ANALYSIS OF TRENDS IN COVID-19 MORTALITY AND CURE BASED ON STATISTICAL DATA BETWEEN PROVINCES IN INDONESIA (2020-2022)

Fachril Dzulfian[1*], Lukman Nur Efendi[1], Ikeu Yulianti[1]
[1]Universitas Teknologi Bandung, Jl. Soekarno-Hatta No.378, Bandung, Indonesia
*Corresponding Author: fachrildz@gmail.com

ABSTRACTS

Employee attrition is a significant strategic concern for organizations as it directly impacts overall performance, productivity, and long-term sustainability. High attrition rates can lead to increased costs in recruitment and training, a loss of skilled and experienced employees, decreased morale among remaining staff, and disruptions to critical business operations. In response to these challenges, many organizations are turning to predictive analytics to anticipate employee turnover and implement effective retention strategies. This study proposes a machine learning-based approach to predict employee attrition using the Logistic Regression algorithm. Logistic Regression is chosen due to its effectiveness in binary classification tasks and its interpretability, which is essential for human resource (HR) professionals when making data-driven decisions. To enhance the model's performance, the SelectKBest feature selection technique is applied in conjunction with the ANOVA F-test. This method allows the identification of the most influential features contributing to attrition, helping reduce noise and computational complexity while improving model accuracy. The IBM HR Analytics Employee Attrition & Performance Dataset is used in this study. The dataset contains a variety of demographic and organizational attributes such as age, monthly income, job role, tenure, and job satisfaction. The data undergoes a comprehensive preprocessing phase that includes numerical transformation, encoding of categorical variables, normalization, and the implementation of feature selection. By combining Logistic Regression with effective feature selection, this research aims to deliver an accurate and interpretable predictive model. The results are expected to help HR departments proactively identify high-risk employees and take strategic actions to reduce attrition, ultimately supporting better workforce planning and organizational stability.

Keywords: Machine learning; Logistic Regression, SelectKBest, ANOVA F-test.

INTRODUCTION

The Coronavirus Disease 2019 (COVID-19) pandemic has become the largest global health crisis in modern history, causing significant impacts on social, economic and healthcare system aspects (Susilo et al. 2020). COVID-19 is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) which was first identified in Wuhan, China, in December 2019 (Yuliana 2020). The disease spread rapidly around the world and was officially declared a global pandemic by the World Health Organization (WHO) on March 11, 2020 (Sohrabi et al. 2020). In Indonesia, the first case was announced on March 2, 2020, and soon all provinces reported cases of infection. The government responded with various policies such as Large-Scale Social Restrictions (PSBB), Enforcement of Restrictions on Community Activities (PPKM), and a national vaccination program (Pramana et al. 2021). However, the effectiveness of these policies varied between regions, influenced by demographic, geographic, and health system capacity differences (Susilo et al. 2022).

In designing a research methodology, the choice of approach greatly influences the direction and output of the resulting analysis. Qualitative analysis is generally used to understand social phenomena in depth through interviews, observations, and document studies, with a focus on meaning, context, and subjective experience. This approach is narrative and does not prioritize numbers (Lewis 2015). Meanwhile, quantitative analysis focuses more on processing numerical data, statistical measurements, and describing trends or patterns based on structured aggregate data (John W 2017). In the context of COVID-19 studies, qualitative analysis is often used to explore public perceptions of health policies or the social impact of the pandemic. In contrast, quantitative analysis is used to observe case trends, calculate mortality and recovery rates, and compare statistics between regions.

Several studies have shown that geographic characteristics and spatial differences contribute to variations in mortality and cure rates, especially in countries with unequal distribution of health resources (Julija Simpson,Viviana Albani, Bambra, and Brown 2021). Indonesia as an archipelago with varying levels of population density and mobility is an important object to study in detail to reveal COVID-19 trends based on administrative areas such as provinces (Putri 2020).

Departing from this background, this study aims to evaluate the dynamics of the death rate (Case Fatality Rate/CFR) and recovery rate (CRR) due to COVID-19 nationally and between provinces in Indonesia during the period 2020 to 2022. The approach used is quantitative descriptive analysis based on secondary data from open statistical sources. By comparing CFR and CRR trends between provinces and analyzing the causes of disparities, this study is expected to contribute to the formulation of health policies that are data-based and in accordance with the needs of each region. This analysis can also serve as a basis for evaluating the sustainability of Indonesia's health system in facing similar crises in the future.

