbreaks that occur within

their boundaries. The rapid development and uptake of modern communication technologies have elevated the role of informal source reporting in public health surveillance, which can enable low income countries to comply with the IHR. According to the WHO, informal information sources already contribute more than 60% of initial outbreak reports[7,8]. The IHR has also clearly outlined that the WHO may use informal sources for outbreak intelligence to complement official government reports[6].

Internet-based surveillance systems have become invaluable tools for public health professionals in the early detection of infectious disease outbreaks[1,5,7,9]. Event -based internet surveillance systems have emerged following the advent of digital data analysis and social media that have influenced people's health-related information seeking behaviour[10]. The gathering of outbreak intelligence through publicly available data could improve the sensitivity and timeliness in detecting health-related events, minimise outbreak impacts, and assist in forecasting emerging infectious disease outbreaks[3,10]. Recent studies[6,11] found that Internet-based surveillance systems showed alerts on average 1-6 days before the formal information was released. However, utilising informal Internet-based public data in outbreak detection requires a necessary trade-off between timeliness and accuracy[1,6]. In general, informal sources correlate well with formal reports, and their main value is in early detection of epidemics[1]. Based on their source of information and data handling methods, some informal information sources are more prone to biases due to heterogeneity, imbalanced sources, and the dynamic, changing nature of news media and outbreak reporting[1,5,10,12].

Internet-based surveillance systems are based on open source data, social media, and news reports. The assumption is that the incidence of disease is correlated with specific terms used in news reports or in open searching for information related to the diseases, symptoms and outbreak[13]. Identifying keywords in any language that strongly correlate with the local disease notifications could potentially be used for rapid intelligence on emerging public health threats[3]. Although the global online media is dominated by certain major languages, initial outbreak reports are often first reported in a local language[14]. In data processing, the data collected from local and international sources in multiple languages require either human linguist experts, machine auto-translation, or natural language processing technology[14]. The different approaches to translation are influenced by the resource and maintenance time availability for each language, and the level of quality required[14]. Out of a total of 50 event-based Internet systems, some of which are not currently online, 34 (68%) systems collected or disseminated their data in a single language, which was primarily English (n=21), while only 16 (32%) were multilingual[10]. This can lead to bias and under reporting of outbreaks.

There are currently 37 event-based Internet surveillance systems that are functioning and available online, including ten collectively known as InfluenzaNet[10]. Of these, four are freely accessible: ProMED-mail, HealthMap, InfluenzaNet and Medical Information System (MedISys)[1,15]. Researchers and public health authorities have developed these systems, which are characterised by: type of information source; data collection, processing and analysis method; format of dissemination; language; target audience; area of service; moderation; and accessibility[1,15].

'Bahasa Melayu', or the Malay language, is a major language spoken by 290 million people across Indonesia, Malaysia, Brunei, Singapore and the southern parts of Thailand and the Philippines[16]. It is the official national language in Indonesia, Malaysia, Singapore and Brunei and exists in different standardised forms in each country, such as Bahasa Indonesia (Indonesia) and Bahasa Melayu (Malaysia, Singapore and Brunei Darussalam)[16, 17]. In this paper, the term 'Malay language' will be used to refer to this group. Based on current literature, there have been no studies published on Internet-based keyword in the Malay language for infectious disease outbreaks.

The aim of this study was to determine the proportion of outbreaks affecting Malay language regions that are missed by common informal surveillance systems.

Methods

Databases

In this study, we conducted a comparison between the following three Internet-based database systems: ProMED-mail, HealthMap and Google News.

Google News is an automated news aggregator that collects information from more than 50000 sources worldwide[18]. For news selection, it uses computer algorithms of selected keywords. Google News has a technology called PageRank to classify resources by frequency and importance. Since 2012, Google News has captured news in 28 different languages from more than 60 regions in the world[18]. This capacity allows Google News to capture epidemic reports in both English and local languages. Because it is believed to be the most sensitive source of epidemic reporting, Google News is considered the gold standard in information mining[18].

