

C A G E

Health inequality and the 1918 influenza in South Africa

CAGE working paper no. 532

January 2021

Johan Fourie
Jonathan Jayes

Health inequality and the 1918 influenza in South Africa *

Johan Fourie[†] and Jonathan Jayes[‡]

January 4, 2021

Abstract

The 1918 influenza – the Spanish flu – killed an estimated 6% of South Africans. Not all were equally affected. Mortality rates were particularly high in districts with a large share of black and coloured residents. To investigate why this happened, we transcribed 39,482 death certificates from the Cape Province. Using a novel indicator – whether a doctor’s name appears on the death certificate – we argue that the unequal health outcomes were a consequence of unequal access to healthcare. Our results show that the racial inequalities in health outcomes that existed before October 1918 were exacerbated during the pandemic. Access to healthcare, as we expected, worsened for black and coloured residents of the Cape Province. Unexpectedly, however, we found that other inequalities were unchanged, or even reversed, notably age, occupation and location. Living in the city, for instance, became a health hazard rather than a benefit during the pandemic. These surprising results contradict the general assumption that all forms of inequality are exacerbated during a crisis. Our analyses suggest explanations for the widening racial gap in healthcare access during the 1918 pandemic, from both the demand and the supply side. We could find, however, no evidence of racial prejudice. Our findings confirm the importance of taking race into account in studying the effects of the 2020 Covid-19 pandemic or other world crises.

Keywords. Spanish flu, health care, inequality, healthcare, influenza, pandemic

JEL code. I14, N37

*We thank Steven Broadberry, Ronelle Burger, Dan de Kadt, Jane Humphries, Johannes Norling, Sheilagh Ogilvie, Howard Phillips, Jan Luiten van Zanden and seminar participants at the universities of Cape Town, Oxford, Stellenbosch and York for their comments. Lauren Coetzee, Kara Dimitruk, Kate Ekama, Jan Greyling, Kelsey Lemon, Elie Murard, Laura Richardson and Young-ook Jang provided valuable research assistance. This article also benefited immensely from a kind and generous referee and supportive editor. A special word of thanks to all our transcribers: Elizabeth Baldwin, Judy Buxton, Kudzai Chidamwoyo, Chris de Wit, Martie du Toit, Anneen Fourie, Hans Heese and Gustav Hendrich. We acknowledge generous funding from the Andrew W. Mellon Foundation through the *Biography of an Uncharted People*-project and the Stellenbosch University Vice-Rector’s Special Covid-19 grant. The usual disclaimer applies.

[†]Department of Economics, Stellenbosch University. ✉johanf@sun.ac.za

[‡]Department of Economics, Stellenbosch University. ✉jonathanjayes@gmail.com

1 Introduction

On 28 September 1918 a train left Cape Town carrying soldiers returning to their homes in South Africa from the trenches of World War I in Europe. Seven disembarked at Queenstown, a town in today's Eastern Cape. It must have been wonderful to see family and friends again and put the horrors of war behind them. Little did they know that the most harrowing month lay ahead: Black October.

It is very likely that Wills Bunu would have known some of these returning soldiers. He was 19 and lived in Cimezile, a district south of Queenstown, where his family farmed. They belonged to the Fingo ethnic group, a Xhosa-speaking clan. Shortly after the soldiers returned, Bunu fell ill. So did most of his extended family. Two weeks later, on 10 October, the first Bunu family member, twenty-five year old Nosamon Bunu, died. Two days later followed Adolphus Bunu, 17, and John Bunu, 40. Another two days later Nomavengi Bunu, 25, died, and the day after, Thomisa Bunu, 20. Five family members in five days. Wills Bunu endured all of this, helpless and deathly ill himself. The most heartbreaking moment must have been when his three-year old sister, Canyiwe, died on 24 October. Two days later, his uncle, Matthais, made a solemn journey to Queenstown to register yet another death. Wills Bunu had died.

The Bunus of Cimezile are just one of many families who felt the devastating consequences of the Spanish flu that swept South Africa in late 1918. Estimates of the number of deaths vary, partly because the scale of the calamity put strain on the system that registered deaths. Phillips (1984) estimates the death toll at between 250,000 and 350,000, or 6% of the population. The burden of the disease was not spread evenly among the population. The mortality rates for black and coloured people were markedly higher than for whites.¹ This may be explained, in part, by the inequalities in access to healthcare that the pandemic exacerbated.

This is the hypothesis we test in this paper. Using a new dataset of 39,482 death certificates from 1915 to 1920 from 15 districts in the Cape Province of South Africa, we identify the factors that explain access to healthcare before, during and after the pandemic. Death certificates in the Cape Colony allowed for the name of a doctor to be recorded on the certificate if the doctor confirmed the death or specified a cause, though just 52% of the certificates in our sample recorded a name. We use the absence of a doctor's name as a proxy for lack of medical access, or an indication that the patient was not treated by a doctor immediately prior to death. Our results show convincingly that black and coloured patients had poorer access to healthcare

¹Following the racial classification introduced during apartheid and still in use in South Africa today, we classify race into three categories: white, coloured and black. In the first census of the unified South Africa in 1911, the population of the Cape Province was classified into 'native' (black, 72%), coloured (13%) and white (15%). The coloured inhabitants of the Cape were a heterogeneous group descended from liaisons between Cape Malays, Khoe pastoralists, San hunter-gatherers, Bantu-speaking (black African) inhabitants, and European immigrants. Indian/Asian is the fourth officially recognized race category in South Africa today, but their numbers were marginal in the Cape Province in 1911.

during the pandemic than they did before. Race is, of course, just one of many dimensions of privilege and exclusion. Our individual-level sources allow us to test other dimensions as well, such as occupation and literacy. These help to explain why race was such an important predictor of access to medical services.

South Africa provides an ideal setting to study determinants of healthcare access. Just as South Africa is rife with inequality today, inequality that remains closely correlated to race, so too was the Cape Province in 1918, with race being a strong predictor of income and social status. We can use the race classifications to measure whether, or to what extent, access to medical services became more unequal during the pandemic. A second reason for using South Africa is that it had one of the highest Spanish flu mortality rates. This was especially true of the Cape Province, where the ‘Black October’ mortality rate was over 100 deaths per 1000 residents in some districts. A third reason is the access to detailed, microlevel data. Death certificates, with important exceptions which we discuss below, were recorded for all inhabitants since Act No. 7 of 1894 – ‘To Provide for the Registration of Births and Deaths’ – came into effect on 1 January 1895. The certificates for the six years from 1 January 1915 to 31 December 1920, preserved in the National Archives in Pretoria, have now been digitised and transcribed.² The nature of this source material allows us to investigate questions where aggregated data would be no help.

Our study makes at least three contributions. First, it questions the common belief that, while access to healthcare is unequally distributed in almost all societies (O’Donnell et al. 2015), these inequalities will worsen during a crisis (Blumenshine et al. 2008). Our rich source material, in a setting with many of the same inequalities as today, enabled us to test which of these inequalities are most likely to be amplified, if any. Our study found that race was the only inequality that increased; we found no effects for gender, immigrant status or literacy. Other inequalities, notably occupation and location (urban or rural), were reversed. Pandemics, it seems, do not exacerbate all inequalities.

Second, our results help to explain why access to healthcare services worsens for some groups during a pandemic but not for others. Having tested several hypotheses on both the demand and the supply side, we found that during a pandemic, the health of the household matters: access to healthcare worsens when more members of the household are infected. Location also matters, but, counterintuitively, we found that the healthcare advantages of living in a town or city reversed during the pandemic. On the supply side, we found no evidence that doctors discriminated, favouring patients from some groups over others.

Third, our study expands the South African historiography on the Spanish flu and, more generally, the role of the medical profession during a pandemic. Thanks to the seminal contributions of Howard Phillips, we already know much about the reasons for and response to the Spanish flu in South Africa (Phillips 2014). Our approach, using newly transcribed sources that Phillips did not have access to and techniques he could

²We have transcribed the death certificates from the collection ‘South Africa, Cape Province, Civil Records, 1840-1972’ available on FamilySearch.org.

not have used, has taught us much about not only the timing and intensity of the disease but also the nature and distribution of the healthcare services that first responded.

Our results not only inform our understanding of past pandemics but also have implications for today. Covid-19 has exposed and exacerbated inequalities in many countries across the world. In the United States, African Americans, only 14% of the population, had higher percentages of cases and deaths than whites (Mein 2020; Dorn et al. 2020). Similar inequality was found during the H1N1 influenza (‘swine flu’) pandemic of 2009 (Quinn et al. 2011). Our study shows that this is not a new phenomenon: when the virus arrived in October 1918, access to healthcare for black South Africans worsened. Wills Bunu and his family were among the thousands who suffered the consequences.

2 Health inequality and the Spanish flu

Many reasons can be found for the large differences in health outcomes in societies today and in the past (Van Doorslaer and Wagstaff 1992; Wagstaff and Van Doorslaer 2000). Income is the most obvious one. If health is a normal good, then the more affluent are likely to consume more of it. Income can not only buy more or better healthcare; it can buy many determinants of good health, such as better food, bigger houses and safer neighbourhoods (O’Donnell et al. 2015). This close correlation between income and health is also why the state often intervenes in the market, either by providing the goods and services itself, such as healthcare, or by subsidising essentials like food and housing.

Health inequalities nevertheless persist, because factors besides income and government intervention also matter. A large literature shows how patients’ race, ethnicity, gender, social class, immigrant status, urban or rural location, and social and cultural traits and beliefs contribute to disparities in healthcare access (Whitehead et al. 2001; Gyasi et al. 2019; Alsan and Wanamaker 2018). On the supply side, user fees, incompetence and negligence, stigma and prejudice, and inefficient collection of vital statistics can increase the disparity (Jaffee et al. 2016; Cuevas et al. 2016; Fisker et al. 2019).

The disparities are often most severe in developing countries where state capacity is weak (Van Doorslaer et al. 2006). South Africa is no exception. Black South Africans are today less likely than white South Africans to report healthcare services as available, affordable or acceptable (Burger and Christian 2020). Race intersects with other dimensions too. Burger and Christian (2020, pp. 51–52) note that ‘rural households and vulnerable groups such as black South Africans, the less educated, the unemployed and the poor tend to be less likely to have adequate access to health care’.

This should not surprise anyone familiar with South Africa’s history of race-based repressive and discriminatory policies during the twentieth century. These policies were reflected in past health outcomes too. Although accurate statistics are hard to find, Natrass and Seekings (2011) estimate that the difference

between life expectancy for blacks and whites in the 1940s was more than 25 years. Mpeta et al. (2018) calculate that the difference between adult heights for those born during the first decade of the twentieth century was just over 5cm; by the end of the century, it had increased to 8cm.

These inequalities were also reflected in the differences in mortality rates across the Union of South Africa during the 1918 pandemic, as shown in Figure 1. Mortality rates as high as 120 deaths per 1000 inhabitants were recorded, notably in the Transkei, a predominantly black area. As we show in this paper, the Spanish flu not only reflected but exacerbated the gap in health outcomes of the time. The difference between white and black death rates during the pandemic was greater than before 1918. But the picture we found is more complicated, as we have mentioned: increased racial inequality, but instead of an increase in other types of inequality, rather the reverse.

Although we are not the first to point out that pandemics affect the vulnerable most severely (Fothergill et al. 1999; Matthews Pillemer et al. 2015), we are, to the best of our knowledge, the first to show that not all subgroups are equally affected. The emphasis so far has been on race or minority status. Blumenshine et al. (2008) describe how race and socioeconomic status may lead to large disparities in healthcare access. Surveys by Quinn et al. (2011) confirm the emergence of these disparities during the H1N1 influenza pandemic. Hutchins et al. (2009, p. 268) say the ‘available data suggest that an increased risk of adverse health outcomes is likely among minority populations during a pandemic: they experience disproportionately poor health outcomes and greater barriers to care during both pandemic and annual pneumonia and influenza illnesses, including ... limited access to health care’. Some have voiced social justice concerns about the lack of preparedness to mitigate the effects of a pandemic on vulnerable groups (Saunders and Monet 2007; Uscher-Pines et al. 2007).

