# DISCUSSION PAPER SERIES

DP19120

## DIGITAL ACCESS AND INFECTIOUS DISEASE SPREAD

Johannes Kunz, Carol Propper and Trong-Anh Trinh

## PUBLIC ECONOMICS

CEPR

ISSN 0265-8003

# DIGITAL ACCESS AND INFECTIOUS DISEASE SPREAD

*Johannes Kunz, Carol Propper and Trong-Anh Trinh*

Discussion Paper DP19120
Published 31 May 2024
Submitted 28 May 2024

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

# DIGITAL ACCESS AND INFECTIOUS DISEASE SPREAD


## Abstract

Digital access may bring important health gains, particularly where physical infrastructure is limited. We examine the impact of internet access in Indonesia on health outcomes using the COVID-19 pandemic as a health shock. We utilize sub-national data on mobile broadband, COVID-19 spread, and an instrumental variable approach using lightning strikes as an exogenous shock to connectivity. Access to 3G internet significantly reduced the transmission of COVID-19. Areas with internet access had approximately 45\% fewer cases. Regions with higher literacy and capacity for telework benefited significantly more. These findings offer novel insights into how digital infrastructure affects public health outcomes.



JEL Classification: I12, I15, I31, O18, L96, H41

Keywords:



Johannes Kunz - johannes.kunz@monash.edu
*Monash University*

Carol Propper - c.propper@imperial.ac.uk
*Imperial College London, Monash University and CEPR*

Trong-Anh Trinh - trong-anh.trinh@monash.edu
*Monash University*

# Digital access and infectious disease spread*

Johannes S. Kunz
Monash University

Carol Propper
Imperial College London
& Monash University

Trong-Anh Trinh
Monash University

May 28, 2024

**Abstract**

Digital access may bring important health gains, particularly where physical infrastructure is limited. We examine the impact of internet access in Indonesia on health outcomes using the COVID-19 pandemic as a health shock. We utilize sub-national data on mobile broadband, COVID-19 spread, and an instrumental variable approach using lightning strikes as an exogenous shock to connectivity. Access to 3G internet significantly reduced the transmission of COVID-19. Areas with internet access had approximately 45% fewer cases. Regions with higher literacy and capacity for telework benefited significantly more. These findings offer novel insights into how digital infrastructure affects public health outcomes.

***Keywords***: Health emergencies; Internet access; Information; COVID-19 Spread; Indonesia
***JEL classification***: I12, I15, I31, O18, L96, H41

* *Addresses for correspondence*: Kunz: Centre for Health Economics, Monash University, 900 Dandenong Road, 3145 Caulfield East, Vic Australia, T: +61 3 9905 0752, johannes.kunz@monash.edu. Propper: Imperial College London, Department of Economics and Public Policy, South Kensington Campus, United Kingdom, SW1A 2AZ, c.propper@imperial.ac.uk. Trinh: Centre for Health Economics, Monash University, 900 Dandenong Road, 3145 Caulfield East, Vic Australia, T: +61 3 9905 0752, trong-anh.trinh@monash.edu.
*Acknowledgments*: We thank Arief Ramayandi and Daniel C. Suryadarma for their outstanding support throughout this project. We thank Kevin Staub, Giovanni van Empel, Paul Raschky, Joachim Winter, and seminar participants at SoDa Lab-Monash, CHE-Monash, and ADB Conference 2023 (Manila) for their useful comments.
Kunz and Trinh gratefully acknowledge financial support from the ADB: S185729 - TA-6989 REG: Asian Development Outlook Update 2023 Theme Chapter: Asia's Preparedness on Health Emergencies (55278-003). Neither author has any conflict of interest to declare.

# 1 Introduction

Digital access to healthcare services has been argued to have the potential to bring large productivity gains in healthcare (e.g., Miller and Tucker, 2011; Lee et al., 2013; Agha, 2014; McCullough et al., 2016). Some of the largest gains may come in lower-income country settings where physical infrastructure provision is relatively low, and there are long-standing issues and disparities in access to services and health outcomes between areas and communities (Barber et al., 2007; Van de Poel et al., 2007). But whether digital access can be a substitute for physical resources is as yet unresolved (Kahn et al., 2010; Chib et al., 2015; Kickbusch et al., 2021). This paper seeks to address one part of this important issue: whether internet access can improve health outcomes. We exploit the biggest health shock in recent years to explore the effect of digital access on health outcomes in a developing country setting.

Our setting is the COVID pandemic and its spread in Indonesia. Our question is whether access to the internet had an impact on the number of COVID cases.[1] This setting has several advantages. First, the COVID pandemic was a worldwide shock for which there was little or no preparedness. Thus, infrastructure investment was not endogenous due to the pandemic. Second, Indonesia is the 6$^{\text{th}}$ most populous country in the world, and both mobile internet access and COVID cases displayed considerable variation across geographical space (c.f. Appendix Figure A.1). Third, the contagious nature of COVID and the global nature of the pandemic meant that COVID cases were well recorded (by international non-profit agencies and the statutory authorities in Indonesia) and timely. This contrasts with other health outcomes, which may be poorly recorded, only recorded in surveys, or take time to improve after medical intervention. Fourth, the rate of COVID infections in Indonesia was high, so data can be analyzed at a relatively small spatial level without encountering measurement issues, which is rare for developing countries and other diseases.

This study provides empirical evidence of the role of internet access during the COVID-19 pandemic by exploiting sub-national data on mobile broadband and COVID-19 spread in Indonesia. At the beginning of 2023, Indonesia was an epicenter of the Southeast Asian pandemic, with a very high number of confirmed cases relative to its peers in the region. Depicted in Figure 1, the association between COVID cases and internet access is clearly visible across countries (left panel) and within Indonesia across regencies (of which there are roughly 500, right panel).[2] The left panel shows that Indonesia stands out among Asian countries with a high number of cases but below-average mobile internet access. In the right panel, regencies are grouped by 20-percentile bins to display the average number of cases within each, revealing a strong and significant negative association across the distribution of internet coverage.[3]

Our empirical strategy builds on the regional variation in 3G mobile broadband access across regencies in Indonesia in December 2019 (just before the outbreak), coupled with an instrumental variable [IV] approach that uses the incidence of lightning strikes in the same month as an exogenous shock to the existing connection network (building on Manacorda and Tesei, 2020; Guriev et al., 2021; Do et al., 2023). We interpret this as an exogenous shock to the accessibility of information in the critical early stage of the pandemic. Additionally, our cross-sectional IV implementation adjusts for a host of potential confounders by including an extensive set of local information often absent in developing country contexts.

We find that access to 3G internet played a vital role in reducing the transmission rate of COVID-19 in Indonesia. Our findings imply that the number of COVID cases could have been reduced by approximately 45% if area connectivity were enhanced by one standard deviation (30% coverage). This

[1] In developing countries with often underdeveloped healthcare systems, internet access is likely crucial in reducing the spread of COVID-19. Yet, the World Bank reports that only 19.1% of the population in low-income countries has access to internet, compared to 87.7% in high-income countries (Kelly and Rossotto, 2011).

[2] The regional government in Indonesia is organized into provinces, with regencies (kabupaten) and cities (kota) nested within provinces at the same level. Our analyses are at this level. We also account for the provincial level as it is the primary level for health policy-making.

[3] Alternative representations (Cattaneo et al., 2024) show similar conclusions, which are available from authors.

reduction is remarkable, particularly when compared to the effectiveness of other non-pharmaceutical interventions (NPIs) in reducing COVID-19 transmission, such as social distancing and wearing masks (Courtemanche et al., 2020; Leech et al., 2022).

We also assess the evolution of the long-run effect of not having access to the internet early on in the pandemic and find that the difference in growth rates diverges early on and persists, meaning that the areas with access grew apart over time. We do not have direct data on mechanisms, but we examine the heterogeneous effects of education level, industry structure, rurality, and economic development. We find that the internet's benefits in mitigating the spread of COVID-19 are less pronounced in settings with a large number of uneducated individuals and where agricultural activities predominate. While graduate education seems to matter less than basic literacy, regions with a higher proportion of individuals capable of teleworking exhibit a more significant reduction in COVID-19 transmission. This underscores the role of information in facilitating social distancing practices among those who can work remotely. Interestingly, these differences are distinct from those of urbanity or socio-economic status of the area. Our findings suggest that 3G internet access during the pandemic likely facilitated critical awareness and adoption of protective measures, emphasizing the importance of enhanced digital connectivity and literacy as key components of public health strategies, furthermore, those with greater scope to adjust appear to benefit more.

Our study builds upon the extensive literature that examines the crucial role of information dissemination in enhancing health awareness and facilitating positive health outcomes in low- and middle-income countries. The transformative potential of accessible, timely, and relevant information has been highlighted, with evidence suggesting its ability to empower communities and individuals to make informed decisions regarding their health and well-being (Jalan and Somanathan, 2008; Aker and Mbiti, 2010; Dupas, 2011). Several factors have been identified as contributing to an effective response to pandemics. Providing information on preventive measures has proven effective in overcoming prevention barriers among impoverished households in Africa, particularly in the fight against malaria (Cohen and Dupas, 2010). In contrast, efforts to empower women and girls or enhance educational outcomes do not necessarily result in improved preventive behaviors (Behrman, 2015; Duflo et al., 2015). Within the COVID-19 context, Chang et al. (2022) and Bargain and Aminjonov (2020) showed that countries with higher levels of political trust and social cohesion were more successful in implementing effective pandemic response strategies. Mendolia et al. (2021) found a significant correlation between government dissemination of pandemic information and reduced human mobility. These findings underscore the importance of effective communication and transparency in preventing and mitigating the spread of infectious diseases such as COVID-19. Despite the underlying mechanism of these measures, which is often internet access, we are (to our knowledge) the first to assess the direct impact of internet access on the spread of COVID-19 in a low-resource setting.

We contribute to the broader literature on the role of the internet in improving health outcomes. Research has shown that the internet can increase the demand for healthcare services (Suziedelyte, 2012; Amaral-Garcia et al., 2022), improve self-reported health status (Lam et al., 2020; Hunsaker et al., 2021; Parys and Brown, 2023), and affect psychological well-being (McDool et al., 2020; Golin, 2022). Particularly in developing countries, internet access has been linked to better health outcomes, especially in low-income and rural communities. For example, Chen and Liu (2022) found that greater internet accessibility decreased the prevalence of overweight individuals in China, facilitating more engagement in health-related activities such as searching for health information and communication with healthcare providers. Collectively, these studies suggest that access to the internet can play a crucial role in improving health outcomes, a factor that becomes even more critical during a pandemic.

