ABSTRACT

This article presents the results of a discursive analysis conducted on a corpus of 12 systematically selected research articles in the discipline of public health surveillance. The analysis focused on the use of complex language and how this affected the interpretation of the article. Complex language was broken down into three types: complex language from public health surveillance, complex language from other disciplines, and common language that takes on a complex meaning when used in the context of public health surveillance. Results showed that all articles contained all three types of complex language usage. Language use from public health surveillance was more common in the Introduction and Discussion sections, while language from other disciplines was more common in the Method and Results sections. However, articles in the corpus were short and detail-oriented which meant phrases containing complex language did not possess high syntactic density. Complex language was generally found only when referring to a specific method or concept, rather than unnecessarily complicating an explanation. Based on these findings, complex language in public health surveillance was found to be common but necessary.

Keywords: public health surveillance, complex language, discursive analysis

Complex Language Usage in Public Health Surveillance

INTRODUCTION

The discipline of public health surveillance contains a substantial amount of complex, technical vocabulary in its research articles. As defined by the World Health Organization (WHO), "public health surveillance is the continuous, systematic collection, analysis and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice" (WHO, 2018). As the field is naturally interdisciplinary, methods, terminology, and procedures are drawn from a diverse spectrum of different fields. Complex, discipline-specific language is therefore varied but prevalent in many research articles in this field. It is often the case that such language in research articles serves as a method to allow the writer to make appeals to shared knowledge (Hyland, 2011). Furthermore, genre theory predicts that increased specificity of a rhetorical situation will employ more distinctive language, and that this will in turn present a greater barrier for those unfamiliar with the discipline (Giltrow et al., 2014).

The question that remains, however, is whether this level of complexity in language usage is a necessary component of the article? Reader opinion regarding the validity and importance of the article plays an inherent role in whether complexity of expression is deemed necessary or unnecessary (Giltrow et al., 2014). An individual's judgement of the process of academic and scholarly research may also play a role in their assessment of the need for complex terminology. Previous research indicates that complicated writing in the research article genre can be viewed as elitist and can even arouse suspicion from those outside the field

3

(Giltrow et al., 2014). Nevertheless, it can be necessary to convey vital prerequisite knowledge, without which the research would lose its originality. However, the degree to which this has an effect in public health surveillance remains unclear, particularly given the interdisciplinary nature of the field.

This research article will be focused around the question of how complex language in public health surveillance insulates research from outside interpretation, and whether this is necessary or unnecessary?

METHODS

This research involved the systematic selection of a corpus of 12 research articles selected from the *Journal of Medical Internet Research-Public Health and Surveillance* (JMIR-JPHS). The selection methodology was the first two original research articles from the six most recent issues of the journal were included in the corpus. This journal was selected with the assistance of an institutional informant from Woodward Library, UBC. Based on the recommendations of the informant, the choice of journals was narrowed to *The American Journal of Epidemiology, The American Journal of Public Health, and the Journal of Medical Internet Research-Public Health and Surveillance.* The expertise of the informant was chosen due to their specialization within the field of interest and familiarity with the journals and publications within the discipline. Based on the recommendations of the informant, this particular journal was then selected for two primary reasons. The journal is of high impact in the discipline and closely aligns with the researcher's current and future areas of both professional and academic interest. Furthermore, the journal publishes original research

articles and thus remains consistent with the proper methodology of corpus construction. Due to the need for a synchronous corpus (Bauer & Aarts, 2000), this selection method was used to ensure a corpus of currently relevant discourse. Additionally, public health surveillance is a constantly evolving discipline and so the preference for recent articles aligns with common disciplinary practice. The selected journal was originally published in 2015 and is published quarterly. Therefore, as 12 issues have not yet been published, the researcher decided to select two articles from the six most recent issues to ensure consistent selection methodology.

Based on the textual data, the need for an operational definition became clear to ensure consistent coding. For this research, complex language was defined as terminology or jargon that is specific to the discipline, which would not be understood by non-specialists. Furthermore, the data and subsequent analysis required for the operational definition to be broken into three parts:

<u>Type 1:</u> Complex terminology that is common to core literature and practice in public health surveillance

<u>Type 2:</u> Complex terminology from other disciplines that is used in this literature due to its inherent interdisciplinarity

<u>Type 3:</u> Common terminology that takes on a specific, complex meaning within the context of public health surveillance

The corpus was colour coded for all instances of each of the three types of complex language. Analysis will focus on the qualitative aspects rather than on quantitative calculations of frequency. As context and syntactic density (Giltrow et al., 2014) are key factors to consider when analyzing complex language, qualitative analysis was deemed to be more appropriate.