METHODS

This study used a quantitative descriptive approach to evaluate the case fatality rate (CFR) and recovery rate (CRR) of COVID-19 among provinces in Indonesia during the period 2020-2022. Data was obtained from open sources on the Kaggle platform, including total cases, recovered cases, and deaths per year. Quantitative

descriptive research is a type of research that aims to describe in detail, factually, and precisely about certain facts and characteristics, as well as trying to reveal phenomena clearly (Helwig, Hong, and Hsiao-wecksler 2021).

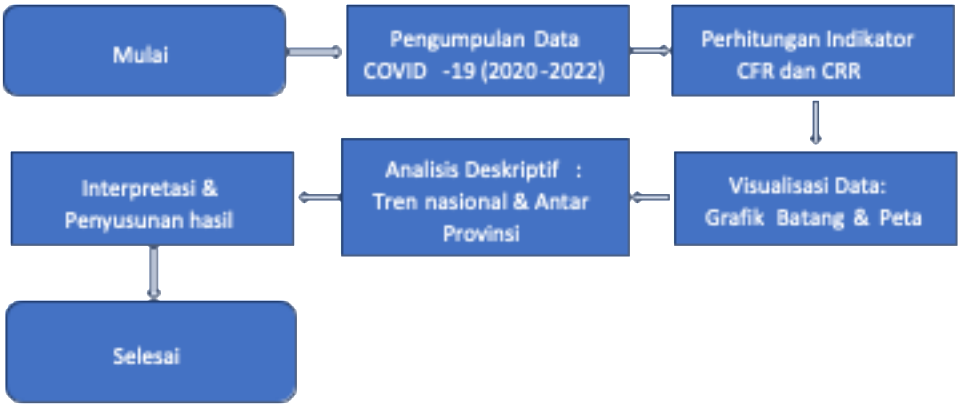

Figure 1. Flowchart

This study begins with the collection of COVID-19 data in Indonesia during the period 2020 to 2022. The data used includes the number of confirmed cases, recovered cases, and deaths per province. All data was taken from the Kaggle platform, which provides official and open datasets, thus supporting the principles of transparency and accountability in scientific research (covid19-indonesia @ www.kaggle.com n.d.).

After data collection, the next step is the calculation of two key epidemiological indicators, namely Case Fatality Rate (CFR) and Case Recovery Rate (CRR). CFR is calculated by dividing the number of deaths due to COVID-19 by the total number of confirmed cases, while CRR is calculated from the number of recovered cases to total cases. This indicator is widely used in epidemiological studies as a basis for evaluating the risk of death and the success of patient handling (Zhao, Tao, and Zhang 2021).

Furthermore, descriptive analysis of CFR and CRR data was carried out, both nationally and between provinces. This approach aims to identify patterns of distribution and spatial variation of mortality and cure rates. Descriptive analysis is considered appropriate for describing phenomena without conducting inferential testing, and is widely used in studies of the inequality of pandemic impacts between regions (Tu et al. 2020).

The results of the analysis were then expressed through data visualization in the form of bar charts and distribution maps developed using Tableau software. This visualization facilitates the interpretation of comparisons between regions and shows trends of change from year to year. Previous studies have shown that visual representations are very effective in conveying complex information about public health to policymakers (Carter, A., Nguyen, C. N., & Huynh 2020).

The final step in this method is the interpretation of results and the preparation of conclusions. The visualized data are thoroughly analyzed to formulate findings and recommendations. This interpretation is important as a basis for developing health policies that are evidence-based and tailored to the conditions of each region, as emphasized in the data-driven public health approach (Mahendradhata, Y. 2021).

RESULTS AND DISCUSSION

The analysis focused on two key indicators, the Case Fatality Rate (CFR) and Case Recovery Rate (CRR), as measures of the severity and success of COVID-19 response in different provinces in Indonesia over the period 2020 to 2022. These two indicators are important to see how the pandemic impacts differently across regions, as well as to assess the capacity of local health systems to respond to the crisis. To strengthen understanding, the analysis is accompanied by Tableau-based graphical visualizations and thematic maps, which facilitate the identification of spatial and temporal patterns. Through this approach, the discussion is directed towards geographical disparities in caseloads, recoveries and deaths, and how these can inform the development of area-based health strategies (Khazanchi, R., Evans, C. T., & Marcelin 2020).