Program for Monitoring Emerging Diseases (ProMED-mail) is an Internet-based surveillance system developed by the International Society for Infectious Diseases in 1994 to monitor and rapidly disseminate infectious disease outbreak information to healthcare professionals and the community worldwide[1,19]. Sources of information come from the media (73%), formal reports (26%) and other systems (1%)[17] (1,19). Reports are received mainly from ProMED's 70,000 subscribers and healthcare professionals worldwide but are supplemented by searches on the Internet and traditional media by

multilingual personnel[1,19]. Reports are manually moderated by subject matter experts, who review and verify the information before posting to the network[19]. Entries may be rejected by an editor based on their relevance and accuracy[1]. Final reports are coded and shared on the website, social media accounts, mailing lists and listserv software[1,19]. ProMED also provides a platform for the international infectious disease community to engage in discussion on issues of mutual concern, share information, and collaborate on public health outbreak response efforts[19]. It is available in nine languages and has regional networks catering for Latin America (Portuguese and Spanish), independent states of the former Soviet Union (Russian), Mekong Basin region in Southeast Asia (English), Africa (French and English), Middle East (English) and South Asia (English)[19].

Founded by a team of public health researchers in Boston Children's Hospital in 2006, HealthMap is a real-time infectious disease intelligence system which utilises disparate data sources to provide an ongoing view of emerging public health threats and monitor disease outbreaks[20]. Its diverse audience includes health professionals, health and government departments, libraries and even international travellers[20]. HealthMap primarily utilises informal information sources, of which 49% are from media, 39% from other systems, and 12% from official sources[1]. This system is partially moderated, with an automated online querying system using language-specific search terms in fifteen different languages to search internet data[1,6,14]. Data collected from websites, news aggregators and social media platforms are categorised, clustered and filtered into one of five alert levels, with those labelled as 'breaking news' marked onto the website's interactive map[1,21]. Through this automated process, information is disseminated online in nine languages and users opting for pushed information are able to specify their parameters of interest, such as by disease or location[1,20] .

MEDLINE is produced by the U.S. National Library of Medicine in Bethesda, Maryland. It is a bibliographic database from which more than 24 million references to journal articles in life sciences are obtained[22]. A specific feature of MEDLINE is the NLM Medical Subject Headings (MeSH) to index the records[23]. MEDLINE is the part of Pubmed database, one of the NLM National Center for Biotechnology Information (NCBI) databases. The database includes literature published since 1966 from more than five thousand journals worldwide. The scope of topics covered by MEDLINE is biomedicine and health, including clinical and public health studies. Citation from social science are also included if they are linked to a biomedical subject[23]. The publications in MEDLINE database are not only scholarly journals but also case studies, news, article from magazines, and newsletters. The majority of publications are in English[24].

Search strategy

We conducted an initial exploratory review of peer-reviewed journals, grey literature, publicly available epidemiological data, and Malaysian and Indonesian government websites to gain familiarity with outbreak intelligence and surveillance systems in Malaysia and Indonesia.

Using the keyword "Malaysia" and "Indonesia", we conducted a search query in ProMED-mail and HealthMap for entries dated between August 1, 2016 to August 31, 2017. We then compared the entries collected from ProMED-Mail and HealthMap against Google News.

The search query we conducted in Google News used a selection of Malay language keywords based on a list of notifiable infectious diseases, emerging and re-emerging infectious diseases, and common disease signs and symptoms[25-31]. For each search query, we included the first ten relevant and distinct events which occurred in Malaysia and Indonesia respectively within the study time frame and were reported by the local media in the Malay language. Only the first ten unique events were selected due to the volume of results, many of which were redundant. We excluded news items that were either not related to infectious disease outbreaks or duplicates of similar events, and those that were reported in both Internet-based surveillance systems.

We searched MEDLINE to determine the proportion of identified outbreaks which were formally published in peer reviewed literature or other sources captured by Medline. For MEDLINE, we used Medical Subject Headings (MeSH) to create a search strategy within the same time period. The search terms were outbreak*, epidemic*, pandemic* AND Malaysia, Indonesia. We conducted the search query without language restrictions. Publication that were not related to outbreak events were excluded.

