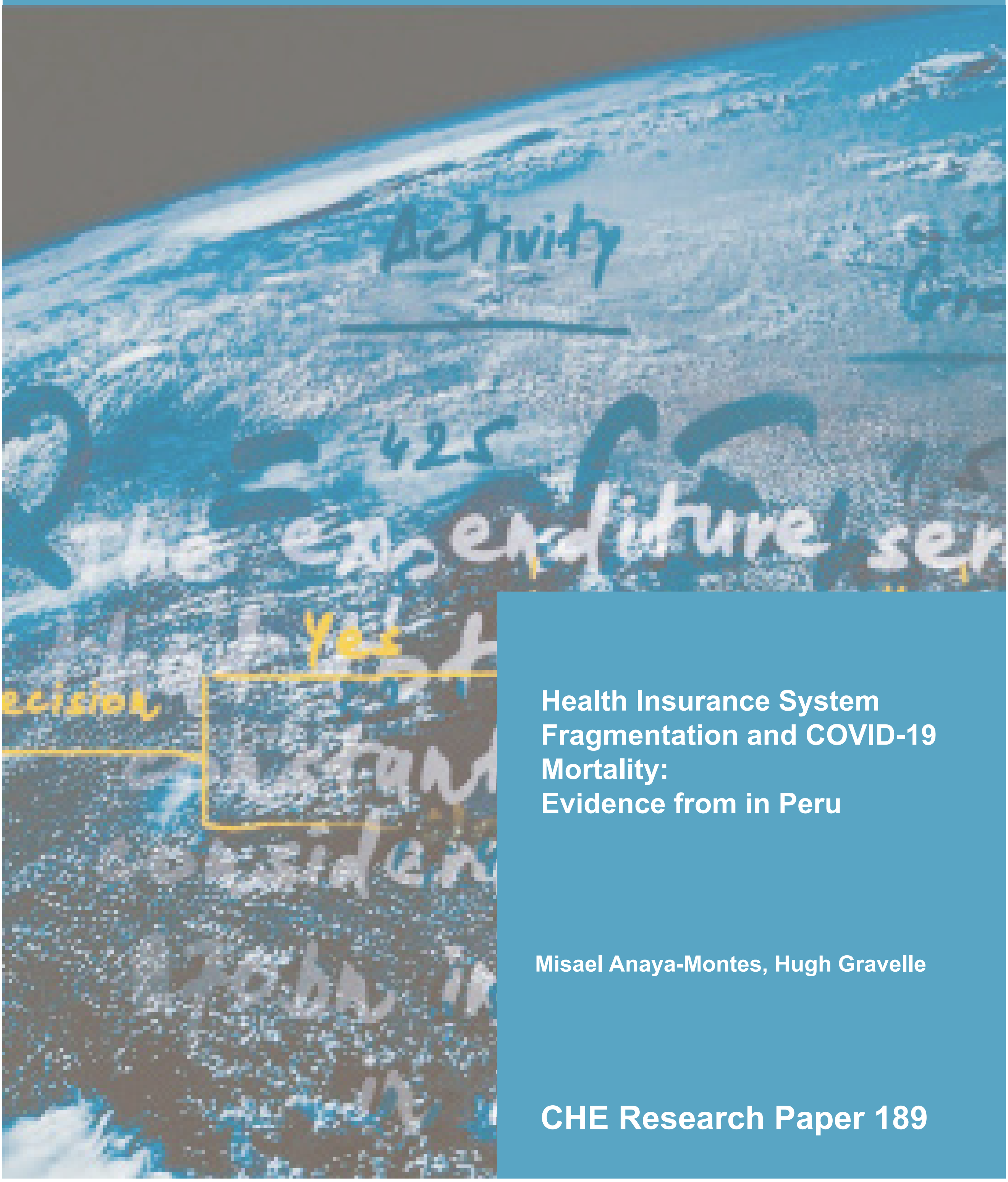
UNIVERSITY of York

RESEARCH

CHE

Centre For Health Economics



Health Insurance System Fragmentation and COVID-19 Mortality: Evidence from in Peru

Misael Anaya-Montes, Hugh Gravelle

CHE Research Paper 189

Health insurance system fragmentation and COVID-19 mortality: evidence from in Peru

[a,b]Misael Anaya-Montes
[b] Hugh Gravelle

[a] Ministry of Economics and Finance, Peru
[b]Centre for Health Economics, University of York, UK

December 2022

Background to series

CHE Discussion Papers (DPs) began publication in 1983 as a means of making current research material more widely available to health economists and other potential users. So as to speed up the dissemination process, papers were originally published by CHE and distributed by post to a worldwide readership.

The CHE Research Paper series takes over that function and provides access to current research output via web-based publication, although hard copy will continue to be available (but subject to charge).

Acknowledgements

We are grateful for comments and suggestions from participants in the Health Econometrics and Data Group seminar in York. Misael Anaya-Montes was employed by the Ministry of Economics and Finance in Peru during early work on the paper. We are grateful to Waldo Mendoza and Betty Sotelo for their encouragement to produce this paper. The views expressed in the paper are those of the authors and not necessarily those of the Ministry of Economics and Finance.

Conflicts of interest

HG has no conflicts of interest to declare. MAM is recognised by the Peruvian State as a victim of the Shining Path terrorist group during 1980-2000, and so has the right to claim dual insurance in SIS and ESSALUD while in formal employment in Peru. No ethical approval was required. All individual level data was pseudonymised.

Further copies

Only the latest electronic copy of our reports should be cited. Copies of this paper are freely available to download from the CHE website www.york.ac.uk/che/publications/. Access to downloaded material is provided on the understanding that it is intended for personal use. Copies of downloaded papers may be distributed to third parties subject to the proviso that the CHE publication source is properly acknowledged and that such distribution is not subject to any payment.

Printed copies are available on request at a charge of £5.00 per copy. Please contact the CHE Publications Office, email che-pub@york.ac.uk, telephone 01904 321405 for further details.

Centre for Health Economics
Alcuin College
University of York
York,
YO10 5DD, UK

www.york.ac.uk/che

Abstract

Background

Peru has a fragmented health insurance system in which most insureds can only access the providers in their insurer's network. The two largest schemes covered 53% and 30% of the population on 5 March 2020. Some individuals have dual insurance: they belong to both schemes and can thereby access a larger set of providers. We investigate whether this greater access to providers for those with dual insurance reduced mortality from COVID-19 between 6 March 2020 (the start of the pandemic in Peru) and 30 June 2021.

Methods

We use data on 24.74M individuals who belonged to one or both schemes for at least 7 consecutive days between 1 January and 5 March 2020. In addition to individual insurance status, age and gender and COVID-19 mortality there is information on the characteristics of the 1874 district in which individuals live: structural quality of, and distances to, the nearest hospitals in each network; occupancy rates for intensive care beds; COVID-19 infections; average deprivation; rurality. To allow for the small probabilities of dual insurance and of mortality from COVID-19 we estimate bivariate probit models. Since it is possible that unobserved factors influence whether an individual has dual insurance and whether they die from COVID-19, we use an instrumental variable: the *difference* in the distances to the nearest hospital in the two insurance schemes. The instrument is a measure of the greater access from having dual insurance and so predicts whether an individual has dual insurance but should not itself directly affect mortality.

Results

5.83% of the sample had dual insurance rather than single insurance for at least one week between 1 January 2020 and 5 March 2020. The effect of dual insurance was to reduce COVID-19 mortality risk by 0.23% compared with the sample mean risk of 0.54%. This implies that the 133,128 COVID-19 deaths in the sample would have been reduced by 56,418 (95%CI: 34,894, 78,069) if all individuals in the sample had dual insurance.

Discussion

The finding that having access to providers in both insurance schemes substantially reduced COVID-19 deaths suggests that policymakers should allow dual insurance for those insureds who potentially qualify for both SIS and ESSALUD under the current rules, and it would be worthwhile experimentally investigating whether removing restrictions on access to providers would be cost-effective in the post-pandemic period for other illnesses.

JEL Nos: I13, I18

Keywords: COVID-19. Mortality risk. Insurance system fragmentation. Provider networks.

1. Introduction

Since the first case reported at the end of 2019 the COVID-19 virus has spread around the world and by 30 June 2021 there had been over 180 million cases and 2 million deaths (WHO, 2022). Mortality from COVID-19 varies with age, gender, ethnicity, morbidity, income, and occupation (Bhaskaran et al., 2021; Egede et al., 2020; Figueroa et al., 2020; Williamson et al., 2020; Yanez et al., 2020). It is also plausible that COVID-19 mortality depends on the resourcing and organisation of the healthcare system, including the way in which the insurance system affects the accessibility of care.

Pre-COVID-19 observational studies either find no or positive effects of health insurance on health outcomes (McWilliams, 2009; Wilper et al., 2009) and experimental studies find that increasing insurance coverage improves health (Finkelstein et al., 2012; Goldin et al., 2021). To date the only study of the implications of insurance coverage for COVID-19 outcomes is Chakrabati et al. (2020) which used a regression discontinuity design to compare counties in US states which had or had not expanded Medicaid coverage under the Affordable Care Act. They found that there were no differences in the logarithms of per capita COVID-19 cases and per capita COVID-19 deaths, though COVID-19 related doctor visits were higher in counties in states which had expanded Medicaid. Rakus and Soni (2021) compared changes in outcomes before and after the start of the COVID-19 pandemic in March 2020 in 35 US states which had expanded Medicaid against 15 which had not. They found that for poor individuals Medicaid expansion improved some, but not all, self-reported health outcomes and health-affecting behaviours. However, none of the outcomes were specific to COVID-19.

Insurance schemes with their own networks of providers typically restrict access to these networks to the individuals they insure. Most previous studies of limited provider networks are based on experience in the US where insurers compete and insureds who chose narrower provider networks can face lower premia (Atwood and Sasso, 2016; Gruber and McKnight, 2016; Polsky et al., 2016). There is little evidence on the effect on health of insurers restricting access to their provider networks. Gruber and McKnight (2016) found no statistically significant effect of choosing a narrow network on health outcomes such as mortality and avoidable hospitalizations, though confidence intervals were very wide. In the English National Health Service, where hospital care is tax funded, the relaxation of constraints on patient choice of hospital for coronary artery bypass grafts led to a reduction in post-operative mortality due to some patients switching to higher quality providers (Gaynor et al. 2016).

