Polytechnic University of Yucatán

Data Engineering

3° A

Data Processing

Proffesor: Dexter Gomez

Student: William Emmanuel Fernández Castillo and Saúl Ruiz Peña

November 23th, 2025

Public Health in Mexico: COVID-19 Lethality and Comorbidity Analysis

Executive Summary

This retrospective analysis examines COVID-19 mortality patterns and associated risk factors in Mexico using integrated national health data. The study combines two datasets totaling 1,185,605 patient records to identify vulnerable populations and inform future pandemic preparedness strategies.

Key findings reveal substantial geographic and demographic disparities in case fatality rates (CFR). Guerrero state exhibited the highest mortality at 9.04%, while Ciudad de México recorded the lowest at 1.40%. Age emerged as a critical determinant, with patients over 60 years experiencing a CFR of 25.07%, nearly 26 times higher than young adults. Chronic kidney disease, COPD, and diabetes constituted the most lethal comorbidities. Among hospitalized patients, 13.93% required intubation, indicating severe respiratory complications. The analysis demonstrates that cumulative comorbidities significantly amplify mortality risk, with patients having three or more conditions facing a CFR of 23.07% compared to just 3.30% for those without underlying health conditions.

Introduction

The COVID-19 pandemic exposed critical vulnerabilities in global health systems and highlighted the urgent need to understand factors influencing disease severity and mortality. Mexico faced particular challenges due to high prevalence of metabolic disorders and uneven healthcare access across regions. This analysis addresses fundamental questions about which patient populations faced the greatest risks and how pre-existing health conditions influenced outcomes.

The Mexican government maintains detailed epidemiological surveillance systems that capture clinical and demographic information for each COVID-19 case. These datasets provide a unique opportunity to conduct large-scale retrospective analysis spanning the pandemic's progression from early 2020 through late 2025. By examining case fatality rates across states, age groups, and comorbidity profiles, this study aims to identify patterns that can guide resource allocation and intervention strategies for future public health emergencies.

Methodology

Data Sources and Integration

Two primary datasets formed the foundation of this analysis:

1. **COVID19MEXICO.csv**: Recent surveillance data (137,030 records) containing detailed clinical variables including PCR results, antigen testing, indigenous language status, and country of origin
2. **Covid Data.csv**: Historical pandemic data (1,048,575 records) from the initial outbreak period with standardized medical classifications

Both datasets originated from Mexican government health surveillance systems tracking confirmed and suspected COVID-19 cases across all 32 federal entities.

Data Preparation

The integration process required careful harmonization of column names, coding schemes, and data formats:

Column Standardization: English column names in the historical dataset underwent translation to Spanish to match the recent data structure. For example, "SEX" became "SEXO", "PATIENT_TYPE" became "TIPO_PACIENTE", and "INTUBED" became "INTUBADO".

Binary Variable Cleaning: Medical coding systems use specific values to indicate unknown or not applicable responses. Code 98 ("not applicable") appeared across multiple comorbidity fields and required conversion to null values to avoid misclassification in statistical analyses. This affected 14 binary variables including diabetes, hypertension, asthma, COPD, immunosuppression, cardiovascular disease, obesity, chronic kidney disease, smoking status, intubation status, pneumonia, ICU admission, and pregnancy.

Date Formatting: Three temporal variables (symptom onset date, hospital admission date, and death date) required conversion to datetime format. The death date field used "9999-99-99" to indicate living patients, which underwent transformation to null values.

Derived Variables:

- **FALLECIO** (deceased): Binary indicator created from death date presence (1 = deceased, 0 = alive)
- **GRUPO_EDAD** (age group): Categorical variable segmenting patients into four brackets: 0-17, 18-35, 36-59, and 60+ years
- **NUM_COMORBILIDADES**: Count of present comorbidities per patient based on ten tracked conditions
- **GRUPO_COMORBILIDADES**: Categorical grouping (0, 1, 2, or 3+ comorbidities)

The final unified dataset contained 1,185,605 records with 44 variables capturing demographic characteristics, clinical presentations, comorbidities, laboratory results, and outcomes.