Results

A total of 371 entries from both ProMED-mail and HealthMap were included for analysis (Figure 1). A search keyword using "Malaysia" and "Indonesia" in the ProMED-mail archive returned 139 and 14 entries respectively, of which 37 were eligible for analysis. In HealthMap, an initial search by location ("Malaysia" and "Indonesia") resulted in a total of 7180 entries for Malaysia and 416 entries for Indonesia. Due to the large number of entries retrieved for Malaysia which included sources other than English and Malay languages, the HealthMap entries source option was filtered to "English language", "Google Bahasa", "Ministry of Health Sites Bahasa", "Ministry of Health Sites Facebook" and "Ministry of Health Sites Twitter". Meanwhile, the results for Indonesia were only filtered to "English Language"; as the use of "Google Bahasa" mostly referred to sources in Bahasa Melayu (Malaysian language). Duplicate reports on outbreak events and animal or plant diseases were excluded for a final total of 334 entries. Entries

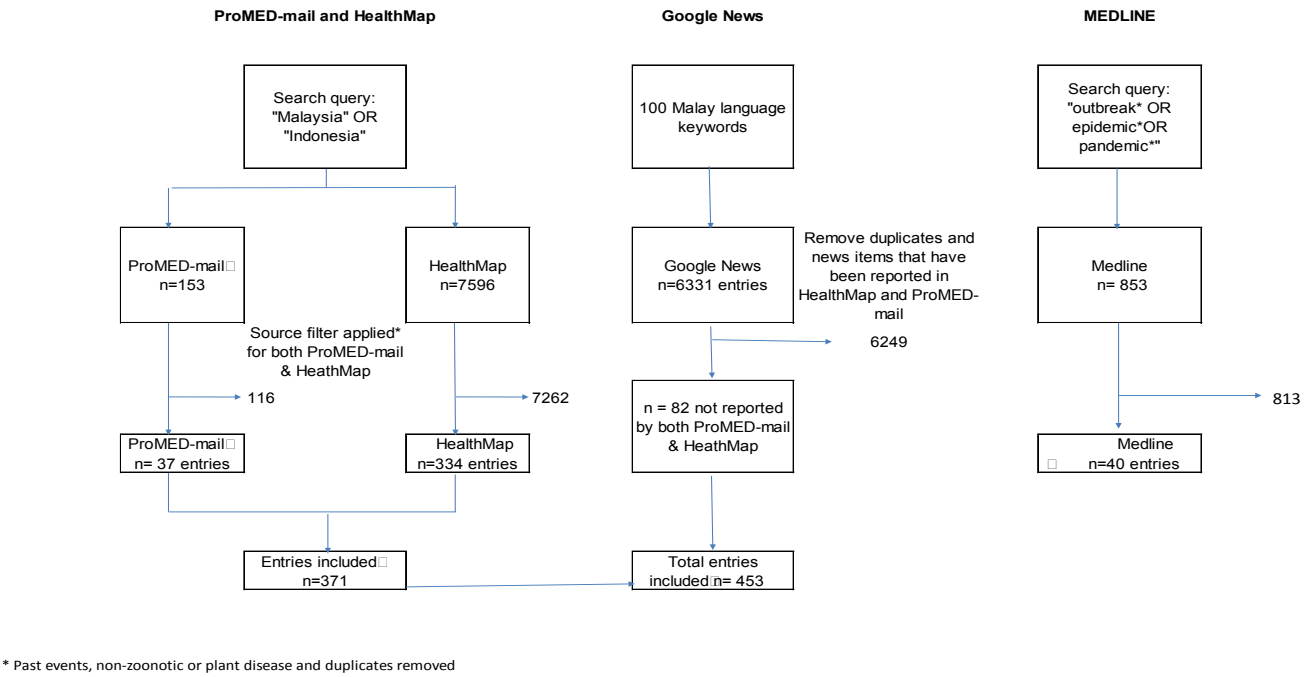
included for analysis from the databases were then categorised according to date of report entry, disease type, location and post headline (Table 1-2 in Appendix). Based on these results, the five highest disease entries were Dengue (140/371, 37.74%), Rabies (63/371, 16.98%), Avian influenza (42/371, 11.32%), Measles (24/371, 6.47%), and Zika (18/371, 4.85%).

More than 100 keywords in the Malay language were identified and categorised by disease and common signs and symptoms of infectious disease conditions (Table 5 in Appendix). Local terminology for specific diseases varied between Indonesia and Malaysia. For example, the term for dengue was "denggi" in Malaysia while it was called "DBD" in Indonesia. The term for avian influenza was "demam selsema burung" in Malaysia and Indonesian called it "flu burung".