Peru has had one of the highest COVID-19 mortality rates in the world (Dyer, 2021; Karlinsky & Kobak, 2021). It also has a fragmented health insurance system in which the two largest schemes, which together insure 83% of the population, only cover care in their separate networks of providers. In this paper we examine whether this fragmentation contributed to COVID-19 mortality.[1]

We use data on 24.74M individuals who were in one or both schemes before the start of the pandemic in Peru (6 March 2020). Of these, 94.17% were in only one of the schemes and 5.83% were in both schemes and so could access providers in both networks. We compare COVID-19 mortality risk of those in only one scheme and those in both schemes (dual insurance).

[1] Soto-Cabezas et al. (2022) report that the COVID-19 age adjusted infection rate was much higher amongst the Amazonian indigenous population in Peru than in the general population but COVID-19 mortality rate was much lower, but did not examine the role of insurance. Two previous studies of health insurance in Peru have examined the effect of changes in each of the health insurance schemes on healthcare use. Neelsen and O'Donnell (2017) found that extension of the tax based insurance scheme (SIS) led to increased receipt of ambulatory care and medication compared to individuals covered by the employment based health insurance (ESSALUD). Bernal et al. (2017) found that individuals just below the SIS income eligibility threshold received more healthcare than those just above it.

We estimate individual level biprobit models to allow for the small probabilities of dual insurance and of COVID-19 mortality in our sample. It is possible that unobserved variables affect both whether an individual has dual insurance and whether they die from COVID-19. We allow for such possible endogeneity bias by using an instrumental variable (IV): the *difference* in the distance to the nearest provider in each of the two networks. The IV is a measure of the improvement in access to care from being in both insurance schemes rather than only one scheme and so is a good predictor of having dual insurance. We argue that, since we control for individual's characteristics and the characteristics of their district of residence (including the structural quality of, and distances to, providers in both networks, and COVID-19 infection rates), the IV satisfies the exclusion requirement that it is not, conditional on the covariates, correlated with COVID-19 mortality risk.

Overall COVID-19 mortality risk between 6 March 2020 and 30 June 2021 was 0.54%. We estimate that the average treatment effect (ATE) of dual insurance was to reduce mortality risk by 0.23% and the average effect on the treated (those with dual insurance) (AETT) was to reduce it by 0.17%.

The next section outlines the characteristics of the Peruvian healthcare system and describes the datasets. Section 3 sets out the empirical model, the rationales for using bivariate probit as the preferred estimation method, and discusses the instrument variable. Section 4 has the results for the biprobit models of COVID-19 mortality risk estimated on 24.74M individuals. We also compare the biprobit model results with those from models which make different assumptions about functional form, the endogeneity of dual insurance, and definition of dual insurance. Section 5 concludes with a discussion of the policy implications of our results.

2. Data

The first case of COVID-19 in Peru was diagnosed on 6th March 2020. On 15th March the government declared a national health emergency, decreed a national lock-down and closed all the frontiers and airports. The government also set up new datasets, linked at patient level, with information on mortality, some types of healthcare use, and insurance status for patients (Health Ministry of Peru, 2021).

2.1 Health insurance

Information on individual insurance status is from the SUSALUD-RAUS database (Superintendencia Nacional de Salud, 2021) which combines data from all insurance schemes. It provides a day-by-day history of insurance status for all citizens from their first insurance enrolment in any health insurance scheme.

Peru has two main healthcare insurance schemes. Seguro Integral de Salud (SIS) mainly covers people with income below a threshold. SIS provides insurance to households – defined as a set of people who live and eat together, including grandparents and siblings regardless of their age. It is financed from general taxation and covered 53% of the 33M population on 5 March 2020. Members of SIS are treated without charge in publicly owned tax-financed providers. These public providers also treat non-SIS members, including the 15% of the population who were uninsured at the start of the pandemic, at subsidised prices. At the start of the pandemic SIS coverage was extended to include uninsured individuals who had been diagnosed with COVID-19, and from 31 July 2021 (after the end of our analysis period) SIS was extended to all the uninsured.

The second largest insurance scheme, covering 30% of the population on 5 March 2020, is El Seguro Social de Salud (ESSALUD) which covers workers in formal employment and their legal dependents (partner and non-adult children). It is financed by compulsory contributions (9% of salary) and provides care free at the point of use.[2] During the pandemic ESSALUD continued to cover those who were temporarily laid off work.

Some individuals are members of *both* SIS and ESSALUD and so entitled to use providers in both schemes' networks at no charge. Members of the household of a SIS member can belong to both schemes if they, their partner (and for children, their parent), are in formal employment. Workers in formal employment, and hence in ESSALUD, but with an income below the SIS maximum threshold can also join SIS. Even if their income subsequently exceeds the SIS threshold they can keep their SIS membership for up to three years. Workers initially in ESSALUD can also claim SIS membership if they are recently unemployed.

An individual initially in SIS is enrolled in ESSALUD when they start formal employment. When the national tax authority deducts 9% of their salary and transfers it to ESSALUD the worker is affiliated automatically in ESSALUD and their membership of ESSALUD is then recorded in the SUSALUD-RAUS insurance database. It may take several weeks before a new worker's details are shared with SIS. If the worker has an income above the SIS income eligibility threshold they will then be disaffiliated from SIS, but until this happens they will be members of both ESSALUD and SIS. When an employee previously in SIS is enrolled in ESSALUD their family (partner and under age children) also have the right to be in ESSALUD.

[2] Members of the police and armed forces (1.7% of the population) are insured in schemes similar to ESSALUD and are treated in networks of providers closed to the rest of the population. Providers belonging to private insurers charge copayments to their insured populations (about 1.5% of the population) and provide care at full price to other patients.

There are also some types of individuals, such as victims of the Shining Path terrorist group between 1980 and 2000, who have permanent membership of SIS and should not be disaffiliated. They will have dual insurance if they are also in ESSALUD by virtue of their formal employment or membership of the family of an ESSALUD member.

2.2 COVID-19 mortality

Information on COVID-19 mortality is from the NOTI-SINADEF COVID-19 mortality database which was established in response to the pandemic. The initial mortality classification system used in Peru until 31 May 2021 led to a substantial undercount of COVID-19 deaths. With the initial system the ratio of excess deaths (all deaths minus the number predicted from all-cause mortality in the five previous pre-pandemic years) to reported COVID-19 deaths was 2.7, with a correlation between the weekly figures of 0.87 (Karlinsky and Kobak, 2021). Revised classification criteria[3] (Health Ministry of Peru, 2021a) were introduced in May 2021 and applied retrospectively to mortality from 1 March 2020 to 31 May 2021. This led to the number of deaths attributed to COVID-19 over this period more than doubling, the ratio of excess deaths to COVID-19 deaths increasing to 1.0 and the weekly correlation between excess deaths and COVID-19 deaths increasing to 1.0 (Karlinsky and Kobak, 2021). This suggests that the COVID-19 mortality data based on the revised definition that we use for 6 March 2020 to 30 June 2021 are reasonably accurate.

We use the insurance and mortality databases to construct the sample for analysis of the effect of dual insurance on COVID-19 mortality. In addition to daily information on insurance status (in SIS, in ESSALUD, or in both) and mortality, the databases record gender, age and district of residence.[4] We use the district of residence to attach measures of district poverty,[5] population, area, district geographical type (capital city, coast, jungle and mountain) and the straight line distance from district centroids to hospitals in the SIS and ESSALUD networks.

After dropping individuals with missing data or who were not in SIS or ESSALUD for at least seven consecutive days during 2019 and for seven consecutive days between 1 January 2020 and 5 March 2020, our sample for the analysis of the effect of dual insurance on mortality from COVID-19 is 24,739,933, of whom 0.54% died from COVID-19 between 6 March 2030 and 30 June 2021.[6]

[3] Details are in Appendix 1

[4] Districts are the lowest administrative unit. The 1874 districts have an average population of 17,600 and a radius of 10.8km. They are grouped within 196 provinces, which in turn are grouped in 25 departments.

[5] Districts are classified by the National Statistics Institution (INEI) into 21 poverty categories, ordered from poorest to least poor. For estimation purposes we collapse these into three categories, ordered from poorest to least poor.

[6] Appendix Table A1 has details of data cleaning.

3. Methods

3.1 Endogeneity of dual insurance

Whether an individual dies from COVID-19 and whether they have dual insurance may both be affected by the same unobservable factors, thereby potentially biasing the estimated effect of dual insurance on COVID-19 mortality risk. Less healthy individuals may be at greater risk of death from COVID-19 but more likely to have dual insurance if eligible because they believe they are more likely to benefit from having better access to providers when ill. This will lead to an underestimate of the reduction in COVID-19 mortality risk from having dual insurance. Conversely more health conscious individuals may take better care of their health, thereby reducing their COVID-19 mortality risk, and be more likely to want dual insurance. This will lead to an overestimate of the beneficial effect of dual insurance. There may also be endogeneity if dual insurance status is affected by decisions by insurance system officials whose interpretation of eligibility rules may depend on their perception of the benefits or cost implications of dual insurance.

We allow for possible endogeneity bias by using an instrumental variable (IV) which predicts having dual insurance and is not correlated with mortality except via its correlation with dual insurance. The IV is the *difference* in the distance from an individual to the nearest provider in each of the SIS and ESSALUD networks. The differential distance IV draws on a literature dating back to McClellan et al. (1994) and uses the fact that distance is a key determinant of patient use of providers (Gutacker et al., 2016).

SIS and ESSALUD have different numbers of insureds, and different numbers of providers with different levels of equipment and facilities. For those eligible for both insurance schemes, the gain from having dual insurance rather than single insurance will depend on the *differences* between the SIS and ESSALUD providers they could access as members of only one scheme. Individual *i* and resident in district *r* who is initially only in SIS but is also eligible for ESSALUD will be more likely to also be in ESSALUD the greater the distance (d_{ir}^{SIS}) to the nearest SIS provider compared with the distance $d_{ir}^{ESS}$to the nearest ESSALUD provider: ($d_{ir}^{SIS} - d_{ir}^{ESS}$). Conversely, those initially in ESSALUD but eligible for SIS will be more likely to get dual insurance the greater is ($d_{ir}^{ESS} - d_{ir}^{SIS}$). Our distance based IV for dual insurance is

$$z_{ir} = S_{ir}^{first}\left(d_{ir}^{SIS} - d_{ir}^{ESS}\right) + \left(1 - S_{ir}^{first}\right) + \ \left(d_{ir}^{ESS} - d_{ir}^{SIS}\right) = (2S_{ir}^{first} - 1)(d_{ir}^{SIS} - d_{ir}^{ESS}) \quad (1)$$

where S_{ir}^{first} is an indicator for the individual initially being in SIS.[7] The larger the value of z_{ir} the greater the reduction in distance from having dual insurance rather than single insurance, and hence the more likely that an individual will opt to have dual insurance if they are eligible for both schemes.