The Covid-19 pandemic, ongoing at the time of writing, has brought such inequities to the fore. Mein (2020, online) notes that in the United States pandemic is ‘highlighting the health disparities that already exist within our communities’. He calls for more transparency on the topic. Van Dorn et al. (2020) also note that ‘pre-existing racial and health inequalities already present in US society are being exacerbated by the pandemic’. Some preliminary evidence supports this. Abedi et al. (2020) show that US counties with more diverse demographics were at a higher risk of COVID-19 infection. But many questions remain. Is this the result of pre-existing inequalities in healthcare? And did these inequalities worsen during the pandemic?

These are the questions we attempt to answer in this paper, for the case of the 1918 Spanish flu pandemic in South Africa’s Cape Province, and using a newly transcribed dataset of death certificates and a novel indicator to measure access to healthcare.

South African 1918 Spanish Flu Mortality

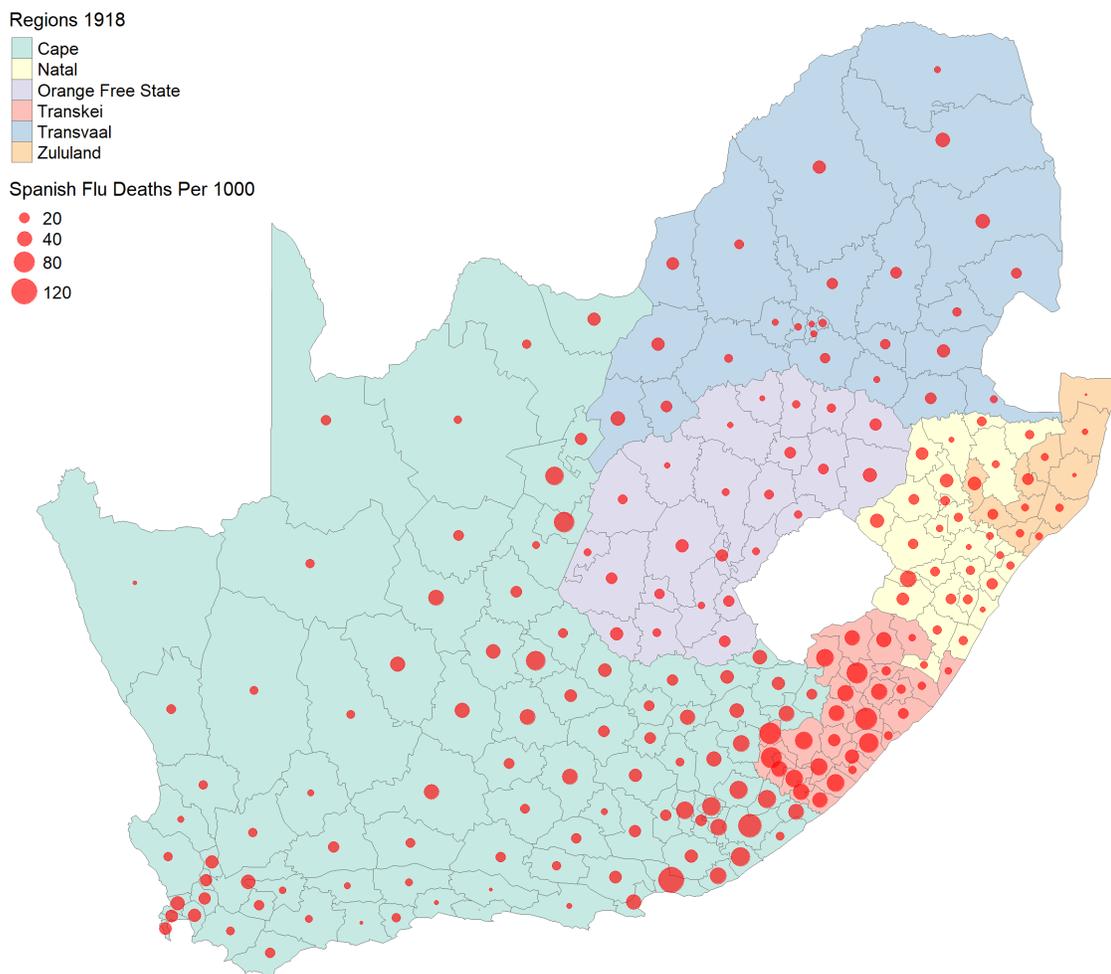


Figure 1: Official mortality rates by magisterial district, 1918

Source: Phillips (1984), South African census 1911; own calculations

3 Death certificates and their research value

The Spanish flu arrived in two waves in South Africa, one mild and one virulent. The milder strain arrived via Durban and spread into the interior, reaching the goldmines of Johannesburg during September 1918. Although more than 14,000 cases were reported within two weeks, and despite the vulnerability of mine workers to pneumonia and the dense living arrangement of the compound system, it initially caused few deaths.³ A Reuters correspondent reported optimistically on 28 September that ‘in view of the fact that such a very large number of people have been affected, the fact that there has been only one death must be considered to be reassuring’ (Phillips 1984, p. 11). It proved to be a poor prediction of what was to come: more than a thousand black mine workers perished as a result of the disease within the following month. On the same day that the correspondent voiced his optimism, the train full of soldiers returning from the Great War left Cape Town, many of them infected with the more virulent strain. Disembarking at almost every station, they spread the virus rapidly across the Cape Province. The Bunu family was one of the unlucky ones to be infected very early.

Deaths like Wills Bunu’s were recorded by the government, as required by the 1894 Act. A family member or friend would visit the appointed official and complete a death certificate. Wills Bunu’s death certificate is shown in Figure 2. It includes personal particulars like name, age, race and occupation, and the date, place and cause of death, ‘Spanish influenza’. It also has the name of the ‘informant’, the person who registered the death and was required to sign the certificate, in this case Wills Bunu’s uncle, Matthais, who could not write so marked the certificate with a cross. We use this, as economic historians do, as an indicator of literacy (A’Hearn et al. 2009). The space for a doctor’s name is left blank. Not all certificates recorded a doctor’s name. This was our variable of interest, a proxy for whether the deceased had access to healthcare. We elaborate below.

We have transcribed 39,482 of these death certificates for 15 Cape Province districts: Cradock, Elliot, Fraserburg, Komga, Kuruman, Mafeking, Oudtshoorn, Paarl, Peddie, Queenstown, Riversdale, Sea Point, Stellenbosch, Stockenstrom and Vryburg. The districts were selected to include all geographic regions of the province, smaller and larger districts, and high and low mortality rates. It is these death certificates that form the basis of our analysis.

The process of registration was, of course, not always perfect. Although no fee was required to register a death, it is likely that, for a variety of reasons, not all deaths were recorded. The sample of deaths as recorded by the certificates may thus be systematically different from the number of all deaths, introducing some selection bias. We account for this bias by introducing district dummies that act as fixed effects in our

³The compound system refers to the hostels in which mine workers lived while employed on the mines.

analysis and hold constant the district-specific rates of under-reporting of deaths.

JP Duminy, a student at the time at the University of Cape Town, explained in an interview in 1978 how haphazard the process of death registrations was during the pandemic:

[W]e read of people dying like flies all around the Peninsula. But my father was a veldkornet, an officer appointed by the government to do registrations of births and deaths among other things. And I could see people coming, I could see through my window, coming in one after one, sometimes in *klompies*, groups of 3, 4, 5 to come and register deaths. At one time there must have been a queue of 16, 17, 18 people outside his little office for the district [of] the Tygerberg. And he told me, I did not see them actually dying while standing in the queue, but he told me that two, I can remember, who had dropped down and he brought them into his office and there they died. How they managed to move I don't know because you feel – well, I felt as weak as a kitten. Anyway, there they were. And it was day after day, day after day, just registering deaths' (Phillips 2018, p. 15).

The speed with which the disease spread was one reason for the overburdened registration system. Queenstown, where the Bunu family lived, is a prime example. On 10 October, the day that Nosamon Bunu died, less than two weeks after the train carrying the soldiers from Cape Town arrived in Queenstown, more than 20 residents of the district also died, as can be seen in panel (a) of Figure 3. Three weeks after the soldiers' arrival, fatalities reached a peak of 80 a day in Queenstown. Also striking is the ages of the deceased. The Spanish flu disproportionately affected young adults, a trend clearly discernible in panel (b) of Figure 3, and in Figure 5. The reason for this 'W' rather than 'U' age-distribution is debated. Most argue it was because the virus triggered an immune over-response – a 'cytokine storm' (Liu et al. 2016) – others have suggested it was the result of exposure to an earlier influenza strain that caused a dysregulated immune response (Gagnon et al. 2013; Oxford and Gill 2018; Woo 2019). Regardless of the scientific explanation, what is important to us is that the age profile of the district would have had an effect on mortality rates.

Whatever the reasons for district-level variations in mortality rate, it is clear that our source allows us to investigate the *individual-level* characteristics that are associated with Spanish flu deaths. This is in contrast to the existing historiography of the Spanish flu in South Africa which has, until this point, only made use of the official statistics released in the Union government's Report of the Influenza Epidemic Commission, and statistics from the Official Year Book of the Union (Phillips 1984). Like ours, these statistics show that the burden of the disease fell more heavily on people living in densely packed living quarters and black and coloured people than on white people. Phillips (1984) cites data on deaths from the Spanish flu in Kimberley showing that seven times as many black workers living in the De Beers compounds died as white residents of the town: 22.4 deaths per 100 of the former as compared with 3.1 per 100 of the latter.

[B. & D. 2.]^R

Form of Information of a Death :  **Act No. 7 of 1894 (Cape).**

WARNING.—The penalties for false statements wilfully made are the same as those for perjury. Anyone who loses a COMPLETED registration form is liable to a penalty not exceeding £5.

DECEASED—

1. Christian Names and Surname (a) Wills
Bunu

2. Sex Male

3. Usual place of Residence Cimazile, Ward 6, Queenstown

4. Age Six Weeks 9 days

5. Race Fingo 5A Birthplace (b) Cape Province

6. Whether Single, Married, Divorced or Widowed (c) Never Married Single

7. Occupation Agriculturist

8. Date of Death Twenty sixth October 1918

9. Place of Death Cimazile, Ward 6, Queenstown

10. Intended Place of Burial Cimazile, Ward 6, Queenstown

11. Causes of Death Spanish Influenza

11A. Duration of last Illness One month

12. Medical Man's Name

INFORMANT—

13. Original Signature [or Mark] Matthews ^{His} Bunu

14. Qualification Uncle present at death

15. Residence Cimazile, Ward 6, Queenstown

Signed in my presence on this day of 19.....

(This space intended for Rural Area Reports) Witness (d)

The following spaces are reserved for the use of Assistants for Urban Areas, and of the Deputy Registrar. No one else should fill them up

When Registered 27 October 1918 Sub-district of Haekney
(Signature) Chas Drumming Asst. to Deputy Registrar (Urban Areas)

WHEN REGISTERED 28th Oct 1918 DISTRICT OF QUEENSTOWN
(Signature) V. A. Clifton DEPUTY REGISTRAR. NO OF ENTRY 1230

N.B.—If the certificate of a Medical Practitioner is produced, the causes of Death and Duration of Illness must be recorded in the Registration Book by the Deputy Registrar and Assistant to the Deputy Registrar (Urban Areas) as stated in such Certificate, which is to be attached to the form.

(a) Insert also in case of a married woman the maiden surname.
(b) Insert the name of the Country, State or Colony where born.
(c) If married, divorced or widowed, state on this line the total number of children deceased has had.
(d) When a Rural Assistant, Field-cornet or Police Officer writes out form for Informant, he should add the words "Form written out by me" and sign as "Rural Assistant," "Field-cornet" or "Police Officer," as the case may be.