Analytical Approach

Case Fatality Rate Calculation: CFR represents the proportion of confirmed cases resulting in death, calculated as (deaths / confirmed cases) × 100. This metric provides standardized comparison across different populations and time periods.

Geographic Analysis: State-level CFR calculations enabled identification of regional disparities in mortality outcomes. Each of Mexico's 32 federal entities received individual assessment.

Age Stratification: Four age brackets facilitated examination of age-related mortality patterns, with particular attention to elderly populations.

Comorbidity Assessment: Ten specific conditions underwent individual evaluation for associated mortality rates. Additionally, the cumulative burden of multiple conditions received analysis through the grouped comorbidity variable.

Temporal Analysis: Daily counts of confirmed cases and deaths from symptom onset dates created time series revealing pandemic progression patterns.

Clinical Severity Indicators: Among hospitalized patients (TIPO_PACIENTE = 2), the proportion requiring mechanical ventilation served as a proxy for severe respiratory complications.

Results

Geographic Distribution of Mortality

Case fatality rates varied dramatically across Mexican states, revealing distinct geographic patterns (Table 1).

Table 1: Case Fatality Rates by State

State	CFR (%)
Guerrero	9.04
Baja California	8.69
Quintana Roo	8.46
Tamaulipas	8.01
Michoacán	7.63
Morelos	6.54

State	CFR (%)
Nuevo León	6.42
Colima	6.34
San Luis Potosí	5.77
Sinaloa	5.59
...	...
Durango	1.41
Ciudad de México	1.40

The Pacific coast state of Guerrero recorded the highest mortality at 9.04%, followed by the northern border state of Baja California at 8.69%. Tourist destinations including Quintana Roo (8.46%) also showed elevated rates. In contrast, Ciudad de México, despite being the epicenter of initial outbreaks, exhibited the lowest CFR at 1.40%, potentially reflecting concentrated healthcare resources and better access to intensive care facilities.

A clear divide emerged between southern states (generally higher mortality) and central Mexico (generally lower mortality), though notable exceptions existed. This pattern suggests influence from multiple factors including healthcare infrastructure density, socioeconomic conditions, population age structures, and timing of outbreak waves.

Age-Stratified Mortality

Age demonstrated the strongest association with COVID-19 mortality (Figure 1).

Figure 1: Case Fatality Rate by Age Group

Age Group	CFR (%)
60+	25.07
36-59	5.82
18-35	0.97
0-17	1.33

Elderly patients (60+ years) experienced a CFR of 25.07%, meaning one in four infected individuals in this age bracket died. Middle-aged adults (36-59 years) faced substantially lower but still significant mortality at 5.82%. Young adults (18-35 years) showed remarkable resilience with a CFR of just 0.97%, while children and adolescents (0-17 years) had a slightly higher rate at 1.33%, possibly reflecting different reporting patterns or specific pediatric risk factors.

The 26-fold difference in mortality between elderly and young adult populations underscores the age-dependent nature of COVID-19 severity. This gradient remained consistent even after accounting for higher comorbidity prevalence in older age groups.

Comorbidity-Associated Mortality

Pre-existing health conditions significantly elevated mortality risk. The five most lethal comorbidities showed striking patterns (Table 2).

Table 2: Top 5 Comorbidities by Associated Case Fatality Rate

Comorbidity	CFR (%)
Chronic Kidney Disease	26.58
COPD	23.18
Diabetes	20.99
Cardiovascular Disease	19.70
Hypertension	18.46

Chronic kidney disease emerged as the deadliest comorbidity with a CFR of 26.58%, indicating that more than one in four COVID-19 patients with kidney disease died. COPD (23.18%) and diabetes (20.99%) followed closely. These three conditions share metabolic and inflammatory characteristics that likely exacerbate COVID-19 pathophysiology.