RESULTS AND DISCUSSION

Using the coding method outlined above, each article in the corpus was found to contain multiple occurrences of each of the three different types of complex language usage. The following are examples from the corpus of each of the three types, with emphasis added on the complex language. For additional examples, please see Appendix 2.

<u>Type 1:</u> The system was originally developed to primarily capture monthly individual level malaria case data from government-supported health facilities in *malaria-endemic* provinces (Fu et al., 2017)

<u>Type 2:</u> The scale asked 5 questions about the usefulness and ease of Twitter on a 5point *Likert* scale. 'Using Twitter improves my job performance' has a *mean* of 3.80, *sd* of 0.86, *n*=49 (Hart et al., 2017)

<u>Type 3:</u> HIV *surveillance* data exchange between the public health departments of DC, Maryland, and Virginia reduced the number of cases misclassified as DC residents. (Hamp et al., 2018) Beginning with the first example, we see the authors use the word endemic to refer to a region in which malaria is regularly found, as opposed to only being found in unique outbreak events. Professionals in the field of public health, particularly epidemiologists, rely heavily on the concept of endemicity in order to study and predict disease trends, as well as inform public health intervention. However, despite the concept being part of the core knowledge of the discipline, it is specific to the discipline and thus non-specialists may not know the meaning, which would affect their ability to interpret the results. This may lead to a gap in understanding of the purpose of the surveillance system described in the article.

Data management and study design techniques are often described with references to terminology from the disciplines of statistics and data science. In the second example, Hart et al. (2017) described their data analysis by using statistical terms such as mean and standard deviation (sd). As data analysis has a close-knit relationship with epidemiology and public health, these terms are commonly found in articles in this discipline, despite being concepts native to statistics. By including these terms in a public health research article, the author is presuming the reader, whether they are a public health professional or not, has a minimum level of prerequisite knowledge in statistics. Thus, this is an example of type 2 of the definition of complex language.

A very prominent example of type 3 of the definition was the use of the word surveillance in the third example. Even appearing in the name of the discipline, this word was found to be widely used across all articles in the corpus, and even in articles titles (Rosenthal et al., 2017). While this word is commonly used within both conversational and written English, it takes on a very specific meaning in the context of this discipline. When used in public health, this term refers to the monitoring of disease trends or factors affecting the health of the population in question. In the case of the above example, it refers to a specific type of data collected for HIV cases in northeastern USA. Contrary to its common use, surveillance in public health does not refer to monitoring individuals' behaviour. Therefore, readers who do not possess background knowledge in public health may be unable to interpret aspects of how the research was conducted as they may lack knowledge of the meaning of the word in this context.

In addition to the frequent occurrences of each type of complex language, general patterns were also observed throughout the corpus. Occurrences of type 1 complex language were more common in the Introduction and Discussion sections, while type 2 occurred more often in the Methods and Results sections. A possible explanation is that Introduction and Discussion sections are more likely to contain interpretation and contextual information from the discipline, as well as claims of centrality (Samraj, 2002). Therefore, this information would require grounding within the core knowledge of the discipline, which would result in an increased amount of type 1 complex language being used. Conversely, the Methods and Results sections typically indicate how the research was conducted or how calculations were done using quantitative data. Since research in public health usually involves sampling and statistical tests, this would result in more interdisciplinary complex language (type 2) occurring in these sections.

Articles in the corpus were short and detail-oriented, which is common in the discipline rather than due to specific journal requirements. The article length prevented extraneous information and explanation about the research. As a result, phrases containing complex language generally had low syntactic density, which allows for easier interpretation by the reader (Giltrow et al., 2014). Use of complex language was found to typically occur only when referencing specific concepts or methods, as opposed to needlessly complicating a simple explanation. Complex language occurs, in most cases, to eliminate ambiguity with regard to what the author is describing. Such ambiguity would otherwise be present with the use of more common terminology.