NOTE.—If Informants in Municipalities, V.M. Boards and Special Urban Areas, do not appear personally before D.R. or A.D.R., proper Declarations must be completed and attached hereto. In Rural Areas, Informants can report in three ways—one being under Declaration. Medical Certificates essential in Urban Areas.

(*) As amended by Government Notice 527 of 26 May, 1905.

DP: 148 00 000 4/18 S.A. Electric DS: 12

Figure 2: Death certificate of Wills Bunu

Source: Image 1101, Queenstown, 1918, South Africa, Cape Province, 'Civil Deaths, 1895-1972,' database with images, FamilySearch; National Archives Pretoria

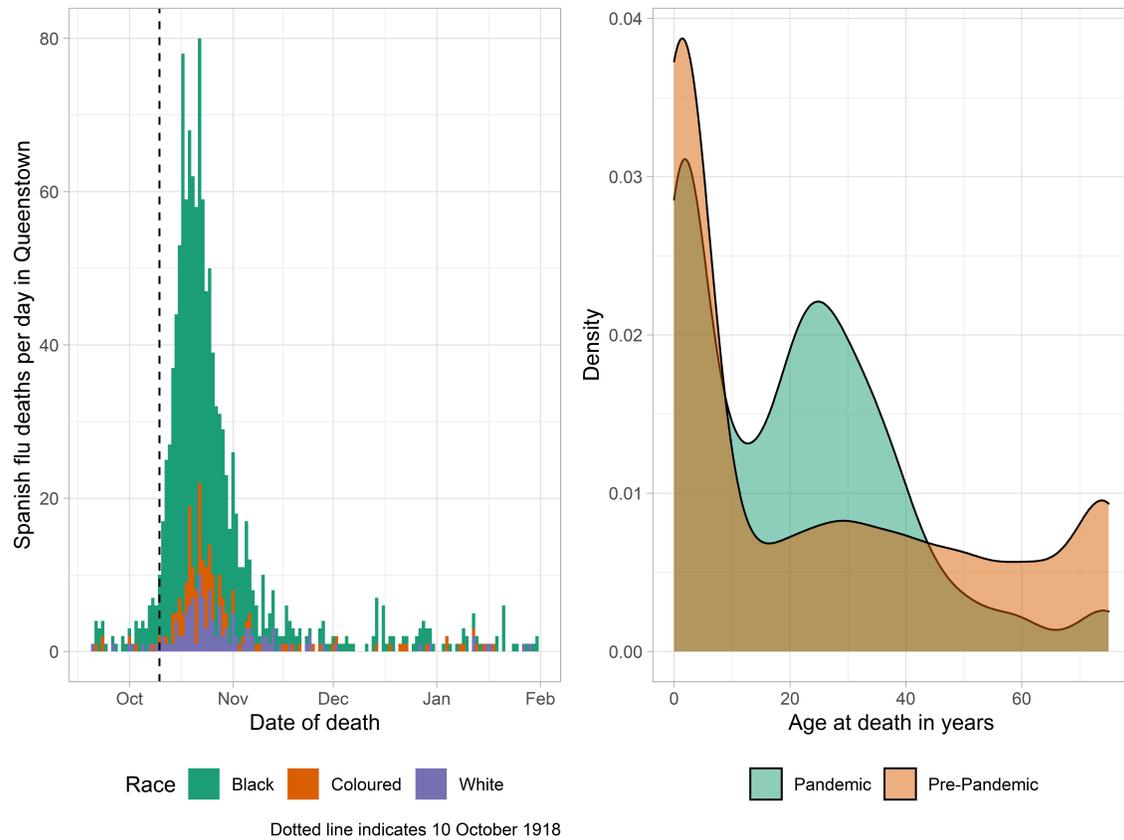


Figure 3: Panel (a): Spanish flu deaths in Queenstown by date. Panel (b) Deaths in Queenstown by age

Source: Cape Province, 'Civil Deaths, 1895-1972,' database with images, FamilySearch; National Archives Pretoria; own calculations, see text.

District	Total Deaths		Deaths per 1000 residents	
	Phillips	Death Certificates	Phillips	Death Certificates
Cradock	499	511	29.49	24.61
Elliot	370	414	45.56	47.84
Fraserburg	134	112	13.10	16.64
Komga	718	552	43.83	33.75
Kuruman	226	255	12.52	14.55
Mafeking	793	597	30.90	24.30
Oudtshoorn	397	471	11.04	12.78
Paarl	840	814	26.48	26.77
Peddie	1368	786	65.46	37.53
Queenstown	1312	1094	38.97	33.25
Riversdale	65	44	2.89	4.30
Stellenbosch	714	684	29.50	32.60
Stockenstrom	663	663	63.10	62.33
Vryburg	332	257	15.06	11.89

Table 1: Comparison of official records and death certificate data on Spanish flu deaths in 14 districts of the Cape Province. Note that the figures quoted by Phillips include Sea Point within greater Cape Town and hence Sea Point is excluded from the comparison above.

Source: Phillips (1984); Cape Province, ‘Civil Deaths, 1895-1972,’ database with images, FamilySearch; own calculations, see text.

The numbers of deaths in the official records do not match the numbers in our death certificates. Table 1 compares the absolute number of deaths and mortality rates as collected by Phillips (1984) with the same figures calculated from the death certificates for 14 of our 15 districts. This is probably a function of under-reporting of black deaths, particularly in rural areas, throughout the sample period. In the eastern part of the Cape Province, the death certificates show fewer deaths than the official records show. For Paarl and Oudtshoorn, in the western part of the Cape Province, the official records and death certificate data are largely congruent. The discrepancies are not surprising. Contemporary observers, such as JP Duminy, quoted above, were aware of the large-scale underreporting of deaths. Even the official report noted that the ‘omission to register deaths is not to be wondered at when one considers the widespread incidence of the disease, when whole families, and in many cases the greater portion of whole communities were incapacitated ... and the organization of civil administration was to a large extent suspended’ (Phillips 1984, p. 288).

Our records suggest that the selection bias is towards black and rural inhabitants. This has two implications. First, we include district fixed-effects in all our specifications. These effects should not only pick up district-specific responses to the virus, but should also include district-specific variation in reporting rates. Indeed, the coefficients on our estimates support the latter interpretation. Second, where there may be additional selection bias based on unobservable characteristics (distance from the civil official’s office,

for example), we note that our coefficients are lower-bound estimates of the true size of the effect of the pandemic. The bias would always tend to deflate rather than inflate our coefficients.

There are, however, advantages to using death certificates rather than the official, district-level, mortality rates, the most obvious being that the individual characteristics available to us in the death certificates help to explain variation both between and within districts. Race, gender, occupation, literacy and age are all important in predicting mortality. The Spanish flu deaths as recorded by the 39,482 death certificates across 15 Cape Province districts are shown on the map in Figure 4, along with the railway lines of 1907.⁴ The size of each bubble corresponds to the number of deaths in the district. In our records, Queenstown recorded the most: 1,094 during pandemic.

Race is reported in the death certificates, although it requires substantial standardisation. We classified the groups reported in the death certificates into three categories: white, coloured and black. In the western part of the province coloured deaths were in the majority, while in the eastern and north-eastern parts most were black. This broadly reflects the population shares in those areas.

Figure 5 shows the mortality rates from Spanish flu by age, gender and race group, aggregated by geography. The 1911 census data does not distinguish between black and coloured residents when reporting age, so we grouped these two and compared the mortality rates with that of white residents.

In the eastern parts of the Cape Province, black and coloured females of working age suffered more severely. This contrasts with the claim by Phillips (1988) that men had higher mortality rates than women. He offers at least two reasons for this: as men left the home for work they were likely to come into more frequent contact with people carrying the disease, and as breadwinners they could not afford to take time off work for bed rest. These are plausible reasons, but our findings suggest they may have been true only for the white population – or for black and coloured residents in the eastern and north-eastern parts of the province. Figure 5 also shows that child mortality was significantly higher for black and coloured children. White children between the ages of 0 and 4 had a mortality rate of 1 death per 100; for black and coloured children it was above 6. We conjecture that this indicates large differences in access to healthcare.

4 Access to healthcare

A major advantage of using death certificates was that it enabled us to construct a novel way of estimating healthcare access: whether a doctor’s name is present on the death certificate. Not all certificates recorded a name. The certificates provide some clue as to why this is the case. The small print on the certificates

⁴In classifying Spanish flu deaths we use a simple text-based classification system that includes as Spanish flu deaths during the pandemic period (‘Black October’, November, December, 1918, and January, 1919) where cause of death is listed as ‘Spanish influenza’, or other terms which are highly correlated with words ‘Spanish’ and ‘influenza’, such as ‘pneumonia following on an attack of influenza’ and ‘epidemic influenza’.

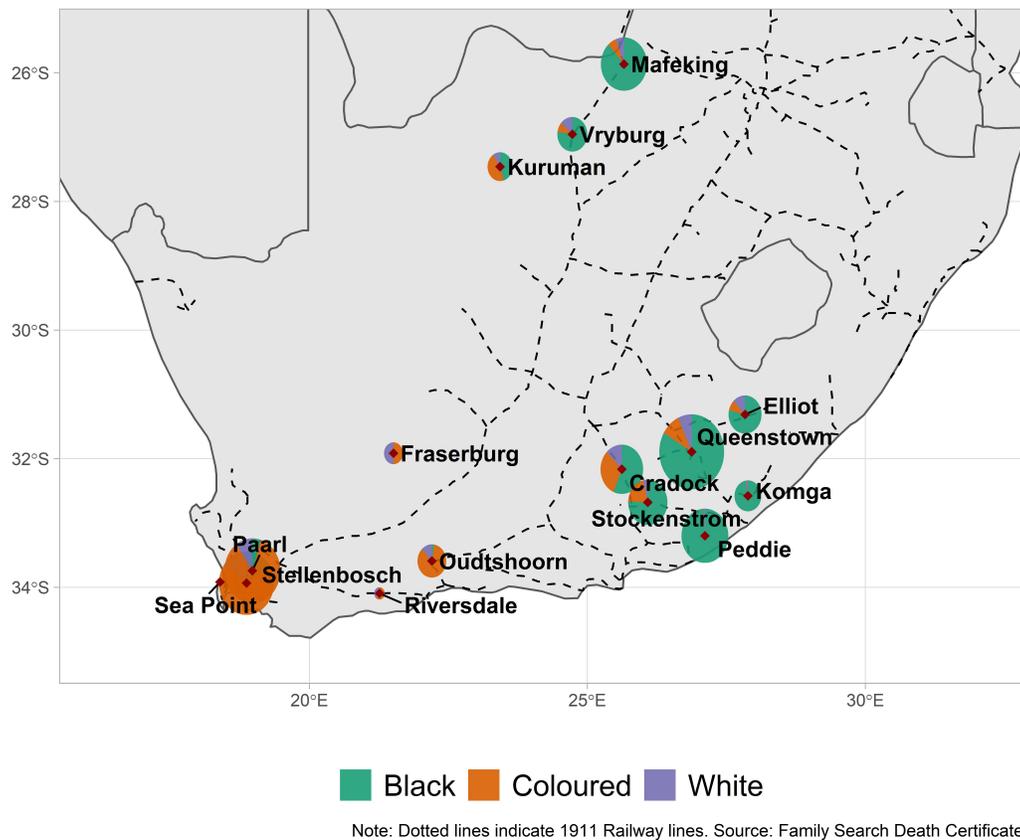


Figure 4: Spanish flu deaths by race in the 15 districts in our sample

Source: Cape Province, 'Civil Deaths, 1895-1972,' database with images, FamilySearch; own calculations, see text.