Finally, our paper intersects with the broader discourse on the impacts of infrastructure on health outcomes. Infrastructure, such as roads and railways, is traditionally viewed as a facilitator of improved access to healthcare services, thus potentially enhancing public health (Adhvaryu and Nyshadham, 2015; Aggarwal, 2021). However, such physical infrastructure may increase the spread during pandemics via increased physical contact across areas. Digital infrastructure presents a more adaptable and potentially cost-effective alternative by facilitating telemedicine and supporting remote working. Our examination of heterogeneity in the effects of internet access contributes to the literature on differential impacts across various groups, including by education and occupation (Hjort and Poulsen, 2019; Couture et al., 2021; Zuo, 2021).

# 2 Data

We start by describing our data sources; more information can be found in Appendix Table A.1.

## 2.1 Main data

### COVID-19 data

We used COVID-19 data provided by the Indonesian COVID-19 Task Force, established in March 2020, to manage, control, monitor, create, and implement strategic policies for the acceleration of the national COVID-19 response. The dataset includes a range of measures at the regency (district) level, such as the weekly number of COVID-19 cases and the incidence rate per 100,000 population. Our unit of analysis is the second-level administrative divisions comprising 514 units: 416 regencies and 98 cities in Indonesia (we refer to these as regencies in what follows); these are located in 38 provinces of Indonesia. We aggregated the weekly data at this regency level, covering 1 March 2020 to 18 February 2023.[4] Appendix Figure A.1 presents the fluctuations in COVID-19 transmission rates across regencies in Indonesia, showing substantial variations across different areas. Populated regions of the country, such as Jakarta and Surabaya, report significantly higher infection rates than their less-populated, rural counterparts, as evidenced by the number of confirmed cases.

### Internet data

Following the literature cited above, we concentrate on the coverage of the 3G network.[5] We utilize digital maps of global mobile network coverage provided by Collins Bartholomew's Mobile Coverage Explorer. The data contains information about signal coverage at a 1-by-1-kilometer grid level.

We use data from December 2019, just before the pandemic started in Indonesia. Our measure of internet access follows previous studies (Manacorda and Tesei, 2020; Guriev et al., 2021; Do et al., 2023). First, we determine if a particular 1-by-1-kilometer grid has a 3G (or 2G) signal. We then compute the proportion of areas within each regency with network access by evaluating the grid cells corresponding to that regency. A visual example is provided in Figure 2. This does not provide a direct measure of internet usage. Instead, it provides a measure of potential access, which is arguably less endogenous to people's preferences and other behaviors. Appendix Figure A.1 shows access to 3G is widespread

[4] In Figure A.2, we plot the aggregated number of cases at the national level using data from both the Government of Indonesia and Johns Hopkins University (Dong et al., 2022). Overall, both sources demonstrate a similar trend in the emergence of COVID-19 in Indonesia, particularly during the early stages of the pandemic. However, the local data in our analysis excludes unidentified areas; therefore, the reported figures are lower than the national trend. The areas not covered in our data comprise 9.5% of the Indonesian population. We conduct several robustness tests to ensure that this issue does not compromise the validity of our findings.

[5] The adoption of 3G technology marks a significant milestone in the evolution of mobile communication, enabling users to access and interact with online content, including social media, much more efficiently. The more basic 2G networks are limited to voice calls, SMS text messaging, and multimedia messages (MMS). During the COVID-19 pandemic, 2G networks, therefore, may be limited in their utility for public health communication and remote services, while 3G networks significantly enhance the capability for timely and effective dissemination of health advisories, telehealth consultations, and support for remote education and work.

in highly developed regions such as Sumatra and Java. In contrast, internet coverage in the Eastern areas, such as Maluku and Papua, is relatively low. Analogous to the number of cases above, the figure indicates substantial variations in access to mobile internet.

#### Lightning strike data

As an instrument for internet coverage, we employ data on lightning strikes sourced from the World Wide Lightning Location Network (WWLLN) dataset (Kaplan and Lau, 2021). This dataset provides the exact coordinates and time of all detected cloud-to-ground lightning strikes for the entire globe.[6] We aggregate these data from the grid level to the regency level. Our measure of lightning strikes is the mean stroke power in megawatts (MW), aggregated at the monthly basis. This measure provides a unique perspective on the intensity and energy of lightning activity, which we hypothesize to be a relevant proxy for assessing internet coverage in the region (discussed in detail in Section 3).

Following previous studies, we control for weather conditions using temperature and precipitation from ERA5, which is the fifth generation of reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) (Manacorda and Tesei, 2020; Do et al., 2023). Reanalysis data, which combine information from ground stations, satellites, weather balloons, and other inputs with a climate model to estimate weather variables across a grid, are particularly valuable in settings where weather stations' spatial and temporal coverage is limited, as is often the case in many developing countries (Dell et al., 2014). Unlike traditional weather station data, reanalysis data cover a larger geographical area and are available over an extended period. Furthermore, these data may resolve issues related to endogeneity concerns associated with weather station placement and variations in the quality and quantity of data collection (Auffhammer et al., 2013; Donaldson and Storeygard, 2016). We aggregate monthly temperature (measured in Celsius) and precipitation (measured in millimeters) at the regency level for Indonesia using the inverse-distance weighting approach (Deschênes and Greenstone, 2011).

The right panel of Figure 2 presents an example of the global distribution of lightning strikes. Two observations are noteworthy. First, lightning strikes are predominantly concentrated in countries around the equator, including Indonesia, aligning with the geographical predisposition to higher thunderstorm activity in these regions. Second, even within Indonesia, there is significant variation in lightning strike frequency (see also Appendix Figure A.1), providing further empirical support for the use of lightning strikes as an instrument in our analysis.

## 2.2 Covariates

Internet access is associated with local circumstances that might also be associated with the spread of the pandemic. Data on these can be hard to obtain in developing country settings. Indonesia, however, is an outlier, allowing us to collect various characteristics which we now discuss.[7]

#### Demographics

To account for demographic characteristics, we aggregate data at the regency level from multiple sources. First, we use our COVID-19 database for data on population density and counts at the regency level. Second, we supplement this data with several key indicators from the 2010 Indonesian population census

[6] The raw lightning observations are recorded with latitude, longitude, and time stamps (see Kaplan and Lau, 2021). Individual lightning stroke observations are summed on a geographic grid at the desired spatial resolution of 5km, which serves as our primary measure. For robustness, we also employ the 10km measure grid data to capture any spatial effects.

[7] The authors thank the ADB team for supporting this data collection.

conducted by Statistics Indonesia in May and June 2010, which surveyed 237.6 million people. Specifically, we include the proportion of the population aged 65 and older, the proportion of individuals without any formal education, the proportion of individuals with graduate degrees, the average household size, and the proportion of the population living in rural areas. We follow Bazzi et al. (2019) and Alesina and La Ferrara (2005) to calculate the ethnic fractionalization and polarization at the regency level.[8] These variables are included to capture the potential impact of social cohesion and cultural diversity on the effectiveness of public health responses and the spread of COVID-19, as these factors can influence community compliance with health directives and the dissemination of information within diverse populations.

To measure the characteristics of the workforce in the pre-COVID-19 period, we utilize data from the National Labor Force Survey (SAKERNAS) for 2019. SAKERNAS is a survey designed specifically to gather employment data and is representative at the province level (Suryadarma et al., 2007). We focus on several indicators, including the proportion of workers in agriculture, the share of workers with a long commute (more than 1 hour), and those who use public transportation. Additionally, we construct an index to assess the ability to work from home (also known as telework ability) by sector (see Dingel and Neiman, 2020; Asian Development Bank, 2021).

#### Healthcare availability

The healthcare data utilized in our analysis are derived from Open Street Map (OSM), an open-source, collaboratively constructed global mapping project that provides detailed geographic data quantifying the number of healthcare facilities across various geographic locations, specifically clinics (that provide outpatient care) and hospitals (which provide inpatient care, specialized surgeries, and emergency services). These data provide a comprehensive spatial distribution of healthcare infrastructures across Indonesia at the regency level. We also incorporate local healthcare spending data at the same level.

#### Economic status

Perhaps the most critical factors tying internet access to the pandemic spread are those associated with economic activity. To estimate local economic activity in Indonesian regencies, we utilize satellite nightlight data from the Visible Infrared Imaging Radiometer Suite (VIIRS) administered by the National Oceanic and Atmospheric Administration (NOAA) (see, e.g., Henderson et al., 2012; Hodler and Raschky, 2014; Martinez, 2022).[9] We calculate the nightlight density for all regencies by aggregating satellite images from daily grids to yearly data.[10] We also use the human development index compiled for Indonesia in 2019, capturing economic status, human well-being, and flourishing more broadly, at the regency level.

## 3 Empirical model

We investigate the effects of internet access, proxied by pre-pandemic ($t_0$) 3G availability ($\text{G3access}_{r,t_0}$), on reducing COVID-19 cases at the regency level $r$ in Indonesia, at different stages of the pandemic. We

[8] The fractionalization and polarisation indices are calculated using the following formulas: $fractionalization = 1 - \sum_{j=1}^{J} \pi_j^2$ and $polarisation = 4 \sum_{j=1}^{J} \pi_j^2 (1 - p_j)$, where $p_j$ is the share of the ethnic group $j$.

[9] In Indonesia Gibson et al. (2021) find a positive relationship between regency-level VIIRS data and gross domestic product (GDP).

[10] Atmospheric conditions may impact the ability of satellite sensors to capture night lights. We follow the Copernicus program's recommendation to exclude results from pixels with above 10 percent cloud fraction by performing cloud masking to address this. For more details, see: https://atmosphere.copernicus.eu/flawed-estimates-effects-lockdown-measures-air-quality-derived-satellite-observations?q=flawed-estimates-effects-lockdown-measures-air-quality-satellite-observations

assess the over-time association between COVID-19 outcomes and pre-pandemic access to the internet via

$$\mathrm{E}(y_{r,t}|X) = \exp(\alpha_t + \tau_t \mathrm{G3access}_{r,t_0} + x'_{r,t_0}\beta_t + \delta_{p,t}), \tag{1}$$

where $y_{r,t}$ represents the cumulative cases in regency (or city) $r$ at time $t$, adjusted per 10,000 population. We employ a Poisson specification (Chen and Roth, 2023).[11] Because our model is cross-sectional – emphasizing the critical role of information availability early in the pandemic – we control for the large set of covariates ($x'_{r,t_0}$) discussed above, at the regency level, measured before the onset of the pandemic. Additionally, we include fixed effects for provinces ($\delta_p$, of which there are 38) to mitigate unobservable factors affecting our outcomes at the province level (the main level of local policymaking). Finally, we employ population weights to allow for areas with larger populations to get higher weights – but show that these are immaterial to our results – and use robust standard errors to account for potential heteroskedasticity.