We measure the straight line distances to the nearest provider in each insurance network. We have the address for 62% of individuals and use the district centroid as the location for the remainder. One in 12 districts has a hospital available to insureds in SIS, one in 29 has an ESSALUD hospital and one in 44 have both SIS and ESSALUD providers. Although the district centroid is an approximate measure of an individual's location, the fact that only a small proportion (2%) of districts have hospitals in both networks means that for most individuals the distances from their district centroid to the nearest SIS or ESSALUD hospital is likely to be a reasonable measure of travel time to that hospital. We also control for the rurality of the district and its geographical characteristics (jungle,

[7] We use individual insurance history from the insurance database SUSALUD-RAUS to determine which scheme the individual was in first.

mountain, coastal region, capital city) which may be correlated with transport facilities and ease of access.

We argue that the IV satisfies the exclusion requirement that it does not influence mortality except via its effect on whether the individual has dual insurance. If an individual infected with COVID-19 seeks hospital treatment for the disease, the distance to the provider they use may affect their mortality risk. However, *differences* in distances to providers in the two networks will not be correlated with the individual's mortality risk conditional on the individual and district covariates in the mortality model. Although we do not observe in which, if any, hospital the infected individual is treated, the covariates in the mortality model include characteristics of local providers: quintiles of the average distance to the nearest SIS and ESSALUD provider, the average structural quality of the nearest SIS and ESSALUD providers, and the proportion of the local population infected with COVID-19.

3.2 Definition of dual insurance

Insurance status can vary over time as individuals claim dual insurance and insurance administrators attempt to remove the ineligible.[8] **Figure 1** plots the daily number of those with dual insurance over the period 1 January 2016 to 30 June 2021.

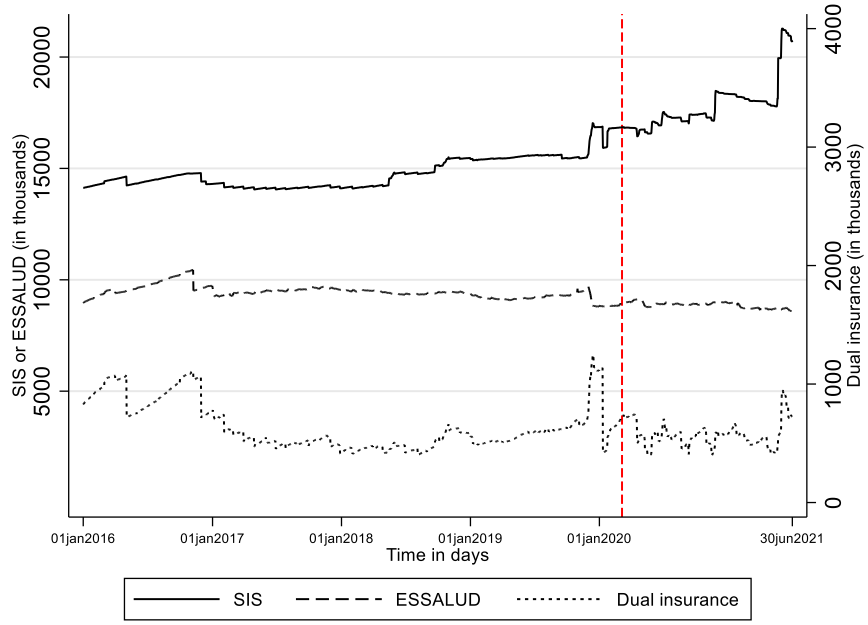


Figure 1: Daily number of insureds by ESSALUD, SIS and dual insurance status 1st January 2016 - 30 June 2021
Note. Figure plots the daily number in SIS, ESSALUD, and in both insurance schemes.

Using information on insurance status *after* the start of the pandemic creates the risk of reverse causality. Becoming infected with COVID-19 (and hence at risk of death from COVID-19) may make eligible individuals more likely to claim dual insurance if they do not already have it or to contest attempts to remove them from one of the schemes if they do. We therefore use a definition of dual insurance based on insurance status *before* the start of the pandemic. We argue that having dual insurance for at least seven consecutive days between 1 January 2020 and 5 March 2020 is a good predictor of being able to get dual insurance if the individual is infected with COVID-19.

[8] **Appendix Figures A1, A2 and A3** plot the daily number of additions and removals from the set of those with dual insurance, cumulative affiliations and disaffiliations to dual insurance before and after the start of pandemic, and the evolution of insurance status, SIS, ESSALUD and dual insurance 100 days up to date of death, respectively.

A narrow definition of dual insurance (for example having dual insurance for *all* days in a particular pre-pandemic period) risks classifying some people who had dual insurance for only part of this period but who were able to get dual insurance during the pandemic period as not having dual insurance. Similarly, defining dual insurance as having dual insurance on a particular day also risks classifying those who had dual insurance on other days and were able to get dual insurance during the pandemic as not having dual insurance. Conversely, a wide definition (for example, having dual insurance on *any* day within a long pre-pandemic period) risks classifying some individuals who could not get dual insurance during the pandemic period as having dual insurance. Both types of misclassification error will lead to an underestimate of the effect of dual insurance on mortality risk (Lewbel, 2007). If some of those classified as not having dual insurance pre-pandemic did get it during the pandemic, this will reduce the average mortality risk for those classified as not having dual insurance compared to those classified as having dual insurance. If some of those classified as having dual insurance pre-pandemic did not have it during the pandemic, this will increase the average mortality risk for those classified as having dual insurance compared to those classified as not having it. We examine the sensitivity of estimates of the effect of dual insurance on mortality risk to using a variety of different pre-pandemic definitions of dual insurance.

3.3 Model specification

Two stage least squares (2SLS) is simple and provides a linear approximation to the conditional expectation outcome function (Angrist and Pischke, 2009). When the IV for dual insurance is continuous the 2SLS coefficient on the dual insurance indicator in the mortality model is an average of local average treatment effects (LATEs) for individuals whose dual insurance status is affected by the instrument. 2SLS can yield estimates of the average effect of treatment (ATE) and the average effect of treatment on the treated (AETT) but these require stronger assumptions (Andresen and Huber, 2018; Brave and Walstrum, 2014; Cornelissen et al., 2016). However, our outcome (mortality) and potentially endogenous treatment (dual insurance) are binary variables with small probabilities and, as we show, this can lead to linear models producing predicted probabilities for individuals outside the [0,1] interval. It is also possible for linear models to produce estimates of average treatment effects (ATE) with the wrong sign (Baum, et al., 2012; Lewbel et al., 2012).

The obvious alternative to 2SLS is the non-linear recursive bivariate probit (biprobit) specification (Greene, 2020; Maddala, 1983, p.123). Biprobit requires stronger assumptions than 2SLS but estimates the average effect of treatment (ATE) and the average effect of treatment on the treated (AETT). A simulation study by Chiburis et al. (2011) suggests that biprobit is more efficient than 2SLS, especially in models with covariates and when, as in our case, the probability of treatment (dual insurance) is low. Other simulations also support the use of biprobit rather than 2SLS (Bhattacharya et al., 2006; Denzer, 2019). Although Chiburis et al. (2011) warn that biprobit can be biased if the error distribution is misspecified, Li et al. (2019) suggest that the use of instrumental variables increases the robustness of estimates of the ATE in such cases.

We estimate recursive biprobit models as our preferred specification. Although it is possible to rely on the non-linear biprobit functional form for identification (Wilde, 2000) our inclusion of the distance difference IV may reduce problems due to misspecification (Li et al., 2019). We also report estimates of treatment effects from biprobit with no instrument, 2SLS, and single equation linear, probit, and logit mortality models.

Our biprobit specification is

$$D_{ir}^{*} = \lambda_0 + z_{ir}\lambda_1 + x_{ir}'\lambda_2 + x_r'\lambda_3 + \upsilon_{ir},\, D_{irt} = 1(D_{irt}^{*} > 0) \qquad (2)$$

$$M_{ir}^{*} = \beta_0 + \beta_1 D_{ir} + x_{ir}'\beta_2 + x_r'\beta_3 + \varepsilon_{ir},\, M_{irt} = 1(M_{irt}^{*} > 0) \qquad (3)$$

D_{ir} is an indicator for individual *i* resident in district *r* having dual insurance (belonging to SIS and ESSALUD) for at least seven consecutive days between 1 January 2020 and 5 March 2020. M_{ir} is an indicator for the individual dying from COVID-19 between 6 March 2020 and 30 June 2021. υ and ε have a joint standard normal distribution with correlation ρ.

x_{ir} is vector of characteristics of individual *i*: gender and age (≤ 19, 20-39, 40-59, 60-79, ≥ 80). x_r is a vector of the characteristics of the district of residence of the individual: indicators for the 25 departments in which the district is located; the geography of the region (capital city, coast, jungle, mountain) in which the district is located; rurality (indicator based on the number of households in the district); and three district poverty categories.[9] It also includes controls for local hospital facilities: indicators for quintiles of the average distance from the district centroid to the nearest SIS and ESSALUD provider, and quintiles of the average structural quality[10] of the nearest SIS and ESSALUD providers. We also include the proportion of the district population recorded as infected with COVID-19 between 6 March 2020 and 30 June 2021 to control for local unobserved factors affecting the risk of COVID-19 infection and thus COVID-19 mortality risk.

We use Stata 17 and the user written *cmp* package (Roodman, 2011) to estimate the average treatment effect (ATE) and the average effect of treatment on the treated (AETT) of having dual rather than single insurance on COVID-19 mortality risk.

There is disagreement about whether and how standard errors should be clustered (Abadie et al., 2020; MacKinnon and Webb, 2019). Abadie et al. (2020) argue that clustering is not necessary when the estimation sample is not a draw from a larger population. MacKinnon and Webb (2019) suggest that one should think of observations as draws from a meta sample of individuals and so clustering should be allowed for. Our sample is all individuals in SIS, ESSALUD, or both at the beginning of the pandemic. We use individual level measures of outcome and treatment and control for individual characteristics such as age and gender. However, some of our explanatories are measured at district level and the instrument for dual insurance is based on the distance from the district centroid for the 38% of our sample for whom we do not have an exact address. We therefore use the more conservative approach of clustering robust standard errors at district level.

Our sample is very large, creating the risk that applying conventional levels of statistical significance will result in rejection of the null hypothesis of no effect of dual insurance even if the estimated effect is tiny and of trivial policy relevance (Deaton, 2018; Leamer, 1978). As we will see, the estimated ATE of dual insurance is large relative to the sample mean COVID-19 mortality risk as well as being very precisely estimated. We also find that the ATE is highly statistically significant after adjusting the required probability levels downward to allow for sample size as suggested by Giles (2019).

