UNIVERSITY of York

HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

THE UNIVERSITY of York

WP 21/20

Explaining the ethnic gaps in COVID-19 outcomes in Mexico

Andrea Salas-Ortiz

November 2021

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Explaining the ethnic gaps in COVID-19 outcomes in Mexico

Andrea Salas-Ortiz*

version: November 3, 2021

Abstract. Indigenous groups are one of the most socially vulnerable groups across societies. Concerns have been raised about the possibility of greater health disparities when the Covid-19 pandemic interacts with non-communicable diseases in contexts of high socioeconomic inequalities (Horton, 2020). Using national and administrative public data on Covid-19, this study investigates this hypothesis by explaining differences in Covid-19 health outcomes (hospitalisations, admissions to intensive care unit, and mortality) between indigenous and non-indigenous groups in Mexico. The analysis uses an adaptation of the Oaxaca decomposition method to account for nonlinear responses. This allows to identify and characterise the factors behind ethnic disparities. Results indicate that indigenous people have worse Covid-19 health outcomes. These differences are mainly attributable to differences in people's characteristics. Disentangling the contribution of each individual and contextual circumstances to the observable differences, we found that underlying health conditions, household and municipal socioeconomic characteristics are the main drivers of observable inequalities in hospitalisations and deaths due to Covid-19. These findings highlight that this pandemic is exacerbating the pre-existing and longstanding health inequalities between indigenous and non-indigenous people in Mexico.

Key words: Health inequalities, Covid-19, Oaxaca decomposition; Indigenous groups; Mexico

*Department of Economics and Related Studies, University of York. E-mail: aso512@york.ac.uk. The author acknowledges funding from the National Council for Science and Technology in Mexico (CONACYT).

1 Introduction

Higher health disparities could be observed when Covid-19 interacts with high prevalence of non-communicable diseases (NCDs) among social and economically unequal populations (Horton, 2020).This hypothetical situation could be aggravated if, within societies, a high number of vulnerable groups exist. Based on past experiences, the World Health Organisation (WHO) warned that epidemics have disproportionate effects on vulnerable populations, such as indigenous people, and perpetuate the pre-existing and longstanding social, economic and health inequalities (Sachs et al., 2020). There is evidence that these conditions could be holding for the Mexican case. First, before Covid-19, Mexico was already facing a public health crisis mainly driven by non-communicable diseases: 75% of the adult population was either obese or overweight (National Institute of Public Health, 2018) and the type II diabetes prevalence was one of the highest globally (13-22%) (Meza et al., 2015; Saeedi et al., 2019). Second, Latin-America is one of the most unequal regions in the world, and Mexico is not the exception; economic inequality, measured via the Gini index, is one of the highest globally, at 0.45 in 2016 (Lambert and Hyunmin, 2019). Third, Mexico is a multi-ethnic country. According to the National Institute of Statistics and Geography (INEGI), 21.5% of the population self-identify as indigenous (INEGI, 2020). The geographic distribution of indigenous people across Mexico can be found in Figure A.1 in the Electronic Supplementary Material.

Unequal impacts of an epidemic on vulnerable populations is closely related with the social and economic circumstances, as well as people's health conditions, since such circumstances and conditions influence a wide range of health risks and outcomes (Tai et al., 2020, p. 2). Indigenous people are particularly vulnerable to the Covid-19 pandemic due to the impoverished social and economic characteristics they face. The National Council for the Evaluation of Social Development Policy in Mexico (CONEVAL, in Spanish) estimated that, in 2016, 15.1% of indigenous people did not have access to health services and 56.3% did not inhabit a household with basic standards, such as proper walls, roofs, floor materials, available running water, a toilet, drainage system or electricity (National Council for the Evaluation of Social Development Policy, 2018). Furthermore, disparities between indigenous and non-indigenous people also exist in other spheres: indigenous people face poorer academic performance, higher levels of poverty, lower life expectancy and health insurance coverage is still insufficient (Servan-Mori et al., 2014;

Leyva-Flores, Infante-Xibille, et al., 2013; Leyva-Flores, Servan-Mori, et al., 2014).

Given this context, this study investigates whether the hypothesis regarding the perpetuation of health disparities among indigenous and non-indigenous people hold in the light of the Covid-19 pandemic and, if so, the extent to which inequalities have widened. Previous studies have already analysed ethnic inequalities in Covid-19 outcomes, such as hospitalisation and mortality in Mexico. These studies have relied on estimating predicted probabilities to observe an outcome, conditional of a set of individual, social and economic characteristics and comparing these estimations across groups. Findings from these studies highlight that differences in access and the quality of care have played a crucial role in higher mortality rates among indigenous people compared to the general population (Ibarra-Nava et al., 2021). Also, that living in municipalities with high social deprivation associates with a higher risk of hospitalisation and an early death due to Covid-19 and that the presence of underlying health conditions increase the probability of hospitalisation and death among indigenous patients (Serván-Mori et al., 2021). This study contributes to the current literature about ethnic inequalities and Covid-19 by analysing inequalities using an appropriate method to do so. This will allow us to better identify, measure and characterise inequalities. To follow our aim, we identify ethnic gaps in Covid-19 outcomes and breakdown this gap into two components, one that explains the differences due to observed characteristics and another that explains them given differences in the link between characteristics and outcomes, the latter differences could potentially be attributable to discrimination towards indigenous people. The remaining part of the paper proceeds as follows: Section 2 explains the decomposition strategy. Section 3 describes the data, while Section 4 offers more detail about the key variables used for the empirical analysis. Section 5 reports the main results of the analysis and the last section discusses them.

2 Methods

To investigate the extent of Covid-19 outcome differences between indigenous and non-indigenous people, we make use of the nonlinear version of the Oaxaca-Blinder decomposition method, which estimates the impact of individual and contextual characteristics for each group on outcomes and decomposes the average inter-group gap due to differences in *observable characteristics* and differences in *the effects of coefficients*. The standard Oaxaca-Blinder decomposition is based on a regression model where a health outcome is a function of a set of covariates, which in this analysis are: individual's health conditions, household deprivation, health infrastructure individuals are exposed to and the geographical economic characteristics of where people live. This model is run separately for indigenous and non indigenous people.

To decompose the average inter-group difference, the method relies on a counterfactual that depicts what would happen if the characteristics of one group were interchanged with the coefficients of the other group. By applying this counterfactual, two components are obtained, this is known as the *two-fold* Oaxaca decomposition. These are known as the *explained* and *unexplained* components. The former shows a counterfactual comparison of the expected difference in outcomes if non indigenous were given the indigenous distribution of covariates. It is explained because this part of the difference can be attributable to differences in the *observed characteristics*. In contrast, the unexplained component reflects a counterfactual comparison of the expected difference if indigenous people experienced the non indigenous response to the set of covariates, thus shows the *effect of coefficients*. While the explained component might justify the group disparities due to differences in people's characteristics, the unexplained part has been labelled, as *discrimination*, since there is no economic justification for group differences (Blinder, 1973; Rahimi and Hashemi Nazari, 2021). As expected, this claim has been controversial, as the concept of discrimination cannot be simply reduced to what a model cannot explained and, at most, it should be labelled as *observed*, incorporating the fact that results are limited to those factors included in the model(Jann, 2008). Another relevant point regarding the interpretation of this component is that of Fortin et al. (2011), who explain that the link between the decomposition methods with the impact evaluation literature is that the *unexplained component* of the Oaxaca decomposition can be interpreted as the *population treatment effect on the treated* (PATT) if selection on observables is assumed to identify the treatment effect. This vision has also been shared by Słoczyński, 2015. Jann (2008) also mentions that the unexplained part captures all the potential effects of differences in *unobserved* variables.

The two-fold decomposition previously described is known as **the aggregate Oaxaca decomposition**. But, for many purposes, policy-related included, it is relevant to further identify and measure the main factors contributing to the explained and unexplained parts of the ethnic gap. This extension of the aggregate decomposition is known as **the detailed Oaxaca-Blinder**

decomposition and consists on subdividing each component and estimating the contribution of each explanatory variable (Fortin et al., 2011). The Oaxaca-Blinder decomposition is possible if average group differences in outcomes exist. So, we examine this by using tests on the equality of proportions, and then we estimate nonlinear models for each group and outcome. We also make us of Linear Probability Models (LPM) aiming to check the robustness of the results, although only results from the nonlinear models are reported in the main text. Differences in βs across groups are tested as well, using a Wald-type test of nonlinear hypotheses for the estimated parameters. Detailed explanations of the method, results of the tests and robustness checks are provided in the Electronic Supplementary Material.

3 Data

This analysis uses open administrative data on Covid-19 from the General Directorate of Epidemiology (Dirección General de Epidemiología, (DGE) in Spanish), from the 2020 National Census and from the General Directorate of Health Information (DGIS).

3.1 Covid-19 data

The Mexican government did not follow a universal Covid-19 testing and only those with symptoms were eligible for a test. Thus, the data collection process (who, where, when and how data is collected) is explained in Figure A.2 in the Electronic Supplementary Material. The publicly available dataset is updated every day and some of the variables might have reporting delays (Giannouchos et al., 2020), for this analysis the version released on the fourth of April 2021 is used. Results about testing, hospitalisation and patient follow-up (discharge, or worsening condition where patients are admitted to ICU (intensive care unit) or passing away) are directly uploaded by the diagnostic facility or hospital.

The dataset contains information about the patient's birthplace, place of residence, age, sex, nationality and ethnicity (whether a patient identifies as indigenous language speaker). Information about the patient's migratory status is also included, as well as patient's health institution affiliation and clinical information. This includes type of medical attention, for an inpatient: admission date, symptom onset date, whether admission to ICU and/or date of death, PCR result (positive, negative, or pending), and if the person is pregnant. Other clinical information about underlying condition such as: pneumonia, chronic obstructive lung disease (COPD), asthma, immunosuppression, diabetes, obesity, hypertension, chronic renal and cardiovascular disease, and other comorbidities are included, as well as whether the patient is a smoker. These are all indicator variables and no further information is provided. Since there is no variable available in the dataset to identify a patient, we matched cases with same information about demographics and clinical history data and eliminated the duplicate observations. A similar approach is followed by Mancilla-Galindo et al., 2020 in their modelling about Covid-19 deaths in Mexico. Patients with incomplete (pending results) or missing information about testing results and ethnicity were also excluded, as well as non-Mexican patients, resulting in a final sample of 4,797,854 observations.

3.2 2020 National Census Data

For our analysis, aggregated information at the municipality level is used and includes: the number of people affiliated to health services, that attend school or are literate, and number of people unemployed. Using this information, household and economic municipal characteristics indicators are constructed by getting the percentage of people with the j characteristic living in the m municipality. Information about the number of health facilities in January 2020 comes from the DGIS.

4 Key variables

4.1 Ethnic groups

Ethnicity is a binary classification with the two groups identified according to whether or not individuals speak an indigenous language.

4.2 Health variables

Three measures of health outcomes are used to reflect a worsening condition of people who contract Covid-19:

- To be hospitalised due to Covid-19
- To be admitted to an intensive care unit given Covid-19,
- To die because of Covid-19 related complications

All outcome variables are binary indicators and take the value of 1 if the event is true and 0 otherwise. For some individuals, the events are conditional on a prior event being true. For example, a person is observed to be in ICU that has previously observed to be hospitalised. However, not all dead patients were either hospitalised or required ICU.

4.3 Individual and structural variables

People are vulnerable to epidemics not only because of their particular health conditions (overall health status, comorbilities or the risky health behaviours they adopt), but also because of their social, economic and household conditions. Covid-19 has made explicit that individual and contextual factors matter. A recent study showed that comorbilities such as obesity, diabetes, hypertension, coronary heart disease, and heart failure were closely related with severe Covid-19 cases (Hernández-Garduño, 2020). It has also been found that socioeconomic factors or structural conditions play a role in worsening the impact of the pandemic within communities (Hawkins et al., 2020). In particular, a recent study found that poor access to water; language barriers; household characteristics; lack of health insurance; and underlying health conditions such as hypertension, type II diabetes, chronic pulmonary diseases and respiratory tract infections are risk factors hampering the ability of indigenous communities to avoid contracting Covid-19 (Díaz de León-Martínez et al., 2020).Based on this evidence, the key explanatory variables used in this analysis are divided in two categories: individual-level characteristics and socioeconomic circumstances, which have been aggregated at the municipal level. Data at the municipality-of-residence level is used since it is an indirect but reliable way to proxy the social and economic aggregated deficiencies that can be correlated with health outcomes. A further description of the individual level variables can be found in Table A.1 in the Electronic Supplementary Material.