As with any other infrastructure, the expansion of mobile network coverage is likely to be influenced by endogenous factors. For example, network service providers tend to prioritize areas with high economic activity and potential demand, which may be associated with other factors that could influence the spread of the pandemic. As a result, despite including a large set of covariates, the simple conditional association between mobile internet penetration and COVID-19 spread might not accurately measure the causal impact.

To address this, we use an instrumental variable strategy using lightning strikes (based on Andersen et al., 2012; Manacorda and Tesei, 2020; Guriev et al., 2021; Do et al., 2023). The underlying rationale is that lightning strikes can cause significant damage to digital infrastructure, leading to localized black-outs. This IV is ideally suited to our context, as countries around the equator experience a high frequency of lightning strikes (c.f. Figure 2). This is compounded by the fact that there are fewer towers than in more developed countries to compensate for those affected by lightning. However, lightning strikes can also impact other types of infrastructure, such as power grids, telecommunications, and emergency services, which could indirectly influence the spread of COVID-19 during the pandemic. These infrastructural components are vital for disseminating information, maintaining healthcare services, and coordinating pandemic response efforts. While the exclusion condition is not directly testable, we mitigate against potential confounding factors in our models, by including controls for weather conditions, health facilities, and economic activities. For example, adverse weather at the onset of the pandemic may lead people to stay indoors more, even without information about the new pathogen. By controlling for temperature and precipitation, we ensure comparisons are made between regions with similar weather conditions, some of which experience lightning activity and others that do not.

We implement a Poisson-IV model (as Graff Zivin et al., 2023), with first stage:

$$G3access_{r,t_0} = \beta + \gamma_1 f(Lightning_{r,t_0}) + \gamma_2 temperature_{r,t_0} + \gamma_3 precipitation_{r,t_0} + x'_{r,t_0}\beta + \delta_p + \varepsilon_{r,t_0} \tag{2}$$

where the instrumental variable, $Lightning_r$, is measured by the average lightning stroke power at the regency level in December 2019, visualized in Appendix Figure A.4. We additionally provide two alternative IVs based on de-trended (residualized) lightning strikes, using larger regions $R$ - 10km, i.e., $Lightning_{r,t_0} = \gamma Lightning_{R,t_0} + \varepsilon_{r,t_0}$ and over time, the whole year of 2019, i.e. $Lightning_{r,t_0} = \gamma Lightning_{r,T_0} + \varepsilon_{r,t_0}$, thus using only variation in excess of the larger area and yearly lightnings strikes. Tailored to our context of a developing country near the equator, we further assess robustness using

[11] This approach accounts for the heavily skewed nature of the case data (cf. Appendix Figure A.3). To verify the robustness of our findings, we explore alternative model specifications, including log-transformed OLS and negative binomial regressions.

humidity and the number of cell towers.

In our primary analysis, we focus on the last date in our dataset (i.e., 18 February 2023) and conduct a single cross-sectional regression.[12] However, to capture the evolution of the lightning-induced lack of internet access over various stages of the pandemic leading up to our end-line date, we adjust the cumulative number of cases per population at different stages of the pandemic $t$ (as in Kunz and Propper, 2022).

We explore the heterogeneous effects of internet access on COVID-19 transmission by considering a range of local-area characteristics, including educational level, labor force characteristics, rurality, and economic development. To this end, we modify Equation (1) to include interaction terms between internet access and these demographic and socio-economic variables:

$$\mathrm{E}(y_{r,t}|X) = \exp(\alpha_t + \tau_t \mathrm{G3access}_{r,t_0} + \tau_t^x \mathrm{G3access}_{r,t_0} \times 1[x \geq med_x] + x'_{r,t_0}\beta_t + \delta_{p,t}) \qquad (3)$$

$1[x \geq med_x]$ represents an indicator for whether variable $x$—for instance, the share of population working in agriculture—is greater than or equal to its median. We similarly use $f(Lightning_{r,t_0}) \times 1[x \geq med_x]$ as an additional instrument for exploring these interactions.

# 4 Results

By 18 February 2023, Indonesia had recorded a cumulative total of approximately 6.5 million confirmed COVID-19 cases.[13] These cases were disproportionately concentrated in areas with limited to no mobile internet access. At the median split of 3% of exposure to the internet (which is notably low), regions with minimal to no internet exposure accounted for 4 million cases, compared to 2.5 million in regions above the median exposure. This stark difference corresponds to a 37.5% decrease. Disaggregating this disparity at the regency level, which aligns with our estimation specification, the difference in the mean case count is 18,000 for areas with low access versus 7,500 for areas with high access, representing a 58% decrease. Figure 1, panel B, demonstrates that this gap is not solely attributable to binary distinctions of no access versus full access but rather decreases in a more linear fashion.

The results of the unconditional effects of internet access on COVID-19 spread using our specification 1 are shown in Appendix Table B.1. The association is 1.495, corresponding to a 77% reduction in cases. This difference could be influenced by various factors, including rurality and economic activity. We examine the conditional association in Table 1. Panel A of this table illustrates the impact of 3G internet access on the COVID-19 transmission rate.[14] The conditional association, presented in Column 1 and adjusted for covariates likely to affect the pandemic spread and province-level fixed effects, reveals a large, negative, and statistically significant coefficient. This implies a 48% reduction in cases when moving from no internet exposure to full exposure (i.e., $exp(-0.66)-1$). For a typical change in exposure (one standard deviation, approximately 30%), this corresponds to a 24% decrease.

Next, we address the potential endogeneity between internet access and COVID-19 transmission rates. The reduced form leveraging lightning strikes as an exogenous factor affecting internet connectivity is presented in Column 2 of Table 1. The coefficient indicates a significant positive association between lightning strikes and COVID-19 cases. This association conditions on the full battery of covariates, including economic status and weather conditions in the local region; thus, lightning strikes are arguably unlikely to affect COVID-19 cases through channels other than internet access. Consequently, we use lightning as the instrument in the Poisson-IV model in Column 3. The first-stage results are compellingly

[12] Hatte et al. (2021) also used this approach cross-sectionally to identify a specific point in time/news.
[13] See: https://www.statista.com/statistics/1103469/indonesia-covid-19-total-cases/
[14] Appendix Table B.1 provides a detailed step-wise inclusion of these factors.

strong, aligning with findings from previous studies. The association between lightning strikes and internet access is highly significant, with an F-statistic of 61. The estimated impact of a one standard deviation increase in internet exposure is a reduction of 45%. Thus, lifting the areas with very low exposure above to 34% coverage would imply a reduction in 8,100 cases (from 18,000) at the mean and 1.8 million cases across the population (from 4 million). The IV results imply that the conditional association is biased towards zero. This is consistent with any explanation where any unobserved factor $x$ is positively (or negatively) correlated with the internet but negatively (or positively) correlated with COVID-19 cases. For example, unmeasured local factors like political efficiency, which could enhance internet accessibility while mitigating case numbers, might explain the observed pattern.

In Columns 4 and 5, we explore alternative implementations of the IV by using detrended lightning strikes. In Column 4, we account for lightning activity in the region before December 2019, using the anomaly of strikes in that month. Analogously, in Column 5, we assess the anomaly of strikes relative to a larger area, aiming to address potential spill-over effects. Although the magnitude of the effect size varies across these specifications, our core finding remains robust: internet access has a strong and significant impact on mitigating the spread of COVID-19.

### Evolution over time

To examine the impact of internet access across different stages of the pandemic, including the emergence of new variants and vaccine rollouts, we estimate the model for cumulative cases at various pandemic dates up to February 2023. Figure 3 displays these temporal dynamics, with the final coefficients corresponding to those in Table 1, Columns 1 and 3. Notably, the disparity in COVID-19 cases between areas with and without internet access began to diverge early in the pandemic, with subsequent growth rates stabilizing.[15] This pattern indicates that the case levels continued to diverge in a consistent manner, unaffected by new variants or other developments over three years, suggesting a critical role for initial information access.

### Robustness checks

As in many developing country settings, there are issues with missing values. Although the satellite-based internet measures and the high-quality COVID data are largely complete, some of the covariates we include at the regency level contain missing values. We address this issue in Appendix Table B.1, columns 8 and 9. Our approach involves setting the missing values to zero and incorporating corresponding dummy indicator variables. This has a negligible effect. Additionally, we apply the multiple imputation approach of missing covariates from Rubin (1996). Our results are virtually unchanged.

In Table B.2, we conduct a variety of robustness checks to address general estimation issues, using Poisson regressions that are not instrumented. First, we validate our Poisson model specification by showing that log-transformed and negative binomial models yield similar results (Columns 2 and 3). Next, we evaluate alternative 3G measures: raw 3G data without extrapolation for areas with missing data (Column 4) and setting missing 3G data to zero while adding a dummy indicator for these values (Column 5). These alternative measures suggest a larger effect of the internet on COVID-19 cases.[16] Next, employing unweighted regressions reveals larger coefficients and marginal effects for a one standard deviation change, reinforcing our findings. We also address the challenge of missing COVID-19 data for 10 regencies, applying zero imputation for these and performing a comprehensive imputation for all

[15] This is consistent with restrictions and realized reductions in mobility early on in the pandemic (c.f. Figure A.5).

[16] We also estimated the effect of access to only 2G internet. We found minimal effects on COVID-19 transmission. This is as expected, as 2G primarily supports basic services (.e.g. SMS and voice calls). In contrast, 3G facilitates the more advanced data services that are required to disseminate health information.

values (including COVID-19 cases, internet, and lightning strikes) using Rubin (1996)'s method. The latter technique yields very little alteration in the estimates (Columns 6 and 7).[17]

Table B.3 presents results of other robustness tests, with Poisson results displayed in the top panel and IV-Poisson results in the bottom panel. We undertake additional robustness checks to account for factors that may be particularly relevant in our context, such as the actual internet accessibility and infrastructure characteristics. We start with an alternative measure of internet access that considers the population within the coverage area, not just the presence of a signal (Column 2).[18] While this population-based approach gives an alternative view of exposure, it may introduce endogeneity—for example, if individuals most benefiting from internet access gravitate towards areas closer to towers. Therefore, our preference is the non population-weighted measure of access. Next, we further control for humidity (via temperature and precipitation interaction), a critical factor in the tropics influencing both lightning occurrence and severity, and, by extension, internet service quality and virus transmission (Column 3). Column 4 includes controls for the number of cell towers in the area, reflecting the extent of mobile network infrastructure. In developed countries, tower density likely precludes significant variation in the instrument. However, sparse tower distribution in our setting means lightning can severely disrupt internet access. Finally, we also repeat the residualized instrumental variables for the number of towers in the area. Across these tests, our findings remain stable, showing that our main conclusions about the role of the 3G internet in reducing COVID-19 spread are highly robust.