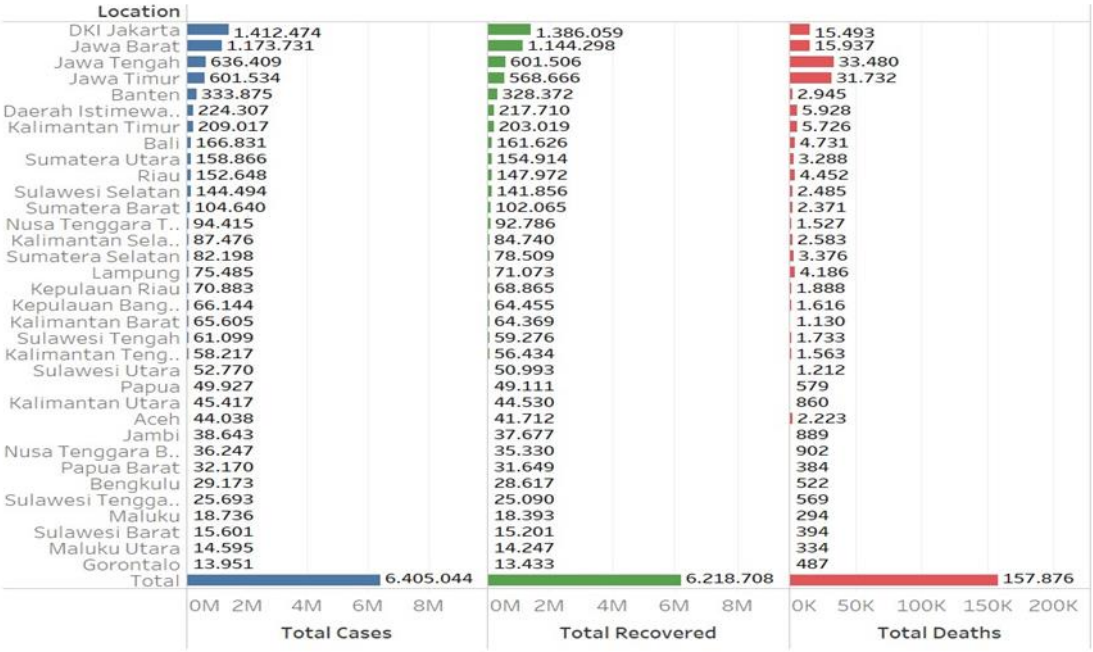

Figure 2. Statistics of COVID-19 Cases in Indonesia by Province

Figure 2 shows the distribution of total cases, recoveries and deaths from COVID-19 across all provinces in Indonesia during 2020-2022. Nationally, there were 6.4 million cases with more than 6.2 million recoveries and 157,000 deaths. Provinces in Java - such as DKI Jakarta, West Java, Central Java, and East Java - recorded the highest number of cases and deaths, while regions outside Java such as Gorontalo, West Papua, and North Maluku showed much lower caseloads. This disparity shows that population density, high mobility and pressure on the healthcare system greatly affect the severity of the pandemic in a region. In addition, variables such as hospital capacity and speed of data reporting also influence the results of each province. These findings are in line with the literature showing that pandemics exacerbate geographical disparities in health systems, particularly in developing countries (Rijanta.R 2020).