Socioeconomic circumstances refer to those characteristics that people cannot change in the short-run and that occur within a geographical area, in this case, the municipality. Among these, we include the number of households that live in deprived conditions within a municipality (average number of people per household; percentage of households with low-quality-material walls, ceilings, floors; percentage of households without electricity; percentage of households without running water or toilet, drainage, electricity or the all these; percentage of households with fridges, radio, TV, mobile phone or internet); and the level of health coverage and medical infrastructure (percentage of people without health insurance and number of health facilities

Individual-level characteristics	**Socioeconomic circumstances**
-Demographics -Underlying health conditions -Risky behaviours -Institution where individuals received medical attention*	-Household deprivation characteristics -Health coverage and medical infrastructure -Economic characteristics

* This variable is relevant because the Mexican health system is fragmented into eight health institutions and a recent analysis found relevant contrasts in Covid-19 mortality rates within public institutions and between the public and private sectors. For example, up to August of 2020, 45% of hospitalised patients in the IMSS died, versus 31% of the patients hospitalised in SSA hospitals and 16% in the private sector (Sánchez T., 2020)

available in the municipality in January 2020, when the pandemic started), and municipality's economic conditions (percentage of unemployed people and with no formal education)

5 Results

5.1 Ethnic differences in Covid-19 outcomes and covariates

Table 1: Ethnic differences in proportions of people hospitalised, in intensive care and dead due to Covid-19 in Mexico

	N_NI	Mean NI	N_I	Mean I	Diff	p-val
People hospitalised	4,797,854	0.129	31,273	0.249	-0.120	0.000
People in ICU	612,291	0.073	7,655	0.085	-0.013	0.000
People dead	4,797,854	0.051	31,273	0.098	-0.048	0.000

Notes: Analysis period Jan 2020- March 2021. NI=Non-indigenous. I=Indigenous Diff=Difference. Two-sided p-value

Table 1 shows the size of the sample for each outcome and group, as well as the proportion of people hospitalised, admitted to ICU and dead. The average differences between groups in Covid-19 outcomes are, respectively, 0.12, 0.013 and 0.048. This shows that indigenous people have worse outcomes as all mean values are higher among this population. The highest difference is observed in hospitalisations and the smallest in admissions to ICU. All differences are statistically significant.

Table 2 shows the mean value of the individual characteristics for all health outcomes for each group. Overall, there are statistically significant differences across the groups for most of the characteristics and outcomes. Exemptions can be found some demographics. Across the majority of the outcomes, the proportion of people with a comorbidity is larger for indigenous than for non-indigenous people, albeit these differences are statistically significant for hospitalisations and deaths. Whereas, for NCD, the non-indigenous group have a higher proportion than indigenous. The difference is not statistically significant for diabetes (in hospitalisations) and cardiovascular disease (for admissions to ICU). A higher proportion of people with obesity and who are smokers is observed in the non-indigenous group, but the difference is not statistically significant for people in ICU and deaths for obesity. Table 2 shows that diabetes, hypertension and obesity have the highest proportions values across all outcomes and these proportions are greater among non indigenous people. In relation to medical care, across all groups IMSS and SSA are the two institutions where most of the people are affiliated and were treated. Most of the hospitalisations among non-indigenous people took place at IMSS hospitals, while SSA hospitals treated indigenous people. This is something expected as SSA hospitals admit most of the people enrolled to "INSABI" (formerly known as the *Seguro Popular*) programme. For cases admitted to ICU, most the admissions were done at SSA hospitals, both for indigenous and non-indigenous people, and among those who died, most of the non-indigenous individuals were affiliated to IMSS, and most of the indigenous that died were affiliated to SSA institutions.

Table 3 describes the average values of the structural socioeconomic circumstances of the areas where individuals live at the municipality level. There are significant differences across all variables. Overall, indigenous people live in municipalities with a lower percent of urban localities and in municipalities where the percentage of households in poor physical condition (no floor, less spacious homes, NO water, no electricity, drainage, motorcycle, home appliances, TV or Radio, telephone, computer or internet) is larger than where non-indigenous people live. With regards to the health insurance coverage, the percentage of people not covered is slightly higher in municipalities where non-indigenous people live, but the number of health facilities is, on average, higher in these municipalities. The economic characteristics of municipalities are overall better in municipalities where non-indigenous people live. The average percentage of illiteracy is lower, but the percentage of unemployment is slightly higher. All differences between the groups are statistically significant.

Table 2: Ethnic differences in individual characteristics for all outcomes

	Hospitalised			**In ICU**			**Died**		
	Mean NI	Mean I	p-val	Mean NI	Mean I	p-val	Mean NI	Mean I	p-val
Demographics									
Age	54.85	54.63	0.33	53.34	53.08	0.76	63.08	63.73	0.02
Female	0.43	0.44	0.01	0.38	0.37	0.49	0.38	0.38	0.72
Comorbidities									
COPD	0.04	0.07	0.00	0.04	0.05	0.08	0.05	0.08	0.00
Asthma	0.02	0.03	0.00	0.02	0.03	0.25	0.02	0.03	0.00
Immunosuppression	0.03	0.03	1.00	0.03	0.04	0.83	0.03	0.02	0.04
Renal D.	0.06	0.05	0.00	0.05	0.04	0.22	0.08	0.06	0.00
Pneumonia	0.58	0.64	0.00	0.81	0.84	0.08	0.70	0.78	0.00
Other Comorb.	0.06	0.05	0.00	0.05	0.05	0.92	0.06	0.05	0.00
NCD									
Diabetes	0.31	0.30	0.52	0.32	0.28	0.06	0.38	0.36	0.01
Hypertension	0.36	0.30	0.00	0.35	0.30	0.00	0.45	0.38	0.00
Cardiovascular D.	0.05	0.04	0.02	0.06	0.05	0.48	0.06	0.05	0.03
Risky Behaviours									
Obesity	0.19	0.18	0.01	0.24	0.22	0.26	0.21	0.22	0.22
Smoker	0.08	0.06	0.00	0.08	0.05	0.02	0.08	0.07	0.02
Medical Attention									
Test waiting-time	4.62	4.50	0.01	5.05	4.65	0.01	5.08	5.00	0.23
Private	0.03	0.01	0.00	0.11	0.04	0.00	0.02	0.01	0.00
IMSS	0.52	0.27	0.00	0.12	0.09	0.07	0.60	0.30	0.00
ISSSTE	0.08	0.07	0.00	0.09	0.06	0.00	0.07	0.07	0.65
SSA	0.31	0.60	0.00	0.57	0.70	0.00	0.27	0.58	0.00
State	0.02	0.01	0.00	0.02	0.01	0.02	0.02	0.01	0.00
PEMEX	0.02	0.00	0.00	0.02	0.02	0.27	0.01	0.00	0.00
SEDENA	0.02	0.04	0.00	0.04	0.08	0.00	0.01	0.04	0.00
SEMAR	0.00	0.00	0.00	0.01	0.00	0.03	0.00	0.00	0.01

Notes: Analysis period Jan 2020-March 2021. NI=Non-indigenous. I=Indigenous. Two-sided p-value
NCD=Non-Communicable Diseases. D=Disease. Comorb=Comorbidities

Table 3: Ethnic differences in municipal socioeconomic characteristics

	Mean NI	Mean I	p-val
Percentage of Urban Localities in the municipality	28.11	15.15	0.00
Household Characteristics			
Average number of people per household	3.49	3.75	0.00
Percentage of households with low-quality-material floors	0.50	1.75	0.00
Percentage of households with only one sleeping room	8.46	10.41	0.00
Percentage of households with only one room	1.46	2.57	0.00
Percentage of households without water	0.57	1.48	0.00
Percentage of households without electricity	0.10	0.51	0.00
Percentage of households with a latrine	0.39	2.52	0.00
Percentage of households without drainage	0.44	3.21	0.00
Percentage of households without E,W,D*	0.02	0.17	0.00
Percentage of households without car or motorcycle	13.17	15.46	0.00
Percentage of households without appliances	0.16	1.08	0.00
Percentage of households without TVs or Radio	0.83	2.53	0.00
Percentage of households without telephone or mobile phone	1.46	4.37	0.00
Percentage of households without computer nor internet service	9.12	14.14	0.00
Percentage of households without ICT technologies	0.28	1.48	0.00
Health Insurance Coverage and Infrastructure			
Percentage of people not affilated to a health institution	25.68	24.09	0.00
Number of health facilities in January 2020	148.48	81.12	0.00
Economic Municipal Characteristics			
Percentage of people from 12-14 years that you did not attend school	0.36	0.53	0.00
Percentage of people from 8-14 years who can not read nor write	0.21	0.46	0.00
Percentage of people above 15 that is illiterate	2.02	6.23	0.00
Percentage of people above 15 without schooling	2.51	5.73	0.00
Percentage of unemployed people	1.05	0.79	0.00

Notes: Data from the 2020 National census. *E,W,D=No electricity, water and drainage.
NI=Non-indigenous. I=Indigenous. Two-sided p-value

5.2 Oaxaca-Blinder decomposition

5.2.1 Aggregate Decomposition

Table 4 shows the results of the average gap decomposition for all outcomes. The explained component, which depicts the extent to which differences between groups is due to differences in observable characteristics, accounts for most of the average difference, 82%, 50% and 88% for hospitalisations, admissions to ICU and Covid-related deaths. The unexplained component, which measures the extent to which average differences between groups is due to the link between characteristics and outcomes, contributes positively to the ethnic gap in magnitudes of approximately 18%, 50% and 12%, respectively. Comparison of linear and nonlinear models are found in the Electronic Supplementary Material, but, overall there are very small differences in the decomposition results.

Table 4: Aggregate Oaxaca Decomposition. Nonlinear models

	Hospitalisations	%	Admissions to ICU	%	Deaths	%
Non Indigenous	0.127***		0.072***		0.050***	
	(0.00)		(0.00)		(0.00)	
Indigenous	0.245***		0.084***		0.097***	
	(0.00)		(0.00)		(0.00)	
Mean Difference	-0.118***		-0.012***		-0.047***	
	(0.00)		(0.00)		(0.00)	
Explained	-0.096***	81.709***	-0.006	49.546	-0.041***	87.634***
	(0.00)	(1.89)	(0.00)	(32.42)	(0.00)	(2.88)
Unexplained	-0.021***	18.291***	-0.006	50.454	-0.006***	12.366***
	(0.00)	(1.89)	(0.00)	(32.42)	(0.00)	(2.88)
Observations	4,796,808	4,796,808	612,427	612,427	4,796,808	4,796,808

Notes: Bootstrapped standard errors in parenthesis (500 replications)

Models fitted using an ANOVA-type normalisation and weights from a first-order Taylor linearisation

% share of each component to the overall gap. + $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

5.2.2 Detailed Decomposition

Figures 1 and 2 show the relative contribution of each sub-set of variables to the explained and unexplained components, respectively. Contributions are shown as percentage of the overall difference. Positive contributions indicate that if the distribution of a characteristic was swapped between indigenous and non-indigenous people a reduction in the ethnic gap would be expected. Likewise, a negative contribution indicates that if the counterfactual is observed, the ethnic gap is expected to increase. For all results, we provide the uncertainty of our estimations. In particular, we report bootstrapped standard errors based on 500 replications with replacement. Tables with detailed results are shown in the Electronic Supplementary Material.

From Figure 1, it can be seen that for hospitalisations: demographics, underlying conditions, risky behaviours, health infrastructure, household characteristics and municipal economic characteristics positively contribute to the ethnic gap, medical attention contributes negatively. All contributions are statistically significant, except for municipal economic characteristics. For admissions to ICU, all set of variables, except household conditions, positively contribute to the outcome gap. Nevertheless, these contributions are not statistically significant. For deaths related to Covid-19, demographics, underlying conditions, risky behaviours and municipal economic characteristics have positive contributions, while medical attention, health infrastructure and household characteristics contribute negatively to the gap. All these contributions are statistically significant, except for health infrastructure, household characteristics and municipal economic characteristics.

The presence of underlying conditions is one of the main drivers of the explained ethnic differences. This means that if indigenous were equal to non-indigenous in the distribution of their comorbidities, the ethnic gap in hospitalisations, admissions to ICU and deaths would be expected to reduce by 46%, 15% and 51%, respectively. Although, these estimates are only statistically significant for hospitalisations and deaths. For hospitalisations, household characteristics are the second driver of the explained differences between the ethnic groups. If indigenous people had the same household conditions as non-indigenous people, the ethnic gap would decrease by 33%. Individual demographics are important drivers of explained differences in Covid-19 deaths, by shifting the age and sex distributions of non-indigenous to match the indigenous distribution, the difference in deaths between groups would decrease by around 30%.

While a positive effects indicates a reduction in the gap, a negative sign denotes a potential increase in differences between groups. For hospitalisations and deaths, the health institution where individuals received medical attention is a factor that increases the indigenous-non indigenous differential. If indigenous people were affiliated to the same health institutions to which non-indigenous people are affiliated, the ethnic difference in hospitalisations and deaths due to Covid-19 would increase by 11% and 8%. The estimations of the detailed decomposition of the unexplained component show a lot of uncertainty as the confidence intervals are very large. Only the set of demographic variables are statistically significant for hospitalisations and deaths and their contribution are negative. This means that if the link between sex and age with these outcome is the same across groups, the ethnic differences would increase by 19% and 10%, respectively.

6 Discussion

Using administrative data on Covid-19, this analysis identifies and measures the average differences in hospitalisations, admissions to ICU and deaths due to Covid-19 between indigenous and non-indigenous people in Mexico. This study uses a nonlinear version of the Oaxaca-Blinder decomposition method and finds four main results. First, differences due to individual characteristics account for most of the observed ethnic gap in hospitalisations and deaths, although half of the differences in admissions to ICU are only explained by observable factors. Once accounting for these characteristics, a non-trivial part of the ethnic gap remains unexplained. Second, people's underlying conditions (comorbidities and non-communicable diseases) are the main driver of the explained differences in hospitalisations and deaths due to Covid-19 between indigenous and non-indigenous people. Third, the health institutions where people received care explains differences in the ethnic gap of hospitalisations and deaths. Fourth, if household conditions were equalised across municipalities, the ethnic gap in hospitalisations due to Covid-19 would decrease by 32%.