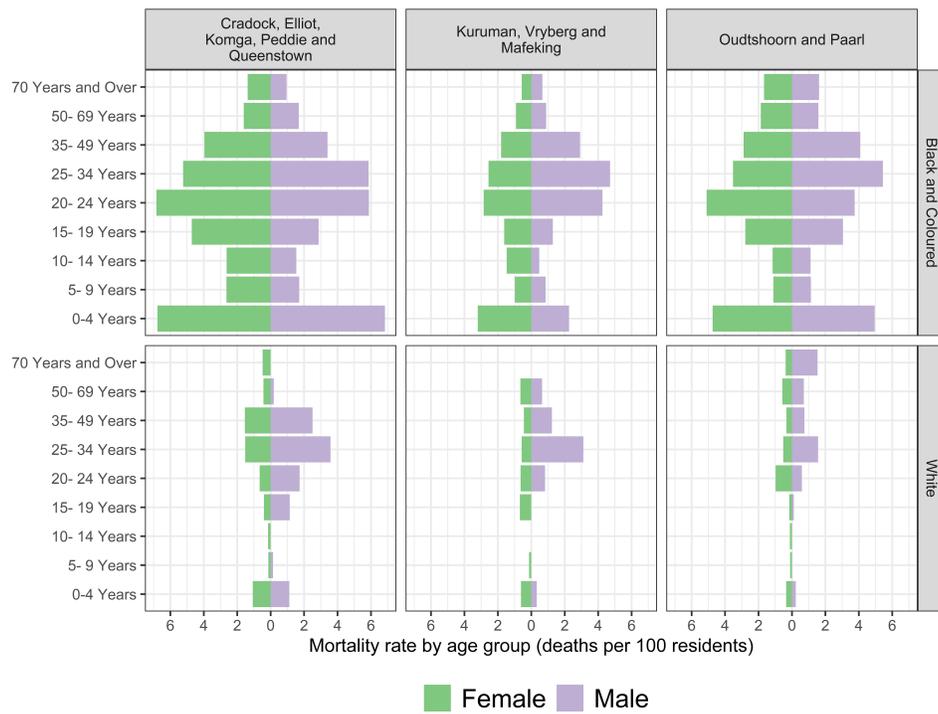


Figure 5: Spanish flu deaths by age, gender and region

Source: Cape Province, 'Civil Deaths, 1895-1972,' database with images, FamilySearch; South African census 1911; own calculations, see text.

District	Race		
	Black	Coloured	White
Cradock	52.87	68.55	90.60
Elliot	21.32	57.14	51.18
Fraserburg	33.33	40.35	48.89
Komga	22.41	66.67	88.24
Kuruman	10.41	11.89	46.97
Mafeking	9.71	57.14	63.71
Oudtshoorn	63.70	64.14	82.82
Paarl	87.18	86.84	96.84
Peddie	23.15	40.00	85.71
Queenstown	43.10	69.78	90.37
Riversdale	25.00	50.54	60.00
Sea Point	90.85	87.50	100.00
Stellenbosch	86.79	88.51	95.14
Stockenstrom	13.64	20.43	43.06
Vryburg	33.92	51.35	60.87

Table 2: Percentage of death certificates with doctor named, January 1915-September 1918

Source: Cape Province, ‘Civil Deaths, 1895-1972,’ database with images, FamilySearch; own calculations, see text.

says: ‘If the certificate of a Medical Practitioner is produced, the causes of Death and Duration of illness must be recorded in the Registration Book by the Deputy Registrar and Assistant to the Deputy Registrar (Urban Areas) as stated in such Certificate, which is to be attached to the form’. We therefore assume that the lack of a doctor’s name implies that the deceased was not attended to by a doctor immediately before death. The certificates, of course, give us no information about the doctor or whether he treated the patient, or the kind and duration of the treatment, if any. It only tells us how long the deceased was ill. The proxy is therefore, at best, an indicator of access to healthcare and not a proxy for its quality or intensity.

There are at least two selection biases in using this proxy. First, we have records only of those who died during the pandemic and none on those who survived. Access to healthcare is presumed to decrease the chance of mortality, and mortality is the very feature that includes a person in our data set. While we acknowledge the selection into death that the presence (or absence) of a doctor can cause, our comparison of pandemic with pre-pandemic death certificates goes some way to mitigating this bias. Our comparative approach avoids making the strong assumption that selection bias is unimportant; instead, it only requires us to make the less drastic assumption that healthcare during a pandemic is not more (or less) likely to decrease the chance of death than in the pre-pandemic period. This may still be an assumption that causes concern. Because the disease was new, doctors had little experience or knowledge to draw on for its treatment. If anything, then, doctors may have offered poorer healthcare during the pandemic than before and after. Put

differently, the presence of a doctor’s name on a certificate during a pandemic would be less likely to indicate that the person received good healthcare than it would in normal times. More people with a doctor’s name on their death certificates should therefore be selected into the sample during the pandemic. This is not what we observe, as we show below, suggesting, again, that our estimates are, at best, underestimates of the true size of the coefficients.

The second selection bias is caused by not distinguishing between a doctor who provided care to a patient until their death and a doctor who determined the cause of death but did not examine or treat the patient while alive. The same may be true for nurses – with some patients receiving healthcare and others none. Clearly, in the first case the doctor’s signature would indicate healthcare had been received, but it would not indicate this in the second case. A way to mitigate this bias would be to classify all death certificates which recorded the name of the district surgeon as a death in which the doctor was not present. However, in the pre-pandemic period, district surgeons did also treat patients, as is evident in the death certificates with, for example, the inscription ‘Medicine From Assistant District Surgeon O’Hock’ for a patient who was treated for whooping cough in Kuruman in 1917. This, again, supports our approach of comparing pandemic with pre-pandemic death certificates.⁵

Table 2 shows the percentage of the sample for which a doctor’s name is present on the death certificates from the 15 districts, by race, between January 1915 and just before the pandemic began in September 1918. It is evident that in every district except Elliot, white residents had a higher likelihood of a death certificate with a doctor’s name present, indicative of superior access to healthcare before and after the pandemic. There is large variation across the districts: the 97% of all white residents in Paarl, whose certificates bear doctors’ names suggesting they had access to healthcare, stands in sharp contrast to the corresponding 10% of black residents of Mafeking. Even within districts there could be large variation. Compare, for example, the 86% for white residents to the 23% for black residents in Peddie. In other cases, like Stellenbosch, differences are much smaller.

Comparing the pre-pandemic period with the pandemic period, we found that the Spanish flu exacerbated these racial inequalities. During the pandemic, strain on the healthcare systems of the districts meant that fewer death certificates were signed by a doctor. We interpret this to mean that the Spanish flu caused a higher demand for health services that could not be met by the health system. Access to healthcare declined. In Figure 6, we show the change in access to healthcare by age, race and gender, outside and during the pandemic. All groups had worse access to healthcare during the pandemic than before, but it is particularly striking how black and coloured people, who had worse access than the other groups before the pandemic, lost access during the pandemic. Further, there is a difference by sex and age: black and coloured women,

⁵Image number 89, Kuruman, 1917, South Africa, Cape Province, ‘Civil Deaths, 1895–1972,’ database with images, FamilySearch; National Archives Pretoria.

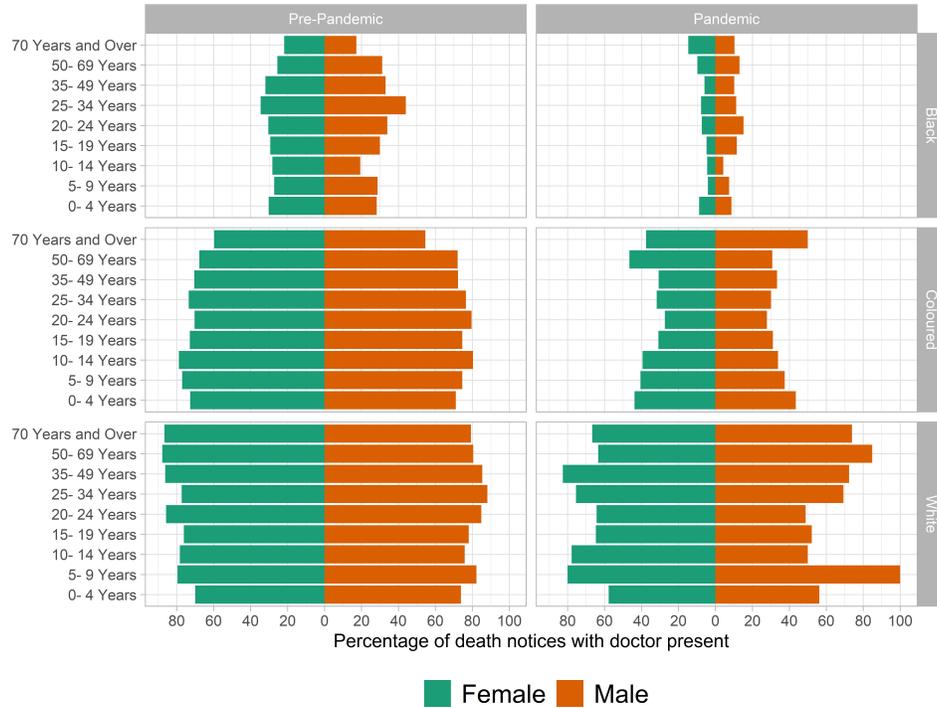


Figure 6: Presence of a doctor's signature on death certificate

Source: Cape Province, 'Civil Deaths, 1895-1972,' database with images, FamilySearch; own calculations, see text.

and to a lesser extent white women, of working age had some of the highest figures for access before the pandemic, but this reversed during the pandemic. Coloured men of working age had high rates of access before the pandemic but this reversed during the pandemic. Black men saw a significant drop in access across the age distribution, but picture is more mixed for white men.

To understand the widening gap in access to healthcare during the pandemic, we pooled the sample and used a logistic regression with the dependent variable as our proxy for access to healthcare – the presence of a doctor's name on the death certificate. We examined the effect on all deaths from 1915 to 1920. Our variables-of-interest are the interactions between the explanatory variables and the dummy for the pandemic period. The pandemic dummy is a time dummy which is equal to one for all deaths between 1 October 1918 and 31 January 1919.

We ran three specifications to further understand the ways in which access to healthcare changed. First, as a baseline, we included only race as an explanatory variable. In this regression, 'white' is the reference category. We included district dummies and clustered the standard errors at the district level. As explained earlier, these control, inter alia, for district-specific rates of under-reporting.