[9] Note that with 24.7M observations and at most 25 dummy variables for each district characteristic, the estimated coefficients on the characteristics dummies will not suffer from the incidental parameters problem (Lancaster, 2000).

[10] We compute the average level by interpreting the numerical label (1 to 3) for the structural quality level category of each hospital as a real number.

4. Results

4.1 Summary statistics

Figure 2 plots COVID-19 mortality rate for groups of individuals defined by insurance status at date of death: insured with SIS or ESSALUD but not both, insured with both SIS and ESALUD, uninsured (no insurance of any kind: SIS, ESSALUD, private, police, or armed services) at the date of death. The COVID-19 mortality rate for the uninsured was much higher than for those with insurance during the first wave of the pandemic, despite membership of SIS being extended to the uninsured when they were diagnosed with COVID-19. During the second wave mortality amongst the uninsured was slightly below average, possibly because of mortality displacement, with the more susceptible being most likely to die in the first pandemic wave.[11]

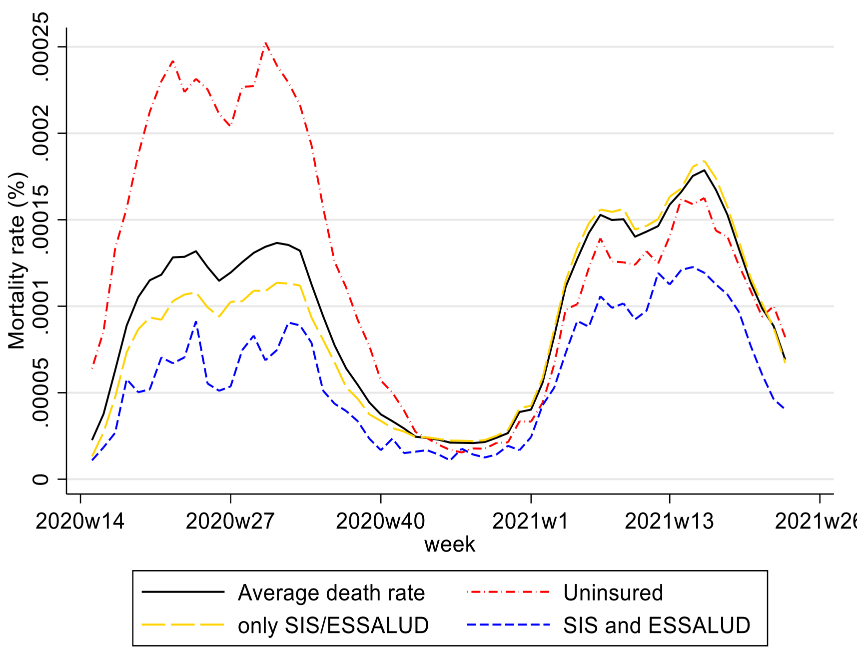


Figure 2: COVID-19 mortality rates by insurance status

Note: Insurance status at 31st December 2019, mortality by Covid-19 followed up 30th June 2020. Daily information averaged by week.

Our research question is whether having dual insurance (being in both SIS and ESSALUD) and thus having access to providers in both networks reduced COVID-19 mortality risk compared to being in only one of the schemes and being able to access providers in only one of the networks. In Figure 2 the mortality rate for those with dual insurance (blue line) is consistently below that of those with single insurance (yellow line). Figure 2 suggests that dual insurance is protective against COVID-19, but it is possible that the lower mortality risk of those with dual insurance is due to them being more likely to have characteristics other than insurance status which protect against COVID-19 mortality.

[11] Figure 2 also shows the COVID-19 mortality rate for individuals who were uninsured (not covered by SIS, ESSALUD, police, military, or private insurance). At the start of the pandemic SIS was extended to the uninsured when they were diagnosed with COVID-19 and on 28 July 2021, the bicentenary of the Peruvian republic, membership of SIS was extended to all uninsured individuals (PCM, 2021).

Table 1. Summary statistics: patient characteristics, dual insurance, COVID-19 mortality

		Proportion of sample	Proportion with dual insurance	Mortality rate	Mortality rate by insurance status		Difference in mortality rate by insurance status
					Single insurance	Dual insurance	
Full sample							
All			0.0583	0.0054	0.0055	0.0037	0.00180***
Gender	Female	0.512	0.0570	0.0038	0.0038	0.0027	0.00111***
	Male	0.488	0.0597	0.0071	0.0072	0.0047	0.00259***
Age	0-19	0.335	0.0619	0.0001	0.0001	0.0001	-0.00001
	20-39	0.296	0.0702	0.0007	0.0007	0.0008	-0.00013***
	40-59	0.229	0.0559	0.0054	0.0054	0.0059	-0.00055***
	60-79	0.113	0.0331	0.0238	0.0236	0.0290	-0.00545***
	> 79	0.026	0.0088	0.0463	0.0463	0.0522	-0.00589*
ESSALUD		0.337	0.1036	0.0087	0.0093	0.0044	0.00487***
SIS		0.663	0.0354	0.0037	0.0037	0.0026	0.00107***
Region or residence	Lima	0.312	0.0802	0.0080	0.0083	0.0043	0.00405***
	Coast	0.249	0.0643	0.0061	0.0062	0.0040	0.00226***
	Jungle	0.146	0.0354	0.0028	0.0028	0.0025	0.00030*
	Mountains	0.293	0.0414	0.0033	0.0033	0.0026	0.00075***
Rurality of district of residence	Urban area	0.790	0.0661	0.0064	0.0066	0.0039	0.00265***
	Rural area	0.210	0.0290	0.0015	0.0015	0.0015	0.00002
Deprivation (more to least)	Group 1	0.366	0.0363	0.0022	0.0022	0.0020	0.00030***
	Group 2	0.426	0.0656	0.0064	0.0066	0.0038	0.00279***
	Group 3	0.208	0.0823	0.0088	0.0092	0.0048	0.00434***
Observations		24,739,933	1,443,205	133,128			

Note. Individuals alive at 5 March 2020 and with at least consecutive 7 days in ESSALUD or SIS between 1 January 2020 and 5 March 2020. Mortality followed from 6 March 2020 to 30 June 2021. ESSALUD (SIS): only have ESSALUD (SIS) or have dual insurance with ESSALUD (SIS) as first insurance. *, **, ***: p < 0.05, < 0.01, <0.001.

Table 1 has summary statistics for our estimation sample of 24,739,933 individuals who were in SIS, ESSALUD, or both for at least seven consecutive days in 2019 and at least seven consecutive days between 1 January 2020 and 5 March 2020, with COVID-19 mortality recorded up to 30 June 2021. We define insurance status by whether the individual had dual insurance for at least seven consecutive days between 1 January and 5 March 2020. 0.55% of the sample (133,128 individuals) died from COVID-19 between 6 March 2020 and 30 June 2021. 5.8% of the sample had dual insurance and they had a lower mortality rate (0.37%) than those in only one of the insurance schemes (0.55%). Males had a much higher mortality rate than females (0.71% vs 0.38%) and were slightly more likely to have dual insurance than females (5.97% vs 5.70%). Dual insurance was associated with a bigger reduction in mortality risk for men than for women (0.26% vs 0.11%). Mortality risk was greater (0.87%) for those initially in ESSALUD than for those initially in SIS (0.37%), possibly because those in ESSALUD are older and located mainly in big cities where the pandemic spread more rapidly.

Mortality risk increased with age whilst dual insurance decreased. The age specific difference in mortality risk provides a nice example of Simpson's paradox[12] in that, although mortality risk was

[12] https://en.wikipedia.org/wiki/Simpson%27s_paradox . As we will see, results from the estimated models which take account of a range of covariates correlated with age, dual insurance, and mortality, indicate that dual insurance *reduces* mortality risk for each age group.

smaller overall for those with dual insurance, mortality risk was higher in *every* age group for those with dual insurance.

Geography is strongly associated with dual insurance and with mortality, and individuals in districts with higher dual insurance (such as those in the capital city or an urban area) also have higher mortality. Both dual insurance and mortality rates are *inversely* related to deprivation, with those in the least deprived districts (Group 3) being more than twice as likely to have dual insurance as those in the most deprived districts (Group 1) and nearly four times as likely to die from COVID-19. This startling reversal of the usual relationship between deprivation and health is likely to be due to the lower risk of COVID-19 infection in rural areas with lower population density and greater deprivation.

Table 2. SIS and ESSALUD provider networks

	ESSALUD	SIS	SIS and ESSALUD
Level 1	39	96	135
Level 2	21	46	67
Level 3	11	23	34
Total	71	165	236
Insureds per provider	158,286	119,886	131,439
Total hospital beds	13,670	16,689	30,359
Total ICU beds	171	238	409
Mean distance to level 1 providers	36.58	35.9	25.18
Mean distance to level 2 providers	76.74	45.49	34.23
Mean distance to level 3 providers	63.92	115.31	89.13

Note. Level 1 hospitals: have 24-hour emergency service and operating theatre; Level 2 as level 1 plus general Intensive Care Unit; Level 3 as level 2 plus Specialist Intensive Care Unit and research capacity. Insureds per SIS (ESSALUD) provider: individuals insured with SIS (ESSALUD) or both at 31 December 2019/total number of SIS (ESSALUD) providers. Mean distance: mean straight line distance (kms) from district centroid weighted by individuals in the district who were in SIS, ESSALUD or both at 31 December 2019. Number of hospital and ICU beds are at 30 April 2020.

Table 2 compares providers available in the SIS and ESSALUD insurance schemes. SIS has more hospitals than ESSALUD, of all levels of structural quality, with 165 in total versus 71. Level 1 hospitals have 24-hour emergency services and operating theatres. Level 2 hospitals additionally have general intensive care units, and level 3 hospitals additionally have specialist intensive care units and research capacity. ESSALUD has fewer general and intensive care beds.

ESSALUD has fewer insureds who are more geographically concentrated because they are more likely to live in large urban areas where formal employment is concentrated. Level 3 hospitals also tend to be located in these areas. Hence, although ESSALUD has fewer hospitals per insured, the average distance of ESSALUD insureds to level 3 ESSALUD hospitals is less than the average distance of SIS insureds to level 3 SIS hospitals. Level 1 providers in both networks are more dispersed and so average distances to them are similar for SIS and ESSALUD insureds. Distances to level 2 providers in their network are much greater for those in ESSALUD.