Figure 1: Detailed Oaxaca Decomposition. Explained Component

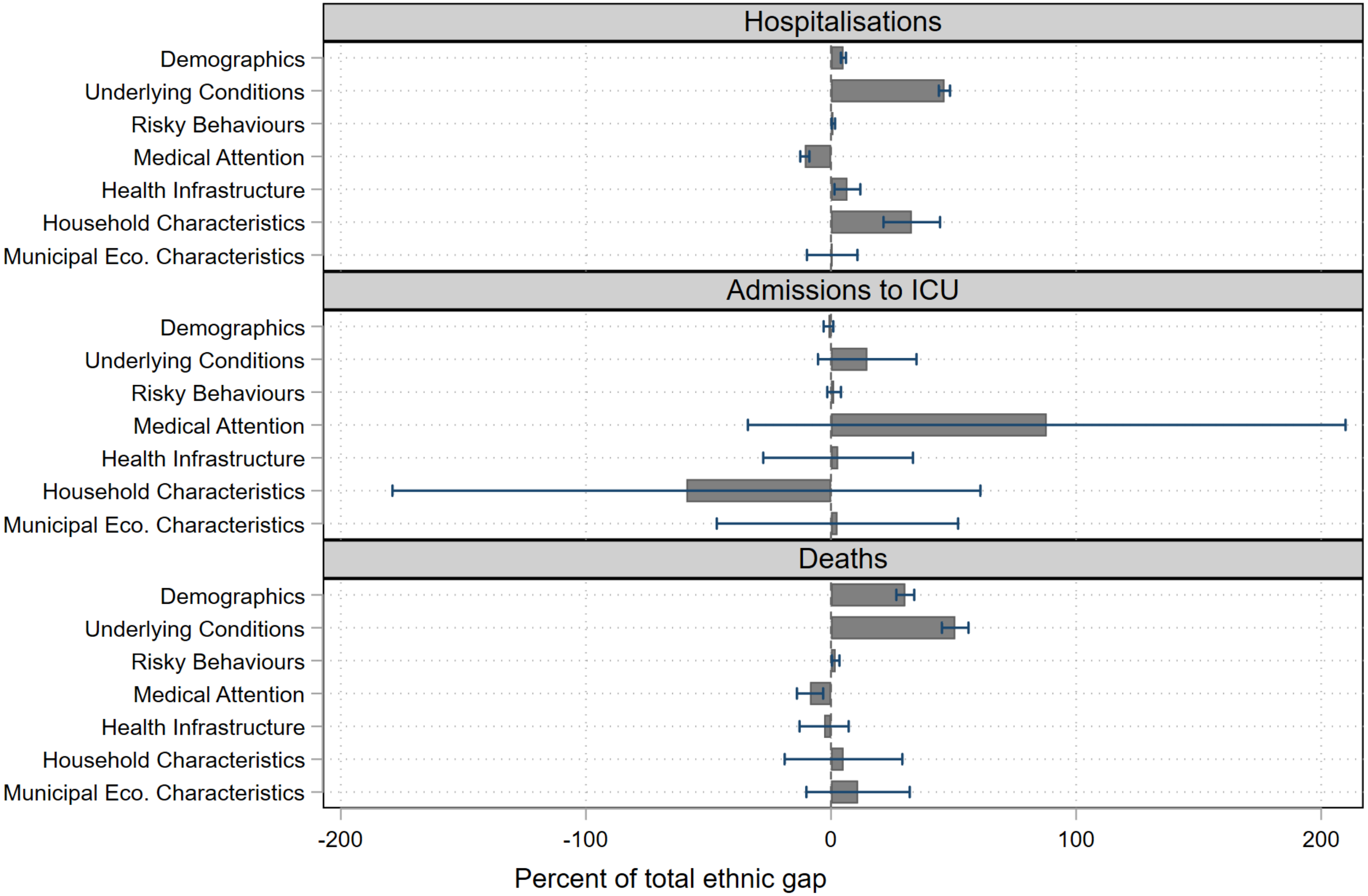
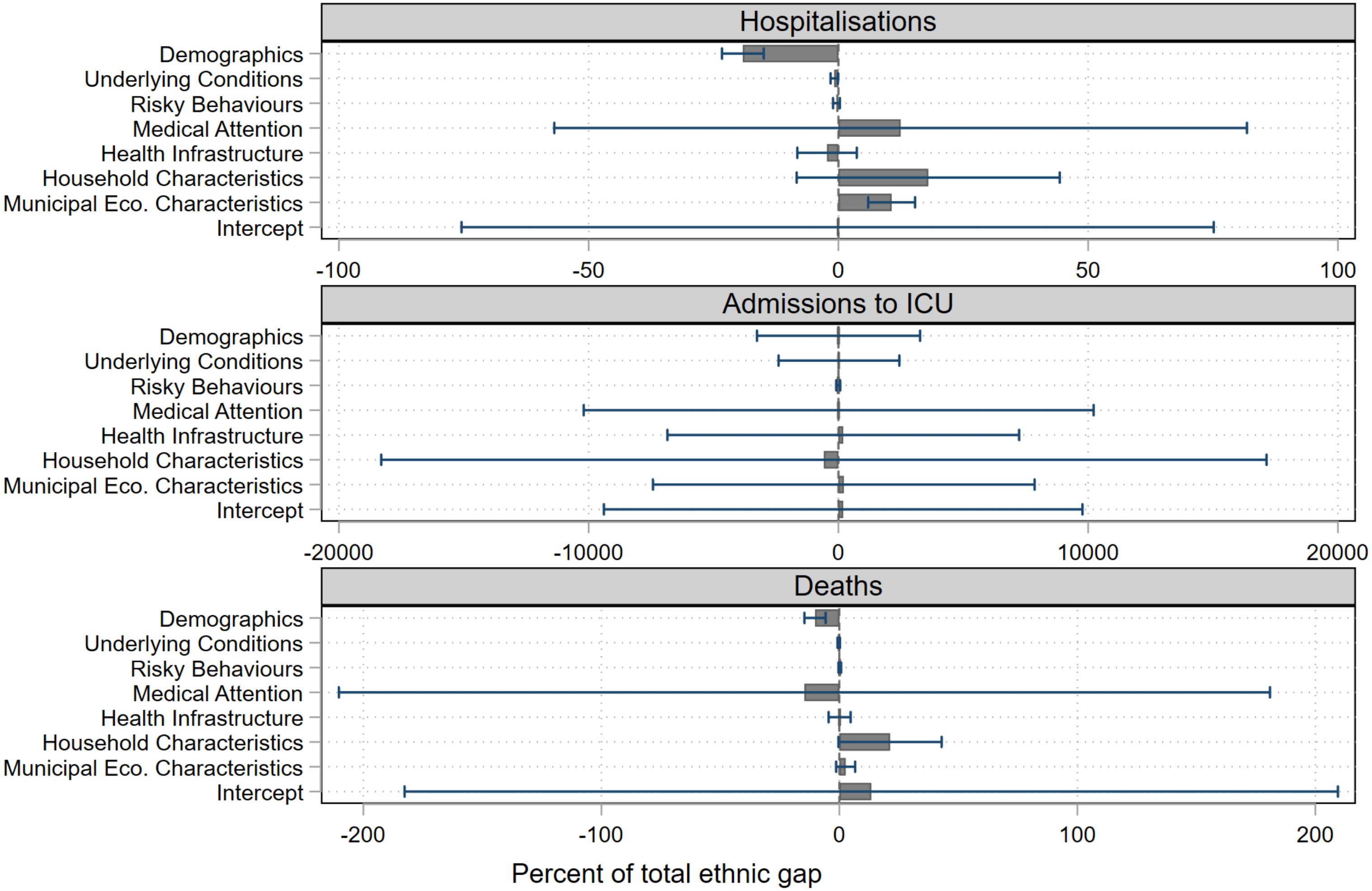


Figure 2: Detailed Oaxaca Decomposition. Unexplained Component



Once differences between groups have been explained given observable characteristics, there remains a part that has no economic explanation. This, for example, accounts for half of ethnic differences in admissions to ICU. This component indicates that inter-group differences in the relationship between characteristics and admissions to ICU are the main driver of the ethnic gap. Indeed, this can be depicted in the results from regression models (Table A.3), that shows differences in magnitude and significance and direction of the coefficients. This unexplained component has been framed as *observed discrimination*, but this needs to be taken with further caution. We argue that although discrimination cannot be defined by a part of a model that cannot be explained, this cannot ruled out the possibility that discrimination against indigenous people actually exist.

The fact that an unequal provision of health and other public services across municipalities exist underpins evidence about a systematic unequal treatment that disadvantages indigenous communities. This is particularly worrying since, in 1990, Mexico signed the *International Labour Organisation (ILO) Indigenous and Tribal Peoples Convention*, better known as ILO Convention #169. This convention highlights the aspirations of indigenous people to develop and maintain their identities, languages and religions while exercising their fundamental human rights to the same degree as the rest of the population, and that this must be guaranteed by the State. Thus, in this sense the ILO convention represents a benchmark for policy-making, since it reinforces the rights that indigenous peoples have besides those they are entitled to by the Mexican Constitution. With regards to health, articles 24 and 25 of the #169 convention state that social security and health services *should be extended* progressively to reach full coverage, that the delivery of health services *should be community-based* and that the health system *should prioritise* the delivery of primary health care services (International Labour Organisation, 2009). In Australia for example, public policies focused on prevention and primary care have shown to be effective to reduce ethnic health disparities (Davis, 2004; McIntyre and Menzies, 2005). Nevertheless, a study for Mexico showed that indigenous people did not utilise primary care due to the the lack of confidence, mistreat to indigenous people, and unavailability or facility's remoteness (Servan-Mori et al., 2014)

Underlying health conditions are a major factor of the explained differences in hospitalisations and deaths due to Covid-19 between indigenous and non-indigenous people. This result is in line with previous studies that pointed out that non-communicable diseases such as, diabetes, obesity and hypertension were positively associated with Covid-19 outcomes (Gutierrez and Bertozzi, 2020; Hernández-Galdamez et al., 2020; Monterrubio-Flores et al., 2020; Serván-Mori et al., 2021). Prior to the pandemic, Mexico was already facing an acute obesity crisis, a health problem that has not been entirely addressed by the government (Barquera and Rivera, 2020). Thus, our findings highlight that unsolved public health problems make indigenous people more vulnerable. In the light of health shocks, the pre-existing and longstanding health inequalities between indigenous and non-indigenous people magnify, persist and can further perpetuate.

This analysis also evidences that disparities in outcomes could potentially come from the Mexican health system itself. The type of health institution where people received medical care for Covid-19 is relevant to explain differences in hospitalisations and deaths due to Covid-19. If medical attention would have been the same between indigenous and non-indigenous people across the health system's institutions, the ethnic differences would have increased. This means that differences in care attention within the health system exist. This coincides with previous studies, Puig et al. (2009) found high levels of heterogeneity in healthcare quality and that users rated better healthcare attention received in SSA institutions than in IMSS facilities (ibid.). Sánchez T. (2020) found that Covid-19 mortality variation across the institutions of the health system was due to structural differences in hospital infrastructure, equipment availability and training of the staff, as well as the use of care protocols and that the pandemic only exhibited these deep-rooted inequalities (ibid.). Indeed, Table 3 shows that, at the beginning of the pandemic, the number of health facilities was, on average, larger in non-indigenous municipalities. This also highlights the lack of an indigenous-prioritising policy regarding health facilities availability .

Household and municipal socioeconomic conditions matter. This is relevant for contexts where a federal political system prevails. The federal system in Mexico has led to different levels of efficiency, efficacy and quality in the provision of health services across the federal States. Therefore, where people live conditions the services to which they have access. Historically, indigenous settlements have experienced a relatively higher scarcity of health facilities along with low quality of healthcare services (Leyva-Flores, Infante-Xibille, et al., 2013; Leyva-Flores, Servan-Mori, et al., 2014; Juárez-Ramírez et al., 2014; Servan-Mori et al., 2014). Furthermore, a study found that to live in areas with low healthcare resources was associated with a higher risk of hospitalisation for Covid-19 (Serván-Mori et al., 2021). This highlights the need to consolidate a coordinating and responsive federal system that can guarantee universal health insurance coverage and access to basic medical care for all citizens, regardless of their ethnicity or post code.

Lastly, in terms of the methods used, our results corroborate previous conceptions about similarities in results between linear and nonlinear models when the outcome variable is binary, although nonlinear models are better when the gaps are located in the tails of the distribution (Fairlie, 2005). This study is not without limitations. The most challenging is the under-representation of the *real* number of deaths. Since barriers to access the health system exist, many people died in their homes and therefore were not registered in the administrative dataset we used (Soberanes, 2021). Further analysis are needed to investigate whether this event increased the mortality ethnic gap and who were affected the most.