In the second specification, we included a dummy variables for male sex, age group, and occupation. Age 25 to 34 is the reference category. We classified the occupation variable into five HISCO (historical

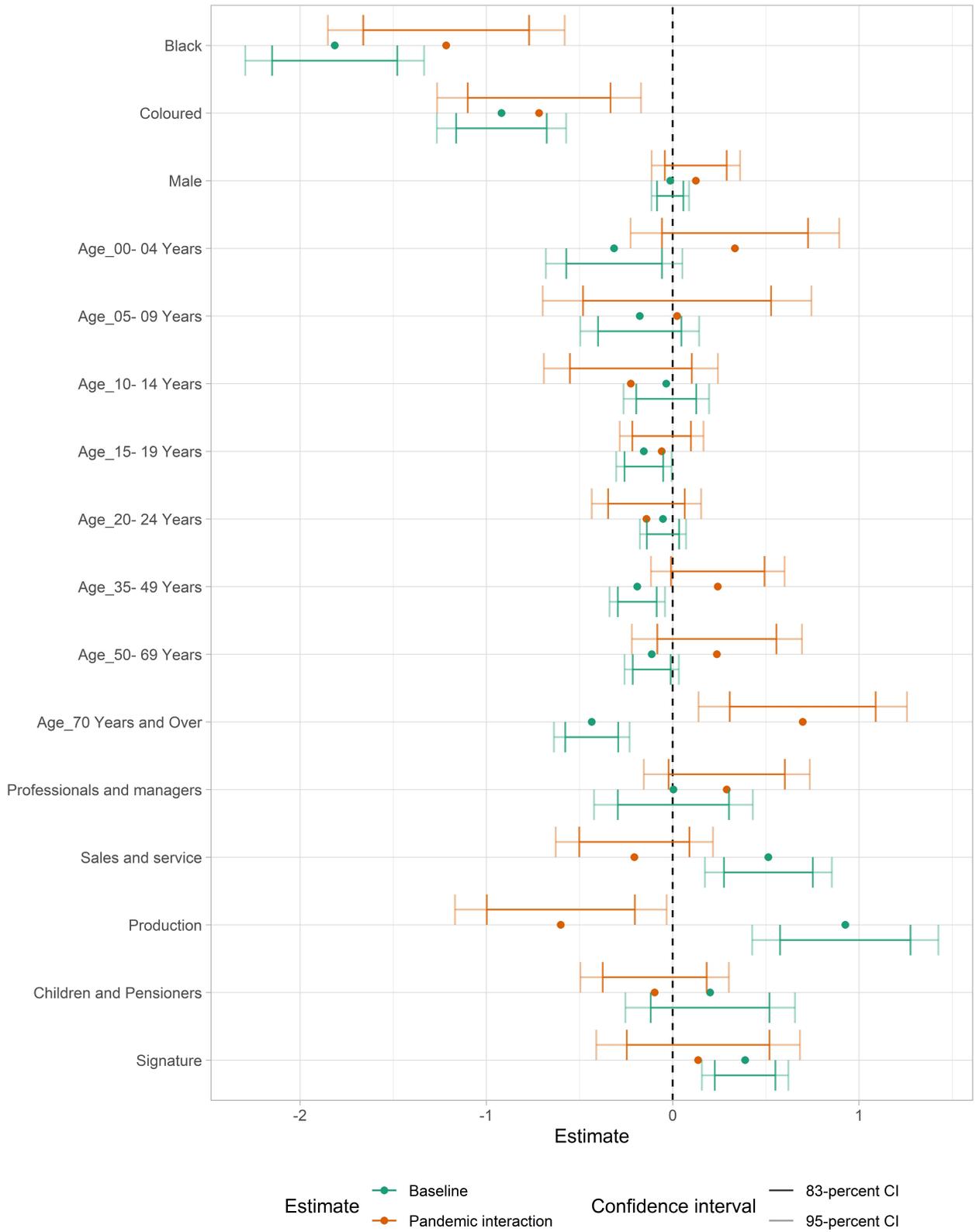


Figure 7: Coefficient plot of second regression specification

Source: Cape Province, 'Civil Deaths, 1895-1972,' database with images, FamilySearch; own calculations, see text.

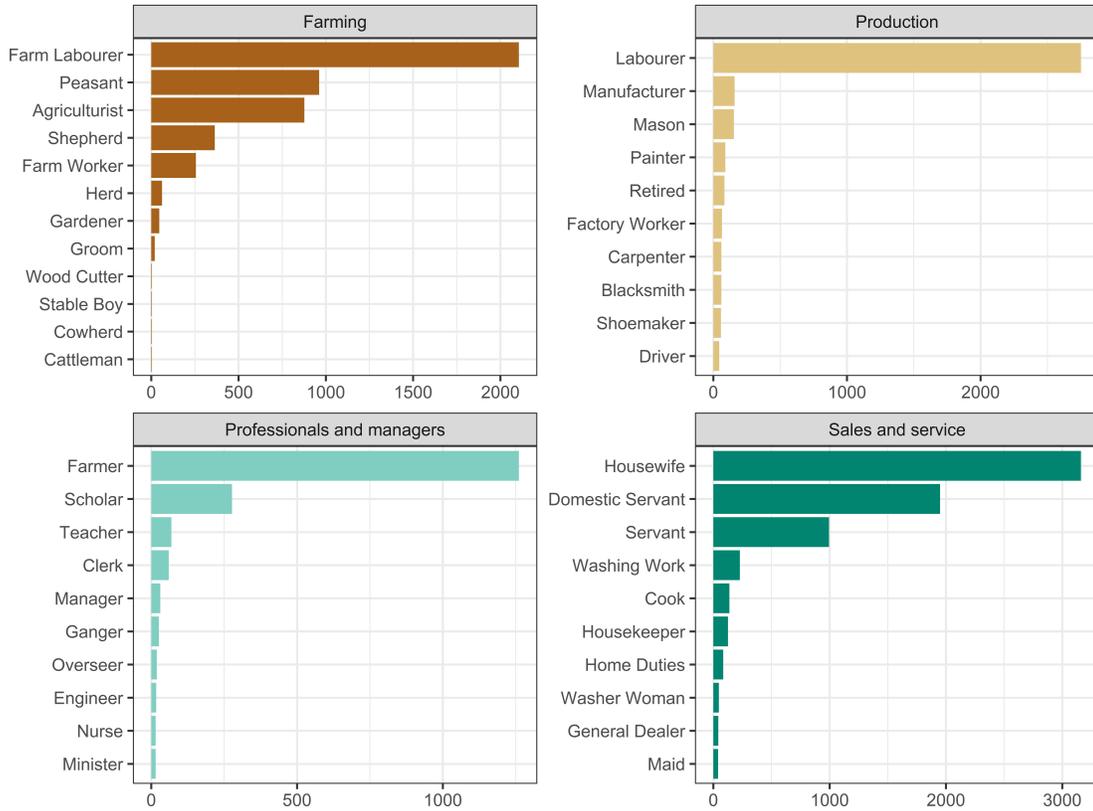


Figure 8: Occupations by HISCO group

Source: Cape Province, ‘Civil Deaths, 1895-1972,’ database with images, FamilySearch; Van Leeuwen (2002); own calculations, see text.

international classification of occupations) codes: professionals and managers, sales and service workers, farming, production, and children and pensioners (Van Leeuwen et al. 2002). Figure 8 shows the occupations in each category. Farming is the reference category.⁶ We also included a proxy for the literacy of the informant of a person’s death – based on whether the informant signed the death certificate or made a mark. The motivation is that the literacy of the family member or person in the deceased’s social circle who is the informant is likely to be correlated with socioeconomic status of the family or neighbourhood. We suspect that a signature on the certificate, rather than a mark, is likely to indicate better access to healthcare for the deceased and their family.

In Figure 9, we present a coefficient plot for the estimates from the second regression specification. The first key result is that the coefficient estimates for black and coloured in the baseline period as well as for the pandemic interaction are negative, large in magnitude, and significant at the 1% level. This means that

⁶Note that Farmers are included in the Professionals and Managers group because they are farm owners, while Agriculturists – the term predominantly used for black farmers – are either small farmers, sharecroppers, or workers on farms.

not only are these people worse off in terms of access to healthcare than white people with comparable characteristics in the baseline period, but, because the pandemic interaction is also negative in sign and significantly different from zero, their access to healthcare is reduced during the pandemic. This result holds constant for sex, age, occupation and literacy of the informant.

During the pandemic, with strained healthcare services, the probability of access to healthcare dropped for all black and coloured people more than for white people. We found evidence of this racial gap widening during the pandemic when we compared the averages in healthcare access from the logistic regression specifications for different groups. For example, white people living in Paarl, aged between 25 and 34, working on a farm, and whose informant was illiterate, saw greater access in the baseline period than both black and coloured people with the same characteristics. The average presence of a doctor, based on the presence of a doctor's signature on the death certificate, was 93.4% for whites, 85% for coloureds and 69.8% for blacks. During the pandemic the differences increased, such that average healthcare access was now 57.7% for whites, 21% for coloureds and 6.2% for blacks.

Not all healthcare inequalities worsened, though. Examining the effect of age, we found that the baseline coefficient estimates for the very young (four years and below) and the elderly (70 and above) indicated worse access to healthcare than for the reference category – those of working age. However, during the pandemic this effect was reversed. Holding constant the other covariates, we conclude that during the pandemic infants and the elderly had better access to healthcare. This is evidenced by the lack of overlap between the orange and green 83% confidence interval bars in Figure 7. For example, in the baseline period white females aged 25 to 34, working on a farm in Paarl, whose informant was illiterate, saw a probability of access to healthcare of 3.2 percentage points higher than those aged 70 and above, 93.4% as opposed to 90.2%. In contrast, during the pandemic and holding all other variables constant, those aged 70 and above saw a 6.3 percentage point higher probability of access to healthcare than those aged 25 to 34: 64% as opposed to 57.7%. This is significant at the 5% level.

Perhaps somewhat surprisingly given the way historians have emphasised the importance of gender, we found no gender effect for healthcare access during the pandemic. Interacted with race and the pandemic dummy, the coefficient on male is not significantly different from zero in the resulting regression.

In terms of occupation, workers in sales and service and in production (largely made up of labourers) had higher access to healthcare before and after the pandemic than the reference category – farming. We posit that the large magnitude of the coefficient on production is driven by the fact that many of these labourers were employed by the municipalities, often as firemen or general labourers, which could have given them better access to healthcare than farm workers, who lived further from the hospital, doctor's practice or house. It is interesting, therefore, to find that the largest *negative* coefficient during the pandemic is for

production workers. For the representative inhabitant of Paarl, employed in production, access to healthcare had decreased from 93% to just 22% during the pandemic. This sizeable decline suggests that class or social status may have been even more important than race in access to healthcare services. This sizeable change is statistically significant – the 83% confidence intervals do not overlap. Similarly, sales and service workers saw lower access during the pandemic, again with 83% confidence intervals that do not overlap.

As regards literacy in the family or social circle, we found evidence of a positive effect on access to healthcare before and after the pandemic. The coefficient on signature, the proxy for literacy of the informant, is positive and significant at the 1% level, indicating that those whose informant was literate were more likely to have received healthcare. Holding constant the observable characteristics noted above, moving from a death certificate with an informant’s mark to one with the informant’s signature increases the probability of a doctor’s name appearing on the death certificate from 21% to 31%. In other words, those whom we can assume to have had literate family members or friends were more likely to have sought or accessed healthcare prior to death. During the pandemic, however, we find no additional effect for literacy, shown by the fact that the error bars in Figure 7 overlap the dotted line at zero. A possible interpretation is that during the pandemic, with health services under strain and many people sick and dying, families were simply unable to obtain health care. The gap between those with literate informants and those without did not widen during the pandemic.

After controlling for age, sex, race, occupation and the literacy of the informant, we conclude that racial disparity in access to healthcare increased during the pandemic. Class or social status seems to have mattered even more than race: those who worked in production experienced a greater decrease in access to healthcare than workers on farms. This may be because production workers lived in densely populated towns rather than on farms. A significant finding is that for the very young and the elderly, access improved rather than worsened.

5 Reasons for the widening gap in healthcare access

Observing the individual observable characteristics that are correlated with the decline in healthcare access for some groups during the pandemic does not tell us *why* the pandemic had such a pronounced effect. There are many possible reasons, on the demand and the supply side. Immigrant status, death of a family member, exposure due to a particular occupation, and location are four likely demand-side reasons that our source material allows us to test. Finally, we tested one possible supply-side reason: discrimination by doctors.

