

**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**

J. Mancilla-Galindo¹

J. M. Vera-Zertuche²

A. R. Navarro-Cruz³

O. Segura-Badilla⁴

G. Reyes-Velázquez⁵

F. J. Tepepa-López^{6,7}

P. Aguilar-Alonso³

J. de J. Vidal-Mayo⁸

A. Kammar-García^{7,8}

¹.- Unidad de Investigación UNAM-INC, Instituto Nacional de Cardiología Ignacio Chávez, México.

².- Departamento de Endocrinología, Clínica de Obesidad. Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, México.

³.- Facultad de Ciencias Químicas, Departamento de Bioquímica y Alimentos, Benemérita Universidad Autónoma de Puebla, México.

⁴.- Facultad de Ciencias de la Salud y de los Alimentos, Departamento de Nutrición y Salud Pública, Programa UBB Saludable, Universidad del Bío-Bío, Chile.

⁵.- Departamento de Cardiología. Centro Médico Nacional La Raza. Instituto Mexicano del Seguro Social, México.

⁶. Servicio de Endoscopia, Hospital Juárez Centro. México.

⁷.- Sección de Estudios de Posgrado e Investigación, Escuela Superior de Medicina, Instituto Politécnico Nacional, México.

⁸.- Departamento de Atención Institucional Continua y Urgencias. Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán. México.

*Corresponding author: A. Kammar-García. Vasco de Quiroga 15, Tlalpan, Col. Belisario Domínguez Sección XVI, CP 14080 Ciudad de México. Tel: +52 5554870900, ext. 5010, e-mail: kammar_nutrition@hotmail.com

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Running Title: The Patient History COVID-19 score

1 **Summary**

2

3 We sought to develop and validate a multivariable prediction model of death in Mexican
4 patients with COVID-19, by using demographic and patient history predictors. We
5 conducted a national retrospective cohort in two different sets of patients from the
6 Mexican COVID-19 Epidemiologic Surveillance Study. To develop the model, we
7 included 264,026 patients tested for SARS-CoV-2 between February 28 and May 30,
8 2020. To validate the model, 592,160 patients studied between June 1 and July 23,
9 2020 were included. Patients with a positive RT-PCR for SARS-CoV-2 and complete
10 unduplicated data were eligible. Demographic and patient history variables were
11 analyzed through Multivariable Cox regression models to evaluate predictors to be
12 included in the prognostic scoring system called PH-Covid19. 83,779 patients were
13 included to develop the model; 100,000, to validate the model. Eight predictors (age,
14 sex, diabetes, COPD, immunosuppression, hypertension, obesity, and CKD) were
15 included in the PH-Covid19 scoring system (range of values: -2 to 25 points). The
16 predictive model has a discrimination of death of 0.8 (95%CI:0.796-0.804). The PH-
17 Covid19 scoring system was developed and validated in Mexican patients to aid
18 clinicians to stratify patients with COVID-19 at risk of fatal outcomes, allowing for better
19 and efficient use of resources.

20

21 **Key words:** COVID-19; scoring system; Mexico; SARS-CoV-2; prediction model

22

23

24

25

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Introduction

The coronavirus disease (COVID-19) pandemic is the global phenomenon which is shaping modern societies in the year 2020, a reason why the severe acute respiratory coronavirus 2 (SARS-CoV-2) has been named the once-in-a-century pathogen that scientists and global leaders had been worrying for [1]. Demographic and patient history risk factors for fatal outcomes in patients with COVID-19 have been characterized in large national cohorts [2-5], and broadly include old age, sex (men), comorbidities, deprivation (a correlate of poverty), and belonging to certain ethnic groups. Other clinical, radiological, and laboratory parameters at presentation have also been studied as risk factors for disease progression and death [5,6].

Low and middle-income countries often suffer from inadequate healthcare due to the lack of equipment, poor organization, and scarce qualified healthcare professionals. Thus, what works in high-income countries may not work in low-income countries [7]. Mexico has been one of the most affected countries by COVID-19, and disparate differences in patient outcomes have been noted be related to inequalities in healthcare access. Therefore, there is a pressing need to develop accessible and simple tools to aid clinicians providing medical attention in the most unfavored regions of Mexico.

Several diagnostic and prognostic models have been developed to be used in patients with COVID-19 [8]. However, most models include laboratory and radiographic variables which would be nearly impossible to collect in low-resource settings. Furthermore, these models have seldom been validated, are often inadequately reported, or are overfitted

1 due to a large predictor to outcome ratio [8-10], reasons that may limit their usefulness
2 in real-world settings. Developing and validating models that only require demographic
3 and patient history data, by using large national or multinational cohorts, may be a way
4 to overcome these shortcomings to provide useful tools to clinicians in low-resource
5 regions.