This analysis identified that indigenous people in Mexico face worse Covid-19 outcomes than the general population and found the existence of systematic barriers that affect indigenous groups in a distinct and exclusionary manner. Hence, since Covid-19 is exacerbating the pre-existing, deep-rooted and longstanding health inequalities between indigenous and non-indigenous people, it is imperative to design programmes that prioritise and target indigenous people and to enhance the current social and health policies if the disproportionate impact of this pandemic is aimed to be mitigated.

References

Barquera, S. and J. A. Rivera (2020). “Obesity in Mexico: rapid epidemiological transition and food industry interference in health policies”. In: *The Lancet Diabetes & Endocrinology* 8.9, pp. 746–747. DOI: 10.1016/S2213-8587(20)30269-2.

Blinder, A. S. (1973). “Wage Discrimination: Reduced Form and Structural Estimates”. In: *The Journal of Human Resources* 8.4. Publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System], pp. 436–455. DOI: 10.2307/144855.

Cotton, J. (1988). “On the Decomposition of Wage Differentials”. In: *The Review of Economics and Statistics* 70.2. Publisher: The MIT Press, pp. 236–243. DOI: 10.2307/1928307.

Davis, M. M. (2004). “Race-Based Immunization Recommendations and the Potential to Reduce Health Disparities”. In: *JAMA* 291.18, pp. 2253–2255. DOI: 10.1001/jama.291.18.2253.

Díaz de León-Martínez, L. et al. (2020). “Critical review of social, environmental and health risk factors in the Mexican indigenous population and their capacity to respond to the COVID-19”. In: *Science of The Total Environment* 733, p. 139357. DOI: 10.1016/j.scitotenv.2020.139357.

Fairlie, R. W. (1999). “The Absence of the African-American Owned Business: An Analysis of the Dynamics of Self-Employment”. In: *Journal of Labor Economics* 17.1. Publisher: [The University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago], pp. 80–108. DOI: 10.1086/209914.

— (2005). “An extension of the Blinder-Oaxaca decomposition technique to logit and probit models”. In: *Journal of Economic and Social Measurement* 30.4. Publisher: IOS Press, pp. 305–316. DOI: 10.3233/JEM-2005-0259.

Fortin, N., T. Lemieux, and S. Firpo (2011). “Chapter 1 - Decomposition Methods in Economics”. In: *Handbook of Labor Economics*. Ed. by O. Ashenfelter and D. Card. Vol. 4. Elsevier, pp. 1–102. DOI: 10.1016/S0169-7218(11)00407-2.

Giannouchos, T. V. et al. (2020). “Characteristics and risk factors for COVID-19 diagnosis and adverse outcomes in Mexico: an analysis of 89,756 laboratory–confirmed COVID-19 cases”. In: *European Respiratory Journal*. Publisher: European Respiratory Society Section: Original article. DOI: 10.1183/13993003.02144-2020.

Gutierrez, J. P. and S. M. Bertozzi (2020). "Non-communicable diseases and inequalities increase risk of death among COVID-19 patients in Mexico". In: *PLOS ONE* 15.10. Publisher: Public Library of Science, e0240394. DOI: 10.1371/journal.pone.0240394.

Hawkins, R. B., E. J. Charles, and J. H. Mehaffey (2020). "Socio-economic status and COVID-19–related cases and fatalities". In: *Public Health* 189, pp. 129–134. DOI: https://doi.org/10.1016/j.puhe.2020.09.016.

Hernández-Galdamez, D. R. et al. (2020). "Increased Risk of Hospitalization and Death in Patients with COVID-19 and Pre-existing Noncommunicable Diseases and Modifiable Risk Factors in Mexico". In: *Archives of Medical Research* 51.7, pp. 683–689. DOI: 10.1016/j.arcmed.2020.07.003.

Hernández-Garduño, E. (2020). "Obesity is the comorbidity more strongly associated for Covid-19 in Mexico. A case-control study". In: *Obesity Research & Clinical Practice* 14.4, pp. 375–379. DOI: 10.1016/j.orcp.2020.06.001.

Horton, R. (2020). "Offline: COVID-19 is not a pandemic". In: *The Lancet* 396.10255, p. 874. DOI: 10.1016/S0140-6736(20)32000-6.

Ibarra-Nava, I. et al. (2021). "Ethnic disparities in COVID-19 mortality in Mexico: A cross-sectional study based on national data". In: *PLOS ONE* 16.3. Publisher: Public Library of Science, e0239168. DOI: 10.1371/journal.pone.0239168.

INEGI (2020). "Presentación de resultados. Estados Unidos Mexicanos". In: p. 116.

Trabajo, O. I. del, I. Arbeitsorganisation, and ILO, eds. (2009). *Convenio Núm. 169 de la OIT sobre pueblos indígenas y tribales en países independientes: Declaración de las Naciones Unidas sobre los Derechos de los Pueblos Indígenas*. OCLC: 931094126. Lima: OIT. 105 pp.

Jann, B. (2008). "The Blinder–Oaxaca Decomposition for Linear Regression Models:" in: *The Stata Journal*. Publisher: SAGE PublicationsSage CA: Los Angeles, CA. DOI: 10.1177/1536867X0800800401.

— (2018). "Decomposition methods in the social sciences". info:eu-repo/semantics/conferenceObject. Workshop at the Bamberg Graduate School of Social Sciences. University of Bamberg.

Juárez-Ramírez, C. et al. (2014). "La desigualdad en salud de grupos vulnerables de México: adultos mayores, indígenas y migrantes". In: *Revista Panamericana de Salud Pública* 35. Publisher: Organización Panamericana de la Salud, pp. 284–290.

Lambert, F. and P. Hyunmin (2019). "Income Inequality and Government Transfers in Mexico". Working Paper. International Monetary Fund.

Leyva-Flores, R., C. Infante-Xibille, et al. (2013). “Inequidad persistente en salud y acceso a los servicios para los pueblos indígenas de México, 2006-2012”. In: *Salud Pública de México* 55, S123–S128.

Leyva-Flores, R., E. Servan-Mori, et al. (2014). “Primary Health Care Utilization by the Mexican Indigenous Population: The Role of the Seguro Popular in Socially Inequitable Contexts”. In: *PLoS ONE* 9.8. DOI: 10.1371/journal.pone.0102781.

Mancilla-Galindo, J. et al. (2020). “Development and validation of the patient history COVID-19 (PH-Covid19) scoring system: a multivariable prediction model of death in Mexican patients with COVID-19”. In: *Epidemiology and Infection* 148. DOI: 10.1017/S0950268820002903.

McIntyre, P. B. and R. I. Menzies (2005). “Immunisation: reducing health inequality for Indigenous Australians”. In: *Medical Journal of Australia* 182.5. _eprint: https://onlinelibrary.wiley.com/doi/pdf/1 5377.2005.tb06667.x, pp. 207–208. DOI: 10.5694/j.1326-5377.2005.tb06667.x.

Meza, R. et al. (2015). “Burden of Type 2 Diabetes in Mexico: Past, Current and Future Prevalence and Incidence Rates”. In: *Preventive medicine* 81, pp. 445–450. DOI: 10.1016/j.ypmed.2015.10.015.

Monterrubio-Flores, E. et al. (2020). “Impact of a Double Epidemic in Mexico: Non-Communicable Diseases Increase the Case Fatality Rate with Covid-19”. In: *Research Square. Preprint.* DOI: 10.21203/rs.3.rs-80669/v1.

National Council for the Evaluation of Social Development Policy (2018). *Población indígena con carencias en todos sus derechos sociales.* Mexico City: National council for the evaluation of social development policy.

National Institute of Public Health, M. (2018). *ENSANUT 2018. Results.*

Neumark, D. (1988). “Employers’ Discriminatory Behavior and the Estimation of Wage Discrimination”. In: *The Journal of Human Resources* 23.3. Publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System], pp. 279–295. DOI: 10.2307/145830.

Powers, D. A., H. Yoshioka, and M.-S. Yun (2011). “Mvdcmp: Multivariate Decomposition for Nonlinear Response Models”. In: *The Stata Journal* 11.4. Publisher: SAGE Publications, pp. 556–576. DOI: 10.1177/1536867X1201100404.

Puig, A., J. A. Pagán, and R. Wong (2009). “Assessing Quality across Health Care Subsystems in Mexico”. In: *The Journal of Ambulatory Care Management* 32.2, pp. 123–131. DOI: 10.1097/JAC.0b013e31819942e5.

Rahimi, E. and S. S. Hashemi Nazari (2021). "A detailed explanation and graphical representation of the Blinder-Oaxaca decomposition method with its application in health inequalities". In: *Emerging Themes in Epidemiology* 18.1, p. 12. DOI: 10.1186/s12982-021-00100-9.

Sachs, J. D. et al. (2020). "Lancet COVID-19 Commission Statement on the occasion of the 75th session of the UN General Assembly". In: *The Lancet* 0.0. Publisher: Elsevier. DOI: 10.1016/S0140-6736(20)31927-9.

Saeedi, P. et al. (2019). "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition". In: *Diabetes Research and Clinical Practice* 157. Publisher: Elsevier. DOI: 10.1016/j.diabres.2019.107843.

Sánchez T., M. (2020). *La letalidad hospitalaria por covid-19 en México: desigualdades institucionales*. Taller de datos. URL: https://datos.nexos.com.mx (visited on 02/10/2021).

Servan-Mori, E. et al. (2014). "An explanatory analysis of economic and health inequality changes among Mexican indigenous people, 2000-2010". In: *International Journal for Equity in Health* 13, p. 21. DOI: 10.1186/1475-9276-13-21.

Serván-Mori, E. et al. (2021). "Hospitalisation and mortality from COVID-19 in Mexican indigenous people: a cross-sectional observational study". In: *J Epidemiol Community Health*. Publisher: BMJ Publishing Group Ltd Section: Original research. DOI: 10.1136/jech-2020-216129.

Słoczyński, T. (2015). "The Oaxaca–Blinder Unexplained Component as a Treatment Effects Estimator". In: *Oxford Bulletin of Economics and Statistics* 77.4. _eprint: https://onlinelibrary.wiley.com/doi/p pp. 588–604. DOI: 10.1111/obes.12075.

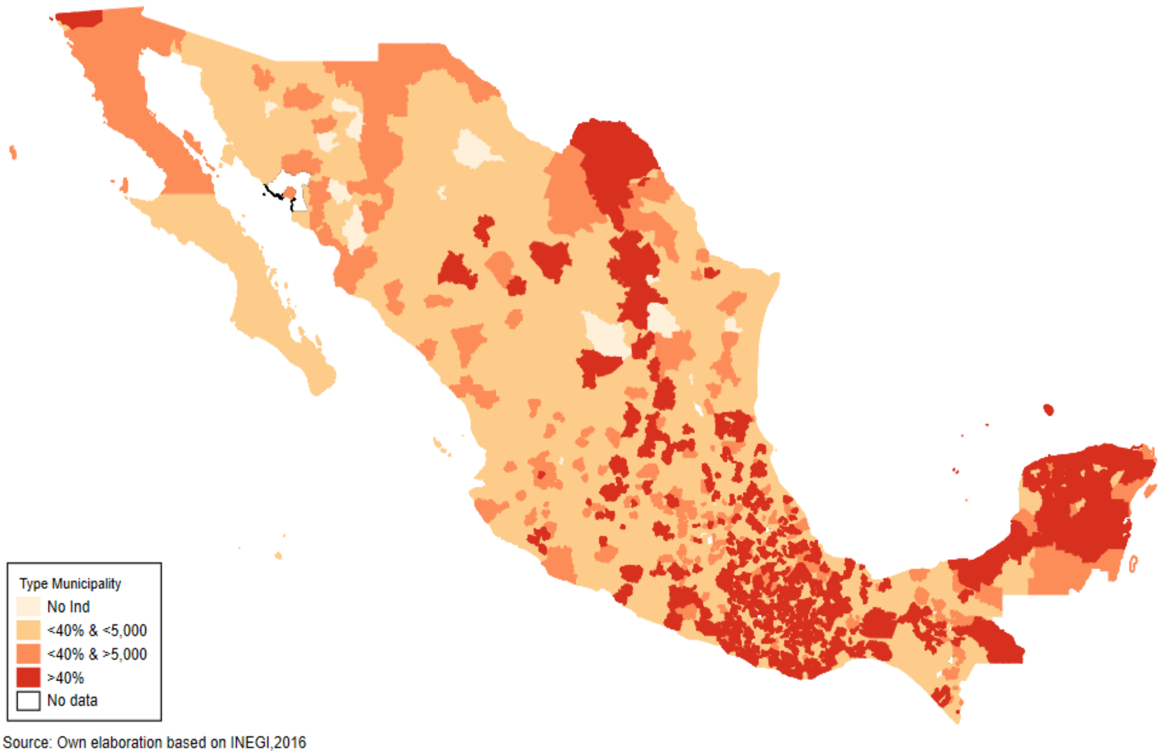
Soberanes, R. (2021). *Las muertes ocultas de los indígenas chiapanecos que temen al Estado*. Mexico: Mexicanos contra la corrupción y la impunidad.

