COVID-19 Is Connected with Lower Health Literacy in Rural Areas

Melody L. GREER[a,1], Steven SAMPLE[b], Hanna K. JENSEN[c], Sacha MCBAIN[c,d], Riley LIPSCHITZ[e], and Kevin W. SEXTON[c]

[a] *Department of Biomedical Informatics, University of Arkansas for Medical Sciences*
[b] *Pinnacle Data Strategies*
[c] *COM Surgery Trauma Surgery, University of Arkansas for Medical Sciences*
[d] *Psychiatric Research Institute, University of Arkansas for Medical Sciences*
[e] *COM Internal Medicine*

Abstract. The relationship between social determinants of health (SDoH) and health outcomes is established and extends to a higher risk of contracting COVID-19. Given the factors included in SDoH, such as education level, race, rurality, and socioeconomic status are interconnected, it is unclear how individual SDoH factors may uniquely impact risk. Lower socioeconomic status often occurs in concert with lower educational attainment, for example. Because literacy provides access to information needed to avoid infection and content can be made more accessible, it is essential to determine to what extent health literacy contributes to successful containment of a pandemic. By incorporating this information into clinical data, we have isolated literacy and geographic location as SDoH factors uniquely related to the risk of COVID-19 infection. For patients with comorbidities linked to higher illness severity, residents of rural areas associated with lower health literacy at the zip code level had a greater likelihood of positive COVID-19 results unrelated to their economic status.

Keywords. Data Integration, Data Quality, Social Determinants of Health, Health Literacy, Pandemic

1. Introduction

The relationship between social determinants of health and health is established. Many studies have demonstrated the role of SDoH, especially socioeconomic status, in driving health disparities in both acute and chronic illnesses. For example, SDoH has an impacted incidence rate and severity for influenza [1-4] and COVID-19.[5] Education level, race, rurality, and socioeconomic status have all been associated with COVID-19 infection likelihood.[6] Although many aspects of SDoH are intertwined with higher risk of COVID-19, it is unclear how each factor uniquely impacts risk.

Health literacy encompasses the abilities and skills required to gather, comprehend and apply information for health-related decisions and has been suggested as a critical factor with regard to health outcomes. [7; 8] The global COVID-19 pandemic has made it clear that health literacy is an essential driver of the public's adoption of public health

---

[1] Corresponding Author, Melody L. GREER, Department of Biomedical Informatics, University of Arkansas for Medical Sciences; E-mail: Mlgreer@uams.edu.

recommendations and therefore disease prevention. [9-10] Furthermore, as public health information becomes increasingly disseminated digitally, technical or digital literacy may also contribute to disease prevention. [11] The relationship between digital competency, literacy, and health will need to be better understood to address infection risk. Despite the known relationship between SDoH and health outcomes, the vast majority of SDoH data is still not part of patients' standard data collected during the care process. [12] For these reasons, we resolved to integrate averaged zip code geographical data, not present in the electronic health record (EHR), with patient clinical information. We have endeavored to examine this relationship using composite variables derived from self-reported surveys and purchasing patterns.

2. Methods

The integrated data set was generated by appending zip level data and Rural-Urban Commuting Area Codes (RUCA)[13] to approximately 55,000 University of Arkansas for Medical Sciences EHR records of COVID-19 positive and negative patients with comorbidities linked to high illness severity (e.g., diabetes, heart disease). We searched for publicly available data, but there were no excellent surrogates for health literacy. The patient zip code was matched to the commercially available aggregate zip code Health Interest Index and Technology Interest indicator. Although these data points are not a direct replacement for foundational health literacy (ability to comprehend versus health interest) we know that seeking health information is strongly related to literacy.[14] This relationship supports using these two index values together as a surrogate for health literacy, and we will define these terms as equivalent in the context of this work. This appended zip level data was selected from a national compiler of consumer data. The compiler's data updates monthly and comprises hundreds of different sources, including consumer surveys, public records, purchase transactions, real estate data, offline and online buying behavior, and warranty information.

The Health Interest Index indicates a level of interest in health (research, exercising, better dieting, preventive care, etc.). It is ranked from 0 (lowest level of interest) to 9 (highest level of interest). For each zip code, the index values were represented by a percentage of that population. For example, a zip code of 11111 might have 20% at the 1 level, 50% at the 5 level, and 30% at the 9 level. In reviewing the Health Interest Index data, the 0's comprised a large percentage of the population. The 0's included individuals who reflected very low interest plus individuals with not enough information to score accurately on the level of health interest. Because of the potentially poor quality of the 0 values, these were removed from the analysis. The Technology Interest indicator is a binary flag indicating technology adopters. The Health Literacy measure was constructed using an overlay of the two factors by zip code and patient then was binned into upper and lower interest groups. Both of these lifestyle factor data points were obtained from survey data and purchasing patterns using transactional and response information from self-reported sources.