**Heterogeneity**

We have established a robust and consistent relationship between internet access and reducing COVID-19 transmission. We now examine whether certain areas derived greater benefits from internet exposure than others. We examine the differential effects of internet access on COVID-19 transmission across population educational levels, labor force characteristics, urbanicity, and economic activity. Our approach utilizes regressions in the form of Equation (3). Figure 4 presents the coefficient estimates for these interactions (other estimates are in Appendix Table B.4).

Two findings emerge as particularly significant. First, our analysis reveals the benefits of internet access on COVID-19 transmission are more pronounced among populations with higher educational levels. Such populations may be more receptive to the emergence of new information via 3G, which is often text-based. Based on the median split of the local population with no education versus some education, the difference is both highly significant (p=0.022) and large in magnitude. This difference, coupled with the statistically insignificant difference for higher education, suggests that basic literacy is required to derive benefits.

Next, we explore variations in industry structure. Regions with a higher potential for remote employment (telework) exhibit greater benefits from internet access (p=0.069). The ability to telework has been essential for containing COVID-19's spread, as it reduces the necessity for commuting and face-to-face interactions, thereby decreasing transmission risks. Furthermore, telework often requires higher digital literacy, potentially leading to increased awareness and compliance with public health guidelines disseminated through digital channels. On the other hand, a large agricultural sector is associated with diminished benefits from internet access (p=0.065), likely due to the infeasibility of remote work in such settings. While these findings should not be interpreted causally, given the probable high correlation between labor force characteristics and educational levels, they are consistent with the notion that leveraging timely information for pandemic mitigation also requires the local population to be able to adapt

[17] The distinction in the last two approaches stems from our ability to observe province and population for some records, while for others, we impute population (see Column 8).

[18] This measure, illustrated in Figure 2, aggregates all grids with internet access, regardless of the population density in each grid.

to shelter-in-place and social distancing policies/recommendations.

Finally, we investigate whether the observed effects associated with educational level and labor force characteristics merely reflect underlying difficulty density or economic development. Interestingly, our findings indicate no such difference. Moreover, these factors seem to operate in opposing directions: regions with higher urbanicity derive more benefit (despite potentially increased population density), whereas areas with greater economic development—proxied by nightlight intensity—experience marginally (and insignificantly) fewer benefits.[19]

# 5 Conclusion

We assess the relationship between internet coverage and health outcomes, leveraging the COVID-19 pandemic as an unexpected health shock to examine the role of digital connectivity in mitigating its spread. Our setting is Indonesia, a country characterized by varied access to broadband services and varying rates of COVID-19 transmission. Its geographical location, particularly susceptible to frequent lightning strikes, provides a unique context for employing an IV framework to isolate the effect of internet access on public health outcomes. Utilizing rich data at the subnational level we find that access to 3G broadband internet reduced COVID-19 cases at the local level by approximately 45%. This implies that improving connectivity by 30% (1 sd) could have prevented 1.8 million cases over the course of the first three years of the pandemic. This would essentially have eradicated the gap between Indonesia and its peers, such as the Philippines, Thailand, and Japan (c.f. Figure 1). Moreover, the effect observed in our study is large compared to the effectiveness of other non-pharmaceutical interventions (NPIs) intended to reduce COVID-19 transmission. Although the impact of specific NPIs varies, the 45% reduction attributed to enhanced internet access is on par with, or even surpasses, the outcomes associated with measures such as social distancing, mask-wearing, and travel restrictions (c.f. VoPham et al., 2020; Leech et al., 2022; Kwok et al., 2021). We also find regions with higher literacy levels and greater capacity for telework experienced larger benefits from access, emphasizing the role of educational and infrastructural readiness in maximizing the public health advantages of digital connectivity. Finally, our finding of a sustained divergence in growth rates between regions with and without early internet access demonstrates the ongoing importance of internet access in curbing virus spread. This enduring effect, evident despite potential other influences like vaccine distribution or initial data inconsistencies, reaffirms the important role of digital connectivity in bolstering public health interventions amid evolving challenges, such as new variants and changing guidelines.

[19] Exploratory analyses indicate significant benefits in relatively light-poor urban areas. But dividing the data by two median splits leads to a limited number of observations, so results should be interpreted with caution. Results available on request.

## References


Adhvaryu, A. and A. Nyshadham (2015): "Return to treatment in the formal health care sector: Evidence from Tanzania," *American Economic Journal: Economic Policy*, 7, 29–57.

Aggarwal, S. (2021): "The long road to health: Healthcare utilization impacts of a road pavement policy in rural India," *Journal of Development Economics*, 151, 102667.

Agha, L. (2014): "The effects of health information technology on the costs and quality of medical care," *Journal of health economics*, 34, 19–30.

Aker, J. C. and I. M. Mbiti (2010): "Mobile phones and economic development in Africa," *Journal of Economic Perspectives*, 24, 207–232.

Alesina, A. and E. La Ferrara (2005): "Ethnic diversity and economic performance," *Journal of Economic Literature*, 43, 762–800.

Amaral-Garcia, S., M. Nardotto, C. Propper, and T. Valletti (2022): "Mums Go Online: Is the Internet Changing the Demand for Health Care?" *Review of Economics and Statistics*, 104, 1157–1173.

Andersen, T. B., J. Bentzen, C.-J. Dalgaard, and P. Selaya (2012): "Lightning, IT diffusion, and economic growth across US states," *Review of Economics and Statistics*, 94, 903–924.

Asian Development Bank (2021): "COVID-19 and Labor Markets in Southeast Asia: Impacts on Indonesia, Malaysia, the Philippines, Thailand, and Viet Nam," Tech. rep., Asian Development Bank, Manila, Philippines.

Auffhammer, M., S. M. Hsiang, W. Schlenker, and A. Sobel (2013): "Using weather data and climate model output in economic analyses of climate change," *Review of Environmental Economics and Policy.*

Barber, S. L., P. J. Gertler, and P. Harimurti (2007): "Differences In Access To High-Quality Outpatient Care In Indonesia: Lower quality in remote regions and among private nurses is a manifestation of the educational, policy, and regulatory frameworks upon which the Indonesian health system is based." *Health affairs*, 26, w352–w366.

Bargain, O. and U. Aminjonov (2020): "Trust and compliance to public health policies in times of COVID-19," *Journal of public economics*, 192, 104316.

Bazzi, S., A. Gaduh, A. D. Rothenberg, and M. Wong (2019): "Unity in diversity? How intergroup contact can foster nation building," *American Economic Review*, 109, 3978–4025.

Behrman, J. A. (2015): "The effect of increased primary schooling on adult women's HIV status in Malawi and Uganda: Universal Primary Education as a natural experiment," *Social science & medicine*, 127, 108–115.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng (2024): "On binscatter," *American Economic Review*, 114, 1488–1514.

Chang, D., X. Chang, Y. He, and K. J. K. Tan (2022): "The determinants of COVID-19 morbidity and mortality across countries," *Scientific reports*, 12, 5888.

Chen, J. and J. Roth (2023): "Logs with Zeros? Some Problems and Solutions," *The Quarterly Journal of Economics*, qjad054.

Chen, L. and W. Liu (2022): "The effect of Internet access on body weight: Evidence from China," *Journal of Health Economics*, 85, 102670.

Chib, A., M. H. van Velthoven, and J. Car (2015): "mHealth adoption in low-resource environments: a review of the use of mobile healthcare in developing countries," *Journal of health communication*, 20, 4–34.

Cohen, J. and P. Dupas (2010): "Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment," *The Quarterly Journal of Economics*, 125, 1–45.

Courtemanche, C., J. Garuccio, A. Le, J. Pinkston, and A. Yelowitz (2020): "Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate: Study evaluates the impact of social distancing measures on the growth rate of confirmed COVID-19 cases across the United States." *Health affairs*, 39, 1237–1246.

Couture, V., B. Faber, Y. Gu, and L. Liu (2021): "Connecting the countryside via e-commerce: evidence from China," *American Economic Review: Insights*, 3, 35–50.

Dell, M., B. F. Jones, and B. A. Olken (2014): "What do we learn from the weather? The new climate-economy literature," *Journal of Economic literature*, 52, 740–798.

Deschênes, O. and M. Greenstone (2011): "Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US," *American Economic Journal: Applied Economics*, 3, 152–185.

Dingel, J. I. and B. Neiman (2020): "How many jobs can be done at home?" *Journal of Public Economics*, 189, 104235.

Do, Q.-T., N. Gomez-Parra, and B. Rijkers (2023): "Transnational terrorism and the internet," *Journal of Development Economics*, 103118.

Donaldson, D. and A. Storeygard (2016): "The view from above: Applications of satellite data in economics," *Journal of Economic Perspectives*, 30, 171–198.

Dong, E., J. Ratcliff, T. D. Goyea, A. Katz, R. Lau, T. K. Ng, B. Garcia, E. Bolt, S. Prata, D. Zhang, et al. (2022): "The Johns Hopkins University Center for Systems Science and Engineering COVID-19 Dashboard: data collection process, challenges faced, and lessons learned," *The Lancet Infectious Diseases*.

Duflo, E., P. Dupas, and M. Kremer (2015): "Education, HIV, and early fertility: Experimental evidence from Kenya," *American Economic Review*, 105, 2757–2797.

Dupas, P. (2011): "Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya," *American Economic Journal: Applied Economics*, 3, 1–34.

Gibson, J., S. Olivia, G. Boe-Gibson, and C. Li (2021): "Which night lights data should we use in economics, and where?" *Journal of Development Economics*, 149, 102602.

Golin, M. (2022): "The effect of broadband internet on the gender gap in mental health: Evidence from germany," *Health Economics*, 31, 6–21.

Graff Zivin, J., M. Neidell, N. J. Sanders, and G. Singer (2023): "When externalities collide: Influenza and pollution," *American Economic Journal: Applied Economics*, 15, 320–51.

Guriev, S., N. Melnikov, and E. Zhuravskaya (2021): "3G internet and confidence in government," *Quarterly Journal of Economics*, 136, 2533–2613.

Hatte, S., E. Madinier, and E. Zhuravskaya (2021): "Reading Twitter in the newsroom: Web 2.0 and traditional-media reporting of conflicts," *Available at SSRN 3845739*.