Tai, D. B. G. et al. (2020). "The Disproportionate Impact of COVID-19 on Racial and Ethnic Minorities in the United States". In: *Clinical Infectious Diseases*. DOI: 10.1093/cid/ciaa815.

Yun, M.-S. (2004). "Decomposing differences in the first moment". In: *Economics Letters* 82.2, pp. 275–280. DOI: 10.1016/j.econlet.2003.09.008.

A Electronic Supplementary Section

Figure A.1: Distribution of indigenous people in Mexico



A.1 Oaxaca-Blinder decomposition

A.1.1 Linear model

Aggregate decomposition

Following Jann, 2018's notation, the standard Oaxaca-Blinder decomposition starts with the following structural function

$$Y_i^g = m^g(X_i, \epsilon_i) = \beta_0^g + \beta_1^g X_{1i} + ... + \beta_k^g X_{ki} + \epsilon_i^g, \qquad g = 0, 1 \tag{1}$$

Where Y^g represents the health outcome for group g. X_k depicts different factors that influence the outcome, i indexes individuals and g represents the comparison and reference groups and ϵ_i^g is the idiosyncratic error term of the model. It assumes *additive linearity*: $m(X, \epsilon) = X\beta^g + \epsilon^g$. This implies that the effect of observed and unobserved characteristics are additively separable in $m()$. It further assumes *zero conditional mean independence*, $E(\epsilon \mid X, G) = 0$.
Thus, the average group difference can be expressed as:

$$\begin{aligned}
\Delta^\mu =& \mu(F_{Y|G=0}) - \mu(F_{Y|G=1}) = E(Y^0 \mid G = 0) - E(Y^1 \mid G = 1) \\
=& E(X\beta^0 + \epsilon \mid G = 0) - E(X\beta^1 + \epsilon \mid G = 1) \\
=& (E(X\beta^0 \mid G = 0) + E(\epsilon \mid G = 0)) - (E(X\beta^1 \mid G = 1) + E(\epsilon \mid G = 1)) \\
=& E(X\beta^0 \mid G = 0) - E(X\beta^1 \mid G = 1)
\end{aligned}$$

$$\Delta^\mu = E(X \mid G = 0)\beta^0 - E(X \mid G = 1)\beta^1 \tag{2}$$

The Oaxaca-Blinder decomposition requires a counterfactual that illustrates what would happen if characteristics of one group were interchanged with coefficients of the other group. Thus this counterfactual could be $F_Y^0 \mid G = 1$, which depicts the average expected outcome for group 1 if they had the characteristics of group 0.

$$\begin{aligned}
\mu(F_{Y^0} \mid G = 1) =& E(X\beta^0 + \epsilon \mid G = 1) \\
=& E(X\beta^0 \mid G = 1) \\
=& E(X \mid G = 1)\beta^0
\end{aligned}$$

Subtracting and adding $E(X \mid G = 1)\beta^0$ in Equation (2):

$$\begin{aligned}
\Delta^{\mu} =& E(X \mid G = 0)\beta^0 - E(X \mid G = 1)\beta^1 \\
=& E(X \mid G = 0)\beta^0 - E(X \mid G = 1)\beta^0 + E(X \mid G = 1)\beta^0 - E(X \mid G = 1)\beta^1 \\
=& (E(X \mid G = 0) - E(X \mid G = 1))\beta^0 + E(X \mid G = 1)(\beta^0 - \beta^1) \\
\Delta^{\mu} =& \Delta^{\mu}_X + \Delta^{\mu}_{\beta}
\end{aligned}$$

β^g can be estimated using a linear regression on the $G = g$ sub-sample and $E(X \mid G = g)$ is the vector of means of X in the same sub-sample. If $\hat{\beta}^g$ is the estimate of β^g and $\bar{X}^g = \hat{E}(X \mid G = g)$ of $E(X \mid G = g)$, the decomposition estimate can be written as follows:

$$\hat{\Delta}^{\mu} = \hat{\Delta}^{\mu}_X + \hat{\Delta}^{\mu}_{\beta} = \underbrace{(\bar{X}^0 - \bar{X}^1)\hat{\beta}^0}_{\text{Explained}} + \underbrace{\bar{X}^1(\hat{\beta}^0 - \hat{\beta}^1)}_{\text{Unexplained}} \tag{3}$$

The decomposition depicted in 3 is seen from group 1's perspective, as this is taken as the reference group. If this was changed, the results of the decomposition would change. This issue is known as the *indexing problem* (Neumark, 1988; Cotton, 1988) and implies that results are not unique and depend on the group chosen as reference. The decision of which group to take as reference should be made based on a preconception of discrimination, if this exist. In our case, given the consistent evidence about the unequal treatment between indigenous and non-indigenous people in Mexico (Servan-Mori et al., 2014; Leyva-Flores, Infante-Xibille, et al., 2013; Leyva-Flores, Servan-Mori, et al., 2014; National Council for the Evaluation of Social Development Policy, 2018), we believe that the assumption of discrimination against indigenous people holds and therefore, we undertake the decompositions using indigenous people as reference group.

Detailed decomposition

For policy purposes, it is relevant to further identify and measure the main factors contributing to the explained and unexplained part of the ethnic gap. Thus, given the assumption of *additive linearity*, both the explained and unexplained part can be further decomposed in order to disentangle the contribution of the k^{th} explanatory variable to the ethnic gap. Thus, from Equation (3) the explained part can be decomposed as:

$$\begin{aligned}\hat{\Delta}^{\mu}_{X} =&(\bar{X}^0 - \bar{X}^1)\hat{\beta}^0 \\ =&\sum_{k=1}^{K} \hat{\beta}^0_k(\bar{X}^0_k - \bar{X}^1_k) \\ =&\hat{\beta}^0_1(\bar{X}^0_1 - \bar{X}^1_1) + \hat{\beta}^0_2(\bar{X}^0_2 - \bar{X}^1_2) + ... + \hat{\beta}^0_k(\bar{X}^0_k - \bar{X}^1_k)\end{aligned}$$

and the unexplained part can be decomposed as:

$$\begin{aligned}\hat{\Delta}^{\mu}_{\beta} =&\bar{X}^1(\hat{\beta}^0 - \hat{\beta}^1) \\ =&\underbrace{(\hat{\beta_0}^0 - \hat{\beta_0}^1)}_{\text{Intercepts}} + \sum_{k=1}^{K}(\hat{\beta_k}^0 - \hat{\beta_k}^1)\bar{X}^1_k \\ =&(\hat{\beta_0}^0 - \hat{\beta_0}^1) + (\hat{\beta_1}^0 - \hat{\beta_1}^1)\bar{X}^1_1 + (\hat{\beta_2}^0 - \hat{\beta_2}^1)\bar{X}^1_2 + ... + (\hat{\beta_k}^0 - \hat{\beta_k}^1)\bar{X}^1_k\end{aligned}$$

Moreover, within each component, variables of k can be aggregated into subsets, for example:

$$\hat{\Delta}^{\mu}_{X} = \sum_{k=1}^{a} \hat{\beta}^0_k(\bar{X}^0_k - \bar{X}^1_k) + \sum_{k=a+1}^{b} \hat{\beta}^0_k(\bar{X}^0_k - \bar{X}^1_k) + ...$$

and for the unexplained part:

$$\hat{\Delta}^{\mu}_{\beta} = \underbrace{(\hat{\beta_0}^0 - \hat{\beta_0}^1)}_{\text{Intercepts}} + \sum_{k=1}^{a}(\hat{\beta_k}^0 - \hat{\beta_k}^1)\bar{X}^1_k + \sum_{k=a+1}^{b}(\hat{\beta_k}^0 - \hat{\beta_k}^1)\bar{X}^1_k + ...$$

A.1.2 Nonlinear model

Aggregate decomposition

The set up is the same, we are interesting in an average decomposition. Nevertheless, the first challenge in the nonlinear model is that $E(Y \mid X) = F(X\beta) \neq F(\overline{X}\beta)$. With the nonlinear case it is not possible to insert in $E(X)$ into $F(.)$ to get $E(Y)$. Therefore, the difficulty is to generate $\hat{Y} = \hat{E}(Y \mid X) = F(X; \hat{\beta})$ and this implies knowing the functional form for $F(.)$. In this respect, Yun (2004) states that any aggregate Oaxaca-Blinder decomposition is feasible as long as the function $F(.)$ is once-differentiable. Fairlie (1999) for example, proposed an extension of the Oaxaca decomposition using the logit function. According to Fairlie (2005), the decomposition of a nonlinear equation such as $Y = F(X; \hat{\beta})$ can also be written as:

$$\hat{\Delta}^{\mu} = \underbrace{\left[\frac{1}{N^0}\sum_{G_i=0} F(X_i^0\hat{\beta}^0) - \frac{1}{N^1}\sum_{G_i=1} F(X_i^1\hat{\beta}^0)\right]}_{\text{Explained}} + \underbrace{\left[\frac{1}{N^1}\sum_{G_i=1} F(X^1\hat{\beta}^0) - \frac{1}{N^1}\sum_{G_i=1} F(X^1\hat{\beta}^1)\right]}_{\text{Unexplained}} \tag{4}$$

Applied to the logit function, $\hat{\Delta}^{\mu}$ denotes the predicted average difference in the coefficients of the binary outcome of interest and $F(.)$ represents the cumulative distribution function from the logistic distribution: $\frac{1}{1+e^{-X\beta}}$. Fairlie (2005) also points out the useful property of the logit regression in that by including a constant term, the average of the predicted probabilities must equal the proportion of the sample. Equation (4) shows that the difference in the predicted average observed outcomes can be decomposed into the explained and unexplained components.

Detailed decomposition

One issue that the nonlinear case faces when estimating a detailed decomposition is the *path dependence problem* (Yun, 2004; Fortin et al., 2011; Powers et al., 2011). Although, there are different ways to tackle this problem, we follow the solution proposed by Yun (2004) which is simple, but robust: a linearisation around $E(X)\beta$ using a set of weights from a first-order Taylor linearisation around Equation (4). This allows to get the contribution of the covariates to Δ_X^{μ} and Δ_{β}^{μ} as relative contributions fixed at the level of the linear predictor(Jann, 2018). For this, let $\hat{E}(X|G = g) = \bar{X}^g$ and $\hat{E}(F(X\beta)|G = g) = \overline{F(X\beta)}^g$. Thus, the aggregate decomposition can be expressed as:

$$\hat{\Delta}^{\mu} = \left\{\overline{F(X\hat{\beta}^0)}^0 - \overline{F(X\hat{\beta}^0)}^1\right\} + \left\{\overline{F(X\hat{\beta}^0)}^1 - \overline{F(X\hat{\beta}^1)}^1\right\} = \hat{\Delta}_X^{\mu} + \hat{\Delta}_{\beta}^{\mu}$$

The individual contribution of each covariate to the characteristics and coefficients effects can be estimated as (ibid.):

$$\hat{\Delta}_{X,X_k}^{\mu} = \frac{(\bar{X}_k^0 - \bar{X}_k^1)\hat{\beta}_k^0}{(\bar{X}^0 - \bar{X}^1)\hat{\beta}^0}\hat{\Delta}_X^{\mu}$$

and

$$\hat{\Delta}_{\beta,\beta_k}^{\mu} = \frac{\bar{X}_k^1(\hat{\beta}_k^0 - \hat{\beta}_k^1)}{\bar{X}^1(\hat{\beta}^0 - \hat{\beta}^1)}\hat{\Delta}_{\beta}^{\mu}$$

such that $\sum_{i=1}^{K}\hat{\Delta}_{X,X_k}^{\mu} = \hat{\Delta}_X^{\mu}$ and $\sum_{i=1}^{K}\hat{\Delta}_{\beta,X_k}^{\mu} = \hat{\Delta}_{\beta}^{\mu}$. Thus, Yun (2004) proposes to approximate $\hat{\Delta}^{\mu}$ by first evaluating the function $F(.)$ at the means of the covariates,

$$\hat{\Delta}^{\mu} \approx \left[F(\bar{X}^0\hat{\beta}^0) - F(\bar{X}^1\hat{\beta}^0)\right] + \left[F(\bar{X}^1\hat{\beta}^0) - F(\bar{X}^1\hat{\beta}^1)\right]$$

and then linearising the differences around $\bar{X}^0\hat{\beta}^0$ and $\bar{X}^1\hat{\beta}^1$ using a first order Taylor expansion (Jann, 2018), as follows:

$$\begin{aligned}\hat{\Delta}^{\mu} &\approx \left[F(\bar{X}^0\hat{\beta}^0) - F(\bar{X}^1\hat{\beta}^0)\right] + \left[F(\bar{X}^1\hat{\beta}^0) - F(\bar{X}^1\hat{\beta}^1)\right] + R_M \\ &\approx \left[(\bar{X}^0 - \bar{X}^1)\hat{\beta}^0)\right] \cdot d^0 + \left[\bar{X}^1(\hat{\beta}^0 - \hat{\beta}^1)\right] \cdot d^1 + R_M + R_T\end{aligned}$$