A Google News search query using selected keywords in the Malay language resulted in 6331 news items. These news items were then removed based on the exclusion criteria. This resulted in 453 outbreak reports including 371 entries that were reported in HealthMap and ProMED-mail. Of these 453 reports in Google News, 82 entries were missed by both Internet-based surveillance systems. There were also 16 entries missed either by ProMED-mail or HealthMap which made a total of 98 reports found in Google News that was not captured by one or both systems (Figure 2). Overall, ProMED-mail missed more Malay-language news items than HealthMap (416/453, 91.8% vs 109/453,24.1%). The top 5 most frequently items from the 98 missed outbreak news were rabies (22/98, 22.45%), dengue (13/98, 13.27%), meningitis (8/98, 8.16%), measles (7/98, 7.14%) and foodborne illness (7/98, 7.14%) (Table 1). Of the 98 missed news report, 76.5% were outbreaks from Indonesia, while the rest 23.5% occurred in Malaysia.

In MEDLINE there were 853 publications possibly related to outbreak events in Malaysia and Indonesia found within the study time frame. After removal of non-outbreak events, duplications, and animal or plant diseases, only 40 reports remained. Figure 3 shows total outbreak reports captured by each database.

Figure 1. Flowchart of the inclusion and exc



lusion progression of Internet-based databases system entries.

Table 1. The top 5 missed infectious disease outbreaks in Malaysia and Indonesia August 1, 2016 to August 31, 2017.

Missed diseases	Number	Proportion of missed news items by Internet-based surveillance systems
Rabies	22	22.45%
Dengue	13	13.27%
Meningitis	8	8.16%
Measles	7	7.14%
Foodborne illness	7	7.14%

Figure 2. Missed outbreak reports in HealthMap and Promed-Mail from Indonesia and Malaysia, August 1, 2016 to August 31, 2017

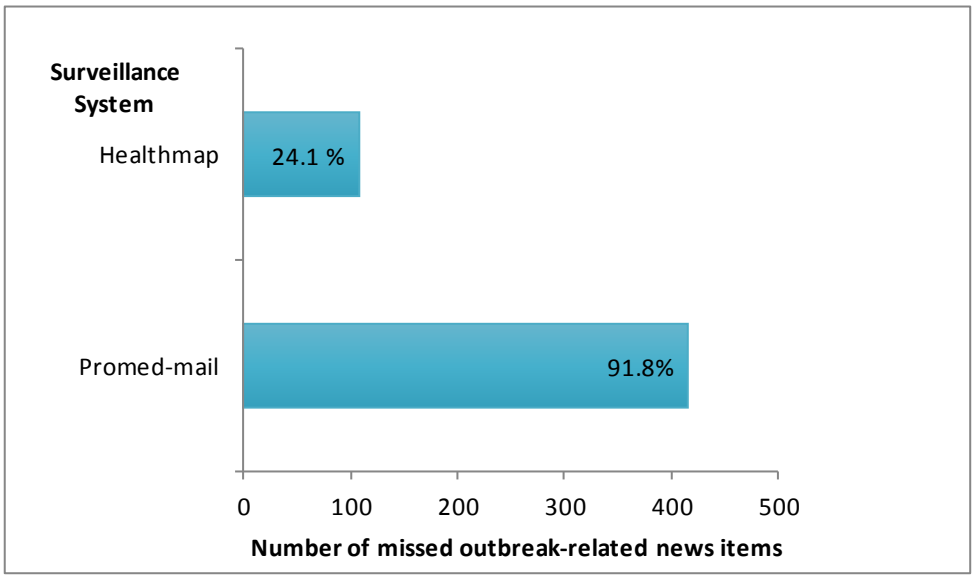
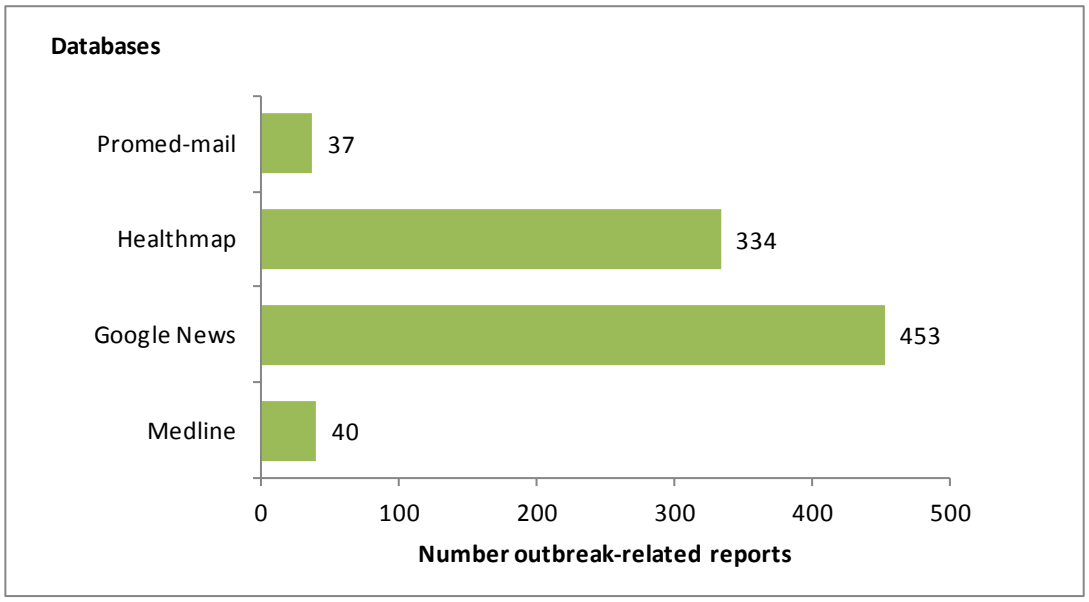