Henderson, J. V., A. Storeygard, and D. N. Weil (2012): "Measuring economic growth from outer space," *American economic review*, 102, 994–1028.

Hjort, J. and J. Poulsen (2019): "The arrival of fast internet and employment in Africa," *American Economic Review*, 109, 1032–1079.

Hodler, R. and P. A. Raschky (2014): "Regional Favoritism," *Quarterly Journal of Economics*, 129, 995–1033.

Hunsaker, A., E. Hargittai, and M. Micheli (2021): "Relationship between internet use and change in health status: panel study of young adults," *Journal of Medical Internet Research*, 23, e22051.

Jalan, J. and E. Somanathan (2008): "The importance of being informed: Experimental evidence on demand for environmental quality," *Journal of Development Economics*, 87, 14–28.

Kahn, J. G., J. S. Yang, and J. S. Kahn (2010): "'Mobile'health needs and opportunities in developing countries," *Health affairs*, 29, 252–258.

Kaplan, J. O. and K. H.-K. Lau (2021): "The WGLC global gridded lightning climatology and timeseries," *Earth System Science Data Discussions*, 2021, 1–25.

Kelly, T. and C. M. Rossotto (2011): *Broadband strategies handbook*, World Bank Publications.

Kickbusch, I., D. Piselli, A. Agrawal, R. Balicer, O. Banner, M. Adelhardt, E. Capobianco, C. Fabian, A. S. Gill, D. Lupton, et al. (2021): "The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world," *The Lancet*, 398, 1727–1776.

Kunz, J. S. and C. Propper (2022): "JUE insight: is hospital quality predictive of pandemic deaths? Evidence from US counties," *Journal of Urban Economics*, 103472.

Kwok, W.-C., C.-K. Wong, T.-F. Ma, K.-W. Ho, L. W.-T. Fan, K.-P. F. Chan, S. S.-K. Chan, T. C.-C. Tam, and P.-L. Ho (2021): "Modelling the impact of travel restrictions on COVID-19 cases in Hong Kong in early 2020," *BMC public health*, 21, 1–8.

Lam, S. S. M., S. Jivraj, and S. Scholes (2020): "Exploring the relationship between internet use and mental health among older adults in England: longitudinal observational study," *Journal of medical Internet research*, 22, e15683.

Lee, J., J. S. McCullough, and R. J. Town (2013): "The impact of health information technology on hospital productivity," *The RAND Journal of Economics*, 44, 545–568.

Leech, G., C. Rogers-Smith, J. T. Monrad, J. B. Sandbrink, B. Snodin, R. Zinkov, B. Rader, J. S. Brownstein, Y. Gal, S. Bhatt, et al. (2022): "Mask wearing in community settings reduces SARS-CoV-2 transmission," *Proceedings of the National Academy of Sciences*, 119, e2119266119.

Manacorda, M. and A. Tesei (2020): "Liberation technology: Mobile phones and political mobilization in Africa," *Econometrica*, 88, 533–567.

Martinez, L. R. (2022): "How much should we trust the dictator's GDP growth estimates?" *Journal of Political Economy*, 130, 2731–2769.

McCullough, J. S., S. T. Parente, and R. Town (2016): "Health information technology and patient outcomes: the role of information and labor coordination," *The RAND Journal of Economics*, 47, 207–236.

McDool, E., P. Powell, J. Roberts, and K. Taylor (2020): "The internet and children's psychological wellbeing," *Journal of Health Economics*, 69, 102274.

Mendolia, S., O. Stavrunova, and O. Yerokhin (2021): "Determinants of the community mobility during the COVID-19 epidemic: The role of government regulations and information," *Journal of Economic Behavior & Organization*, 184, 199–231.

Miller, A. R. and C. E. Tucker (2011): "Can health care information technology save babies?" *Journal of Political Economy*, 119, 289–324.

Parys, J. V. and Z. Y. Brown (2023): "Broadband Internet Access and Health Outcomes: Patient and Provider Responses in Medicare," Tech. rep., National Bureau of Economic Research Cambridge.

Rubin, D. B. (1996): "Multiple imputation after 18+ years," *Journal of the American statistical Association*, 91, 473–489.

Suryadarma, D., A. Suryahadi, S. Sumarto, et al. (2007): *Reducing unemployment in Indonesia: Results from a growth-employment elasticity model*, SMERU Research Institute.

Suziedelyte, A. (2012): "How does searching for health information on the Internet affect individuals' demand for health care services?" *Social Science & Medicine*, 75, 1828–1835.

Van de Poel, E., O. O'Donnell, and E. Van Doorslaer (2007): "Are urban children really healthier? Evidence from 47 developing countries," *Social science & medicine*, 65, 1986–2003.

VoPham, T., M. D. Weaver, J. E. Hart, M. Ton, E. White, and P. A. Newcomb (2020): "Effect of social distancing on COVID-19 incidence and mortality in the US," *MedRxiv*.

Zuo, G. W. (2021): "Wired and hired: Employment effects of subsidized broadband Internet for low-income Americans," *American Economic Journal: Economic Policy*, 13, 447–482.

# Figures & Tables

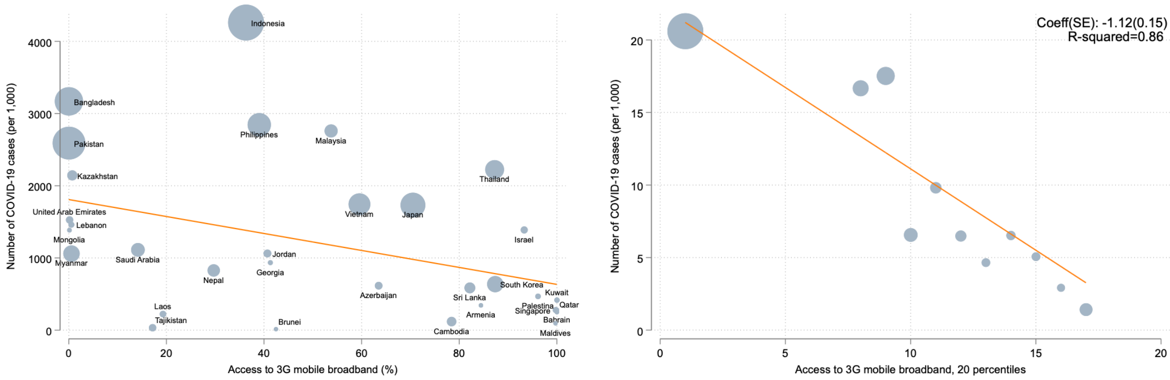


**Figure 1:** Mobile broadband access and COVID-19 cases - Asia region and within Indonesia

*Notes:* COVID-19 case data were measured as of February 18$^{\text{th}}$, 2023, and adjusted for population. Mobile broadband speed (3G access) measurements were taken in 2019 - before the pandemic started. In the left panel, the data were based on 32 countries in Asia for which data were available. The right panel shows an average number of cases from Indonesia's 454 regencies (districts) aggregated to 20 percentiles (several regencies have no access or full access, which are subsumed in the lowest and highest percentile), the size of the bubbles indicates the population living in these regencies, line denoting the population-weighted linear fit, with a correlation coefficient of -1.12 (standard error 0.15) and an $R^2$ of 89%. *Sources:* COVID-19 global data from the Johns Hopkins cross-national COVID-19 file; district-level data for Indonesia from the Indonesian COVID-19 Task Force. Mobile internet data from Collins Bartholomew's GSMA Mobile Coverage Explorer database, own calculations.

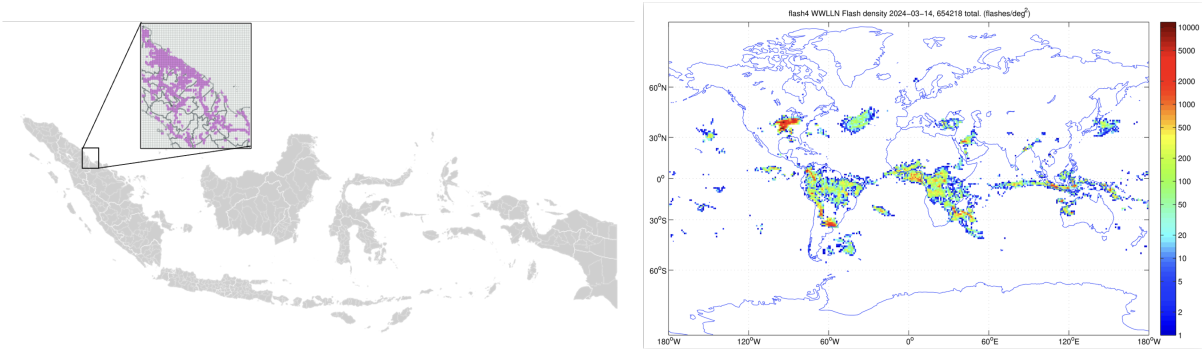


**Figure 2:** Regency boundaries in Indonesia and global lightning strike distribution

*Notes:* The left panel represents the regency boundary in Indonesia and provides an example of 3G grid data. See Appendix Figure A.1 for the complete distribution. The right panel displays world lightning strike data accessed on March 14$^{\text{th}}$, 2024. *Sources:* Mobile internet data from Collins Bartholomew's GSMA Mobile Coverage Explorer database. Lightning strike data is from the World Wide Lightning Location Network (WWLLN), own representation.

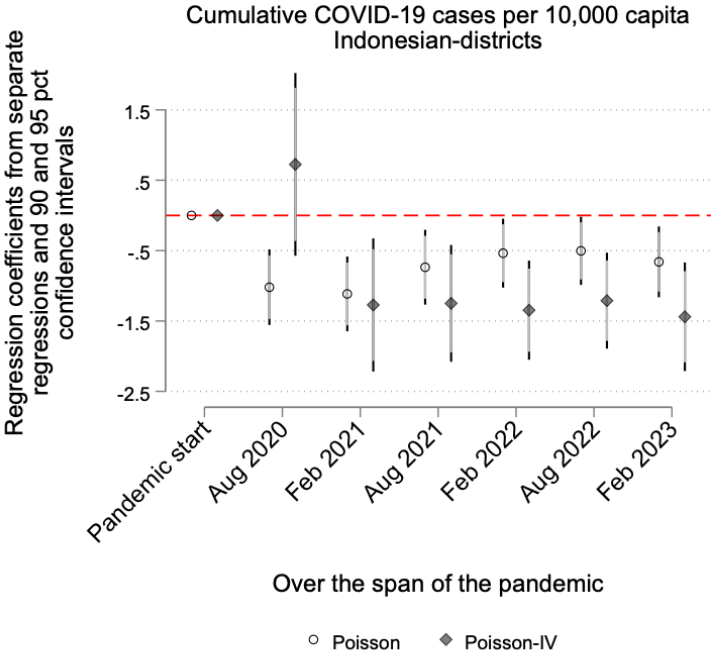


**Figure 3:** Mobile broadband access and COVID-19 cases – Effects throughout the pandemic

*Notes:* Figure presents regression coefficients of regressions (95-dark and 90-light level confidence intervals) presented in Table 1 (see notes therein) Columns 1 - circles and 2 - diamonds in the last date, separately for various dates throughout the pandemic covering various waves of new Covid variants. *Source:* COVID-19 data, adjusted for population, was sourced from the Indonesian COVID-19 Task Force. Mobile internet data was obtained from Collins Bartholomew's GSMA Mobile Coverage Explorer database, own calculations.