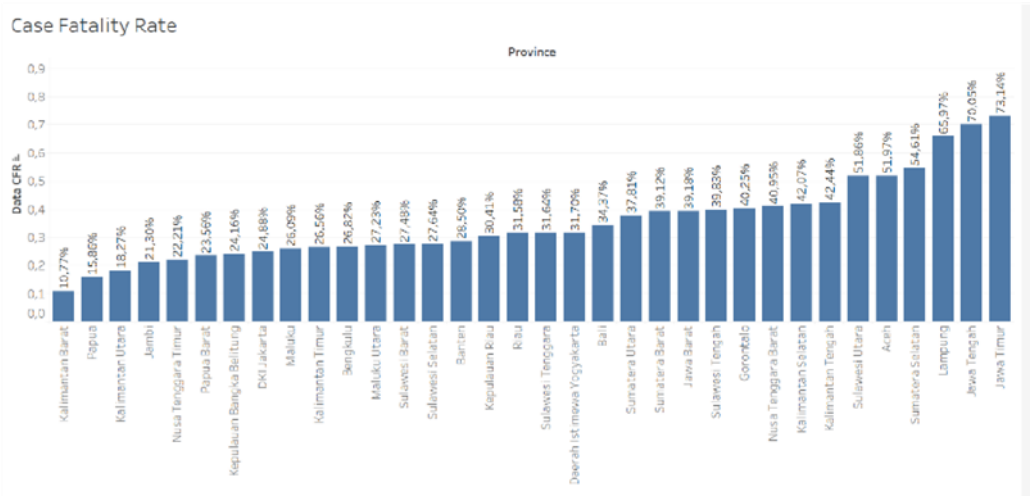

Figure 3. Data on Disparities in Case Fatality Rate (CFR) between Provinces

Figure 3 illustrates the comparison of Case Fatality Rate (CFR) between provinces in Indonesia during the period 2020-2022. There is a striking disparity in the distribution of fatality rates. The province with the highest CFR is East Java (73.4%), followed by Central Java (70.05%) and Lampung (65.97%). In contrast, West Kalimantan (10.77%), Papua (15.86%) and North Kalimantan (18.27%) recorded the lowest CFRs. This pattern shows that provinces in Java, which also recorded the highest number of cases, face a much greater mortality burden than other regions. This is most likely influenced by the high population density, limited capacity of health facilities, and potential delays in case management. Previous studies have shown that a surge in caseloads in an already overcrowded system can worsen quality of care and increase mortality, especially if not accompanied by increased medical capacity (Seet et al. 2021). These findings reinforce the importance of area-based strategies in pandemic response, especially in resource allocation and strengthening referral systems in provinces with high mortality risk.

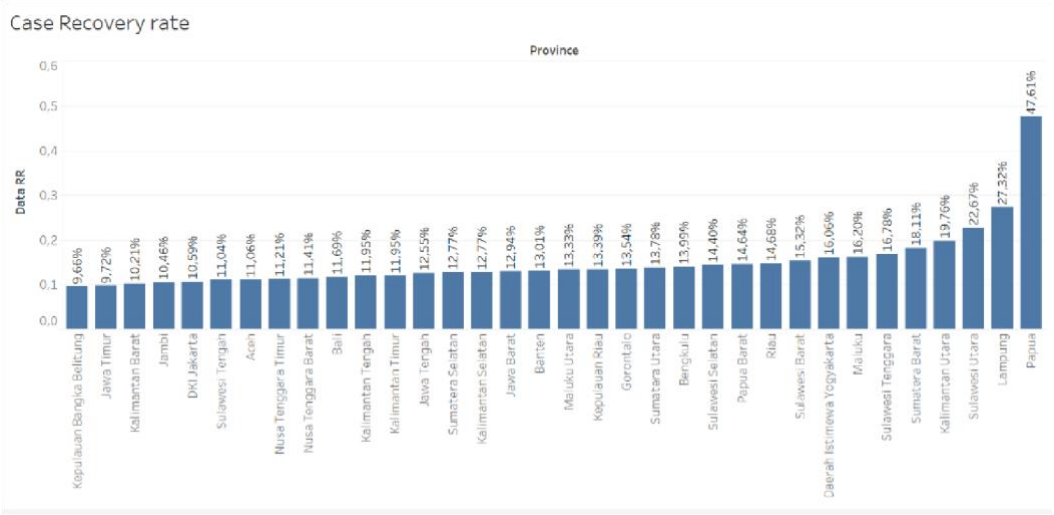

Figure 4. Case Cure Rate (CRR) Data

Figure 4 shows the case recovery rate (CRR) of COVID-19 by province in Indonesia. This data reveals significant variation between regions, with the lowest CRR recorded in Bangka Belitung Islands (9.66%), East Java (9.72%), and Central Java (10.21%). In contrast, Papua (47.61%), North Sulawesi (27.29%), and Lampung (22.67%) were the highest in terms of recovery rates. This pattern shows an inverse relationship between CRR and Case Fatality

Rate (CFR) in some provinces, especially in Java, which recorded high CFR but very low CRR. This phenomenon may indicate severe pressure on the health system, delays in medical treatment, and constraints in reporting and recording the recovery status of patients. A study by (Abdullatif Khafaie and Rahim 2021) states that calculation methods and reporting time greatly affect the accuracy of CFR and CRR estimates, especially in developing countries with suboptimal health reporting systems. Therefore, improving health data management systems and post-infection patient monitoring is essential, especially in areas with high caseloads.