CONCLUSION

Based on the results of this analysis, it is apparent that complex language is highly prevalent in research articles within public health surveillance. While this may be exclusionary in some capacity to non-specialists, low syntactic density prevents entire topics or sections of the article from becoming completely intractable to those outside the field. Therefore, it is the view of this researcher that the use of complex language in this case primarily serves to invoke vital prerequisite knowledge, without which original research would be impossible. Due to an emphasis on technical process and a high level of education required to work in the field, a certain amount of complex language is necessary for authors to align themselves with, and contribute knowledge to, their academic and professional community (Giltrow et al., 2014).

Limitations of this research include only having a solo researcher when deriving the operational definition, as well as in the subsequent coding process. To avoid embedding subjectivity into the results, it would beneficial to have multiple researchers. Furthermore, given that the chosen journal was specific to public health surveillance, this prevents the generalization of results to the entire field of public health. Future research could be done with a similar research question applied to a corpus of articles from a different discipline, to see if the results had similarities. It would be especially interesting to conduct the research on a less interdisciplinary field than public health in order to see if this affected the frequency of type 2 complex language. This research could also serve as a starting point for more detailed research into whether specific groups of people are disproportionately affected by complex language. This could include students in the field, professionals in other fields, or specialists with a different native language.

APPENDIX 1

Corpus Articles

- Al Manir, M. S., Brenas, J. H., Baker, C. J., & Shaban-Nejad, A. (2018). A surveillance infrastructure for malaria analytics: Provisioning data access and preservation of interoperability. *JMIR Public Health and Surveillance*, 4(2), e10218. doi:10.2196/10218
- de Lusignan, S., Correa, A., Pebody, R., Yonova, I., Smith, G., Byford, R., . . . Jones, S. (2018). Incidence of lower respiratory tract infections and atopic conditions in boys and young male adults: Royal college of general practitioners research and surveillance centre annual report 2015-2016. *JMIR Public Health and Surveillance*, 4(2), e49. doi:10.2196/publichealth.9307
- Fu, C., Lopes, S., Mellor, S., Aryal, S., Sovannaroth, S., & Roca-Feltrer, A. (2017). Experiences from developing and upgrading a web-based surveillance system for malaria elimination in cambodia. *JMIR Public Health and Surveillance*, 3(2) doi:10.2196/publichealth.6942
- Hamp, A. D., Doshi, R. K., Lum, G. R., & Allston, A. (2018). Cross-jurisdictional data exchange impact on the estimation of the HIV population living in the district of columbia: Evaluation study. *JMIR Public Health and Surveillance*, 4(3), e62. doi:10.2196/publichealth.9800
- Hart, M., Stetten, N. E., Islam, S., & Pizarro, K. (2017a). Twitter and public health (part 1): How individual public health professionals use twitter for professional development. *JMIR Public Health and Surveillance*, 3(3), e60. doi:10.2196/publichealth.6795
- Hart, M., Stetten, N., Islam, S., & Pizarro, K. (2017b). Twitter and public health (part 2): Qualitative analysis of how individual health professionals outside organizations use microblogging to promote and disseminate health-related information. *JMIR Public Health and Surveillance*, 3(4), e54. doi:10.2196/publichealth.6796
- Ip, D. K., Lau, E. H., So, H. C., Xiao, J., Lam, C. K., Fang, V. J., . . . Cowling, B. J. (2017). A smart card-based electronic school absenteeism system for influenza-like illness surveillance in hong kong: Design, implementation, and feasibility assessment. *JMIR Public Health and Surveillance*, 3(4), e67. doi:10.2196/publichealth.6810
- Lindsay, A. C., Mesa, T., Greaney, M. L., Wallington, S. F., & Wright, J. A. (2017). Associations between maternal depressive symptoms and nonresponsive feeding styles and practices in mothers of young children: A systematic review. *JMIR Public Health and Surveillance*, 3(2), e29. doi:10.2196/publichealth.6492
- Mejova, Y., Weber, I., & Fernandez-Luque, L. (2018). Online health monitoring using facebook advertisement audience estimates in the united states: Evaluation study. *JMIR Public Health and Surveillance*, 4(1), e30. doi:10.2196/publichealth.7217
- Meyers, L., Ginocchio, C. C., Faucett, A. N., Nolte, F. S., Gesteland, P. H., Leber, A., . . . Poritz, M. A. (2018). Automated real-time collection of pathogen-specific diagnostic data:

Syndromic infectious disease epidemiology. *JMIR Public Health and Surveillance*, 4(3), e59. doi:10.2196/publichealth.9876

- Rosenthal, M., Anderson, K., Tengelsen, L., Carter, K., Hahn, C., & Ball, C. (2017). Evaluation of sampling recommendations from the influenza virologic surveillance right size roadmap for idaho. *JMIR Public Health and Surveillance*, 3(3), e57. doi:10.2196/publichealth.6648
- Victoria Wan, Lorraine McIntyre, Debra Kent, Dennis Leong, & Sarah B Henderson. (2018). Nearreal-time surveillance of illnesses related to shellfish consumption in british columbia: Analysis of poison center data. *JMIR Public Health and Surveillance*, 4(1), e17. doi:10.2196/publichealth.8944

APPENDIX 2

Additional Examples of Complex Language Usage

<u>Type 1:</u>

Populations in sub-Saharan African countries are most **susceptible**, with 80% of observed cases and recorded deaths worldwide. (Al Manir et al., 2018)

National influenza **virologic** surveillance is conducted by 60 National Respiratory and **Enteric** Virus Surveillance System laboratories. (Rosenthal et al., 2017)

<u>Type 2:</u>

A total of 15,236 tweets were collected and then analyzed using the **constant comparison method**. (Hart et al., 2017)

Some Facebook interests such as *plus-size clothing* show encouraging **levels of correlations** (r=.74) across the 50 US states. (Mejova et al., 2018)

<u> Type 3:</u>

This platform is used to **extract, clean**, and merge all necessary **raw** data tables into a single data file. (Wan et al., 2018)

Due to the range of study designs, assessment of **exposure**, and outcomes, conducting a metaanalysis of the data was not appropriate. (Lindsay et al., 2017)

REFERENCES

- Bauer, M.W., & Aarts, B. (2000). Corpus construction: A principle for qualitative data collection.
 In M.W. Bauer and G. Gaskell (Eds.), *Qualitative researching with text, image and sound:* A practical handbook (pp. 19-37). London: Sage.
- Fu, C., Lopes, S., Mellor, S., Aryal, S., Sovannaroth, S., Roca-Feltrer, A., (2017) Experiences From Developing and Upgrading a Web-Based Surveillance System for Malaria Elimination in Cambodia. *Journal of Medical Internet Research-Public Health* and Surveillance, 3(2), 1-12.
- Giltrow, J., Gooding, R., Burgoyne, D., Sawatsky, M., (2014). *Academic Writing An Introduction*. Peterborough, ON: Broadview Press.
- Hamp, A., Doshi, R., Lum, G., Allston, A., (2018). Cross-Jurisdictional Data Exchange Impact on the Estimation of HIV Population Living in the District of Columbia: Evaluation Study. *Journal of Medical Internet Research-Public Health and Surveillance, 4*(3), 1-9.
- Hart, M., Stetten, N., Islam, S., Pizarro, K., (2017). Twitter and Public Health (Part 1): How Individual Public Health Professionals Use Twitter for Professional Development. *Journal* of Medical Internet Research-Public Health and Surveillance, (3)3, 1-6.
- Hyland, K. (2011). Disciplines and discourses: Social interactions in the construction of knowledge. In D. Starke-Meyerring, A. Paré, N. Artemeva, M. Horne, and L. Yousoubova (Eds.), Writing in knowledge societies (pp. 193-214). West Lafayette, IN: The WAC Clearinghouse.
- Rosenthal, M., Anderson, K., Tengelsen, L., Carter, K., Hahn, C., & Ball, C. (2017). Evaluation of sampling recommendations from the influenza virologic surveillance right size roadmap for idaho. *JMIR Public Health and Surveillance*, 3(3), e57. doi:10.2196/publichealth.6648
- Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes*, 21(1), 1-17.
- World Health Organization (2018) *Public Health Surveillance*. Retrieved from http://www.who.int/topics/public_health_surveillance/en/