Notably, obesity (not shown in top 5) appeared less directly lethal than other metabolic conditions, though it frequently co-occurred with diabetes and hypertension. Asthma, despite being a respiratory condition, showed lower associated mortality than COPD, potentially reflecting younger average age of asthma patients or differences in disease mechanisms.

Cumulative Comorbidity Burden

The number of concurrent comorbidities dramatically amplified mortality risk (Figure 2).

Figure 2: Case Fatality Rate by Number of Comorbidities

Comorbidity Group	CFR (%)
0 comorbidities	3.30
1 comorbidity	8.51
2 comorbidities	15.90
3+ comorbidities	23.07

Patients without any tracked comorbidities maintained a relatively low CFR of 3.30%. The presence of a single condition more than doubled this rate to 8.51%. Each additional comorbidity compounded risk: two conditions elevated CFR to 15.90%, while three or more conditions resulted in 23.07% mortality.

This seven-fold increase from zero to multiple comorbidities reveals a dose-response relationship between disease burden and COVID-19 outcomes. The pattern suggests synergistic rather than merely additive effects, where multiple conditions interact to create disproportionate vulnerability.

Hospitalization and Respiratory Support

Among the 262,695 hospitalized patients, 36,592 required intubation, representing 13.93% of all hospital admissions. This substantial proportion indicates that roughly one in seven hospitalized COVID-19 patients developed respiratory failure severe enough to necessitate mechanical ventilation.

The intubation rate serves as a proxy for critical illness and healthcare system strain. These patients required intensive care unit beds, specialized respiratory therapy equipment, and extended hospital stays, placing immense pressure on medical infrastructure during surge periods.

Temporal Patterns

Time series analysis of confirmed cases and deaths revealed multiple distinct pandemic waves (data available in exported CSV files). The temporal distribution showed:

- Initial outbreak period (early 2020) with limited testing capacity
- Major surge periods corresponding to variant emergence
- Gradual decline in case fatality rate over time as treatment protocols improved
- Periodic fluctuations reflecting seasonal patterns and public health interventions

Daily CFR calculations demonstrated high volatility during low-case periods but stabilized during major waves, reaching mean values consistent with the overall analysis.

Discussion

Principal Findings

This analysis of 1.2 million COVID-19 cases in Mexico reveals three critical vulnerability factors: advanced age, chronic kidney disease and metabolic disorders, and cumulative comorbidity burden. The 25% mortality rate among elderly patients and the seven-fold increase in CFR from zero to multiple comorbidities provide quantitative targets for intervention strategies.

Geographic disparities in mortality rates point to systemic inequalities in healthcare access and quality. States with the highest CFRs generally possess lower healthcare infrastructure density and higher poverty rates. The contrast between Guerrero (9.04% CFR) and Ciudad de México (1.40% CFR) illustrates how healthcare capacity directly influences survival.

Clinical Implications

The dominant role of chronic kidney disease in COVID-19 mortality deserves particular attention. Patients with impaired renal function face multiple vulnerabilities: altered immune response, fluid and electrolyte imbalances, difficulty managing inflammatory responses, and complications from dialysis therapy during acute illness. Healthcare systems should prioritize early identification and aggressive management of COVID-19 in this population.

The 13.93% intubation rate among hospitalized patients establishes a planning benchmark for future respiratory pandemics. Healthcare systems must maintain surge capacity to handle scenarios where approximately 15% of admissions require mechanical ventilation, including adequate ICU beds, ventilators, and trained respiratory therapy staff.

The strong age gradient in mortality aligns with global patterns observed throughout the pandemic. However, the specific CFR values for each age bracket provide Mexico-specific data for risk communication and vaccination prioritization strategies. The relatively low mortality among young adults (0.97%) contrasts sharply with elderly outcomes (25.07%), supporting age-stratified public health approaches.