Figure 3. Total outbreak-related reports from Internet-based sources, August 1, 2016 to August 31, 2017.



Discussion

Commonly used outbreak surveillance systems may miss reports in the Malay language, which is spoken by over 290 million people in Indonesia, Malaysia, Brunei, Singapore and the southern parts of Thailand and the Philippines. The region is important for emerging infectious diseases of pandemic potential, including H5N1 avian influenza[32]. The most accessible rapid intelligence sources such as HealthMap, whilst they do search in a range of languages, are biased toward English language reporting and may miss some non-English language reports. Both Internet-based surveillance systems in this study had missed a large proportion of infectious disease outbreaks reported in Malay languages. ProMED-mail missed 91.8% of the Malay-language news items, while HealthMap missed 24.1%. Both systems use common data sources and will often feed off one another for information[1,10]. These data sources include Google News and official WHO and government reports, with HealthMap additionally collecting data from social media platforms[1]. Final reports were influenced by the method adopted for data collection and processing and keyword language used in searching for outbreak-related information[1]. HealthMap, which missed fewer Malay-language news items than ProMED-mail, automatically aggregates data from over 200,000 sources using language-specific terms in fifteen different languages[6]. ProMED-mail is

dependent on the reports it receives from its contributors (crowd sourced) and multilingual personnel who search the internet for data[1]. Moreover, manual moderation introduces the risk for selection bias as one editor filters the reports ProMED-mail receives for relevancy before being forwarded on to expert moderators[1].

Final reports considerably overlap as both HealthMap and ProMED-mail use each other as a source of data[10]. These overlaps and mutual reliance for final reports would potentially lead to a similar editorial bias in all three systems to favour English-language data. As they similarly rely on online news articles for outbreak-related events information, data collection could also be influenced by both human resource constraints in providing comprehensive media coverage and the 'crowd out' effect of 'sensational' diseases over other outbreak-related events[33]. A study by Scales et al[33] on English-language data sourced from Google News by HealthMap showed that compared to other weekdays, there was a drop of 58.3% to 14.7% in reported outbreak events on Saturday, Sunday and Monday. A 'sensational' disease, which can potentially cause widespread fear, could also predominate media coverage and lead to the decline of other disease events reported in Google News[31]. As seen with the recent rabies outbreak in Sarawak, Malaysia between July and August of 2017, 77% (47/61) of HealthMap entries were related to rabies in comparison to only 9.8% (6/61) on dengue, which is the most common infectious disease in Malaysia with an incidence rate of 392.96 per 100,000 population[29].

Compared to the total of 453 entries reported in HealthMap, ProMED-mail, and Google News, MEDLINE only published 40 epidemics, which was less than 10% of outbreak reports captured by Internet-based outbreak databases. This means that only a small proportion of epidemic reports are published, highlighting the importance of informal Internet sources of outbreak intelligence.