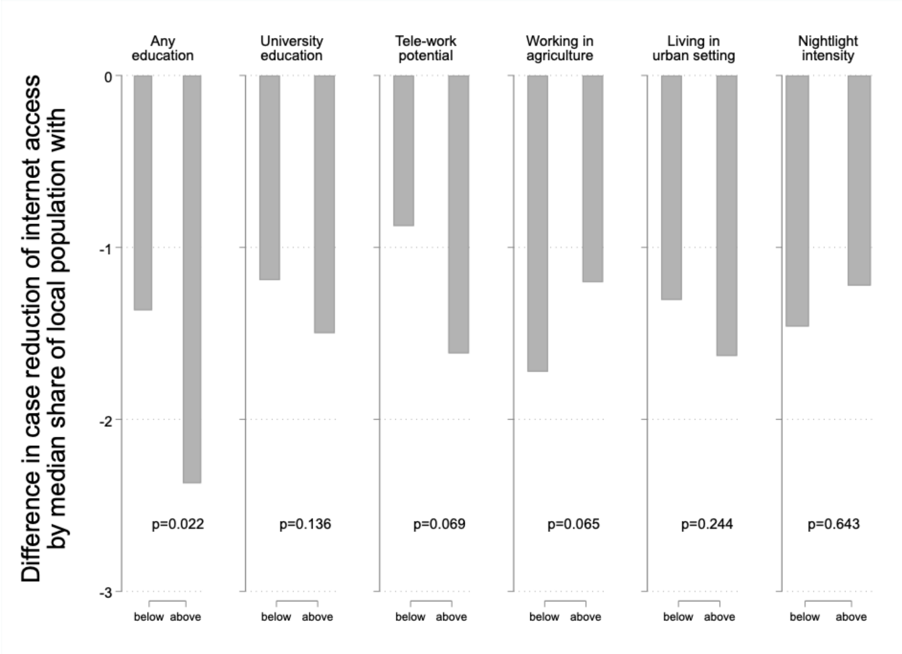


**Figure 4:** Heterogeneity analysis

*Notes:* Figure shows the effect heterogeneity for selected area characteristics: above median share - of any education (first panel) and university education (second), occupations with telework potential (third) and working in agriculture (fourth), living in urban settings (fifth) and above the median level of nightlight intensity (sixth). The respective left bars show the main coefficients, and on the right, the main plus interaction coefficients $\tau^x$ from Equation (3). The p-values indicate whether the difference is significantly different from zero. See Appendix Table B.4 for coefficient estimates. *Source:* COVID-19 data, adjusted for population, was sourced from the Indonesian Covid-19 Task Force, using measurements taken as of February 18th, 2023. Mobile internet data was obtained from Collins Bartholomew's GSMA Mobile Coverage Explorer database, own calculations.

**Table 1:** Mobile broadband access and COVID-19 cases – Main results

| Dependent variables: Cumulative cases by 10T population, 18 Feb 2023 | | | | | |
|---|---|---|---|---|---|
| | Poisson | | Poisson-IV | | |
| | Covars | Reduced form | 5 km rad. dec 2019 | Detrending time | Detrending area |
| | (1) | (2) | (3) | (4) | (5) |
| *Panel A. Cases* | | | | | |
| Internet exposure (3G) | -0.660 | | -1.440 | -0.892 | -2.011 |
| | (0.257) | | (0.394) | (0.397) | (0.555) |
| Semi-elasticity (1sd) | -0.24 | | -0.45 | -0.31 | -0.57 |
| Lightning strike frequency (5km, Dec 2019) | | 7.919 | | | |
| | | (2.429) | | | |
| *Panel B. First stage* | | | | | |
| Lightning strike frequency (5km, Dec 2019) | | | -4.050 | -5.920 | -3.203 |
| | | | (0.518) | (0.767) | (0.579) |
| $N$ | 454 | 454 | 454 | 454 | 454 |
| Difference per mil. for 1sd. increase | -0.96 | | -1.80 | -1.24 | -2.28 |
| Fstat | | | 61.05 | 59.54 | 30.56 |
| Province FEs | ✓ | ✓ | ✓ | ✓ | ✓ |
| Demographics | ✓ | ✓ | ✓ | ✓ | ✓ |
| Health facilities | ✓ | ✓ | ✓ | ✓ | ✓ |
| Economic status | ✓ | ✓ | ✓ | ✓ | ✓ |
| Ethnic composition | ✓ | ✓ | ✓ | ✓ | ✓ |
| Labour force | ✓ | ✓ | ✓ | ✓ | ✓ |
| Weather controls | ✓ | ✓ | ✓ | ✓ | ✓ |

*Notes*: The Table presents Poisson coefficients estimates from equation (1) and IV-Poisson with control function using (2). Column 1 uses the Poisson regression conditional on the full array of fixed effects and covariates; the stepwise inclusion and further robustness tests are presented in Appendix Table B.1. Column 2 shows the reduction estimates based on the lightning strikes in December 2019 – contemporaneously to the internet exposure in a 5 km radius, and Column 3 shows the corresponding IV estimates. Columns 4 and 5 present alternative IV definitions using in 4 the residualised lightning strikes conditional on the lightning strike frequency in the year 2019, and in 5 residualised conditional on the 10 km radius frequency. The semi-elasticity is calculated for a 1 standard deviation change in the internet measure, i.e., 0.34 - 34% exposure, via $\exp(\tau * sd(\text{internet})) - 1$. Panel B presents the IV estimates' corresponding first stage (control function). $N$ - number of observations, the projected difference in cases, and the first stage F-statistic. Further robustness on the IV estimates is presented in Appendix Table B.3. *Source:* COVID-19 data, adjusted for population, was sourced from the Indonesian COVID-19 Task Force, using measurements taken as of February 18$^{\text{th}}$, 2023. Mobile internet data was obtained from Collins Bartholomew's GSMA Mobile Coverage Explorer database, own calculations.

# Online Appendix

## A Additional data information

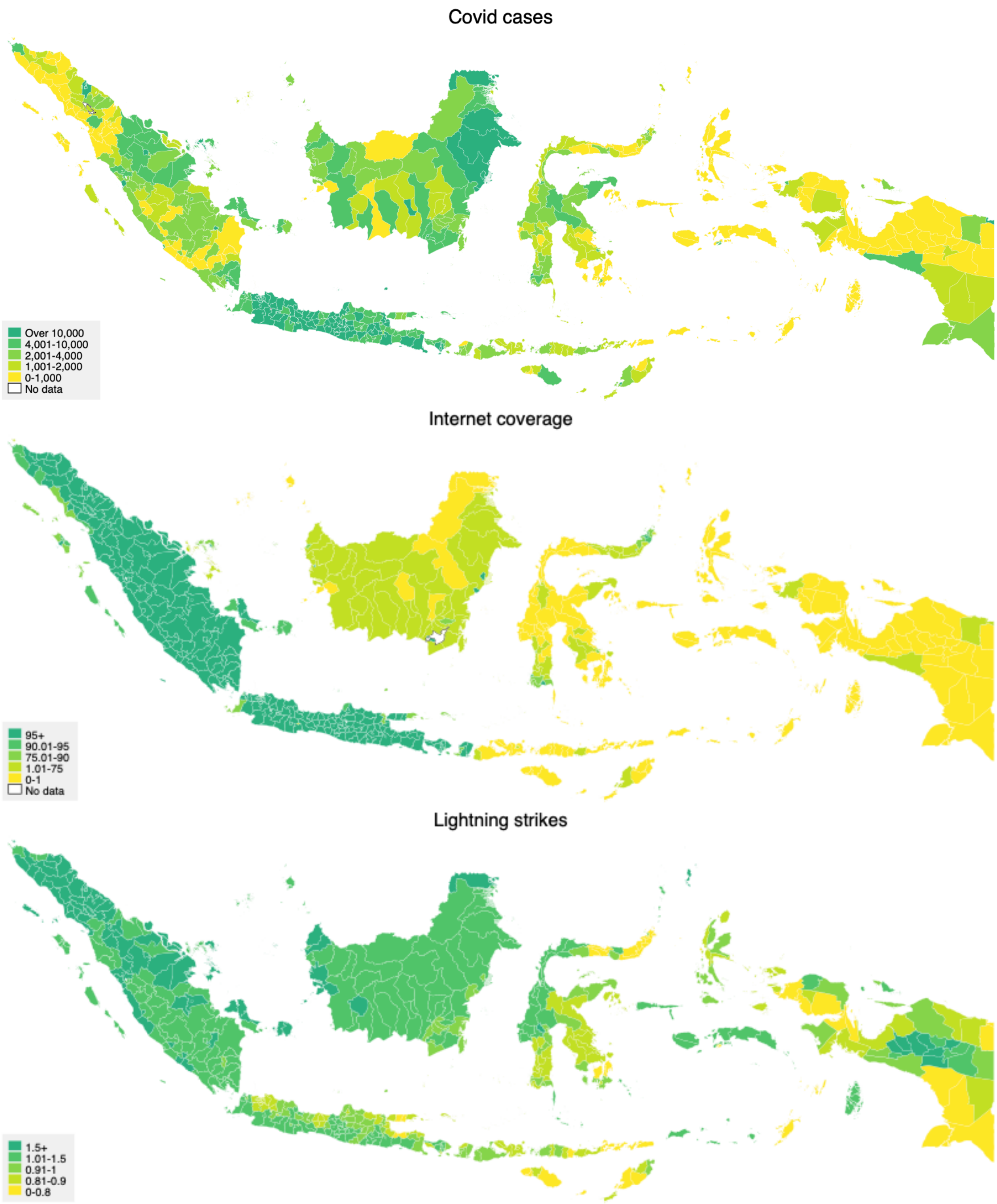


**Figure A.1:** Number of Covid cases, mobile coverage, and lightning strikes - Indonesia

*Notes:* Figure displays cumulative Covid cases on 18. Feb 2023 (top), 3G mobile network exposure (middle), and density of lightning strikes in December 2019 (bottom) across regencies in Indonesia. *Sources:* Covid data come from the Indonesian Covid-19 Task Force. Mobile internet data from Collins Bartholomew's GSMA Mobile Coverage Explorer database. Lightning strike data is from the World Wide Lightning Location Network (WWLLN), own calculations.