Table 3 has summary statistics for the distance difference IV we use to predict whether an individual has dual insurance. The distance IV measures the *reduction* in distance to the nearest provider an individual initially insured with one scheme would have if they were also a member of the other scheme. We expect that individuals in only one scheme are more likely to have a smaller improvement in access from dual insurance than those with dual insurance. This is so, on average, for individuals in ESSALUD. Those who have dual insurance and were initially in ESSALUD have a larger distance reduction (7.096km) from belonging to both schemes than those who are only in ESSALUD (4.563km). Because there are more SIS providers than ESSALUD providers the average distance to the nearest SIS provider is *smaller* than the distance to the nearest ESSALUD provider. However, it still the case that individuals with dual insurance who were initially in SIS but joined ESSALUD as well have a greater distance reduction from joining ESSALUD as well as SIS than the potential reduction for those who remain in SIS: −14.043 > −17.183. This suggests that for those initially in ESSALUD there is, on average, a bigger distance reduction from dual insurance than for those initially in SIS. However, these are averages, so that some individuals with dual insurance and initially in SIS will have a positive reduction if they could also access providers in ESSALUD. We get similar patterns of results with a binary indicator version of the IV: $1(2S_{irt}^{first} - 1)(d_{ir}^{SIS} - d_{ir}^{ESS})$.

Table 3. Summary statistics: instrumental variables

	Estimation sample	With only SIS	With dual insurance, initially in SIS	With only ESSALUD	With dual insurance, initially in ESSALUD
Observations	24,739,933	15,829,375	580,188	7,467,353	863,017
Proportion		0.640	0.023	0.302	0.035
Distance difference IV (km)					
Mean	-9.70	-17.18	-14.04	4.56	7.10
SD	36.49	37.53	33.44	30.08	26.07
Min	-208.22	-208.22	-208.22	-208.22	-80.65
Max	208.22	208.22	80.65	208.22	208.22
Binary distance difference IV					
Mean	0.384	0.274	0.283	0.594	0.655

Note. The distance difference IV is $(2S_{irt}^{first} - 1)(d_{ir}^{SIS} - d_{ir}^{ESS})$ where S_{irt}^{first} is an indicator for a patient in the estimation sample initially being only in SIS and d_{ir}^{SIS}, d_{ir}^{ESS} are distances from the district centroid to the nearest SIS (ESSALUD) provider of any level. The binary version is $\mathbf{1}(2S_{irt}^{first} - 1)(d_{ir}^{SIS} - d_{ir}^{ESS})$. The statistics for IVs are weighted by the number of the relevant type of patients within the district who are in the estimation sample.

4.2 Model results

Table 4 compares the estimates of the effect of dual insurance on COVID-19 CFR from six models which differ in whether they attempt to allow for endogeneity of dual insurance, for the binary nature of dual insurance and mortality, and in assumptions about functional form.

Table 4. Estimated effects of dual insurance on COVID-19 mortality rate

	(1)	(2)	(3)	(4)	(5)	(6)
	LPM	Logit	Probit	2SLS	Biprobit	Biprobit
Instrument	No IV	No IV	No IV	Distance difference	No IV	Distance difference
ATE	-0.00044***	0.00001	-0.00002	0.03505***	-0.00445***	-0.00227***
	(0.0001)	(0.0001)	(0.0001)	(0.0052)	(0.0003)	(0.0005)

Note. 24,739,933 observations in all models. Dual insurance: membership of SIS and ESSALUD for at least 7 days between 1 January 2020 and 5 March 2020. All models for mortality risk and dual insurance contain the full set of explanatory variables. F stat on IV equal to 257. Robust standard errors clustered at district level. *, **, ***: p < 0.05, < 0.01, <0.001.

The single equation linear probability, logit, and probit mortality models reported in columns (1) to (3) do not take account of the potential endogeneity of dual insurance. The single equation linear probability model (LPM) (column (1)) suggests that having dual insurance reduces mortality risk by 0.043%, compared to the sample mortality risk of 0.538%. Although the average estimated mortality risk is very similar to the sample mean risk, the LPM produces *negative* estimates of mortality risk for over 31% of the sample (**Appendix Table A3**).

The logit and probit single equation mortality models cannot produce impossible estimates of individual mortality risk but yield very small (0.001% and −0.001%) and statistically insignificant estimates of the average effect of dual insurance on mortality risk.

The linear 2SLS model (column (4)) allows for the potential endogeneity of dual insurance by using the distance difference IV in the first stage model for dual insurance. The IV is a strong predictor of dual insurance (F statistic 258), comfortably above the conventional threshold of 10, so avoiding the weak instrument problem in 2SLS (Stock and Yogo, 2005). The estimated ATE of dual insurance is highly statistically significant but positive, suggesting that the average effect of dual insurance is to *increase* mortality risk by 3.432%. This seems highly implausible given that the sample mortality risk is 0.538%. The average estimated mortality risk is, as with the linear model of column (1), very similar to sample mean risk. But the 2SLS mortality model produces nonsensical negative estimated mortality probabilities for 46% of the sample.

Columns (5) and (6) report the ATEs from biprobit models.[13] In column (5) the biprobit model has no IVs and identification is by functional form. In column (6) the biprobit model includes the distance difference IV. The ATEs are negative and precisely estimated in both biprobit models. Monfardini and Radice (2008) suggest using a likelihood ratio test on the correlation of residuals as a test of exogeneity in recursive biprobit models. A likelihood ratio test on the correlation of residuals (Monfardini and Radice, 2008) from column (5) biprobit model with no IV suggests we reject the null hypothesis of exogeneity of dual insurance ($\chi^2 = 1089.64$, $p < 0.01$) and use the ATE of −0.0227% from the biprobit model in column (6) with the distance IV as our preferred estimate.[14]

4.3 Robustness checks

In **Table 5** we investigate the sensitivity of the biprobit model results to the definition of dual insurance. Column (1) reproduces the results from our preferred definition of having dual insurance for at least 7 days between 1 January 2020 and 5 March 2020. Columns (2) to (4) have results from successively looser definitions of dual insurance, and any individual with dual insurance as defined in column (1) will also have dual insurance as defined in column (4). We suggested in Section 3.2 that as the definition of dual insurance became looser the estimated effect of dual insurance would become smaller, and this is what we observe in columns (2) to (4) irrespective of whether the model includes the IV for dual insurance. For all of the definitions of dual insurance in columns (1) to (4), the ATEs are all large relative to the mean mortality rate of 0.54%. With all four definitions the ATE is larger in models without the IV for dual insurance. This suggests that there are unobserved characteristics which increase the probability that an individual has dual insurance and reduces their mortality risk.

[13] Full results for these two models are in **Appendix Table A2**.

[14] The *t* statistic for the ATE from the biprobit model with the differential distance IV (column (6)) is 5.13. The *t* statistic adjusted for our sample size as suggested by Giles (2021) is 4.13, implying that the significance level of the ATE after allowing for our very large sample is 0.00018.

Table 5. Sensitivity to definition of dual insurance

	(1)	(2)	(3)	(4)
	At least 7 consecutive days 01/01/20 -05/03/20	All 7 days 25/12/19 – 31/12/19	At least 7 consecutive days 01/01/19 – 31/12/19	At least 7 consecutive days 01/01/19 – 05/03/20
ATE (Distance IV)	-0.00227***	-0.00315***	-0.00225***	-0.00175***
	(0.0005)	(0.0004)	(0.0004)	(0.0005)
ATE (no IV)	-0.00445***	-0.00480***	-0.00416***	-0.00403***
	(0.0003)	(0.0002)	(0.0003)	(0.0003)
Mean dual insurance	0.0583	0.0447	0.0763	0.0887

Note. 24,739,933 observations in all models. All models for mortality risk and dual insurance contain the full set of explanatory variables. Robust standard errors clustered at district level. *, **, ***: p < 0.05, < 0.01, <0.001.

4.4 Heterogeneity

Table 6 reports the Average Treatment Effect (ATE) and the Average Treatment Effect on the Treated (AETT) for the full sample and for various patient subgroups. In Table 1 the mortality rates for those with dual insurance were higher than for those with single insurance in every age group, in apparent conflict with the lower mortality rate for those with dual insurance in the whole sample. The ATE and AETT subgroup results in Table 6 show that, after taking account of the endogeneity of dual insurance and the full set of covariates, mortality risk in every age group is *reduced* by dual insurance and there is a clear age gradient with dual insurance being more protective at higher ages.

Table 6. Treatment effects for subgroups

		ATE	(SE)	AETT	(SE)
All		-0.00227***	(0.00045)	-0.00164***	(0.00047)
Gender	Male	-0.00300***	(0.00055)	-0.00215***	(0.00060)
	Female	-0.00154***	(0.00035)	-0.00111**	(0.00036)
Age	Age 0-19	-0.00005***	(0.00001)	-0.00007***	(0.00002)
	Age 20-39	-0.00037***	(0.00009)	-0.00043**	(0.00014)
	Age 40-59	-0.00277***	(0.00061)	-0.00321**	(0.00093)
	Age 60-79	-0.00956***	(0.00259)	-0.01117**	(0.00396)
	Age > 79	-0.01974***	(0.00531)	-0.02155**	(0.00728)
ESSALUD		-0.00855***	(0.00050)	-0.00536***	(0.00059)
SIS		-0.00331***	(0.00015)	-0.00201***	(0.00025)
Region	Lima (capital city)	-0.00284***	(0.0008)	-0.00183**	(0.0007)
	Coast	-0.00113	(0.0008)	-0.00070	(0.0006)
	Jungle	-0.00035	(0.0004)	-0.00027	(0.0004)
	Mountains	-0.00055	(0.0004)	-0.00039	(0.0004)
Rurality	Urban district	-0.00218**	(0.00064)	-0.00145**	(0.00056)
	Rural district	0.00016	(0.00031)	0.00009	(0.00021)
Deprivation of district	Group 1 (most)	-0.00008	(0.0004)	0.00005	(0.0003)
	Group 2	-0.00160*	(0.0007)	-0.00100	(0.0006)
	Group 3 (least)	-0.00288**	(0.0009)	-0.00189*	(0.0008)

Note. 24,739,933 observations in all models. Dual insurance: at least 7 days in 1 January 2020 to 5 March 2020. All models for mortality risk and dual insurance contain the full set of explanatory variables. Robust standard errors clustered at district level. *, **, ***: p value for difference < 0.05, < 0.01, <0.001.

Individuals initially in ESSALUD had a bigger reduction in mortality risk from dual insurance than those initially in SIS, perhaps reflecting the fact that getting dual insurance led to a greater increase in access to providers than for those initially in SIS. This is in line with the greater number of providers in the SIS network and shorter distances to them compared with providers in the ESSALUD network.