3. Results

Comparing areas in Arkansas with a larger percentage of COVID-19 cases, we discovered a pronounced difference between metropolitan and rural areas concerning

interest in health and technology. These differences were most pronounced in rural areas. Patients residing in zip codes where there was less interest in technology and lower interest in health showed a larger percentage of positive COVID cases. The reverse is also true. Living in a neighborhood where typical residents had an interest in health along with technology had fewer positive test results.

The rural residents appear from the distributions in figure 1 appear to be less health literate in the COVID-19 positive group. To confirm the rural group's tendency to be COVID positive when less health literate we compared the rural and metropolitan populations. Because the data did not follow normal distribution, we used the Wilcoxon signed-rank test. We determined that the rural distributions were statistically different with a p-value = 0.0003924 when evaluating low health literacy rates and COVID-19 for rural residents. This difference was true regardless of the neighborhood income level. The surrogate health literacy variable (combined Health Interest Index and Technology Interest Index) was not correlated with income at the zip code level.
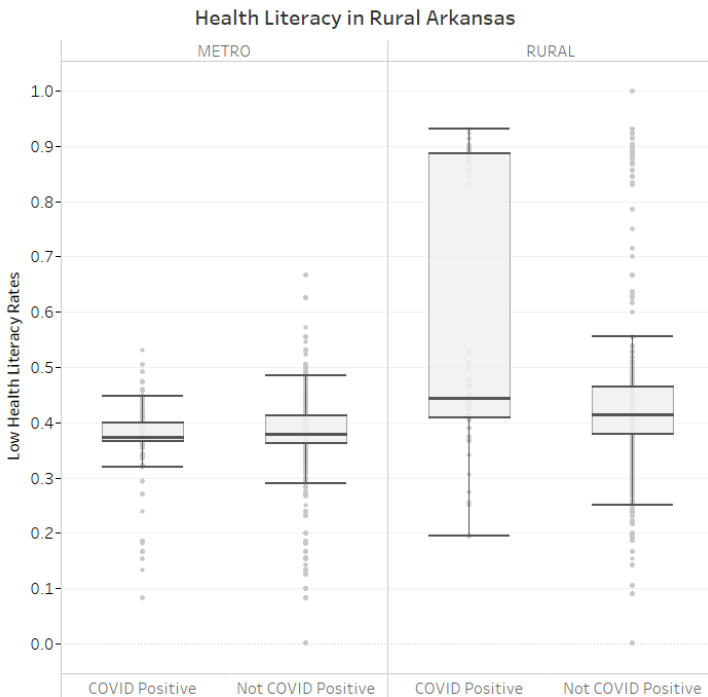

Figure 1. COVID-19 is more prevalant where health literacy is low in rural areas.

4. Discussion

The results above describe the relationship between zip code, attitudes towards health and technology, and rates of COVID-19 infection. The large number of residents with low health literacy within the COVID-19 positive rural group shown in figure 1 is striking. Metropolitan areas were found to have higher rates of health literacy (positive health attitudes and interest in technology) which were correlated with lower COVID-19 infection rates, even when controlling for other SDoH like socioeconomic status. The

most unexpected result was the lack of a relationship with income; the relationship between health literacy, interest in technology, and COVID-19 infection rates held at every income level for all zip codes. Further study is needed to determine what underlying factors may be the cause of these associations. Although we eliminated income, older age and less education which are associated with rurality in the United States these may be confounders and warrant further investigation. Importantly, the relationship between health literacy was reflected by the EHR data only after the addition of zip level information. This suggests that to truly understand our patients' health information needs, we will need to improve efforts at data collection, integration, and management. Analysis of this data prior to the addition of surrogate health literacy values revealed the disparity between metropolitan and rural areas for positive COVID-19 results, but with no other associations. Notably, other associations beyond health literacy (e.g. distance to pharmacy or grocery, neighborhood violence, or number of generations in the home) may remain obscured by a lack of information.