Figure 5. Geographical condition of case distribution

Figure 5 presents a map of the geographical distribution of COVID-19 cases in Indonesia based on color intensity, where dark red areas indicate the highest concentration of cases. It is clear that Java Island, especially the provinces of West Java, DKI Jakarta, Central Java, and East Java, is the center of the pandemic burden, marked by the darkest color. Meanwhile, regions outside Java such as Kalimantan, Sulawesi, Maluku and Papua show lighter colors, reflecting a relatively lower number of cases. This spatial distribution corroborates previous findings of geographical disparities in the impact of the pandemic in Indonesia. Factors such as high population density, rapid urbanization, and intense intercity mobility are believed to be the main drivers of making the Java region the epicenter of the virus spread. For example, research by (Widiawaty et al. 2022) shows that COVID-19 cases are statistically closely related to population density, urbanization, and commuting workers at the provincial level in Indonesia. Therefore, this kind of spatial mapping is very important, not only as a means of visualization, but also as a basis for area-based policy making that is adaptive to local risks.

While this article provides a comprehensive overview of COVID-19 mortality and recovery trends in Indonesia, there are some limitations that should be noted in the interpretation of the results. The approach used is descriptive quantitative without inferential tests, limiting the ability to draw causal relationships between the variables analyzed. The data used is annual aggregate data, which is unable to represent daily or weekly dynamics that better reflect fluctuations in case distribution in detail. In addition, the quality of the data is highly dependent on reporting from each region, which is not necessarily consistent and accurate, which can affect the validity of the analysis results. The main indicators used, Case Fatality Rate (CFR) and Case Recovery Rate (CRR), also do not consider other important variables such as patient age, comorbidities, hospital capacity, and speed of medical treatment, which also affect mortality and recovery rates. These limitations mean that the results of

the analysis do not fully capture the complexity of factors that influence disparities between regions. Therefore, the results of this study should be viewed as an initial basis for further in-depth and comprehensive studies with more diverse approaches and more detailed data.

CONCLUSION

This study illustrates that the COVID-19 pandemic impacted unevenly across provinces in Indonesia during 2020 to 2022. Provinces in Java experienced much higher caseloads and deaths than other regions, while some regions outside Java showed better recovery rates. These differences reflect inequalities in the health system, population density, and regional preparedness for the pandemic. Therefore, future efforts need to consider the specific conditions and needs of each region. Future research is recommended to use more detailed data and consider other factors such as population age, health facilities, and socioeconomic conditions to better understand the causes of these differences.

REFERENCES

Abdullatif Khafaie, Morteza, and Fakher Rahim. 2021. "Estimating Case Fatality and Case Recovery Rates of COVID-19: Is This the Right Thing to Do?" *Central Asian Journal of Global Health* 10(1). doi:10.5195/cajgh.2021.489.

Carter, A., Nguyen, C. N., & Huynh, T. 2020. "Spatio-Temporal Visualization of the COVID-19 Pandemic in Canada." *International Journal of Health Geographics* 19(41).

"Covid19-Indonesia @ Www.Kaggle.Com." https://www.kaggle.com/datasets/hendratno/covid19-indonesia.

Helwig, Nathaniel E, Sungjin Hong, and Elizabeth T Hsiao-wecksler. 2021. *Metode Penelitian Kualitatif Dan Metode Penelitian Kuantitatif*.

John W, Creswell. 2017. "John W, Creswell, Research Design: Qualitative, Quantitative, and Mixed Methods Approaches." *Journal of Social and Administrative Sciences* 4(June): 3–5.

Julija Simpson,Viviana Albani, Zoe Bell, Clare Bambra, and Heather Brown. 2021. "Effects of Social Security Policy Reforms on Mental Health and Inequalities: A Systematic Review of Observational Studies in High-Income Countries." *Social Science & Medicine* 272.