Even though Indonesia has similar geographic and demographic patterns to Malaysia, its rate of detected news items is 21.2% lower[34]. Considering the land size, population and digital penetration of Indonesia, the number of the news item picked up by English-based internet surveillance systems such as ProMED-mail and HealthMap is very low. Therefore, many outbreaks reported in the local language are likely being missed.

In Indonesia, English is taught as a foreign language instead of a first or second language, as practiced in Singapore and Malaysia. As a foreign language, English has no official status in domains such as government and education systems, resulting in the low level of English literacy among the majority of Indonesians [35-37]. Consequently, there are limited sources of English news for surveillance purpose. In addition, the media is inclined to focus on political or economic issues rather than on health[38]. Thus, local-scale disease outbreaks are rarely picked up by the media.

Another potential reason for Indonesia's lower media coverage for health related news is that despite having official websites and social media presence, the Indonesian Ministry of Health (MoH) does not regularly publish health situation updates to the public. Press releases on outbreaks are only issued when there is a large-scale outbreak that requires national attention. By contrast, Malaysia has weekly situation reports on a number of infectious diseases, such as dengue, chikungunya and Zika disseminated via the Malaysian MoH's social media accounts and to the media with daily dengue case counts and weekly hotspot locations made available on the Ministry's iDengue website[39]. The Malaysian MoH also routinely uses an internal web-based early outbreak reporting, which combines both indicator- and event-based reporting functions specific for Malaysia, but this is not publicly available. Additionally, when there is a significant public health concern, Malaysia MoH issues press releases in both Malay and English-language[40].

A limitation of this study is the use of grey literature as sources of news items, which may have not been validated prior to publishing. The news items were also manually screened, which may have resulted in errors during news screening and analysis. An additional limitation of this study is the short time period which it covers (thirteen months). Media-driven interest or real-life events may have altered search behaviours, leading to potential language shifts during the study period[3,13]. Internet-based surveillance systems are also limited to individuals or their proxies who seek health-related information on the internet[13]. Nonetheless, Internet penetration in Malaysia and Indonesia is as high as at 71% (24.1 million) and 51% (132.7 million) respectively. Of these, texting (92.7%), seeking information (90.1%) and social networking (80.0%) were the main three activities among online users[40].

Conclusion

The Malay language has 290 million native speakers across the Southeast Asian region in Indonesia, Malaysia, Brunei, Singapore and the southern parts of Thailand and the Philippines. Informal, internet-based surveillance can improve compliance with IHR, particularly for low income countries. However, using search results restricted to only English and European language sources may limit the detection of outbreaks. Results limited to only English and European language sources may miss important and potentially serious outbreaks. Both Internet-based surveillance systems in this study relied heavily on English-language formal and informal sources and missed one or more Malay-language infectious disease outbreak news items. Nuanced surveillance is needed in Malay language to detect subtle differences in disease terminology between countries. The inclusion of keywords in the Malay language improves epidemic intelligence in the Southeast Asia region. There is a need to include local language search terms to improve internet-based outbreak detection, especially in regions at high risk of emerging infectious diseases.