5. Conclusion

To understand the relationship between COVID-19 infection rates, health literacy, interest in technology, and geographic information, we created a more comprehensive representation of our patients. We examined in detail the relationship between the health literacy of patients with COVID-19 compared to those with risk factors, but no indication of infection. The integrated data returned some unexpected results, suggesting income level alone is not necessarily associated with low health literacy and COVID-19 infection. A limitation of this study is that the data is sourced only from one hospital in Arkansas. To confirm these results we would need an independent study in another region. This work has provided a better understanding of factors that affect rates of COVID-19 infection and emphasizes the importance of science communication and public health messages to improve the public's interest in health, especially those at greatest risk for infection such as those in rural areas. From a data standpoint, the results revealed the value of appended data in planning for future infectious disease, making healthcare processes more proactive since understanding the populations we serve gives us a better chance at delivering what they need. We believe that other relationships likely exist but are currently going undetected and that data integration could provide a means to identify other factors related to COVID-19, as well as other conditions. Although this information is not available with EHR data alone, the findings can be used to place particular emphasis on addressing health attitudes in communications with rural residents.

References

[1]   Kivimäki M, Batty GD, Pentti J, Shipley MJ, Sipilä PN, Nyberg ST, Suominen SB, Oksanen T, Stenholm S, Virtanen M, Marmot MG, Singh-Manoux A, Brunner EJ, Lindbohm JV, Ferrie JE, Vahtera J. Association between socioeconomic status and the development of mental and physical health conditions in adulthood: a multi-cohort study. Lancet Public Health. 2020 Mar;5(3):e140-e149.

[2]   Hadler JL, Yousey-Hindes K, Pérez A, Anderson EJ, Bargsten M, Bohm SR, Hill M, Hogan B, Laidler M, Lindegren ML, Lung KL, Mermel E, Miller L, Morin C, Parker E, Zansky SM, Chaves SS. Influenza-Related Hospitalizations and Poverty Levels - United States, 2010-2012. MMWR Morb Mortal Wkly Rep. 2016 Feb 12;65(5):101-5.

[3]   Lowcock EC, Rosella LC, Foisy J, McGeer A, Crowcroft N. The social determinants of health and pandemic H1N1 2009 influenza severity. Am J Public Health. 2012 Aug;102(8):e51-8.

[4]   Quinn SC, Kumar S. Health inequalities and infectious disease epidemics: a challenge for global health security. Biosecur Bioterror. 2014 Sep-Oct;12(5):263-73.

[5]   Anyane-Yeboa A, Sato T, Sakuraba A. Racial disparities in COVID-19 deaths reveal harsh truths about structural inequality in America. J Intern Med. 2020 Oct;288(4):479-480.

[6]   Berkowitz SA, Cené CW, Chatterjee A. Covid-19 and Health Equity - Time to Think Big. N Engl J Med. 2020 Sep 17;383(12):e76.

[7]   Dugravot A, Fayosse A, Dumurgier J, Bouillon K, Rayana TB, Schnitzler A, Kivimaki M, Sabia S, Singh-Manoux A. Social inequalities in multimorbidity, frailty, disability, and transitions to mortality: a 24-year follow-up of the Whitehall II cohort study. Lancet Public Health. 2020 Jan;5(1):e42-e50.

[8]   Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, Brand H; (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health. 2012 Jan 25;12:80

[9]   Paakkari L, Okan O. COVID-19: health literacy is an underestimated problem. Lancet Public Health. 2020 May;5(5):e249-e250.

[10] Benis A, Tamburis O, Chronaki C, Moen A. One Digital Health: A Unified Framework for Future Health Ecosystems. J Med Internet Res. 2021 Feb 5;23(2):e22189

[11] Levin-Zamir D, Bertschi I. Media Health Literacy, eHealth Literacy, and the Role of the Social Environment in Context. Int J Environ Res Public Health. 2018 Aug 3;15(8):1643

[12] Hatef E, Rouhizadeh M, Tia I, Lasser E, Hill-Briggs F, Marsteller J, Kharrazi H. Assessing the Availability of Data on Social and Behavioral Determinants in Structured and Unstructured Electronic Health Records: A Retrospective Analysis of a Multilevel Health Care System. JMIR Med Inform. 2019 Aug 2;7(3):e13802.

[13] Economic Research Service, Rural-Urban Commuting Area Codes, in: Economic Research Service, U.S. Department of Agriculture, 2020.

[14] Feinberg I, Frijters J, Johnson-Lawrence V, Greenberg D, Nightingale E, Moodie C. Examining Associations between Health Information Seeking Behavior and Adult Education Status in the U.S.: An Analysis of the 2012 PIAAC Data. PLoS One. 2016 Feb 16;11(2):e0148751.