The effect of dual insurance depends on the type of geographical region, with the ATE and AETT being greatest for those living in the capital city but with no effect for those in the other types of region. This may be because distances to all types of provider are greater, and transport facilities are worse in the other types of region and so the *improvement* in access from belonging to both schemes has a smaller effect on mortality. Similarly, dual insurance reduces mortality in urban districts but not in rural districts. Dual insurance also reduced mortality risk for those in the least poor districts but had no effect for those in the poorest districts, possibly because poorer areas tend to be more likely to be rural.

5. Conclusion

The main limitation of our analysis is the measure of dual insurance. We suggest that having dual insurance when infected with COVID-19 can reduce the risk of death from COVID-19 because having dual insurance gives access to the larger set of providers available in both networks. But we do not observe whether an individual who did not die from COVID-19 (99.46% of the sample) was infected with COVID-19 infection, and for the 0.54% who died from COVID-19 we do not observe when they were infected. Moreover, whether individuals have dual insurance can change over time as they claim and reclaim dual insurance status as their circumstances change and as the rules defining eligibility change. We argue that whether an individual had dual insurance for at least 7 days in the two months (1 January to 5 March 2020) before the outbreak of the pandemic in Peru is a good predictor of their ability to claim dual insurance when it is likely to be highly beneficial, i.e., when they are infected with COVID-19. Some individuals we define as dual insured did not have dual insurance when infected and some we define as not dual insured did have it when infected. These classification errors will tend to *underestimate* the true effect of having dual insurance when infected as they will increase the observed mortality rate amongst those we classify as dual insured and reduce it amongst those we classify as uninsured. Thus we argue that our finding that having dual insurance in the pre-pandemic period is associated with lower COVID-19 mortality risk is evidence for a genuine effect of having dual insurance when infected.

The probability of death from COVID-19 (mortality risk) is the probability of infection with COVID-19 multiplied by the probability of death from COVID-19 when infected (case fatality risk). With our data we cannot examine the separate effects of dual insurance on these latter two probabilities, only its effect on their product. Access to two sets of providers rather than one seems more likely to affect the COVID-19 case fatality risk rather than the risk of COVID-19 infection. However, mortality is not the only possible policy relevant outcome: a non-fatal infection with COVID-19 can increase the risk of worse health in the long term and increase health service costs.

The relevance of our results for other countries depends on the extent to which they have insurance schemes which restrict access to their provider networks. For example, in the US health insurers cannot refuse to pay for out of network emergency care, and major health insurance companies joined with Medicare and Medicaid and agreed to waive all copayments for treatment of COVID-19 (Roehr, 2020). But other countries in Latin America, such as Brazil, Colombia and Mexico, have had high COVID-19 fatality rates (World Health Organization, 2022b) and fragmented insurance schemes with restricted provider networks (Bossert et al., 2014) and so might have benefited from removing restrictions on access.

Being in both SIS and ESSALUD reduced COVID-19 mortality because it gives access to the hospitals in the networks of both schemes. Time matters in the treatment of COVID-19 since a shorter time from onset to treatment reduces the probability of complications (Han et al., 2020). Rapid access to care is important not only in COVID-19 but also for outcomes in other conditions, such as cancer (Lee et al., 2015), AMI (Stenestrand et al., 2001), stroke (Basu, 2014), and serious mental illness (McGorry, 2015). Giving insureds who are only in one of the insurance schemes access to the provider networks in both schemes may therefore also have health benefits in the post-COVID-19 pandemic period. This suggests that policymakers should allow dual insurance for those insureds who potentially qualify for both SIS and ESSALUD under the current rules, and that consideration should be given to an experimental evaluation of removing the restriction that insureds can only access providers in their network in order to quantify the system wide health effects and possible costs.

References

Abadie, A., Athey, S., Imbens, G., Wooldridge, J. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica,* 88(1), 265-296. https://doi.org/https://doi.org/10.3982/ECTA12675

Andresen, M., Huber, M. (2018). Instrument-based estimation with binarized treatments: issues and tests for the exclusion restriction. Econometrics Journal, 24, 536-558. https://doi.org/10.1093/ectj/utab002

Angrist, J., Pischke, J. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Atwood, A., Sasso, A. (2016). The effect of narrow provider networks on healthcare use. *Journal of Health Economics*, 50, 86-98. doi: 10.1016/j.jhealeco.2016.09.007

Basu, A. P. (2014). Early intervention after perinatal stroke: opportunities and challenges. *Developmental Medicine and Child Neurology*, 56(6), 516-521.

Baum, C. F., Dong, Y., Lewbel, A., Yang, T. (2012). Binary choice models with endogenous regressors. https://www.stata.com/meeting/sandiego12/materials/sd12_baum.pdf

Bernal, N., Carpio, M., Klein, T. (2017). The effects of access to health insurance: evidence from a regression discontinuity design in Peru. *Journal of Public Economics*, 154, 122-136. https://doi.org/https://doi.org/10.1016/j.jpubeco.2017.08.008

Bhaskaran, K., Bacon, S., Evans, S., Bates, C., Rentsch, C., MacKenna, B., Tomlinson, L., Walker, A., Schultze, A., Morton, C. (2021). Factors associated with deaths due to COVID-19 versus other causes: population-based cohort analysis of UK primary care data and linked national death registrations within the OpenSAFELY platform. *The Lancet Regional Health-Europe*, 6, 100109.

Bhattacharya, J., Goldman, D., McCaffrey, D. (2006). Estimating probit models with self-selected treatments. *Statistics in Medicine*, 25(3), 389-413.

Bossert, T., Blanchet, N., Sheetz, S., Pinto, D., Cali, J., Perez Cuevas, R. (2014). Comparative review of health system integration in selected countries in Latin America. Washington, DC: Inter-American Development Bank.

Brave, S., and Walstrum, T. (2014). Estimating marginal treatment effects using parametric and semiparametric methods. *The Stata Journal*, 14(1), 191-217.

Chakrabarti, R., Meyerson, L., Nober, W., Pinkovskiy, M. (2020). The Affordable Care Act and the COVID-19 pandemic: a regression discontinuity analysis. Federal Reserve Bank of New York Staff Report No. 948.

Chiburis, R. C., Das, J., Lokshin, M. (2011). A practical comparison of the bivariate probit and linear IV estimators. Policy Research Working papers, World Bank. https://doi.org/10.1596/1813-9450-5601

Cornelissen, T., Dustmann, C., Raute, A., Schönberg, U. (2016). From LATE to MTE: Alternative methods for the evaluation of policy interventions. *Labour Economics*, 41, 47-60.

Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometrics Approach to Development Policy.* World Bank. John Hopkins Press.

Denzer, M. (2019). Estimating causal effects in binary response models with binary endogenous explanatory variables. A comparison of possible estimators. Gutenberg School of Management and Economics, Johannes Gutenberg-Universität Mainz Working Papers 1916.

Dyer, O. (2021). Covid-19: Peru's official death toll triples to become world's highest. British Medical Journal, 373, 1442. https://doi.org/10.1136/bmj.n1442

Egede, L., Walker, R., Garacci, E., Raymond J. (2020). Racial/ethnic differences in COVID-19 screening, hospitalisation, and mortality on southeast Wisconsin. *Health Affairs*, 39(11), 1926-1934.

Figueroa, J., Wadhera, R., Lee, D., Yeh, R., Sommers, B. (2020). Community-level factors associated with racial and ethnic disparities in COVID-19 rates in Massachusetts. *Health Affairs*, 39(11), 1984-1992.

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J., Allen, H., Baicker, K., Group, O. (2012). The Oregon health insurance experiment: evidence from the first year. *Quarterly Journal of Economics*, 127(3), 1057-1106.

Gaynor, M., Propper, C., Seiler, S. (2016). Free to choose? Reform, choice, and consideration sets in the English National Health Service. *American Economic Review*, 106(11), 3521-3557.

Giles, D. (2019). Everything's significant when you have lots of data. https://davegiles.blogspot.com
Goldin, J., Lurie, I., McCubbin, J. (2021). Health insurance and mortality: experimental evidence from taxpayer outreach. *Quarterly Journal of Economics*, 136(1), 1-49.

Greene, W. (2020). *Econometric Analysis*. Pearson Education.

Gruber, J., McKnight, R. (2016). Controlling healthcare costs through limited network insurance plans: evidence from Massachusetts state employees. *American Economic Journal: Economic Policy*, 8(2), 219-250.

Gutacker, N., Siciliani, L., Moscelli, G., Gravelle, H. Choice of hospital: which type of quality matters? *Journal of Health Economics*. 2016, 50, 230-246.

Han, J., Shi, L., Xie, Y., Zhang, Y., Huang, S., Li, J., Wang, H., Shao, S. (2020). Analysis of factors affecting the prognosis of COVID-19 patients and viral shedding duration. *Epidemiology and Infection*, 148.

Health Ministry of Peru. (2021a). Criterios técnicos para actualizar la cifra de fallecidos por COVID-19 en El Perú. 31 May 2021.
https://www.gob.pe/institucion/pcm/informes-publicaciones/1943691-criterios-tecnicos-para-actualizar-la-cifra-de-fallecidos-por-covid-19-en-el-peru

Health Ministry of Peru. (2021b). Datos Abiertos de COVID-19. [Accessed 17 November 2021].https://www.datosabiertos.gob.pe/group/datos-abiertos-de-covid-19?sort_by=changed&f%5B0%5D=field_tags%3A913&f%5B1%5D=field_tags%3A489

Karlinsky, A., Kobak, D. (2021). Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. *Elife*, 10, e69336.

Lancaster, T. The incidental parameter problem since 1948. *Journal of Econometrics*, 95, 391-413.
Leamer, E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. Wiley, New York.

Lee, D., Suh, G. Y., Ryu, J., Chung, C., Yang, J., Park, C., Jeon, K. (2015). Effect of early intervention on long-term outcomes of critically ill cancer patients admitted to ICUs. *Critical Care Medicine*, 43(7), 1439-1448.

Lewbel, A. (2007). Estimation of average treatment effects with misclassification. *Econometrica*, 75(2), 537-553.

Lewbel, A., Dong, Y., Yang, T. (2012). Comparing features of convenient estimators for binary choice models with endogenous regressors. *Canadian Journal of Economics*, 45(3), 809-829.

Li, C., Poskitt, D., Zhao, X. (2019). The bivariate probit model, maximum likelihood estimation, pseudo true parameters and partial identification. *Journal of Econometrics*, 209(1), 94-113.

MacKinnon, J., Webb, M. (2019). When and how to deal with clustered errors in regression models. Queen's Economics Department Working Paper 1421.

McClellan, M., McNeil, B., Newhouse, J. (1994). Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. JAMA, 272(11), 859-866.