Public Health Policy Recommendations

Targeted Screening Programs: Establishing routine COVID-19 monitoring for patients with chronic kidney disease, diabetes, and COPD would enable early intervention before severe symptoms develop. Community health workers could conduct regular outreach to high-risk patients.

Regional Resource Allocation: States with historically higher CFRs require infrastructure investment including expanded ICU capacity, medical equipment, and specialist training. Guerrero, Baja California, and Quintana Roo should receive priority in health budget distributions.

Elderly Protection Protocols: Age-specific interventions including priority vaccination, assisted isolation during outbreaks, and enhanced healthcare access for patients over 60 would directly address the population segment facing 25% mortality rates.

Comorbidity Management: Strengthening chronic disease management programs reduces baseline health vulnerability. Better control of diabetes, hypertension, and kidney disease through medication access and lifestyle interventions would improve outcomes in future outbreaks.

Healthcare Workforce Development: Training primary care physicians to recognize early COVID-19 complications in high-risk patients could prevent progression to critical illness requiring intubation. Telemedicine expansion would extend specialist consultation to underserved regions.

Limitations

Several constraints affect interpretation of these findings:

Selection Bias: Patients seeking medical care and receiving testing represent a subset of all infections. Asymptomatic or mildly symptomatic individuals often went undetected, potentially inflating calculated CFRs. True infection fatality rates likely sit below reported case fatality rates.

Temporal Coverage: The datasets span different pandemic periods with varying testing availability, treatment protocols, and viral variants. Earlier cases occurred before effective treatments emerged, while later cases benefited from accumulated medical knowledge. This temporal heterogeneity complicates direct comparison across the full timeline.

Comorbidity Recording: The binary coding of comorbidities (present/absent) lacks nuance regarding disease severity. A patient with well-controlled diabetes differs substantially from one with poorly managed diabetes, yet both receive identical classification. Disease duration and treatment adherence remained unrecorded.

Coding Inconsistencies: Despite harmonization efforts, subtle differences in how medical staff recorded information across healthcare facilities and time periods may introduce measurement error. The large sample size likely minimizes this effect but cannot eliminate it entirely.

Socioeconomic Variables: The datasets lack direct socioeconomic indicators such as income, education, occupation, or health insurance status. These factors likely contribute to observed geographic disparities but cannot be directly quantified from available data.

Healthcare Access Measurement: While state-level CFR differences suggest healthcare disparities, the analysis cannot separate the effects of healthcare quality from population health characteristics or timing of outbreak waves in different regions.

Conclusion

This retrospective analysis of 1.2 million COVID-19 cases quantifies the profound impact of age, chronic disease, and healthcare access on pandemic mortality in Mexico. Elderly individuals, patients with chronic kidney disease and metabolic disorders, and those with multiple comorbidities faced dramatically elevated death rates. Geographic disparities revealed that state-level mortality varied more than six-fold, from 1.40% in Ciudad de México to 9.04% in Guerrero.

The findings establish clear priorities for future pandemic preparedness: protecting elderly populations, managing chronic diseases aggressively, equalizing healthcare access across regions, and maintaining surge capacity for respiratory support. The seven-fold increase in mortality from zero to multiple comorbidities underscores the importance of preventing and managing chronic conditions as a public health security measure.

Beyond COVID-19, these results illuminate fundamental vulnerabilities in Mexico's health system and population health profile. The high prevalence and lethality of metabolic disorders signals a looming chronic disease crisis requiring immediate attention. The geographic mortality gradient reflects deeper structural inequalities demanding long-term investment in underserved regions.

Public health authorities should use this evidence base to design targeted interventions addressing identified risk factors. Protecting the most vulnerable populations requires both immediate measures (vaccination prioritization, enhanced screening) and sustained efforts (healthcare infrastructure development, chronic disease prevention programs). The data-driven approach demonstrated here provides a model for evidence-based policy formation applicable to current and future health emergencies.