where

$$R_M = \left[\overline{F(X\hat{\beta}^0)}^0 - \overline{F(X\hat{\beta}^0)}^1\right] + \left[\overline{F(X\hat{\beta}^0)}^1 - \overline{F(X\hat{\beta}^1)}^1\right] - \left[F(\bar{X}^0\hat{\beta}^0) - F(\bar{X}^1\hat{\beta}^0)\right] - \left[F(\bar{X}^1\hat{\beta}^0) - F(\bar{X}^1\hat{\beta}^1)\right]$$

and

$$R_T = \left[F(\bar{X}^0\hat{\beta}^0) - F(\bar{X}^1\hat{\beta}^0)\right] + \left[F(\bar{X}^1\hat{\beta}^0) - F(\bar{X}^1\hat{\beta}^1)\right] - \left[(\bar{X}^0 - \bar{X}^1)\beta^0 \cdot d^0\right] - \left[\bar{X}^1(\beta^0 - \beta^1) \cdot d^1\right]$$

where d^g represents the first derivative of $F(\bar{X}^g\hat{\beta}^g) = \frac{\partial F(\bar{X}^g\hat{\beta}^g)}{\partial(\bar{X}^g\hat{\beta}^g)}$. Yun (2004) also mentions that R_M and R_T are approximation residuals from the evaluation of the function $F(.)$ at the means values and the linearisation. After this, the set of weights for the explained part can be calculated as:

$$W_{\Delta_{Xk}} = \frac{((\bar{X}^0_k - \bar{X}^1_k)\hat{\beta}^0_k)d^0}{((\bar{X}^0 - \bar{X}^1)\hat{\beta}^0)d^0} = \frac{(\bar{X}^0_k - \bar{X}^1_k)\hat{\beta}^0_k}{(\bar{X}^0 - \bar{X}^1)\hat{\beta}^0}$$

and for the unexplained part as:

$$W_{\Delta_{\beta k}} = \frac{((\hat{\beta}^0_k - \hat{\beta}^1_k)\bar{X}^1_k)d^1}{((\hat{\beta}^0 - \hat{\beta}^1)\bar{X}^1)d^1} = \frac{(\hat{\beta}^0_k - \hat{\beta}^1_k)\bar{X}^1_k}{(\hat{\beta}^0 - \hat{\beta}^1)\bar{X}^1}$$

and

$$W_{\Delta_{Xk}} = W_{\Delta_{\beta k}} = 1$$

The weights, $W_{\Delta_{Xk}}$, show the contribution of the k^{th} variable to the linearisation of the explained part according to the magnitude of the mean group difference and accounting for the reference group's effect (Powers et al., 2011). Thus, this detailed decomposition using weights are path invariant. The decomposition can be expressed in terms of the overall components as a sum of weighted sums of the unique contributions, as:

$$\overline{Y}_A - \overline{Y}_B = E + U = \sum_{k=1}^{K} W_{\Delta_{X\,k}} E + \sum_{k=1}^{K} W_{\Delta_{\beta\,k}} U = \sum_{k=1}^{K} E_k + \sum_{k=1}^{K} U_k$$

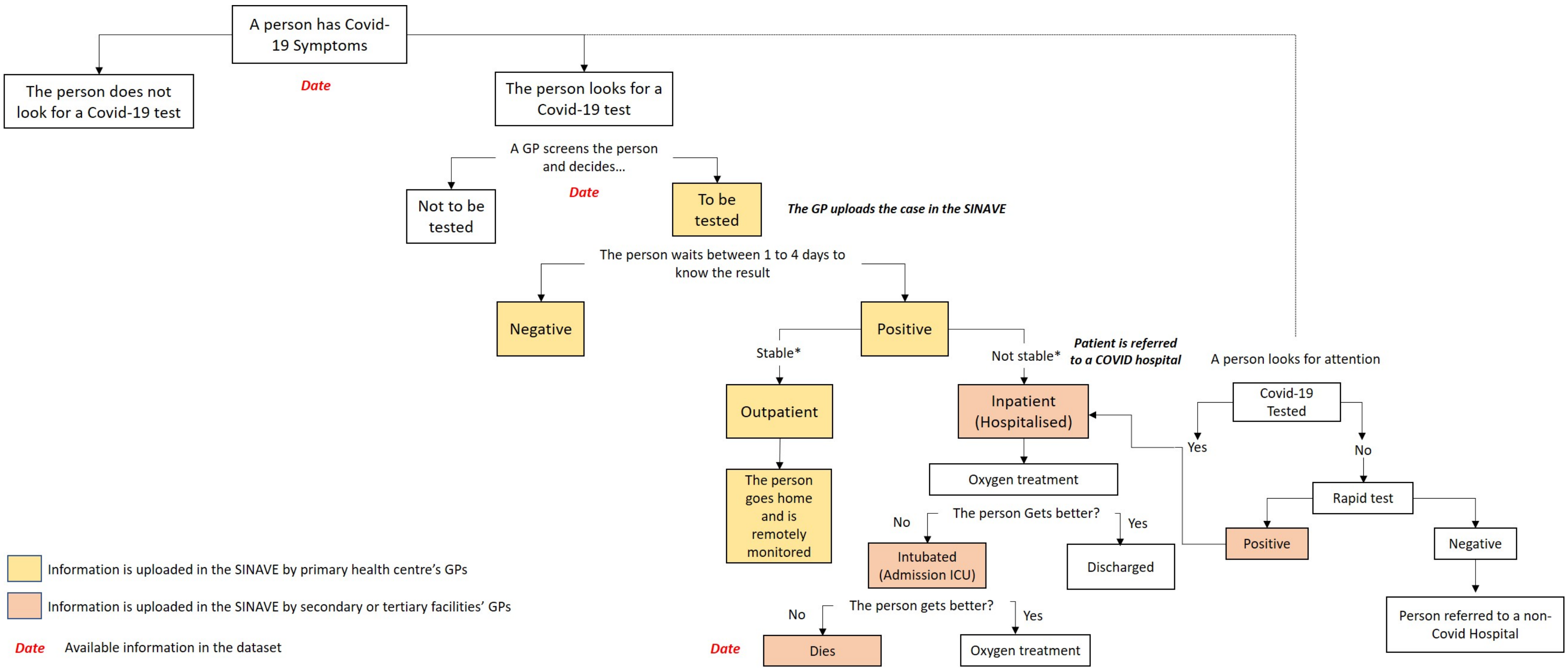
Jann (2018) warns that if the volume of data is in highly nonlinear regions of $F(.)$, or differences in coefficients or means are large, the approximation could be poor.

A.2 Covid-19 procedures and data collection

The testing procedure was as follows: people who have symptoms and sought out a test arrive at the health unit (this assumes that people are physically capable of going to their regular health unit). Once in the health facility, the general practitioner (GP) screens the patient and decides if the patient meets the inclusion criteria to be tested for Covid-19. If patients are tested, GPs capture information about their medical history, the current date and the date when the patient first showed a symptom. This information is recorded in an online platform called SINAVE (National Epidemiological Surveillance System). Cases in the dataset represent both ambulatory (outpatient) and hospitalised (inpatient) individuals. Swabs are obtained from outpatients and samples are sent to the nearest Laboratory of Respiratory Virus (InDRE). This process could take up to four days. If the case is positive, there are two potential paths to follow which depends on the health status of the patient. If the person is clinically assessed and diagnosed as with a mild to moderate infection, the person can remain at home and be remotely monitored. Follow-up of all suspected Covid-19 cases and ambulatory patients is done by the responsible healthcare professional of every Local Health Jurisdiction. This person is also in charge of uploading the data into SINAVE. Due to collection procedures, a patient who is tested more than once in different jurisdictions and at different points in time may lead to duplicate records as there is no unique identification variable available to identify individual patients.

A patient clinically diagnosed with a complicated to severe infection (when the patient has difficulties with breathing or hypoxemia) is admitted to a specialised Covid-19 hospital. In these hospitals, patients immediately receive drug and oxygen treatment. If patients do not respond favourably to the treatment, they can be admitted to the intensive care unit (ICU). It is also possible that patients who never asked for a test when they first felt symptoms could arrive at a hospital seeking medical attention, without any previous test or clinical record. In this scenario, patients are rapidly screened and if the test is positive the patient is admitted to a Covid-hospital; if not, the patient is referred to another hospital to receive care. In the case of patients that for some reason are already intubated, bronchoalveolar lavage sample is obtained and tested for Covid-19. If an inpatient died due to suspected Covid-19, lung biopsies are obtained from an autopsy. Reporting of deaths is obligatory and must be done in less than 48 hours after occurrence. If patients are not able to give details about their medical history, this is retrieved from records. All these data are undertaken by accredited hospital epidemiologists and uploaded in the SINAVE.

Figure A.2: Covid-19 procedures and data collection



A.3 Variable definitions

Table A.1: Definition of individual-level variables

Dimension	Variable	Definition
		Individual-level characteristics
Demographics	Sex	Sex of the individual. 1 if female, 0 male
	Age	Individual years of age
Underlying Health Conditions	Pneumonia	1 if the patient has a diagnosis of pneumonia, 0 otherwise
	Hypertension	1 if the patient has a diagnosis of hypertension, 0 otherwise
	Diabetes	1 if the patient has a diagnosis of diabetes, 0 otherwise
	COPD	Chronic obstructive pulmonary disease. 1 if the person has a diagnosis of a COPD, 0 otherwise
	Asthma	1 if the patient has a diagnosis of asthma, 0 otherwise
	Immunosuppression	1 if the patient has immunosuppression, 0 otherwise
	Renal disease	1 if the patient has a diagnosis of a renal disease, 0 otherwise
	Cardiovascular disease	1 if the patient has a diagnosis of a cardiovascular disease, 0 otherwise
	Other	Other comorbidities
Risky health behaviours	Obesity	To be obese. 1 if the patient has obesity, 0 otherwise. There is no clinical definition available in the dataset
	Smoking	To smoke. 1 if the patient smokes regularly, 0 otherwise
Institution where individuals received medical attention	Testing waiting-time	Number of days the person waited to get tested since the first symptom
	IMSS	Mexican Social Security Institute
	ISSSTE	Civil Service Social Security and Services Institute
	SSA	Health Ministry. SSA hospitals provide health services to people enrolled in the INSABI programme, former known as "Seguro Popular"
	Federal States	Hospitals owned and managed by the Federal States
	PEMEX	Hospitals owned and managed by the state-owned petroleum company "Mexican Petroleum"
	SEDENA	Hospitals owned and managed by the Secretariat of National Defence
	SEMAR	Hospitals owned and managed by the Secretariat of the Navy

A.4 LPM and logit regression models

Tables A.2 and A.3 depict the results from the linear and nonlinear regression models, respectively. There are two columns associated to each outcome, the first shows the coefficients for non-indigenous($G_i = 0$) and second indigenous($G_i = 1$). Table A.3 used the a logit function and coefficients are expressed in log-odds.

Table A.2: Linear regression results for indigenous and non-indigenous people. All outcomes