McGorry, P. (2015). Early intervention in psychosis: obvious, effective, overdue. *Journal of Nervous and Mental Disease*, 203(5), 310.

McWilliams, J. (2009). Health consequences of uninsurance among adults in the United States: recent evidence and implications. *The Milbank Quarterly*, 87(2), 443-494.

Maddala, G. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

Monfardini, C., Radice, R. (2008). Testing exogeneity in the bivariate probit model: a Monte Carlo study. *Oxford Bulletin of Economics and Statistics*, 70(2), 271-282. https://doi.org/https://doi.org/10.1111/j.1468-0084.2007.00486

Neelsen, S., O'Donnell, O. (2017). Progressive universalism? The impact of targeted coverage on healthcare access and expenditures in Peru. *Health Economics*, 26(12), e179-e203. https://doi.org/https://doi.org/10.1002/hec.3492

PCM. (2021). Decreto de Urgencia 046-2021 Decreto de Urgencia que dicta medidas extraordinarias y urgentes en materia economica y financiera para fortalecer el Aseguramiento Universal en Salud en el marco de la emergencia nacional por la Covid-19. Presidencia del Consejo de Ministros. Decreto de Urgencia 046-2021.

Polsky, D., Cidav, Z., Swanson, A. (2016). Marketplace plans with narrow physician networks feature lower monthly premiums than plans with larger networks. *Health Affairs*, 35(10), 1842-1848.

Rakus, A., Soni, A. (2022). Association between state Medicaid expansion status and health outcomes during the COVID-19 pandemic. *Health Services Research*, https://doi.org/https://doi.org/10.1111/1475-6773.14044

Roehr, B. (2020). The health of private insurance in the US during covid-19. *British Medical Journal*, 370, 2 July 2020. BMJ 2020;370:m2606. http://dx.doi.org/10.1136/bmj.m2606

Roodman, D. (2011). Fitting fully observed recursive mixed-process models with *cmp*. *Stata Journal*, 11(2), 159-206.

Soto-Cabezas, M., Reyes, M., Soriano, A., Rodríguez, J., Ibargüen, L., Martel, K., Jaime, N., Munayco, C. (2022). COVID-19 among Amazonian indigenous in Peru: mortality, incidence, and clinical characteristics. *Journal of Public Health*, 44(3), e359-e365. https://doi.org/10.1093/pubmed/fdac058

Stenestrand, U., Wallentin, L., Care, S. (2001). Early statin treatment following acute myocardial infarction and 1-year survival. JAMA, 285(4), 430-436.

Stock, J., Yogo, M. (2005). Testing for weak instruments in linear IV regression. In Andrews, D. (ed.) *Identification and Inference for Econometric Models*. New York: Cambridge University Press ; 2005. pp. 80-108.

Superintendencia Nacional de Salud. (2021). Base de datos del Registro de Afiliados al Aseguramiento Universal en Salud-RAAUS actualizado al 06 de octubre del 2021.

Wilde, J. (2000). Identification of multiple equation probit models with endogenous dummy regressors. *Economics Letters*, 69(3), 309-312. https://doi.org/https://doi.org/10.1016/S0165-1765(00)00320-7

Williamson, E. J., Walker, A. J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C. E., Curtis, H. J., Mehrkar, A., Evans, D., Inglesby, P. (2020). Factors associated with COVID-19-related death using OpenSAFELY. *Nature*, *584*(7821), 430-436.

Wilper, A., Woolhandler, S., Lasser, K., McCormick, D., Bor, D., Himmelstein, D. (2009). Health insurance and mortality in US adults. *American Journal of Public Health*, 99(12), 2289-2295.
World Health Organization. (2022). WHO Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int/ [Accessed 01 October 2022].

Yanez, N. Weiss, N., Romand, J., Treggiari, M. (2020). COVID-19 mortality risk for older men and women. *BMC Public Health*, 20(1), 1-7.

Appendices

Appendix 1. Mortality data

Individual level mortality data is from the NOTI-SINADEF[15] dataset. Criteria for defining deaths from COVID-19 were revised on 31 May 2021 and applied retrospectively to data from 1 March 2021 (Health Ministry of Peru, 2021). The seven criteria are applied hierarchically to classify a death as due to COVID-19:

- Virological criterion: death in a confirmed case of COVID-19 within 60 days after a molecular test (PCR) or reactive antigen for SARS-CoV-2.
- Serological criterion: death in a confirmed case of COVID-19 within 60 days after a positive IgM or IgM / IgG serological test for SARS-CoV-2.
- Radiological criteria: death in a probable case of COVID-19 that presents a radiological, tomographic or nuclear magnetic resonance image compatible with COVID-19 pneumonia.
- Epidemiological link criterion: death in a probable case of COVID-19 that presents an epidemiological link with a confirmed case of COVID-19.
- Epidemiological investigation criteria: death in a suspected case of COVID-19 that is verified by epidemiological investigation of the National Epidemiology Network.
- Clinical criterion: death in a suspected case of COVID-19 that presents a clinical picture compatible with the disease.
- SINADEF Criterion: death with death certificate in which the diagnosis of COVID-19 is presented as the cause of death.

[15] SINADEF is the Sistema Nacional de Defunciones, which is an administrative national system in which death certificates are recorded.

Appendix 2: Additional Tables and Figures

Table A1. Data cleaning

Number with insurance records at 05/03/2020				32,278,940
	Missing data on age, district of residence		61,653	
	Not in SIS or ESSALUD*		7,477,354	
Total exclusions				7,539,007
Estimation sample				24,739,933

* For at least one continuous week in 2019 and one continuous week from 01/01/2020 to 05/03/2020

Table A2. Full results for biprobit models

	Biprobit no IV		Biprobit Distance IV	
	Dependent variable: dual insurance	Dependent variable: mortality	Dependent variable: dual insurance	Dependent variable: mortality
Dual insurance		-0.547***		-0.215***
		(0.048)		(0.052)
Age 2 (20 - 39 years)	0.051***	0.594***	0.054***	0.599***
	(0.006)	(0.015)	(0.006)	(0.015)
Age 3 (40 - 59 years)	-0.095***	1.229***	-0.101***	1.254***
	(0.007)	(0.019)	(0.007)	(0.016)
Age 4 (60 - 79 years)	-0.362***	1.779***	-0.367***	1.825***
	(0.006)	(0.023)	(0.006)	(0.019)
Age 5 (> 79 years)	-0.913***	2.070***	-0.922***	2.135***
	(0.016)	(0.025)	(0.016)	(0.020)
Male	0.029***	0.295***	0.028***	0.296***
	(0.003)	(0.004)	(0.003)	(0.004)
Geography				
Lima Metropolitana	Ref.	Ref.	Ref.	Ref.
Coast	-0.061**	0.009	-0.063**	0.013
	(0.023)	(0.017)	(0.022)	(0.017)
Jungle	-0.254***	-0.117***	-0.213***	-0.106**
	(0.038)	(0.033)	(0.038)	(0.034)
Mountains	-0.180***	-0.123***	-0.141***	-0.116***
	(0.030)	(0.026)	(0.029)	(0.026)
Rural district	-0.110***	-0.255***	-0.099***	-0.255***
	(0.014)	(0.016)	(0.014)	(0.016)
District deprivation				
Group 1 (most deprived)	Ref.	Ref.	Ref.	Ref.
Group 2	0.097***	0.156***	0.082***	0.153***
	(0.016)	(0.019)	(0.016)	(0.019)
Group 3 (least deprived)	0.259***	0.119***	0.246***	0.107**
	(0.028)	(0.034)	(0.028)	(0.035)
Department 1	-0.067	-0.035	-0.110**	-0.033
	(0.036)	(0.045)	(0.036)	(0.046)
Department 2	0.067**	0.053**	0.087***	0.052**
	(0.025)	(0.019)	(0.024)	(0.019)
Department 3	-0.038	0.011	-0.090*	0.013
	(0.038)	(0.036)	(0.039)	(0.036)

Department 4	0.039	0.067*	0.009	0.067*
	(0.028)	(0.032)	(0.027)	(0.032)
Department 5	-0.081**	-0.018	-0.108***	-0.015
	(0.029)	(0.035)	(0.030)	(0.035)
Department 6	-0.008	-0.045	-0.047	-0.045
	(0.054)	(0.025)	(0.053)	(0.025)
Department 7	0.118***	0.088	0.115***	0.082
	(0.030)	(0.049)	(0.027)	(0.048)
Department 8	0.043	0.068**	-0.001	0.068**
	(0.043)	(0.025)	(0.043)	(0.026)
Department 9	0.037	0.040	-0.006	0.039
	(0.047)	(0.029)	(0.046)	(0.028)
Department 10	-0.054	0.034	-0.109***	0.037
	(0.030)	(0.024)	(0.032)	(0.024)
Department 11	-0.031	0.106*	-0.037	0.109*
	(0.033)	(0.046)	(0.032)	(0.046)
Department 12	0.020	0.142***	-0.010	0.144***
	(0.029)	(0.026)	(0.028)	(0.026)
Department 13	-0.005	-0.040	-0.005	-0.041
	(0.023)	(0.022)	(0.022)	(0.022)
Department 14	-0.098**	-0.025	-0.103**	-0.020
	(0.037)	(0.021)	(0.035)	(0.021)
Department 15 (Lima)	Ref.	Ref.	Ref.	Ref.
Department 16	0.085	0.167***	0.033	0.166***
	(0.051)	(0.049)	(0.051)	(0.049)
Department 17	-0.117**	0.035	-0.171***	0.044
	(0.045)	(0.052)	(0.044)	(0.052)
Department 18	0.110*	-0.092	0.101*	-0.100
	(0.046)	(0.061)	(0.043)	(0.062)
Department 19	0.199***	0.093**	0.134**	0.085*
	(0.039)	(0.036)	(0.041)	(0.036)
Department 20	0.121***	0.070**	0.115***	0.065*
	(0.030)	(0.027)	(0.029)	(0.027)
Department 21	-0.046	0.031	-0.076*	0.033
	(0.031)	(0.030)	(0.032)	(0.031)
Department 22	-0.015	-0.073	-0.045	-0.072
	(0.040)	(0.038)	(0.040)	(0.038)
Department 23	-0.073	-0.117**	-0.076	-0.114**
	(0.046)	(0.036)	(0.044)	(0.036)
Department 24	-0.151***	-0.042	-0.155***	-0.035
	(0.042)	(0.023)	(0.040)	(0.023)
Department 25	-0.055	0.117**	-0.100*	0.122**
	(0.039)	(0.041)	(0.039)	(0.041)
Average structural quality nearest SIS and ESSALUD providers				
Average level 1	Ref.	Ref.	Ref.	Ref.
Average level 1.5	0.015	0.002	0.009	0.001
	(0.013)	(0.015)	(0.013)	(0.015)
Average level 2	0.042**	-0.015	0.037*	-0.018
	(0.015)	(0.016)	(0.015)	(0.016)
Average level 2.5	0.051*	-0.041	0.047*	-0.044*
	(0.022)	(0.021)	(0.022)	(0.022)
Average level 3	-0.004	-0.101***	-0.014	-0.102***
	(0.021)	(0.024)	(0.021)	(0.024)