References

1. Yan SJ, Chughtai AA, Macintyre CR. Utility and potential of rapid epidemic intelligence from internet-based sources. Int J Infect Dis. 2017 Oct 1;63:77-87. https://doi.org/10.1016/j.ijid.2017.07.020
2. Declich S, Carter AO. Public health surveillance: Historical origins, methods and evaluation. Bull World Health Organ. 1994;72(2):285-304.
3. Milinovich GJ, Avril SMR, Clements ACA, Brownstein JS, Tong S, Hu W. Using internet search queries for infectious disease surveillance: screening diseases for suitability. BMC Infect Dis. 2014;14(1):690. https://doi.org/10.1186/s12879-014-0690-1
4. Anema A, Kluberg S, Wilson K, Hogg RS, Khan K, Hay SI, et al. Digital surveillance for enhanced detection and response to outbreaks. Lancet Infect Dis. 2014 Nov 1;14(11):1035-7. https://doi.org/10.1016/S1473-3099(14)70953-3
5. Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am J Trop Med Hyg. 2012 Jan 1;86(1):39-45. https://doi.org/10.4269/ajtmh.2012.11-0597
6. Bahk CY, Scales DA, Mekaru SR, Brownstein JS, Freifeld CC. Comparing timeliness, content, and disease severity of formal and informal source outbreak reporting. BMC Infect Dis. 2015 Dec 20;15(1):135. https://doi.org/10.1186/s12879-015-0885-0
7. Brownstein JS, Freifeld CC, Madoff LC. Digital Disease Detection - Harnessing the Web for Public Health Surveillance. N Engl J Med. 2009 May 21;360(21):2153-7. https://doi.org/10.1056/NEJMp0900702
8. World Health Organization. Epidemic intelligence - systematic event detection [Internet]. WHO. WHO; 2016 [cited 2017 Oct 31]. Available from: http://www.who.int/csr/alertresponse/epidemicintelligence/en/
9. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EHY, Olsen JM, et al. Using social media for actionable disease surveillance and outbreak management: A systematic literature review. Vol. 10, PLoS ONE. 2015. https://doi.org/10.1371/journal.pone.0139701
10. O'Shea J. Digital disease detection: A systematic review of event-based internet biosurveillance systems. Int J Med Inform. 2017 May 1;101:15-22. https://doi.org/10.1016/j.ijmedinf.2017.01.019
11. Barboza P, Vaillant L, Mawudeku A, Nelson NP, Hartley DM, Madoff LC, et al. Evaluation of Epidemic Intelligence Systems Integrated in the Early Alerting and Reporting Project for the Detection of A/H5N1 Influenza Events. PLoS One. 2013;8(3). https://doi.org/10.1371/journal.pone.0057252
12. Kamel Boulos MN, Sanfilippo AP, Corley CD, Wheeler S. Social Web mining and exploitation for serious applications: Technosocial Predictive Analytics and related technologies for public health, environmental and national security surveillance. Computer Methods and Programs in Biomedicine. 2010;100(1):16-23. https://doi.org/10.1016/j.cmpb.2010.02.007
13. Milinovich GJ, Williams GM, Clements ACA, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. Vol. 14, The Lancet Infectious Diseases. Elsevier; 2014. p. 160-8. https://doi.org/10.1016/S1473-3099(13)70244-5
14. Hartley DM, Nelson NP, Arthur RR, Barboza P, Collier N, Lightfoot N, et al. An overview of internet biosurveillance. Clin Microbiol Infect. 2013 Nov 1;19(11):1006-13 https://doi.org/10.1111/1469-0691.12273
15. Choi J, Cho Y, Shim E, Woo H. Web-based infectious disease surveillance systems and public health perspectives: a systematic review. BMC Public Health. 2016 Dec 8;16(1):1238. https://doi.org/10.1186/s12889-016-3893-0
16. Jehwahe P. Jurnal Pustaka Budaya. Pustaka Budaya. 2014;1(2):11
17. Alisjahbana ST. The concept of language standardization and its application to the Indonesian language. Linguistic Minorities and Literacy: Language Policy Issues in Developing Countries. 1984;26:77
18. Das AS, Datar M, Garg A, Rajaram S, editors. Google news personalization: scalable online collaborative filtering. Proceedings of the 16th international conference on World Wide Web; 2007: ACM. https://doi.org/10.1145/1242572.1242610
19. International Society for Infectious Diseases. About ProMED-mail [Internet]. International Society for Infectious Diseases. 2010 [cited 2017 Nov 2]. Available from: http://www.promedmail.org/aboutus/
20. HealthMap. About HealthMap [Internet]. Boston Children's Hospital. 2015 [cited 2017 Nov 2]. Available from: http://www.healthmap.org/site/about
21. HealthMap. HealthMap [Internet]. Boston Children's Hospital. 2017 [cited 2017 Nov 2]. Available from: http://www.healthmap.org/en/index.php
22. US National Library of Medicine. Medline https://www.nlm.nih.gov/bsd/medline.html [cited 2018]. Available from: https://www.nlm.nih.gov/bsd/medline.html
23. Suarez-Almazor ME, Belseck E, Homik J, Dorgan M, Ramos-Remus C. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. Controlled clinical trials. 2000;21(5):476-87. https://doi.org/10.1016/S0197-2456(00)00067-2
24. Woods D, Trewheellar K. Medline and Embase complement each other in literature searches. BMJ: British Medical Journal. 1998;316(7138):1166. https://doi.org/10.1136/bmj.316.7138.1166