Khazanchi, R., Evans, C. T., & Marcelin, J. R. 2020. "Racism, Not Race, Drives Inequity Across the COVID-19 Continuum." *JAMA* 324(3)(237–238).

Lewis, Sarah. 2015. "Qualitative Inquiry and Research Design: Choosing Among Five Approaches." *Health Promotion Practice* 16(4): 473–75. doi:10.1177/1524839915580941.

Mahendradhata, Y., et al. 2021. "The COVID-19 Response in Indonesia: Lessons Learned and Challenges for the Future." *The Lancet Global Health* 9.

Pramana, Setia, Yuniarti Yuniarti, Dede Yoga Paramartha, and Satria Bagus Panuntun. 2021. "Mobility Pattern Changes in Indonesia in Response to COVID-19." *Economics and Finance in Indonesia* 67(1): 75. doi:10.47291/efi.v67i1.924.

Putri, Ririn Noviyanti. 2020. "Indonesia Dalam Menghadapi Pandemi Covid-19." *Jurnal Ilmiah Universitas Batanghari Jambi* 20(2): 705. doi:10.33087/jiubj.v20i2.1010.

Rijanta.R, Baiquni.M. 2020. "16 | Rembug Pageblug." *Rembug Pageblug Dampak, Respons Dan Konsekuensi*

Pandemi Covid-19 Dalam Dinamika Wila.

Seet, Raymond Chee Seong, Amy May Lin Quek, Delicia Shu Qin Ooi, Sharmila Sengupta, Satish Ramapatna Lakshminarasappa, Chieh Yang Koo, Jimmy Bok Yan So, et al. 2021. "Positive Impact of Oral Hydroxychloroquine and Povidone-Iodine Throat Spray for COVID-19 Prophylaxis: An Open-Label Randomized Trial." *International Journal of Infectious Diseases* 106: 314–22. doi:10.1016/j.ijid.2021.04.035.

Sohrabi, Catrin, Zaid Alsafi, Niamh O'Neill, Mehdi Khan, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, and Riaz Agha. 2020. "World Health Organization Declares Global Emergency: A Review of the 2019 Novel Coronavirus (COVID-19)." *International Journal of Surgery* 76(February): 71–76. doi:10.1016/j.ijsu.2020.02.034.

Susilo, Adityo, Chyntia Olivia Maurine Jasirwan, Syahidatul Wafa, Suzy Maria, Wulyo Rajabto, Akhmadu Muradi, Ihza Fachriza, Myranda Zahrah Putri, and Stacy Gabriella. 2022. "Mutasi Dan Varian Coronavirus Disease 2019 (COVID-19): Tinjauan Literatur Terkini." *Jurnal Penyakit Dalam Indonesia* 9(1): 59. doi:10.7454/jpdi.v9i1.648.

Susilo, Adityo, Cleopas Martin Rumende, Ceva Wicaksono Pitoyo, Widayat Djoko Santoso, Mira Yulianti, Herikurniawan Herikurniawan, Robert Sinto, et al. 2020. "Coronavirus Disease 2019: Tinjauan Literatur Terkini." *Jurnal Penyakit Dalam Indonesia* 7(1): 45. doi:10.7454/jpdi.v7i1.415.

Tu, Huilan, Sheng Tu, Shiqi Gao, Anwen Shao, and Jifang Sheng. 2020. "Current Epidemiological and Clinical Features of COVID-19; a Global Perspective from China." *Journal of Infection* 81(1): 1–9. doi:10.1016/j.jinf.2020.04.011.

Widiawaty, Millary Agung, Kuok Choy Lam, Moh Dede, and Nur Hakimah Asnawi. 2022. "Spatial Differentiation and Determinants of COVID-19 in Indonesia." *BMC Public Health* 22(1): 1–16. doi:10.1186/s12889-022-13316-4.

Yuliana. 2020. "WELLNESS AND HEALTHY MAGAZINE Literatur." *Parque de los afectos. Jóvenes que cuentan* 2(February): 124–37.

Zhao, Rui, Yusha Tao, and Lei Zhang. 2021. "Can Social Media-Based Self-Testing Be a New Paradigm for Human Immunodeficiency Virus Testing in China?" *Clinical Infectious Diseases* 73(7): E2258–60. doi:10.1093/cid/ciaa817.