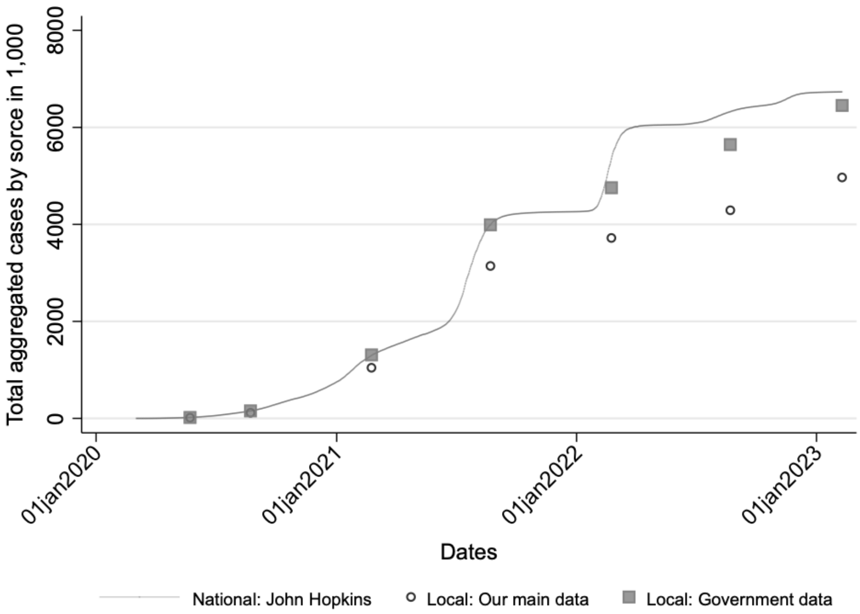


**Figure A.2:** Comparison of data sources - cases throughout pandemic

*Notes:* Figure compares the aggregated local data available (circles), not all cases can be associated with a region, i.e., cruise ships, tourists, etc., including these (squares), and finally, the national data aggregated by John Hopkins. *Source:* COVID-19 data, adjusted for population, was sourced from the Indonesian COVID-19 Task Force. National data comes from John Hopkins cross-national COVID file, own calculations.

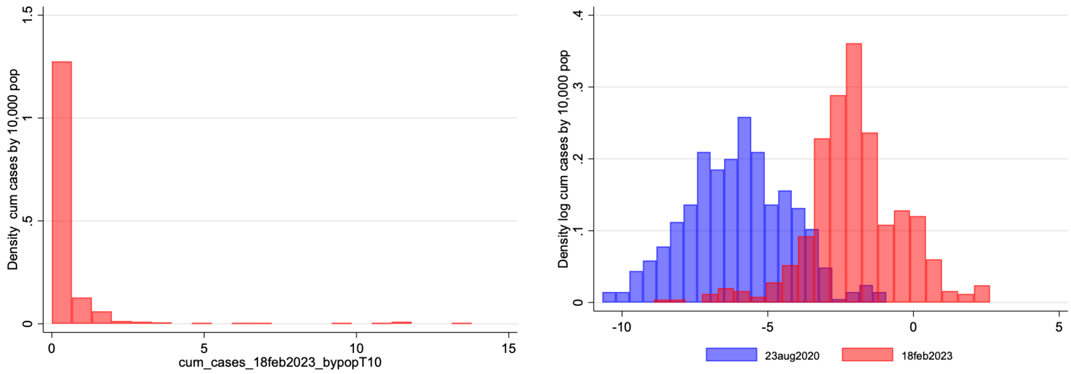


**Figure A.3:** Skewed cases on last day, and logged values beginning and end

*Notes:* The Right figure shows the raw distribution of our outcome variable for the last date available at data extraction. The left figure depicts the log-transformed outcome at the beginning of the pandemic (23 Aug 2020) and the late stage (18 Feb 2023)– our main outcome. *Source:* COVID-19 data, adjusted for population, was sourced from the Indonesian Covid-19 Task Force, own calculations.

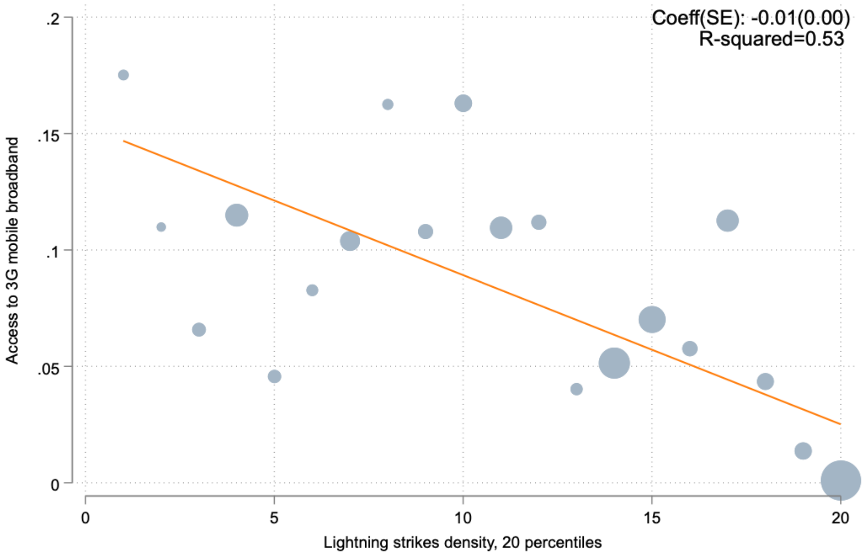


**Figure A.4:** Lightning strikes and 3G internet

*Notes:* The figure shows average internet access from Indonesia's 454 regencies (districts) aggregated to 20 percentiles. The size of the bubbles indicates the population living in these regencies, line denoting the population-weighted linear fit, with a correlation coefficient of -0.006 (standard error 0.001) and an $R^2$ of 53%. *Source:* Mobile internet data from Collins Bartholomew's GSMA Mobile Coverage Explorer database. Lightning strike data is from the World Wide Lightning Location Network (WWLLN), own calculations.

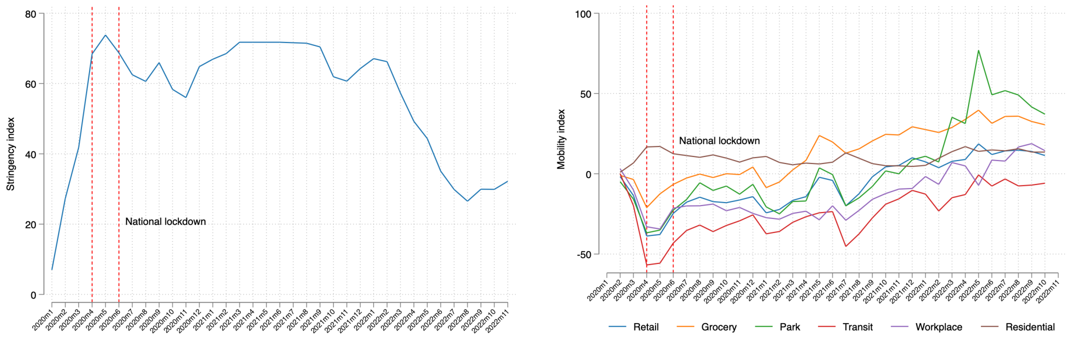


**Figure A.5:** Stringency index and Google mobility

*Notes:* Data from the Oxford COVID-19 Government Response Tracker. Data from COVID-19 Community Mobility Reports.

**Table A.1:** Variable overview

| | N | Mean | SD | Description | Source | Link |
|---|---|---|---|---|---|---|
| Cumulative cases 18 Feb 2023 | 454 | 10942.60 | 31381.30 | Total number of cases at the regency level as of 18 Feb 2023 | Indonesian Covid-19 Task Force | link |
| Cumulative deaths 18 Feb 2023 | 417 | 305.66 | 504.09 | Total number of deaths at the regency level as of 18 Feb 2023 | Indonesian Covid-19 Task Force | link |
| Mobile 3g OCI | 500 | 0.37 | 0.45 | Share of areas at the regency level having access to 3G as of Dec 2019 (OpenCellID) | Collins Bartholomew's Mobile Coverage Explorer | link |
| Mobile 3g MCE | 499 | 62.93 | 44.91 | Share of areas at the regency level having access to 3G as of Dec 2019 (Mobile Coverage Explorer) | Collins Bartholomew's Mobile Coverage Explorer | link |
| Mobile 4g OCI | 482 | 0.76 | 0.36 | Share of areas at the regency level having access to 4G as of Dec 2019 (OpenCellID) | Collins BartholomewÕs Mobile Coverage Explorer | link |
| Mobile 2g OCI | 500 | 0.52 | 0.34 | Share of areas at the regency level having access to 2G as of Dec 2019 (OpenCellID) | Collins Bartholomew's Mobile Coverage Explorer | link |
| Population density | 465 | 0.89 | 2.43 | Population density at the regency level | Indonesian Covid-19 Task Force | link |
| Population | 465 | 2.85 | 2.14 | Total population at the regency level | Indonesian Covid-19 Task Force | link |
| Population over 65 | 510 | 0.08 | 0.03 | Share of population with a university degree or higher at the regency level | 2010 Indonesian population census | link |
| Population with no education | 510 | 0.19 | 0.11 | Share of population with a university degree or higher at the regency level | 2010 Indonesian population census | link |
| Population with high education | 510 | 0.09 | 0.05 | Share of population with a university degree or higher at the regency level | 2010 Indonesian population census | link |
| Household size | 510 | 4.16 | 0.46 | Average household size at the regency level | 2010 Indonesian population census | link |
| Rural areas | 510 | 0.60 | 0.31 | Share of population living in rural areas | 2010 Indonesian population census | link |
| Number of clinics | 363 | 14.00 | 42.60 | Number of clinics in 2019 at the regency level | Open Street Map (OSM) | link |
| Number of hospitals | 363 | 8.00 | 14.03 | Number of hospitals in 2019 at the regency level | Open Street Map (OSM) | link |
| Health spending | 484 | 27.01 | 18.76 | Amount of (realized) health spending in 2019 at the regency level | Open Street Map (OSM) | link |
| Nighttime light | 500 | 2.14 | 5.07 | Average nightlight (Visible Infrared Imaging Radiometer Suite - VIIRS) as of Dec 2019 | National Oceanic and Atmospheric Administration (NOAA) | link |
| Human development index | 465 | 2.97 | 1.42 | Human development index at the regency level in 2019 (quintiles) | Badan Pusat Statistik | link |
| Ethnic fractionalization | 510 | 0.46 | 0.31 | Probability that two randomly selected people in a regency belong to different ethnic groups | 2010 Indonesian population census | link |
| Ethnic polarization | 510 | 0.45 | 0.25 | Probability that a group of individuals in a regency is divided into different ethnic groups | 2010 Indonesian population census | link |
| Long-distance work | 510 | 0.05 | 0.03 | Share of population traveling ¿=1hour to work at the regency level | National Labour Force Survey 2019 | link |
| Public transport work | 510 | 0.06 | 0.05 | Share of population using public transport to work at the regency level | National Labour Force Survey 2019 | link |
| Teleworkability | 510 | 0.22 | 0.05 | Teleworkability by industry at the regency level | National Labour Force Survey 2019 | link |
| Agriculture share | 510 | 0.38 | 0.22 | Share of population working in agriculture | National Labour Force Survey 2019 | link |
| Temperature | 500 | 25.80 | 1.92 | Average monthly temperature as of December 2019 (Celsius degree) | ERA5 reanalysis data | link |
| Rainfall | 500 | 0.21 | 0.08 | Average monthly precipitation as of December 2019 (millimeters) | ERA5 reanalysis data | link |
| Lightning strike | 500 | 0.01 | 0.01 | Average lightning stroke power as of Dec 2019 (5minute resolution) | World Wide Lightning Location Network | link |