Immigrants, those born outside the Cape Province, may have had poorer access to healthcare than locally born residents. To test this, we created five categories, using the birthplace on the death certificate: born in the Cape (the reference category), in one of the other three provinces of the Union, in the rest

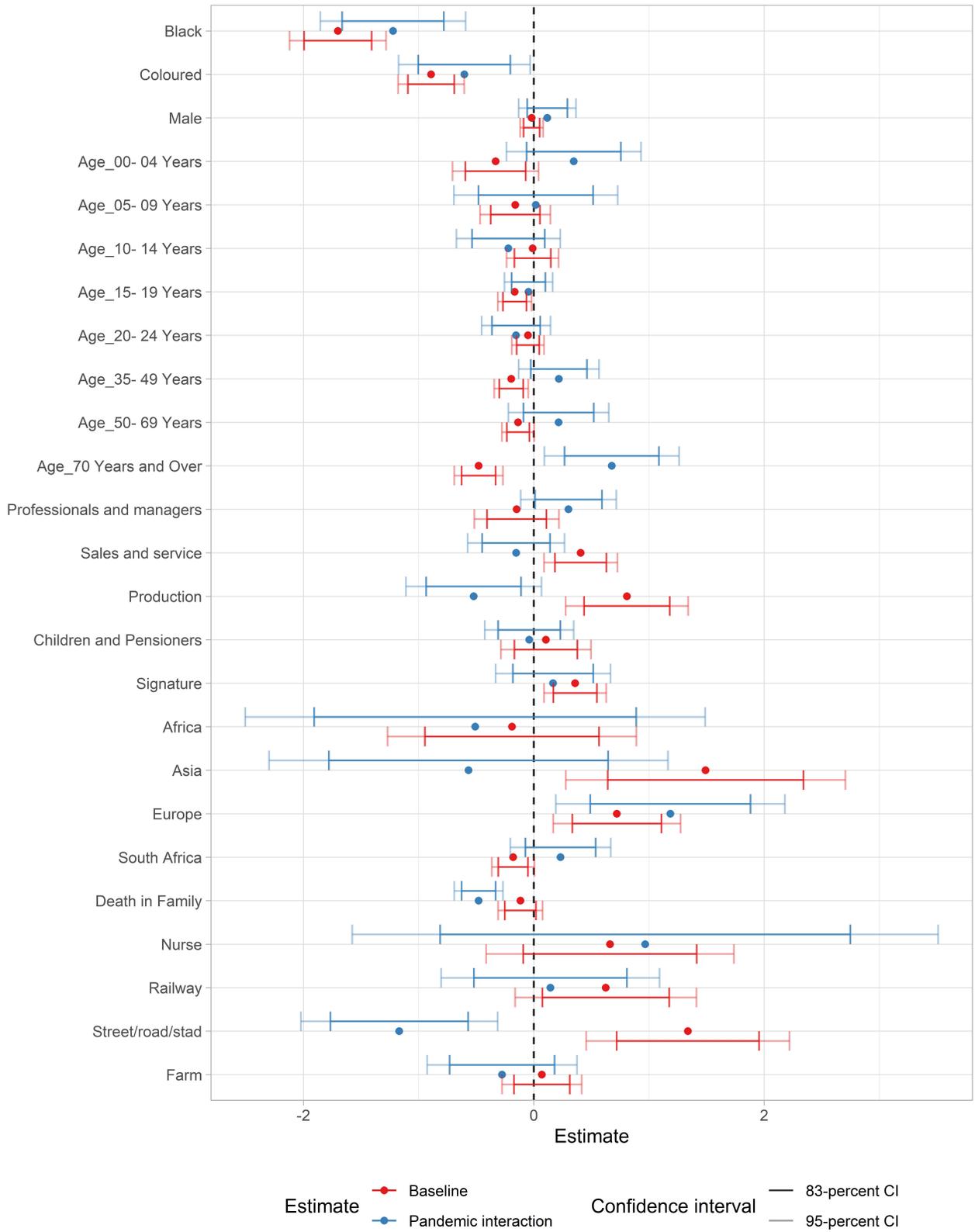


Figure 9: Coefficient plot third regression specification

Source: Cape Province, 'Civil Deaths, 1895-1972,' database with images, FamilySearch; own calculations, see text.

of Africa, in Europe, and in Asia. We hypothesize that for first generation immigrants from the rest of the African continent, social capital and cohesion were limited, inhibiting their access to healthcare. For European immigrants, almost all of whom were white, we hypothesize that since they entered the country through a populous city they had better access to healthcare than those born in the Cape Province.

Death of a family member could have had an effect on access to healthcare. A household is inevitably an important locus of disease transmission, as has been seen during the Covid-19 pandemic (Zhang et al. 2020). We posit that transmission was especially pronounced within households. Larger household sizes – and co-habitation practices – may therefore help to explain high mortality rates. We therefore included a dummy variable that proxies for disease spread within a family. The dummy is equal to one if we find a death of someone with the same surname (last name) within a week before the death of a person named in a death certificate. The implication is that, with infectious diseases, having a family member die is likely to put a person at risk of contracting, and dying from, the disease. Consequently, if the Spanish flu spread through a family, it is less likely that family members would be able to obtain healthcare, resulting in a death without the signature of a doctor appearing on the death certificate.

A particular occupation might have been a health hazard and affected a person’s access to healthcare. To identify the effects of particular occupations, we included dummy variables for if an individual (or parent in the case of a child) was employed as a nurse or on the railways. The hypothesis we test is that working in the healthcare system increased the access to healthcare for nurses both during and before the pandemic. The disease spread along the railway lines. Hence, we hypothesize that as the soldiers travelled along the railway lines and passed through the railway stations, railway employees would be among the first to contract the disease, and thus be more likely to have access to healthcare before they died than those who contracted it later.

A person’s location, rural or urban, would be expected to have an effect. We created some proxies for this. Because address data is not uniformly reported on the death certificates, and place names have changed over time (for example, Elliot is now ‘Khowa’), GIS-mapping was almost impossible. To work around this, we created two location proxies using words in the addresses: a dummy for ‘farm’ and a dummy for ‘street’, ‘road’, or ‘stad’ (town).

Figure 9 shows the coefficient plots for the final specification. The inclusion of these four novel predictors reveals that transmission of the disease within a family did have an effect on access to healthcare during the pandemic. The coefficient for pandemic interaction for ‘death in family’ is negative and significant at the 1% level. The interpretation is that for the representative inhabitant of Paarl, a member of the family dying within the previous week decreased the likelihood of accessing healthcare by one quarter, from 32% to 24%. In contrast, before and after the pandemic, the coefficient on ‘death in family’ was not statistically different

from zero, even at the 10% level. This supports our hypothesis that in normal times another death in the family in the previous week would have no effect on the family obtaining healthcare.

Birthplace was a significant predictor of access to healthcare before and after the pandemic, with a positive and significant coefficient on birth in Europe and Asia, relative to the Cape Province, meaning that immigrants from those countries had better access to healthcare than the Cape-born. Being born in one of the other three provinces of South Africa had no significant effect on access to healthcare, controlling for age, race, gender, occupation, address, and spread of disease within a family. Being born on the African continent, outside South Africa, had no significant effect on access to healthcare, despite negative coefficient estimates both during and before the pandemic. The difference in access between European immigrants and people born in the Cape increased during the pandemic, and was the largest positive effect we found, measured by the magnitude of the coefficient. Being born in Europe raised the probability of healthcare access during the pandemic from 32% to 79%, holding constant the other observables. This was significant at the 5% level, as is evident in the confidence intervals shown in Figure 9. In addition, controlling for place of birth decreased the magnitude of the coefficients on race, both during and before the pandemic.

Working on the railways had no effect on access to healthcare before and after the pandemic. During the pandemic, however, the coefficient on the railway dummy is positive, economically significant in magnitude and statistically significant at the 10% level. It seems therefore that, for the representative resident of Paarl, working on the railways increased the probability of having access to healthcare during the pandemic from 4.3% to 9.5%. As suggested above, this may be because railway workers contracted the disease early and obtained help before healthcare systems were overwhelmed. The median date of death for railway workers during the pandemic, six days earlier than for people with other occupations, supports this explanation.

The coefficient on the nurse dummy during the pandemic is larger in magnitude than the railway dummy, at 0.97, but not statistically significant. The reason for this is probably the small number of nurses in our sample, only 27. Perhaps surprisingly, we find no evidence that the pandemic had an effect on nurses' access to healthcare, but again this is probably an effect of the small sample.

The presence of the term 'farm' in an address, the rural location proxy, is not statistically significant in either outside or during the pandemic. The urban location proxy is highly significant in both periods, with large magnitude and *opposite* signs. Before and after the pandemic, in line with the hypothesis that town dwellers had better access to healthcare by virtue of proximity to the doctor's house or practice, the coefficient on the 'street', 'road', or 'stad' address dummy is positive and significant at the 1% level. For the representative inhabitant of Paarl, the presence of 'street', 'road', or 'stad' on the death certificate increases the likelihood of a doctor's name on the certificate from 62% to 87%. During the pandemic however, the 'street', 'road', or 'stad' dummy is negative, with a coefficient of -1.17, and decreases the likelihood of access

to healthcare for town dwellers by a magnitude of 10%, from 35% to 31%.

The regressions indicate that the death of another family member in the previous week means that the person whose death is on the certificate probably had decreased access to healthcare during the pandemic. The birthplace of an individual had almost no effect in widening or narrowing the gap in access that existed, with Europeans having greater access than those born on the African continent. The presence of an indicator for living in an urban area increased access to healthcare before and after the pandemic, but decreased it during the pandemic.

The increasing racial gap in access to healthcare was probably a function both of demand and supply side factors. On the demand side, we examined the ‘death in family’ dummy and the ‘street, road, stad’ address dummy, interacted with race in order to see how far these two factors were related to race. In Table 3, we show that the number of deaths where a family member had died in the previous week was higher during the pandemic than before and after, for all race groups. For whites this nearly doubled, for coloureds the increase is a factor of more than 3, and for blacks it is a factor of nearly 4. Despite this relatively higher level of clustering for black and coloured individuals, relative to whites, Figure 10 shows that the interaction of race with this variable is not significant (even with an 83% confidence interval) during the pandemic. Although the coefficient for coloureds is negative and the coefficient relatively large – perhaps suggesting that with more statistical power, the interpretation might be different – the most sensible interpretation of this result is that controlling for the same variables as before, access to healthcare is no worse for blacks and coloureds whose families had suffered a death in the previous week, compared with before and after the pandemic.

As regards location, Table 3 shows that our finding that urban location during the pandemic meant poorer access to healthcare was true for whites but not for coloureds and blacks. Figure 10 shows that more blacks in our sample lived in rural areas, which could explain the reduced likelihood of accessing healthcare in the period before and after the pandemic. This situation reverses during the pandemic. One interpretation of this, and a major concern raised earlier, is that those who died before and after the pandemic, whose death certificates are included in our sample, have in effect selected out of the sample of those who died during the pandemic. While this selection obviously reduces the size of the coefficient for black – from -1.7 to -1.1 – it is still large and statistically significant, suggesting that sample selection is not enough to explain the racial difference in access to healthcare.