	Hosp_Gi=0		**Hosp_Gi=1**		**ICU_Gi=0**		**ICU_Gi=1**		**Dead_Gi=0**		**Dead_Gi=1**	
Demographics												
Age	0.00***	(0.00)	0.00***	(0.00)	-0.00***	(0.00)	-0.00*	(0.00)	0.00***	(0.00)	0.00***	(0.00)
Women	-0.02***	(0.00)	-0.01**	(0.00)	-0.01***	(0.00)	-0.02**	(0.01)	-0.02***	(0.00)	-0.02***	(0.00)
Comorbidities												
COPD	0.08***	(0.00)	0.05***	(0.01)	0.00	(0.00)	-0.01	(0.01)	0.04***	(0.00)	-0.01	(0.01)
Asthma	-0.01***	(0.00)	-0.01	(0.01)	-0.01**	(0.00)	-0.00	(0.02)	-0.01***	(0.00)	0.00	(0.01)
Immunosuppression	0.10***	(0.00)	0.14***	(0.02)	0.02***	(0.00)	0.03	(0.02)	0.01***	(0.00)	-0.01	(0.01)
Renal D.	0.16***	(0.00)	0.10***	(0.01)	-0.01***	(0.00)	-0.02	(0.01)	0.08***	(0.00)	0.05***	(0.01)
Pneumonia	0.69***	(0.00)	0.69***	(0.00)	0.05***	(0.00)	0.07***	(0.01)	0.32***	(0.00)	0.35***	(0.00)
Other C.	0.08***	(0.00)	0.09***	(0.01)	0.01***	(0.00)	0.02	(0.02)	0.02***	(0.00)	0.03**	(0.01)
NCD												
Diabetes	0.06***	(0.00)	0.07***	(0.01)	0.00***	(0.00)	-0.01	(0.01)	0.03***	(0.00)	0.01***	(0.00)
Hypertension	0.03***	(0.00)	0.03***	(0.01)	0.00***	(0.00)	0.01	(0.01)	0.02***	(0.00)	0.02***	(0.00)
Cardio D.	0.07***	(0.00)	0.07***	(0.01)	0.01***	(0.00)	0.01	(0.02)	0.02***	(0.00)	0.00	(0.01)
Risky Behaviours												
Smoking	-0.01***	(0.00)	-0.02*	(0.01)	-0.01***	(0.00)	-0.03*	(0.01)	-0.00***	(0.00)	-0.01	(0.01)
Obesity	-0.00***	(0.00)	-0.00	(0.01)	0.02***	(0.00)	0.02**	(0.01)	0.00***	(0.00)	0.02***	(0.00)
Medical Att.												
Wait Test	0.00***	(0.00)	0.00***	(0.00)	0.00***	(0.00)	-0.00	(0.00)	0.00***	(0.00)	0.00***	(0.00)
Private	-0.15***	(0.00)	-0.19**	(0.06)	0.12***	(0.00)	0.15+	(0.08)	-0.02***	(0.00)	-0.03	(0.05)
IMSS	-0.11***	(0.00)	-0.11+	(0.06)	-0.11***	(0.00)	-0.12+	(0.07)	0.05***	(0.00)	0.05	(0.05)
ISSSTE	-0.11***	(0.00)	-0.12*	(0.06)	-0.05***	(0.00)	-0.09	(0.07)	0.01**	(0.00)	0.02	(0.05)
SSA	-0.20***	(0.00)	-0.21***	(0.06)	-0.00	(0.00)	-0.05	(0.07)	0.00	(0.00)	0.00	(0.05)
State	-0.19***	(0.00)	-0.21***	(0.06)	-0.04***	(0.01)	-0.05	(0.08)	-0.01***	(0.00)	-0.04	(0.05)
PEMEX	-0.16***	(0.00)	-0.24***	(0.06)	-0.05***	(0.01)	0.14	(0.09)	-0.03***	(0.00)	-0.05	(0.05)
SEDENA	0.17***	(0.00)	0.28***	(0.06)	0.06***	(0.01)	0.03	(0.07)	0.03***	(0.00)	0.13*	(0.05)
SEMAR	-0.12***	(0.00)	-0.14+	(0.08)	0.09***	(0.01)	0.06	(0.12)	-0.01*	(0.00)	-0.05	(0.06)
Health Infrastructure												
No Affiliated	0.00***	(0.00)	0.00**	(0.00)	-0.00***	(0.00)	0.00*	(0.00)	0.00***	(0.00)	0.00	(0.00)
Health Fac.	-0.00***	(0.00)	-0.00***	(0.00)	-0.00***	(0.00)	-0.00	(0.00)	-0.00	(0.00)	0.00	(0.00)
Urban loc.	-0.00***	(0.00)	-0.00	(0.00)	0.00	(0.00)	-0.00+	(0.00)	-0.00***	(0.00)	-0.00	(0.00)
Household Ch.												
Overcrow	0.00	(0.00)	0.00	(0.01)	-0.02***	(0.00)	-0.07***	(0.01)	-0.00+	(0.00)	0.01*	(0.01)
Perc. Floors	-0.00***	(0.00)	-0.00	(0.00)	0.01***	(0.00)	0.00	(0.00)	-0.00+	(0.00)	-0.00	(0.00)
Perc. Sleeping	0.00**	(0.00)	0.00	(0.00)	-0.00***	(0.00)	0.00	(0.00)	0.00**	(0.00)	0.00	(0.00)
Perc. Room	-0.01***	(0.00)	-0.01*	(0.00)	0.01***	(0.00)	-0.00	(0.00)	-0.00**	(0.00)	0.00	(0.00)
No Water	-0.00**	(0.00)	-0.00+	(0.00)	0.00*	(0.00)	0.01**	(0.00)	0.00**	(0.00)	-0.00	(0.00)
No Elec.	-0.04***	(0.00)	-0.03***	(0.01)	-0.01**	(0.00)	0.02+	(0.01)	-0.01***	(0.00)	-0.01	(0.00)
Latrine	-0.00*	(0.00)	-0.01***	(0.00)	-0.00***	(0.00)	-0.00	(0.00)	-0.00***	(0.00)	-0.00	(0.00)
No Drainage	0.00***	(0.00)	0.01***	(0.00)	-0.00***	(0.00)	0.00	(0.00)	0.00***	(0.00)	0.00	(0.00)
No E,W,D	0.01*	(0.00)	0.05***	(0.01)	0.02**	(0.01)	-0.02	(0.01)	-0.01***	(0.00)	-0.01	(0.01)
Perc. No car	0.00***	(0.00)	-0.00	(0.00)	-0.00***	(0.00)	0.00	(0.00)	-0.00***	(0.00)	0.00+	(0.00)
Perc. No appl.	0.00	(0.00)	0.03***	(0.01)	0.02***	(0.00)	0.00	(0.01)	-0.01***	(0.00)	-0.01	(0.01)
Perc. No TV	0.01***	(0.00)	0.01	(0.00)	-0.01***	(0.00)	0.00	(0.01)	-0.01***	(0.00)	-0.01*	(0.00)
Perc. No phone	-0.01***	(0.00)	-0.01**	(0.00)	-0.00*	(0.00)	0.01+	(0.00)	-0.00***	(0.00)	-0.01***	(0.00)
Perc. No Comp.	0.00***	(0.00)	0.01***	(0.00)	0.00***	(0.00)	-0.01***	(0.00)	0.00***	(0.00)	0.00***	(0.00)
Perc. NO ICT	0.02***	(0.00)	-0.01	(0.01)	-0.00	(0.01)	-0.01	(0.02)	0.02***	(0.00)	0.02*	(0.01)
Economic Ch.												
Perc. No School Ch	0.02***	(0.00)	0.01	(0.01)	0.01***	(0.00)	0.02	(0.01)	-0.00**	(0.00)	-0.01	(0.01)
Perc. Illiterate Ch.	-0.01***	(0.00)	0.00	(0.01)	-0.00	(0.00)	-0.01	(0.01)	0.01***	(0.00)	0.00	(0.01)
Perc. Illiterate Adu.	0.00+	(0.00)	-0.00	(0.00)	0.01***	(0.00)	-0.00	(0.00)	-0.00	(0.00)	0.00	(0.00)
Perc. No School a	-0.00***	(0.00)	0.00	(0.00)	-0.01***	(0.00)	0.00	(0.00)	-0.00	(0.00)	-0.00	(0.00)
Perc. Unemployed	-0.02***	(0.00)	0.00	(0.00)	-0.02***	(0.00)	0.01+	(0.01)	-0.00***	(0.00)	-0.01+	(0.00)
_cons	0.15***	(0.00)	0.14*	(0.07)	0.27***	(0.01)	0.39***	(0.09)	-0.07***	(0.00)	-0.17**	(0.06)
N	4,765,878		30,930		604,894		7,533		4,765,878		30,930	
r2	.499		.526		.0785		.0619		.291		.311	

Notes: standard errors in parenthesis + p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table A.3: Nonlinear regression results for indigenous and non-indigenous people. All outcomes

	Hosp_Gi=0		**Hosp_Gi=1**		**ICU_Gi=0**		**ICU_Gi=1**		**Dead_Gi=0**		**Dead_Gi=1**	
Demographics												
Age	0.03***	(0.00)	0.01***	(0.00)	-0.01***	(0.00)	-0.00*	(0.00)	0.06***	(0.00)	0.05***	(0.00)
Women	-0.37***	(0.00)	-0.12**	(0.04)	-0.14***	(0.01)	-0.28**	(0.09)	-0.54***	(0.01)	-0.39***	(0.05)
Comorbidities												
COPD	0.52***	(0.01)	0.38***	(0.11)	0.03	(0.03)	-0.16	(0.21)	-0.07***	(0.01)	-0.25*	(0.10)
Asthma	-0.22***	(0.01)	-0.10	(0.12)	-0.12***	(0.04)	-0.08	(0.28)	-0.21***	(0.02)	0.06	(0.14)
Immunosuppression	1.04***	(0.01)	1.21***	(0.15)	0.25***	(0.03)	0.27	(0.25)	0.41***	(0.02)	0.19	(0.18)
Renal D.	1.20***	(0.01)	0.85***	(0.12)	-0.13***	(0.03)	-0.33	(0.24)	0.64***	(0.01)	0.40***	(0.12)
Pneumonia	4.11***	(0.01)	3.92***	(0.05)	1.03***	(0.01)	1.28***	(0.12)	2.81***	(0.01)	2.79***	(0.05)
Other C.	0.75***	(0.01)	0.67***	(0.11)	0.26***	(0.03)	0.21	(0.22)	0.25***	(0.01)	0.39**	(0.13)
NCD												
Diabetes	0.63***	(0.01)	0.61***	(0.05)	0.07***	(0.01)	-0.10	(0.10)	0.40***	(0.01)	0.27***	(0.06)
Hypertension	0.30***	(0.01)	0.28***	(0.05)	0.07***	(0.01)	0.13	(0.11)	0.16***	(0.01)	0.17**	(0.06)
Cardio D.	0.55***	(0.01)	0.63***	(0.12)	0.17***	(0.02)	0.17	(0.22)	-0.08***	(0.01)	-0.08	(0.13)
Risky Behaviours												
Smoking	-0.20***	(0.01)	-0.24**	(0.09)	-0.10***	(0.02)	-0.46*	(0.20)	-0.20***	(0.01)	-0.22*	(0.10)
Obesity	0.04***	(0.01)	-0.03	(0.06)	0.25***	(0.01)	0.26*	(0.11)	0.32***	(0.01)	0.40***	(0.06)
Medical Att.												
Wait Test	0.03***	(0.00)	0.03***	(0.01)	0.01***	(0.00)	-0.01	(0.01)	0.04***	(0.00)	0.04***	(0.01)
Private	-1.54***	(0.03)	-1.62**	(0.56)	0.62***	(0.06)	0.68	(0.84)	-0.59***	(0.06)	-1.40+	(0.77)
IMSS	-0.98***	(0.03)	-0.86	(0.52)	-2.34***	(0.06)	-1.91*	(0.80)	1.15***	(0.06)	-0.05	(0.71)
ISSSTE	-1.07***	(0.03)	-0.94+	(0.53)	-0.70***	(0.06)	-1.04	(0.81)	0.21***	(0.06)	-0.47	(0.71)
SSA	-2.42***	(0.03)	-1.87***	(0.52)	-0.24***	(0.06)	-0.50	(0.79)	-0.18**	(0.06)	-0.65	(0.71)
State	-2.14***	(0.03)	-1.82**	(0.58)	-0.61***	(0.07)	-0.55	(0.92)	0.09	(0.06)	-1.06	(0.77)
PEMEX	-1.70***	(0.04)	-2.14***	(0.62)	-0.71***	(0.07)	0.37	(0.89)	-0.34***	(0.06)	-1.40+	(0.80)
SEDENA	0.95***	(0.03)	1.53**	(0.53)	0.59***	(0.06)	0.40	(0.81)	0.80***	(0.06)	0.88	(0.72)
SEMAR	-1.05***	(0.04)	-1.13	(0.71)	0.51***	(0.08)	0.53	(1.15)	-0.01	(0.07)	-1.91	(1.32)
Health Infrastructure												
No Affiliated	0.00***	(0.00)	0.01+	(0.00)	-0.02***	(0.00)	0.02**	(0.01)	0.01***	(0.00)	0.00	(0.00)
Health Fac.	-0.00***	(0.00)	-0.00**	(0.00)	-0.00*	(0.00)	-0.00	(0.00)	-0.00***	(0.00)	0.00	(0.00)
Urban loc.	-0.00***	(0.00)	-0.00***	(0.00)	-0.00*	(0.00)	-0.00+	(0.00)	-0.00***	(0.00)	-0.00	(0.00)
Household Ch.												
Overcrow	-0.03*	(0.01)	0.08	(0.08)	-0.32***	(0.03)	-0.95***	(0.17)	-0.10***	(0.02)	0.09	(0.10)
Perc. Floors	-0.06***	(0.01)	-0.02	(0.02)	0.09***	(0.01)	0.07+	(0.04)	-0.02*	(0.01)	-0.02	(0.02)
Perc. Sleeping	-0.01***	(0.00)	0.02	(0.02)	-0.06***	(0.01)	0.04	(0.04)	-0.03***	(0.00)	0.01	(0.02)
Perc. Room	-0.13***	(0.00)	-0.05*	(0.02)	0.13***	(0.01)	-0.04	(0.05)	0.00	(0.01)	0.02	(0.03)
No Water	-0.02***	(0.00)	-0.03*	(0.01)	0.02**	(0.01)	0.07*	(0.03)	-0.01*	(0.00)	-0.02	(0.02)
No Elec.	-0.49***	(0.03)	-0.26***	(0.06)	-0.18*	(0.07)	0.11	(0.09)	-0.29***	(0.04)	-0.02	(0.08)
Latrine	-0.01***	(0.00)	-0.07***	(0.01)	-0.02**	(0.01)	-0.04	(0.02)	-0.02***	(0.00)	-0.01	(0.01)
No Drainage	0.06***	(0.00)	0.06***	(0.01)	-0.04***	(0.01)	-0.01	(0.02)	0.03***	(0.01)	0.00	(0.01)
No E,W,D	0.16**	(0.06)	0.42***	(0.11)	0.40**	(0.14)	-0.08	(0.16)	0.04	(0.08)	-0.02	(0.14)
Perc. No car	0.01***	(0.00)	-0.00	(0.01)	-0.01***	(0.00)	0.00	(0.02)	-0.00+	(0.00)	0.01	(0.01)
Perc. No appl.	-0.03	(0.03)	0.26**	(0.08)	0.20**	(0.07)	-0.00	(0.17)	0.14***	(0.04)	-0.06	(0.10)
Perc. No TV	0.24***	(0.01)	0.07	(0.05)	-0.08**	(0.03)	0.01	(0.09)	-0.01	(0.02)	-0.10+	(0.06)
Perc. No phone	-0.05***	(0.01)	-0.07*	(0.03)	-0.02	(0.02)	0.10+	(0.05)	-0.03***	(0.01)	-0.11**	(0.04)
Perc. No Comp.	0.02***	(0.00)	0.05***	(0.01)	0.05***	(0.00)	-0.08***	(0.02)	0.03***	(0.00)	0.04**	(0.01)
Perc. NO ICT	0.10**	(0.04)	-0.09	(0.12)	-0.12	(0.08)	-0.07	(0.24)	0.12*	(0.05)	0.27+	(0.16)
Economic Ch.												
Perc. No School Ch	0.26***	(0.02)	0.09	(0.09)	0.12*	(0.05)	0.40*	(0.19)	0.01	(0.03)	-0.00	(0.12)
Perc. Illiterate Ch.	-0.00	(0.03)	0.01	(0.08)	0.05	(0.07)	-0.28+	(0.17)	0.22***	(0.04)	-0.02	(0.12)
Perc. Illiterate Adu.	0.00	(0.01)	-0.02	(0.02)	0.10***	(0.01)	-0.01	(0.05)	-0.06***	(0.01)	0.02	(0.03)
Perc. No School a	-0.02***	(0.01)	0.02	(0.02)	-0.19***	(0.01)	0.03	(0.05)	0.02*	(0.01)	-0.01	(0.03)
Perc. Unemployed	-0.31***	(0.01)	-0.02	(0.05)	-0.33***	(0.02)	0.19*	(0.09)	-0.17***	(0.01)	-0.08	(0.07)
_cons	-2.41***	(0.06)	-2.41***	(0.64)	0.01	(0.14)	0.79	(1.14)	-7.03***	(0.09)	-6.36***	(0.86)
N	4,765,878		30,930		604,894		7,533		4,765,878		30,930	
r2_p	.477		.46		.155		.113		.453		.403	