Average distance to providers in quintiles (lowest to highest)				
q1	Ref.	Ref.	Ref.	Ref.
q2	0.025	0.012	0.025	0.010
	(0.013)	(0.012)	(0.013)	(0.012)
q3	-0.002	0.005	-0.005	0.005
	(0.013)	(0.013)	(0.013)	(0.014)
q4	-0.048**	-0.066***	-0.038**	-0.065***
	(0.015)	(0.013)	(0.015)	(0.013)
q5	-0.125***	-0.110***	-0.084***	-0.105***
	(0.016)	(0.016)	(0.016)	(0.016)
District COVID-19 infection rate	0.113	1.137***	0.109	1.142***
	(0.106)	(0.290)	(0.107)	(0.295)
Distance difference IV			0.003***	
			(0.0001)	
Constant	1.543***	3.994***	1.522***	4.065***
	(0.022)	(0.036)	(0.021)	(0.030)
Number of observations	24,739,933	24,739,933	24,739,933	24,739,933
Rho (correlation of residuals.)		0.284***		0.103***
		(0.029)		(0.025)

Note: Robust standard errors clustered at district level. *, **, ***: p value for difference < 0.05, < 0.01, <0.001.

Table A3. Predictions from linear models

	LPM	LPM	2SLS (diff distance IV)	
Predictions of	Dual insurance	Mortality	Dual insurance	Mortality
Mean	0.05834	0.00538	0.05834	0.00538
SD	0.02863	0.01051	0.02863	0.01320
Inter-quartile range	0.04226	0.00649	0.04226	0.00822
% < 0	2.91%	30.57%	2.91%	46.31%
% > 1	-	-	-	-

Table A4. 2SLS and biprobit models with binary IV

	2SLS	Biprobit
ATE	0.02077***	-0.00262***
	(0.0056)	(0.0005)
N	24,739,933	24,739,933

Note: Robust standard errors clustered at district level. *, **, ***: p value for difference < 0.05, < 0.01, <0.001.

Table A5. Past dual insurance and probability of future dual insurance status

	7 continuous days during 1 Jan to 5 Mar 2020	7 continuous days during 1 Jan to 5 Mar 2020	7 continuous days during 1 Jan to 5 Mar 2020 (Biprobit/2SLS)
Probit regression			
Dual insurance 7 continuous days in 2019	0.1476***	0.1471***	0.5210***
	(0.0001)	(0.0001)	(0.0969)
Linear regression			
Dual insurance 7 continuous days in 2019	0.5857***	0.5853***	1.0485***
	(0.0004)	(0.0004)	(0.0058)
Difference in distance as control variable	No	Yes	
Difference in distance as IV			Yes

Note: Dependent variable: indicator for dual insurance for at least 7 consecutive days in 01/01/2020 to 05/03/2020. Explanatories: all covariates plus indicator for dual insurance for at least 7 consecutive days in 2019. Robust standard errors in parenthesis. *, **, ***: p value < 0.05, < 0.01, <0.001.

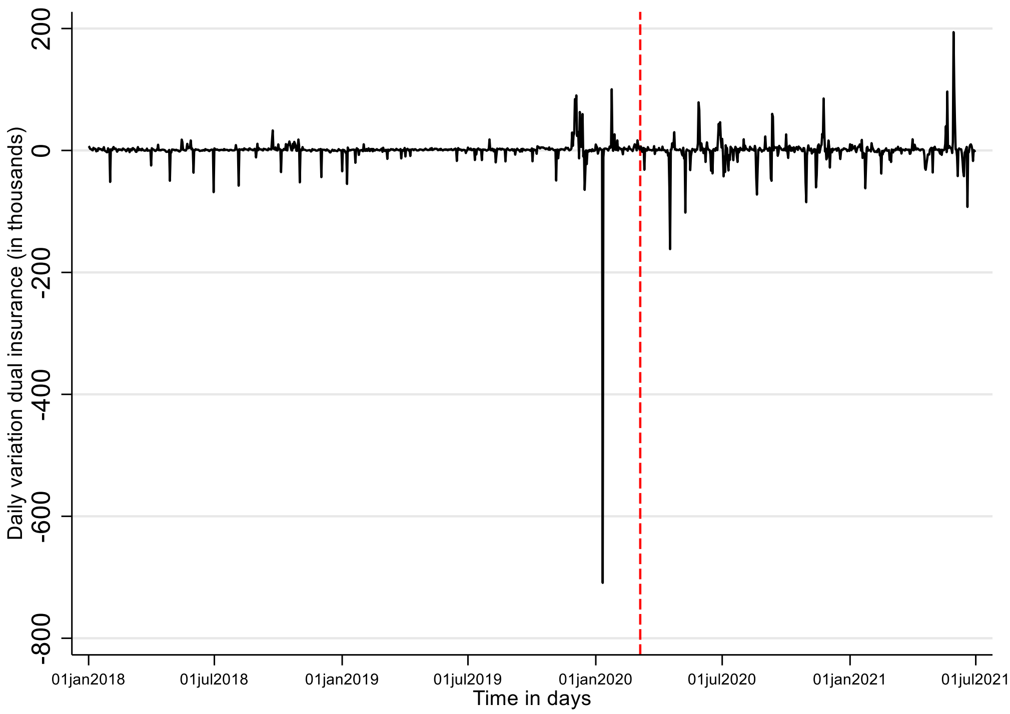


Figure A1: Changes in dual insurance 1 January 2018-30 June 2021

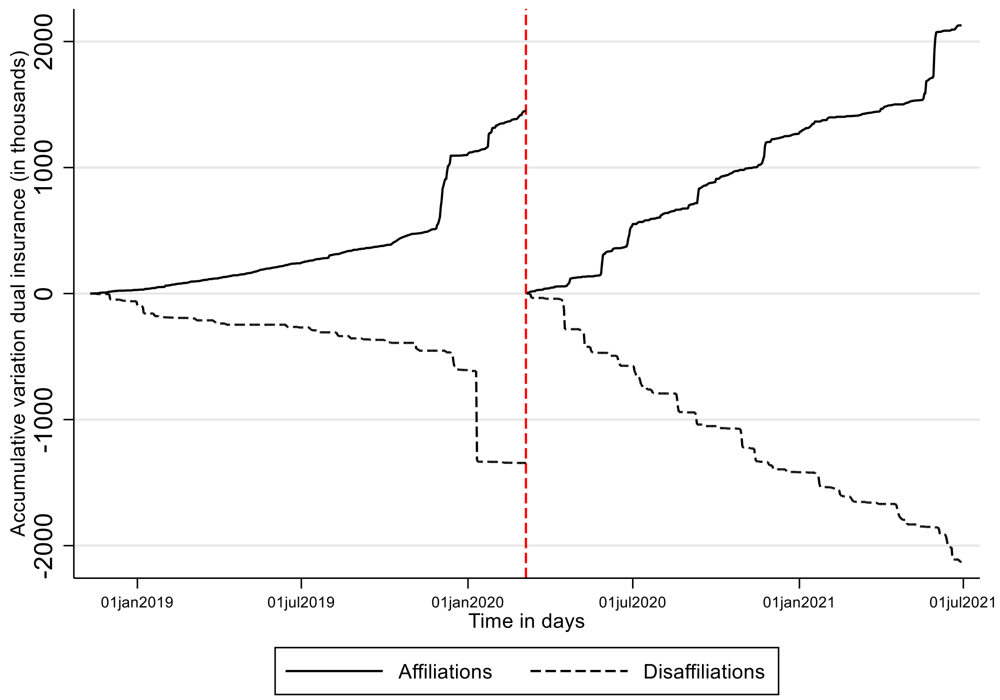


Figure A2. Cumulative affiliations and disaffiliations before and after start of pandemic

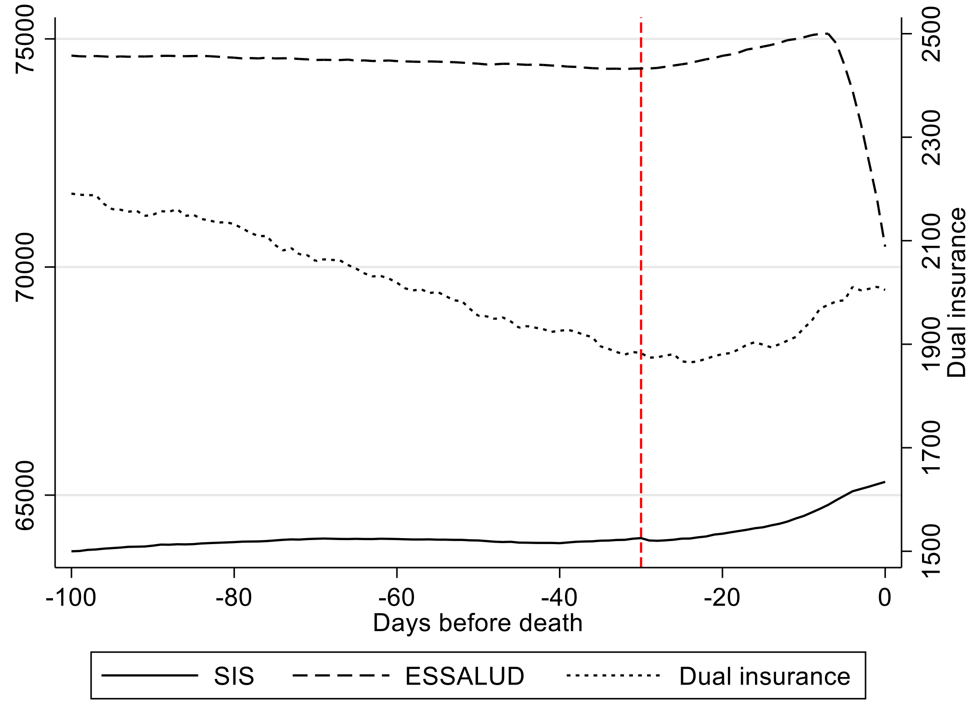


Figure A3: Evolution of insurance status 100 days up to date of death