To look for another explanation for the racial inequality in access to healthcare we considered the supply side. Might doctors have avoided treating the most vulnerable in an attempt to avoid being infected themselves? The death certificates reveal not only the presence of a doctor but also the duration of the deceased’s illness that the doctor recorded. This gives us a way to look for evidence of possible prejudice on

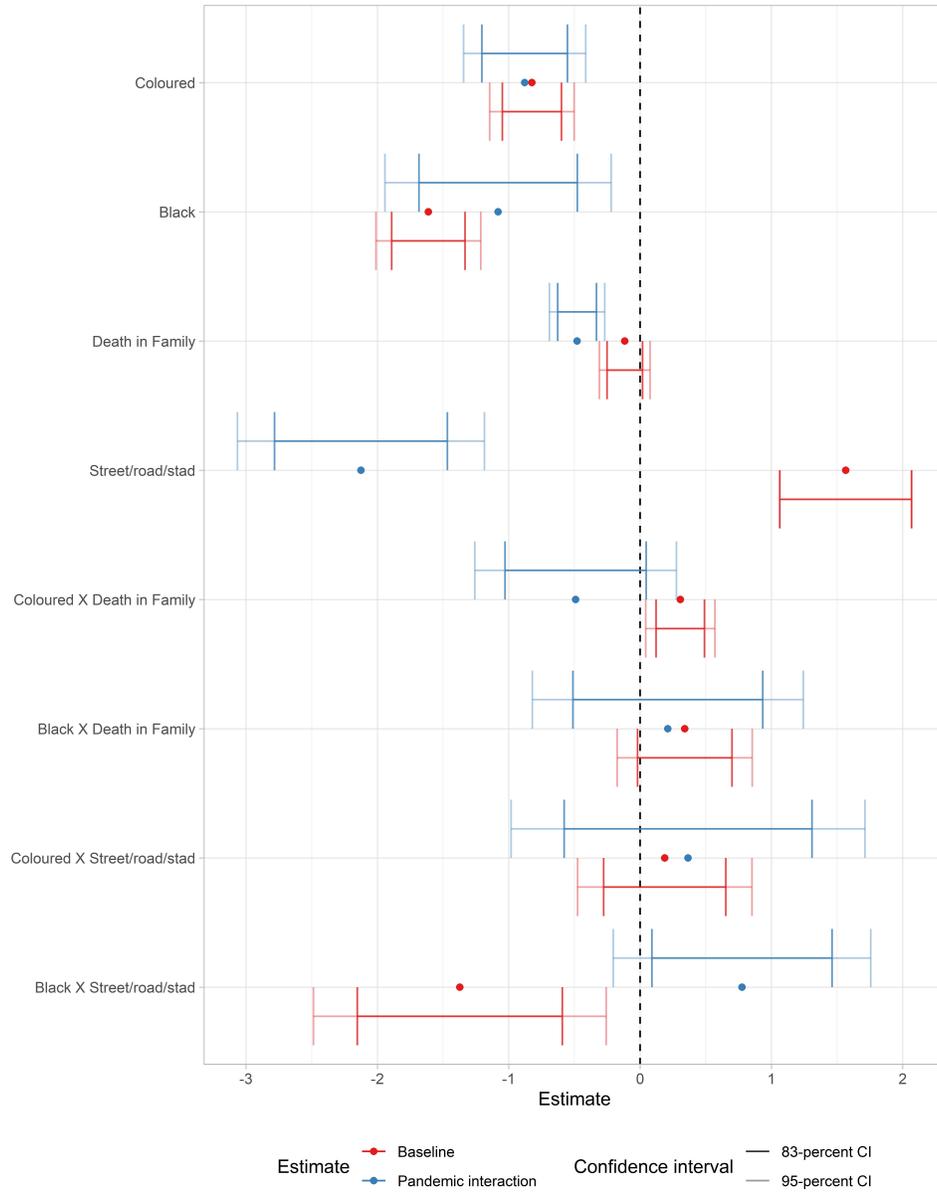


Figure 10: Coefficient plot for race interaction with demand side factors

Source: Cape Province, 'Civil Deaths, 1895-1972,' database with images, FamilySearch; own calculations, see text.

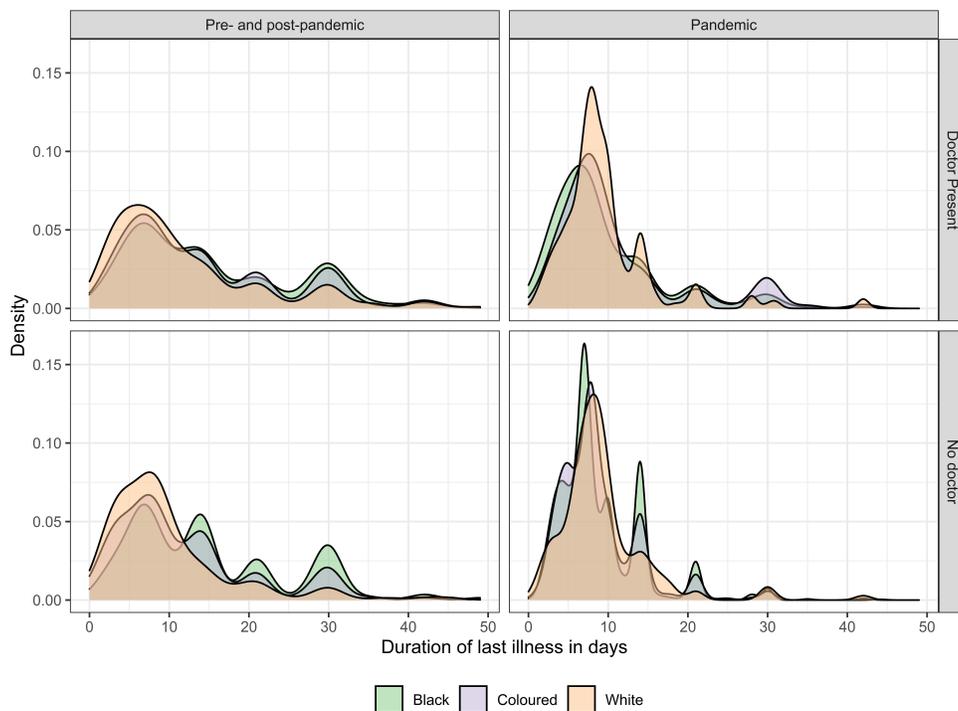


Figure 11: Density plot of duration of last respiratory illness in days by race and period

Source: Cape Province, ‘Civil Deaths, 1895-1972,’ database with images, FamilySearch; own calculations, see text.

Race	Recent death in Family		Street, Road, Stad Address	
	Pre-Pandemic	Pandemic	Pre-Pandemic	Pandemic
White	7.51	13.51	22.18	20.67
Coloured	9.21	28.77	21.28	24.84
Black	4.87	19.20	4.29	5.69

Table 3: Percentage of deaths recorded where dummy variable is equal to one

the part of doctors that might help explain the increased racial disparity in access to healthcare during the pandemic.

In Figure 11 we show the distribution of the reported duration of last illness for individuals dying of respiratory illnesses, by period and the presence of a doctor. We find pronounced lumping of the duration of last illness for blacks and coloureds who died during the pandemic period around the 7, 14, and 21 day marks, shown in the bottom right panel. This could be a function of the informant telling the person registering the death that the deceased has been sick for a week, a fortnight, or three weeks. Alternatively, it could indicate misreporting through discrimination on the part of the registrar, who lumped the duration of last illness figures of for blacks and coloureds rather than being precise. This lumping appears less prominent where a doctor’s signature is on the death certificate, but it is still evident. It is just possible this could

Doctor	Pre-Pandemic		Pandemic	
	Black & Coloured	White	Black & Coloured	White
Bester	31.75	55.56	41.58	77.78
Brodziak	69.23	77.78	24.74	66.67
Cross	41.41	77.78	42.86	47.01
De Jager	31.48	77.78	40.74	55.56
Don	51.52	38.89	42.86	55.56
Du Plessis	22.41	37.04	34.98	55.56
Grobbelaar	32.76	55.56	21.37	37.91
Heynes	30.16	44.44	41.67	58.59
Lewin	35.27	49.57	37.37	30.86
Stusser	39.22	33.33	34.19	66.67

Table 4: Modified Whipple index for recorded duration of last illness

Source: Cape Province, ‘Civil Deaths, 1895-1972,’ database with images, FamilySearch; own calculations, see text.

indicate discrimination on the part of doctors, suggesting they were less willing to treat black and coloured individuals or, more likely, were less willing to give them the best healthcare, and consequently assigned durations of last illnesses for these people heaped at certain values.

One way to quantify the possible discrimination is to use a modified Whipple index to measure digit preference (Szoltysek et al. 2018). This creates an index between 0 and 100, where 100 is absolute preference for one digit, and 0 is an equal distribution among the digits. With a large enough sample size, we might suspect discrimination on the part of doctors if duration of last illness is heaped at specific points rather than distributed more uniformly, as expected.

Table 4 shows our modified Whipple index for duration of last illness as recorded by the ten doctors who signed the most certificates during the pandemic in our sample. We find no convincing pattern that might suggest discrimination by the doctors.⁷ This accords with anecdotal evidence of the time. Following Black October, for example, Dr Abdullah Abdurahman, a leader in the coloured community, was happy to express ‘the thanks of the coloured community to those citizens who had so ungrudgingly and unselfishly helped them during the Epidemic’ (Phillips 1984, p. 58). Both the quantitative and qualitative evidence suggests that prejudice cannot explain the worsening access to healthcare that black and coloured individuals experienced. We are ultimately unable to fully explain the negative and statistically large pandemic and race interaction – evidence of racial inequality in healthcare access that was exacerbated by the pandemic.

⁷One concern is the small sample size in each period. In other words, without enough observations to provide a normal distribution against which to compare the pandemic distribution, we can draw no solid conclusion.

6 Conclusions

It seems likely that not a single member of the Bunu household who died in October 1918 received professional healthcare, since none of their death certificates include the name of a doctor. Had they died before or after the pandemic, the likelihood of access to healthcare would have been substantially higher. This is because the Spanish flu not only exposed the healthcare sector's racial inequalities but exacerbated them.

The 39,482 death certificates we examined reveal a large variation in mortality rates across the Cape Province, before, during and after the pandemic. One reason for this, we argue, was uneven access to healthcare. To test this hypothesis we used a novel indicator of access to healthcare – whether a doctor's name appears on the death certificate. This approach revealed inequalities along age, racial, occupational and educational lines before the arrival of the Spanish flu; early twentieth-century healthcare in South Africa reflected the high inequality within society. We then interacted these inequalities with a dummy for the period of the pandemic. Access to healthcare worsened along racial lines – a finding we expected and which resembles similar findings for the Covid-19 pandemic – but other inequalities reversed during the pandemic, notably age, location and occupation. We found no effect for gender, immigrant status or literacy. These surprising results challenge the notion that all forms of inequality are exacerbated during a crisis.

We attempted to find the reasons for the widening racial gap in access to healthcare during the pandemic. We found that household health matters. If someone with the same family name had died in the week before a death was recorded, a doctor's name was less likely to appear on the death certificate. It may be that a previous death made it more difficult for the family to access healthcare. Despite (or, perhaps, because) the disease spread along the railway lines, railway workers seem to have been more likely to have access during the pandemic than before and after. This may be because they contracted the disease first when healthcare was less constrained or because they were more urbanised and more mobile. We do not find any answers to whether healthcare workers were more or less likely to have access to healthcare, because the number of certificates in our sample recording healthcare as an occupation is too small to make any meaningful inferences. Possibly our most revealing result is that the benefit of urban living in having access to healthcare was reversed during this pandemic. Finally, we attempted to explain the worsening inequalities from the supply side. Using a modified Whipple index to calculate the extent of discrimination by doctors, essentially looking for evidence of careless numerical reporting of length of illness as a possible indication of lack of care, we found no obvious evidence of discrimination, although here the sample size is too small to support any meaningful inferences. Anecdotal evidence of the time do, however, confirm that doctors were not guilty of racial discrimination in treating patients.

Like the Spanish flu of 1918, Covid-19 has exacerbated racial inequalities. In the United States, Covid-19

deaths have been disproportionately higher among African Americans than in the population overall (Dorn et al. 2020; Mein 2020). One likely reason is uneven access to healthcare. This was true of the H1N1 ('swine flu') pandemic of 2009. Quinn et al.(2011) found that the risk of exposure to that virus was highly correlated to race and ethnicity, and that this was largely a result of the disparities in healthcare for Spanish-speaking Hispanics and African Americans. Blumenshine et al. (2008, p. 711) warned that racial disparities would become acute under pandemic conditions: 'the likelihood of substantial disparities in access to timely and appropriate care under influenza pandemic conditions is high, given long-standing and persistent disparities in access to medical care'. Our findings in a setting very different from that of the twenty-first century United States confirm that race matters in explaining health outcomes during a pandemic. But they also contradict the belief that all inequalities are exacerbated.

For future research, our findings indicate that more can be done to study health inequalities, both in the past and today. Innovative sources and methods, as we have demonstrated, can suggest reasons for unequal health outcomes. As a large body of research has pointed out, race has implications for people's health. Our approach may help to explain why this is so. Only when these inequalities are understood can appropriate policies be designed to help deal with a pandemic.