25. Ministry of Health Malaysia. Health Facts 2012 [Internet]. Ministry of Health Malaysia. 2012 [cited 2017 Nov 1]. p. 13. Available from: http://www.moh.gov.my/images/gallery/stats/heal_fact/health_fact_2012_page_by_page.pdf
26. Ministry of Health Malaysia. Health Facts 2013 [Internet]. Ministry of Health Malaysia. 2013 [cited 2017 Nov 1]. p. 20. Available from: http://www.moh.gov.my/images/gallery/publications/HEALTH FACTS 2013.pdf
27. Ministry of Health Malaysia. Health Facts 2014 [Internet]. Ministry of Health Malaysia. 2014 [cited 2017 Nov 1]. p. 19. Available from: http://www.moh.gov.my/images/gallery/publications/HEALTH FACTS 2014.pdf
28. Ministry of Health Malaysia. Health Facts 2015 [Internet]. Ministry of Health Malaysia. 2015. p. 19. Available from: http://www.moh.gov.my/english.php/pages/view/56
29. Ministry of Health Malaysia. Health Facts 2016 [Internet]. Ministry of Health Malaysia. 2016 [cited 2017 Nov 1]. p. 19. Available from: http://www.moh.gov.my/images/gallery/publications/KKM HEALTH FACTS 2016.pdf
30. (Kemenkes) KKR. Pedoman penyelengaraan sistem surveilans epidemiologi penyakit menular dan penyakit tidak menular terpadu. NOMOR 1479/MENKES/SK/X/2003. Pemerintah Republik Indonesia; 2003
31. Kemenkes R. Survey Kesehatan Nasional 2013. Kementerian Kesehatan RI; 2013.
32. Coker RJ, Hunter BM, Rudge JW, Liverani M, Hanvoravongchai P. Emerging infectious diseases in southeast Asia: regional challenges to control. The Lancet. 2011;377(9765):599-609. https://doi.org/10.1016/S0140-6736(10)62004-1
33. Scales D, Zelenev A, Brownstein JS. Quantifying the effect of media limitations on outbreak data in a global online web-crawling epidemic intelligence system, 2008-2011. Emerg Health Threats J. 2013;6:21621. https://doi.org/10.3402/ehtj.v6i0.21621
34. Ding J-A, Koh LC, Surin JA, Dragomir M, Thompson M, Watts G, et al. Mapping digital media: Malaysia. Open Society Malaysia, Kuala Lumpur. 2013.
35. Lauder A. The status and function of English in Indonesia: A review of key factors. Makara Hubs-Asia. 2010;8(3).
36. Dardjowidjojo S. The role of English in Indonesia: A dilemma. Rampai Bahasa, Pendidikan, dan Budaya: Kumpulan Esai Soenjono Dardjowidjojo. 2003:41-.
37. Dardjowidjojo S. The socio-political aspects of English in Indonesia. TEFLIN Journal. 1996;3(1):1-13.
38. Kakiailatu T. Media in Indonesia: Forum for political change and critical assessment. Asia Pacific Viewpoint. 2007;48(1):60-71. https://doi.org/10.1111/j.1467-8373.2007.00330.x
39. CPRC Kementerian Kesihatan Malaysia. Current situation of dengue, chikungunya and zika in Malaysia week 43/2017 [Internet]. 2017 [cited 2017 Nov 2]. Available from: https://www.facebook.com/kkmcprc/posts/728929057299781
40. Malaysian Communications and Multimedia Comission. Internet users survey 2016 [Internet]. Malaysian Communications and Multimedia Commission. 2016 [cited 2017 Nov 2]. p. 75. Available from: https://www.skmm.gov.my/skmmgovmy/media/General/pdf/IUS2015-Appendix_281216_final-20171016.pdf

How to cite this article: Feroza Binti Sulaiman, NK Semara Yanti, Dyah Ayu Shinta Lesmanawati, MJ Trent, CR MacIntyre & AA Chughtai. Language-specific gaps in identifying early epidemic signals – a case study of the Malay language**.** *Global Biosecurity, 2019; 1(3)*

Published: October 2019

Global Biosecurity is a peer-reviewed open access journal published by University of New South Wales.