Notes: standard errors in parenthesis + p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table A.4: Testing Beta coefficients across logit regression models. P-values displayed

	Hosp.	ICU	Deaths
Age	0.00	0.58	0.00
Sex	0.00	0.14	0.01
Urbanity	0.11	0.11	0.11
Diabetes	0.68	0.12	0.02
COPD	0.26	0.37	0.12
Asthma	0.33	0.88	0.09
Immunosuppression	0.31	0.95	0.30
Hypertension	0.69	0.56	0.89
Cardio D.	0.61	0.99	0.99
Renal D.	0.01	0.42	0.09
Pneumonia	0.00	0.06	0.62
Other Comorb.	0.51	0.81	0.38
Smoking	0.68	0.07	0.82
Obesity	0.25	0.94	0.19
Wait Test	0.47	0.13	0.47
Aff.Private	0.89	0.94	0.34
Aff. IMSS	0.81	0.58	0.14
Aff. ISSSTE	0.80	0.67	0.41
Aff. SSA	0.26	0.73	0.56
Aff. State	0.58	0.94	0.19
Aff. PEMEX	0.43	0.22	0.24
Aff. SEDENA	0.25	0.81	0.93
Aff. SEMAR	0.91	0.99	0.20
No Affiliated	0.98	0.00	0.72
Health Fac.	0.16	0.64	0.52
Overcrow	0.17	0.00	0.08
Perc. Floors	0.09	0.62	0.99
Perc. Sleeping	0.18	0.00	0.05
Perc. Room	0.01	0.00	0.65
No Water	0.47	0.07	0.66
No Elec.	0.01	0.02	0.01
Latrine	0.00	0.62	0.40
No Drainage	0.98	0.28	0.09
No E,W,D	0.10	0.04	0.76
Perc. No car	0.05	0.58	0.22
Perc. No appl.	0.00	0.29	0.08
Perc. No TV	0.00	0.37	0.13
Perc. No phone	0.52	0.05	0.06
Perc. No Comp.	0.00	0.00	0.87
Perc. NO ICT	0.20	0.86	0.39
Perc. No School Ado.	0.10	0.17	0.94
Perc. Iliterate Child.	0.89	0.07	0.07
Perc. Illiterate Adu.	0.30	0.03	0.01
Perc. No School Adu.	0.07	0.00	0.36
Perc. Unemployed	0.00	0.00	0.27
Intercept	0.99	0.99	0.99

Note: Null hypothesis $\beta_k^0 = \beta_k^1$
P-values smaller or equal than 0.05 in red

A.5 Oaxaca-Blinder decomposition approach using different models

Tables A.5 to A.8 show the results of the aggregate and detailed Oaxaca decompositions using nonlinear and linear models.

Table A.5: Aggregate Oaxaca Decomposition. Linear models

	Hospitalisations	%	Admissions to ICU	%	Deaths	%
Non Indigenous	0.127***		0.072***		0.050***	
	(0.00)		(0.00)		(0.00)	
Indigenous	0.245***		0.084***		0.097***	
	(0.00)		(0.00)		(0.00)	
Mean Difference	-0.118***		-0.012***		-0.047***	
	(0.00)		(0.00)		(0.00)	
Explained	-0.093***	79.507***	-0.005	41.235	-0.041***	88.159***
	(0.00)	(1.97)	(0.00)	(39.94)	(0.00)	(3.52)
Unexplained	-0.024***	20.493***	-0.007	58.765	-0.006**	11.841***
	(0.00)	(1.97)	(0.01)	(39.94)	(0.00)	(3.52)
Observations	4,796,808	4,796,808	612,427	612,427	4,796,808	4,796,808

Notes: Bootstrapped standard errors in parenthesis (500 replications)
Models fitted using an ANOVA-type normalisation and weights from a first-order Taylor linearisation
% share of each component to the overall gap. + p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table A.6: Detailed Oaxaca Decomposition for Hospitalisations. Linear and Nonlinear models

	Nonlinear		Linear	
E. Demographics	-0.006***	(0.00)	-0.004***	(0.00)
%	5.061***	(0.53)	3.567***	(0.36)
E. UndLyingCond	-0.054***	(0.00)	-0.068***	(0.00)
%	46.282***	(1.15)	57.442***	(1.15)
E. Risky Behav.	-0.001*	(0.00)	-0.001*	(0.00)
%	0.931*	(0.36)	0.494*	(0.21)
E. Med. Attent.	0.013***	(0.00)	0.008***	(0.00)
%	-10.683***	(0.95)	-6.469***	(0.75)
E. Health Infra.	-0.008*	(0.00)	-0.004*	(0.00)
%	6.695*	(2.69)	3.773*	(1.49)
E. Household Ch.	-0.039***	(0.01)	-0.026***	(0.00)
%	32.954***	(5.88)	21.791***	(4.11)
E. Mun. Eco. Ch.	-0.001	(0.01)	0.001	(0.00)
%	0.470	(5.25)	-1.093	(3.92)
Ue. Demographics	0.022***	(0.00)	0.016**	(0.01)
%	-19.132***	(2.13)	-13.434**	(4.43)
Ue. UndLyingCond	0.001*	(0.00)	-0.001	(0.00)
%	-0.835*	(0.38)	0.699	(0.92)
Ue. Risky Behav.	0.000	(0.00)	0.001	(0.00)
%	-0.394	(0.34)	-1.097	(0.74)
Ue. Med. Attent.	-0.015	(0.04)	0.006	(0.07)
%	12.453	(35.37)	-5.434	(62.68)
Ue. Health Infra.	0.003	(0.00)	-0.003	(0.01)
%	-2.272	(3.04)	2.506	(6.89)
Ue. Household Ch.	-0.021	(0.02)	-0.029	(0.04)
%	17.974	(13.44)	24.461	(32.53)
Ue. Mun. Eco. Ch.	-0.013***	(0.00)	-0.022**	(0.01)
%	10.655***	(2.39)	18.410**	(6.12)
Intercept	0.000	(0.05)	0.007	(0.08)
%	-0.157	(38.42)	-5.617	(71.68)
Observations	4,796,808		4,796,808	

Notes: Bootstrapped standard errors in parenthesis (500 replications)
Nonlinear models fitted using an ANOVA-type normalisation
and weights from a first-order Taylor linearisation
% share of each component to the overall gap
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table A.7: Detailed Oaxaca Decomposition for Admissions to ICU. Linear and nonlinear models

	Nonlinear		Linear	
E. Demographics	0.00	(0.00)	0.00	(0.00)
%	-1.04	(1.01)	-2.49	(1.72)
E. UndLyingCond	-0.00	(0.00)	-0.00***	(0.00)
%	14.79	(10.25)	25.63**	(8.64)
E. Risky Behav.	-0.00	(0.00)	-0.00	(0.00)
%	1.26	(1.42)	2.40	(2.27)
E. Med. Attent.	-0.01	(0.01)	-0.01***	(0.00)
%	88.00	(62.21)	109.33***	(31.81)
E. Health Infra.	-0.00	(0.00)	-0.00	(0.00)
%	2.89	(15.59)	20.10	(33.42)
E. Household Ch.	0.01	(0.01)	0.02+	(0.01)
%	-58.98	(61.20)	-125.54	(78.08)
E. Mun. Eco. Ch.	-0.00	(0.00)	-0.00	(0.01)
%	2.62	(25.11)	11.81	(57.89)
Ue. Demographics	-0.00	(0.21)	0.01	(0.01)
%	4.47	(1665.63)	-61.28	(90.68)
Ue. UndLyingCond	-0.00	(0.15)	-0.01*	(0.00)
%	22.81	(1233.80)	82.40*	(37.66)
Ue. Risky Behav.	0.00	(0.01)	0.00	(0.00)
%	-6.62	(40.98)	-12.51	(16.05)
Ue. Med. Attent.	-0.00	(0.64)	0.02	(0.10)
%	9.92	(5208.91)	-184.78	(799.80)
Ue. Health Infra.	-0.02	(0.44)	-0.05***	(0.01)
%	194.47	(3592.40)	383.02*	(151.45)
Ue. Household Ch.	0.07	(1.11)	0.20**	(0.06)
%	-582.05	(9042.57)	-1597.25**	(572.89)
Ue. Mun. Eco. Ch.	-0.03	(0.48)	-0.06***	(0.01)
%	216.09	(3898.11)	491.25***	(147.48)
Intercept	-0.02	(0.60)	-0.12	(0.12)
%	191.37	(4888.27)	957.92	(919.39)
Observations	612,427		612,427	

Notes: Bootstrapped standard errors in parenthesis (500 replications)
Nonlinear models fitted using an ANOVA-type normalisation
and weights from a first-order Taylor linearisation
% share of each component to the overall gap
+ $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table A.8: Detailed Oaxaca Decomposition for Deaths. Linear and nonlinear models

	Nonlinear		Linear	
E. Demographics	-0.014***	(0.00)	-0.008***	(0.00)
%	30.295***	(1.86)	17.464***	(0.86)
E. UndLyingCond	-0.024***	(0.00)	-0.032***	(0.00)
%	50.678***	(2.76)	68.224***	(1.84)
E. Risky Behav.	-0.001*	(0.00)	-0.000	(0.00)
%	1.909*	(0.78)	0.717	(0.53)
E. Med. Attent.	0.004**	(0.00)	0.003***	(0.00)
%	-8.552**	(2.73)	-7.099***	(1.40)
E. Health Infra.	0.001	(0.00)	0.001	(0.00)
%	-2.825	(5.11)	-3.027	(2.92)
E. Household Ch.	-0.002	(0.01)	-0.003	(0.00)
%	5.097	(12.25)	6.587	(7.53)
E. Mun. Eco. Ch.	-0.005	(0.01)	-0.002	(0.00)
%	11.032	(10.76)	5.294	(7.51)
Ue. Demographics	0.005***	(0.00)	-0.021***	(0.00)
%	-10.242***	(2.27)	45.707***	(8.77)
Ue. UndLyingCond	0.000	(0.00)	-0.000	(0.00)
%	-0.276	(0.26)	0.065	(2.39)
Ue. Risky Behav.	-0.000	(0.00)	-0.002+	(0.00)
%	0.171	(0.28)	3.716+	(1.98)
Ue. Med. Attent.	0.007	(0.05)	-0.003	(0.06)
%	-14.691	(99.80)	5.806	(126.97)
Ue. Health Infra.	-0.000	(0.00)	-0.005	(0.01)
%	0.095	(2.35)	9.824	(14.45)
Ue. Household Ch.	-0.010+	(0.01)	-0.087**	(0.03)
%	21.291+	(11.06)	185.661**	(63.66)
Ue. Mun. Eco. Ch.	-0.001	(0.00)	0.003	(0.00)
%	2.631	(2.04)	-6.736	(10.00)
Intercept	-0.006	(0.05)	0.109+	(0.07)
%	13.387	(100.04)	-232.202+	(138.69)
Observations	4,796,808		4,796,808	

Notes: Bootstrapped standard errors in parenthesis (500 replications)
Nonlinear models fitted using an ANOVA-type normalisation
and weights from a first-order Taylor linearisation
% share of each component to the overall gap
+ $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$