References

- Abedi, V. et al. (2020). "Racial, economic, and health inequality and COVID-19 infection in the United States". In: *Journal of racial and ethnic health disparities*, pp. 1–11.
- A'Hearn, B., J. Baten, and D. Crayen (2009). "Quantifying quantitative literacy: Age heaping and the history of human capital". In: *The Journal of Economic History* 69.3, pp. 783–808.
- Alsan, M. and M. Wanamaker (2018). "Tuskegee and the health of black men". In: *The quarterly journal of economics* 133.1, pp. 407–455.
- Blumenshine, P. et al. (2008). "Pandemic influenza planning in the United States from a health disparities perspective". In: *Emerging infectious diseases* 14.5, p. 709.
- Burger, R. and C. Christian (2020). "Access to health care in post-apartheid South Africa: availability, affordability, acceptability". In: *Health Economics, Policy and Law* 15.1, pp. 43–55.
- Cuevas, A. G., K. O'Brien, and S. Saha (2016). "African American experiences in healthcare: "I always feel like I'm getting skipped over"." In: *Health Psychology* 35.9, p. 987.
- Dorn, A. van, R. E. Cooney, and M. L. Sabin (2020). "COVID-19 exacerbating inequalities in the US". In: *The Lancet* 395.10232, pp. 1243–1244.
- Fisker, A. B., A. Rodrigues, and S. Helleringer (2019). "Differences in barriers to birth and death registration in Guinea-Bissau: implications for monitoring national and global health objectives". In: *Tropical Medicine & International Health* 24.2, pp. 166–174.
- Fothergill, A., E. G. Maestas, and J. D. Darlington (1999). "Race, ethnicity and disasters in the United States: A review of the literature". In: *Disasters* 23.2, pp. 156–173.

- Gagnon, A. et al. (2013). “Age-specific mortality during the 1918 influenza pandemic: unravelling the mystery of high young adult mortality”. In: *PLoS One* 8.8.
- Gyasi, R. M., D. R. Phillips, and R. David (2019). “Explaining the gender gap in health services use among Ghanaian community-dwelling older cohorts”. In: *Women & health* 59.10, pp. 1089–1104.
- Hutchins, S. S., K. Fiscella, R. S. Levine, D. C. Ompad, and M. McDonald (2009). “Protection of racial/ethnic minority populations during an influenza pandemic”. In: *American journal of public health* 99.S2, S261–S270.
- Jaffee, K. D., D. A. Shires, and D. Stroumsa (2016). “Discrimination and delayed health care among transgender women and men”. In: *Medical Care* 54.11, pp. 1010–1016.
- Liu, Q., Y.-h. Zhou, and Z.-q. Yang (2016). “The cytokine storm of severe influenza and development of immunomodulatory therapy”. In: *Cellular & molecular immunology* 13.1, pp. 3–10.
- Matthews Pillemer, F., R. J. Blendon, A. M. Zaslavsky, and B. Y. Lee (2015). “Predicting support for non-pharmaceutical interventions during infectious outbreaks: a four region analysis”. In: *Disasters* 39.1, pp. 125–145.
- Mein, S. A. (2020). “COVID-19 and Health Disparities: the Reality of “the Great Equalizer””. In: *Journal of General Internal Medicine*, pp. 1–2.
- Mpeta, B., J. Fourie, and K. Inwood (2018). “Black living standards in South Africa before democracy: New evidence from height”. In: *South African Journal of Science* 114.1-2, pp. 1–8.
- Nattrass, N. and J. Seekings (2011). “The economy and poverty in the twentieth century”. In: *The Cambridge history of South Africa, 1885–1994*. Ed. by R. Ross, A. Mager, and B. Nasson. Vol. 2. Cambridge: Cambridge University Press, pp. 518–572.
- Oxford, J. S. and D. Gill (2018). “Unanswered questions about the 1918 influenza pandemic: origin, pathology, and the virus itself”. In: *The Lancet Infectious Diseases* 18.11, e348–e354.
- O’Donnell, O., E. Van Doorslaer, and T. Van Ourti (2015). “Health and inequality”. In: *Handbook of income distribution*. Vol. 2. Elsevier, pp. 1419–1533.
- Phillips, H. (1984). “Black October: The impact of the Spanish influenza epidemic of 1918 on South Africa”. PhD thesis. University of Cape Town.
- (1988). “South Africa’s worst demographic disaster: The Spanish influenza epidemic of 1918”. In: *South African Historical Journal* 20.1, pp. 57–73.
- (2014). “The recent wave of ‘Spanish’ flu historiography”. In: *Social History of Medicine* 27.4, pp. 789–808.
- (2018). *In a Time of Plague: Memories of the ‘Spanish’ Flu Epidemic of 1918 in South Africa*. Van Riebeeck Society.
- Quinn, S. C. et al. (2011). “Racial disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza pandemic”. In: *American journal of public health* 101.2, pp. 285–293.
- Saunders, G. L. and T. Monet (2007). “Eliminating injustice toward disadvantaged populations during an influenza pandemic”. In: *North Carolina medical journal* 68.1, pp. 46–48.
- Szolysek, M., R. Poniak, and S. Gruber (2018). “Age heaping patterns in Mosaic data”. In: *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 51.1, pp. 13–38.
- Uscher-Pines, L., P. S. Duggan, J. P. Garron, R. A. Karron, and R. R. Faden (2007). “Planning for an influenza pandemic: social justice and disadvantaged groups”. In: *Hastings Center Report*, pp. 32–39.
- Van Doorslaer, E., C. Masseria, X. Koolman, et al. (2006). “Inequalities in access to medical care by income in developed countries”. In: *Cmaj* 174.2, pp. 177–183.
- Van Doorslaer, E. and A. Wagstaff (1992). “Equity in the delivery of health care: some international comparisons”. In: *Journal of health Economics* 11.4, pp. 389–411.
- Van Leeuwen, M., I. Maas, and A. Miles (2002). *HISCO: Historical international standard classification of occupations*. Leuven: Leuven University Press.

- Wagstaff, A. and E. Van Doorslaer (2000). “Equity in health care finance and delivery”. In: *Handbook of health economics* 1, pp. 1803–1862.
- Whitehead, M., G. Dahlgren, and T. Evans (2001). “Equity and health sector reforms: can low-income countries escape the medical poverty trap?” In: *The Lancet* 358.9284, pp. 833–836.
- Woo, G (2019). “Age-dependence of the 1918 pandemic”. In: *British Actuarial Journal* 24.
- Zhang, J. et al. (2020). “Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China”. In: *Science*.

7 Appendix

The regression specifications are as follows: for race, the reference category is white. For age, the reference category is 25–34 years. For sex, the reference category is female. For occupation, the reference category is farming. The signature dummy is equal to one if the informant signed their name, and zero otherwise – the majority of the latter read, ”X - his mark”.

In the following set of regressions, for birthplace, the reference category is the Cape. The ‘death in family’, variable is equal to one if there is a death of a family member (signified by same surname in same town) within the week before an individual’s death. The street, road, stad dummy is equal to one if the usual place of address line contains the words ‘street’, ‘road’ or ‘stad’.

Table 5: Results

	<i>Dependent variable:</i>		
	Presence of a Doctor		
	(1)	(2)	(3)
(Intercept)	1.219*** (0.2698)	2.45*** (0.1971)	2.056*** (0.5663)
Coloured	-0.4664* (0.2603)	-0.9187*** (0.1771)	-0.8913*** (0.1466)
Black	-2.169*** (0.2557)	-1.814*** (0.2449)	-1.701*** (0.2138)
Pandemic	-0.4862** (0.2404)	-2.243*** (0.3982)	-2.576** (1.282)
Coloured \times Pandemic	-0.8441*** (0.2056)	-0.7167** (0.2793)	-0.6029** (0.2915)
Black \times Pandemic	-0.8178* (0.4736)	-1.215*** (0.3244)	-1.222*** (0.3218)
Age_00-04Years		-0.3140* (0.1871)	-0.3319* (0.1907)
Age_05-09Years		-0.1766 (0.1629)	-0.1604 (0.1557)
Age_10-14Years		-0.0341 (0.1172)	-0.0100 (0.1152)
Age_15-19Years		-0.1545** (0.0757)	-0.1659** (0.0746)
Age_20-24Years		-0.0518 (0.0633)	-0.0507 (0.0713)

Continued on next page

Table 5 – continued from previous page

	<i>Dependent variable:</i>		
	Presence of a Doctor		
	(1)	(2)	(3)
Age_35-49Years		-0.1897** (0.0760)	-0.1958*** (0.0755)
Age_50-69Years		-0.1123 (0.0744)	-0.1364* (0.0715)
Age_70Years and Over		-0.4340*** (0.1034)	-0.4794*** (0.1076)
Male		-0.0127 (0.0515)	-0.0186 (0.0509)
Professionals and managers		0.0044 (0.2174)	-0.1482 (0.1875)
Sales and service		0.5140*** (0.1738)	0.4072** (0.1627)
Production		0.9270*** (0.2550)	0.8088*** (0.2711)
Children and Pensioners		0.2014 (0.2321)	0.1055 (0.1993)
Signature		0.3891*** (0.1184)	0.3587*** (0.1379)
Pandemic \times Age_00 – 04Years		0.3346 (0.2858)	0.3469 (0.2983)
Pandemic \times Age_05 – 09Years		0.0239 (0.3679)	0.0174 (0.3626)
Pandemic \times Age_10 – 14Years		-0.2240 (0.2385)	-0.2204 (0.2300)
Pandemic \times Age_15 – 19Years		-0.0588 (0.1148)	-0.0456 (0.1068)
Pandemic \times Age_20 – 24Years		-0.1405 (0.1499)	-0.1537 (0.1526)
Pandemic \times Age_35 – 49Years		0.2423 (0.1830)	0.2177 (0.1780)
Pandemic \times Age_50 – 69Years		0.2375 (0.2330)	0.2153 (0.2225)
Pandemic \times Age_70YearsandOver		0.6985** (0.2854)	0.6768** (0.2986)
Pandemic \times Male		0.1246 (0.1211)	0.1176 (0.1270)
Pandemic \times Professionalsandmanagers		0.2907 (0.2274)	0.3014 (0.2119)
Pandemic \times Salesandservice		-0.2054 (0.2153)	-0.1536 (0.2146)
Pandemic \times Production		-0.6003** (0.2898)	-0.5221* (0.3003)
Pandemic \times ChildrenandPensioners		-0.0963 (0.2034)	-0.0393 (0.1973)
Pandemic \times Signature		0.1370 (0.2789)	0.1674 (0.2544)
Africa			-0.3798 (0.4785)
Asia			1.412**

Continued on next page

Table 5 – continued from previous page

				<i>Dependent variable:</i>		
				Presence of a Doctor		
				(1)	(2)	(3)
						(0.6063)
Europe						0.7213**
						(0.2962)
South Africa						−0.2464
						(0.0993)
Death in Family						−0.1162
						(0.0986)
nurse						0.6616
						(0.5487)
railway						0.6251
						(0.4015)
Street, road and stad						1.338***
						(0.4502)
Farm						0.0703
						(0.1766)
Pandemic × Africa						0.7714
						(0.7214)
Pandemic × Asia						−0.0324
						(0.9357)
Pandemic × Europe						1.094**
						(.5184)
Pandemic × SouthAfrica						0.7245**
						(0.2993)
Pandemic × Deathin family						−0.4795***
						(0.1077)
Pandemic × Nurse						0.9676
						(1.298)
Pandemic × Railway						0.1443*
						(0.4837)
Pandemic × Street, road and stad						−1.168***
						(0.4356)
Pandemic × Farm						−0.2752
						(0.3318)
District Dummies	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>			
<i>Fit statistics</i>						
Observations	39,482	32,555	32,555			
Squared Correlation	0.228393	0.364945	0.381468			
Pseudo R ²	0.177067	0.300112	0.316021			
BIC	45,009.99	32,081.08	31,574.30			
<i>One-way (district) standard-errors in parentheses</i>						
<i>Signif. Codes: ***: 0.01, **: 0.05, *: 0.1</i>						