# B Additional results

**Table B.1:** 3G Internet and COVID-19 spread - Poisson results

| Dependent variables: Cumulative cases by 10T population, 18 Feb 2023 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Adding covariates | | | | | | Imputation | |
| | | | | | | | | via: | |
| | Raw | +Province FE | +Demographics | +Health facilities | +Econ. status | +Ethinc-comp. | +Labour force | Set 0 | Multiple |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Internet exposure | -1.495 | -1.125 | -0.651 | -0.606 | -0.611 | -0.690 | -0.787 | -0.627 | -0.627 |
| | (0.222) | (0.417) | (0.323) | (0.315) | (0.285) | (0.290) | (0.313) | (0.255) | (0.255) |
| Semi-elasticity (1sd) | -0.46 | -0.37 | -0.24 | -0.22 | -0.23 | -0.25 | -0.28 | -0.23 | -0.23 |
| $N$ | 454 | 454 | 454 | 345 | 323 | 321 | 321 | 454 | 454 |
| Mean dep. | 0.50 | 0.50 | 0.50 | 0.40 | 0.42 | 0.42 | 0.42 | 0.49 | 0.49 |
| SD dep. | 1.45 | 1.45 | 1.45 | 0.87 | 0.89 | 0.89 | 0.89 | 1.43 | 1.43 |
| $pR^2$ | 0.05 | 0.41 | 0.51 | 0.37 | 0.38 | 0.38 | 0.38 | 0.53 | 0.53 |
| Province FE | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Demographics | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Health facilities | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Economic status | | | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Ethnic composition | | | | | | ✓ | ✓ | ✓ | ✓ |
| Labour force | | | | | | | ✓ | ✓ | ✓ |

*Notes*: The Table presents coefficients estimates from equation (1), of regressions of cumulative Covid cases per 10,000 population on the regency level (18 February 2023) and increasing various sets of covariates all measure before the pandemic begins, for a description of the covariates see Appendix Table A.1. At the bottom, we present pseudo (ml) Rsquared ($pR^2$), the mean and standard deviation of the outcome variable, and the semi-elasticity for 1 sd change in internet exposure, that is $(\exp(\tau * sd(\text{Internet exposure})) - 1)$. Column (1) depicts the raw association, Col. 2 adds province (38) fixed effects, three demographic variables (population density, population, the share of the population aged 65 and older, the share of people without any education, the share of people with graduate degrees, log household size, rural area), 4 adds health infrastructure measures (number of clinics, number of hospitals, and log of healthcare spending), 5 adds economic indicators (second order polynomial of nightlight, sub-national human development index - score), 6 ethnicity (ethnic diversity, and polarization), and 7 labor force variables (share of workers able to telework, the share of workers with a long distance (>1hour) commute, and that use public transportation, and the share working in agriculture). The final two columns replace missing values in covariates with 0 and an indicator for missing values (8), and (9) implements the multiple imputation approach by Rubin (1996). All regressions are weighted by population and use robust standard errors. *Source:* COVID-19 data, adjusted for population, was sourced from the Indonesian COVID-19 Task Force, using measurements taken as of February $18^{\text{th}}$, 2023. Mobile internet data was obtained from Collins Bartholomew's GSMA Mobile Coverage Explorer database, own calculations.

**Table B.2:** Robustness tests - Poisson regression

| Dependent variables: Cumulative cases by 10T population, 18 Feb 2023 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Log-cases | Neg. | Alt. 3G measure | | Without | Missings | Impute |
| | Base | OLS | Bin. | Plain | Imputed | weights | Outcomes | All |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Internet exposure (3G) | -0.630<br>(0.255) | -0.365<br>(0.200) | -0.630<br>(0.255) | | | -1.005<br>(0.226) | -0.627<br>(0.255) | -0.627<br>(0.255) |
| Internet exposure (3G-MCE) | | | | -1.090<br>(0.491) | | | | |
| Internet exposure (3G-MCE)-imputed | | | | | -1.126<br>(0.496) | | | |
| Semi-elasticity (1sd) | -0.20 | -0.13 | -0.20 | -0.13 | -0.13 | -0.26 | -0.20 | -0.19 |
| $N$ | 454 | 454 | 454 | 365 | 454 | 454 | 465 | 510 |
| Mean dep. | 0.50 | 0.50 | 0.50 | 0.60 | 0.50 | 0.50 | 0.49 | 0.45 |
| SD dep. | 1.45 | 1.45 | 1.45 | 1.60 | 1.45 | 1.45 | 1.43 | 1.37 |
| $pR^2$ | 0.53 | . | 0.46 | 0.54 | 0.53 | 0.57 | 0.53 | . |
| Province FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Demographics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Health facilities | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Economic status | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Ethnic composition | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Labour force | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

*Notes*: The Table additional robustness checks. Table B.1-Column (3) presents the base model, see notes therein, represented in Column (1). Column (2) presents a OLS regression on the logged outcome measure and the semi-elasticity (1sd) which is calculated via $exp(\tau * sd(internet)) - 1$. Column (3) alternatively presents negative binominal model, (4) uses the alternative exposure measure an (Mobile Coverage Explore - MCE) are sourced directly from the network operators and thus incur gaps in coverage, which we impute with 0 and add missing indicator in (5). Column (6) drops the population-weighting and (7) and (8) impute areas with missing observation in the outcome (all variables, respectively) with 0. *Source:* COVID-19 data, adjusted for population, was sourced from the Indonesian COVID-19 Task Force, using measurements taken as of February 18$^{\text{th}}$, 2023. Mobile internet data was obtained from Collins Bartholomew's GSMA Mobile Coverage Explorer database, own calculations.

**Table B.3:** Robustness tests - Poisson and IV-Poisson regressions

| Dependent variables: Cumulative cases by 10T population, 18 Feb 2023 | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | IV | |
| | | | | | Deviation from | |
| | Main | Internet Exposure Population | Humidity | No Tower | time | area |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Panel A. Poisson* | | | | | | |
| Internet exposure | -0.630 | | -0.663 | -0.592 | | |
| | (0.255) | | (0.255) | (0.252) | | |
| Internet exposure (population) | | -0.526 | | | | |
| | | (0.249) | | | | |
| Semi-elasticity (1sd) | -0.23 | -0.20 | -0.24 | -0.22 | | |
| *Panel B. IV-Poisson* | | | | | | |
| Internet exposure | -1.440 | | -1.372 | -0.901 | -0.812 | -1.516 |
| | (0.394) | | (0.384) | (0.407) | (0.408) | (0.570) |
| Internet exposure (population) | | -1.448 | | | | |
| | | (0.401) | | | | |
| Semi-elasticity (1sd) | -0.45 | -0.45 | -0.44 | -0.31 | -0.29 | -0.47 |
| $N$ | 454 | 454 | 454 | 454 | 454 | 454 |
| Mean dep. | 61.05 | 59.00 | 61.53 | 55.01 | 56.44 | 26.06 |
| SD dep. | 1.45 | 1.45 | 1.45 | 1.45 | 1.45 | 1.45 |
| Province FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Demographics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Health facilities | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Economic status | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Ethnic composition | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Labour force | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Weather controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Number of towers | | | | ✓ | ✓ | ✓ |

*Notes*: The Table presents additional robustness. Col. (1) baseline, Col. (2) internet exposure only over populated areas, Col. (3) rain and temperature interaction, Col. (4) controls for towers, Col. (5) adjusts the IV for deviation from yearly average and (6) from the surrounding areas average. *Source:* COVID-19 data, adjusted for population, was sourced from the Indonesian COVID-19 Task Force, using measurements taken as of February 18$^{\text{th}}$, 2023. Mobile internet data was obtained from Collins Bartholomew's GSMA Mobile Coverage Explorer database, own calculations.

**Table B.4:** Heterogeneity - Poisson-IV results

| Dependent variable: Cumulative cases by 10T population, 18 Feb 2023 | | | | | | |
|---|---|---|---|---|---|---|
| | Some educ. | Grad. educ. | Tele-work potential | Agriculture exposure | Urbanity | Econ. status |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Internet exposure | -1.365 | -1.190 | -0.876 | -1.723 | -1.305 | -1.460 |
| | (0.425) | (0.468) | (0.516) | (0.486) | (0.428) | (0.376) |
| Internet exposure $\times 1[x \geq med_x]$ | -1.006 | -0.308 | -0.740 | 0.520 | -0.326 | 0.238 |
| | (0.441) | (0.207) | (0.408) | (0.282) | (0.280) | (0.513) |
| Province FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Demographics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Health facilities | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Economic status | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Ethnic composition | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Labour force | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Weather controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

*Notes*: See Figure 4 notes. *Source:* COVID-19 data, adjusted for population, was sourced from the Indonesian COVID-19 Task Force, using measurements taken as of February 18$^{\text{th}}$, 2023. Mobile internet data was obtained from Collins Bartholomew's GSMA Mobile Coverage Explorer database, own calculations.