6 Therefore, we sought to develop and validate a multivariable prediction model of death
7 in Mexican patients with COVID-19, by using demographic and patient history
8 predictors.

9

10 **Method**

11

12 **Study Design**

13 We conducted a national retrospective cohort in two different sets of patients from the
14 Mexican COVID-19 Epidemiologic Surveillance Study [11] to develop and validate a
15 multivariable prediction model of death in Mexican patients with COVID-19. Patient
16 history variables were used as predictors of death as the outcome of interest. Blind
17 assessment was not required since these are objective variables unlikely subjected to
18 bias.

19 To develop the model, we included 264,026 patients studied between 28 February and
20 30 May, 2020. All patients with a positive RT-PCR for SARS-CoV-2 were included to
21 maximize the power and generalizability of results. Patients with incomplete data were
22 excluded, whereas patients with the same demographic, clinical, and follow-up variables
23 were considered duplicated and only one entry was kept.

1 To validate the model, we included 592,160 patients studied between 1 June and 23
2 July, 2020. Only patients with a positive RT-PCR for SARS-CoV-2 and complete
3 unduplicated data were included to validate the model. We further performed simple
4 random sampling of positive cases to increase statistical power in approximately 15%
5 with respect to the sample used for developing the model.

6

7 **Source of Data**

8 Data is collected and regularly updated by the Mexican Secretariat of Health and is
9 available in the Open Data platform of the Federal Government of Mexico [12]. A
10 historical repository of individual datasets starting on 12 April 2020 is available through
11 the General Directorate of Epidemiology [13]. Patients who met criteria of suspected
12 COVID-19 case and were subsequently tested for SARS-CoV-2 were included in the
13 study, starting on late February 2020 when the first suspected cases arrived in Mexico.
14 Two diagnostic strategies are outlined in the National COVID-19 Epidemiologic
15 Surveillance Plan [11]: 1. testing of 10% of ambulatory patients with mild symptoms of
16 respiratory disease and 100% of patients with respiratory distress at evaluation in one of
17 the 475 monitoring units of viral respiratory disease (USMER, for its acronym in
18 Spanish) which are strategically distributed to be representative of the Mexican
19 population, and 2. testing 100% of patients who meet diagnostic criteria of Severe Acute
20 Respiratory Infection (defined as shortness of breath, temperature ≥ 38 °C, cough, and
21 ≥ 1 of the following: chest pain, tachypnea, or acute respiratory distress syndrome) who
22 seek medical attention in non-USMER units.

1 Healthcare professionals collecting a diagnostic specimen are required to fill out a
2 format containing demographic and patient history variables. Follow-up of all suspected
3 COVID-19 cases is registered by accredited hospital epidemiologists (inpatients) and
4 the responsible healthcare professional of every Local Health Jurisdiction (ambulatory
5 patients), who ultimately upload data into the Respiratory Diseases Epidemiologic
6 Surveillance System. Results of diagnostic RT-PCR for SARS-CoV-2 are directly
7 uploaded by the diagnostic facility; accreditation of diagnostic procedures by the
8 Mexican Institute of Diagnostics and Epidemiological Reference is required to upload
9 results. Reporting of deaths is obligatory and must be done in less than 48 hours after
10 occurrence. One caveat to this reporting method is that patients who are tested more
11 than once in different jurisdictions may be duplicated. No variables that could lead to
12 identification of patients are provided in datasets. Thus, the only way to eliminate
13 duplications is through matching of cases with equal demographic and clinical variables.
14 Specific information of treatments is not released.

15 Variables provided in the datasets are: origin (USMER, non-USMER), healthcare
16 provider, state, birthplace, place of residency, nationality, indigenous language speaker,
17 migratory status, type of medical attention (hospitalization/ambulatory), admission date,
18 symptom onset date, invasive mechanical ventilation (intubation [yes/no]), admission to
19 ICU (yes/no), pneumonia (yes/no), date of death, contact with confirmed COVID-19
20 cases (yes/no), SARS-CoV-2 RT-PCR result (positive, negative, or pending), age, sex,
21 current pregnancy, and the following comorbidities (yes/no): diabetes, hypertension,
22 obesity, cardiovascular disease (CVD), chronic kidney disease (CKD),

1 immunosuppression, asthma, chronic obstructive pulmonary disease (COPD), and
2 smoking.

3

4 **Statistical Analysis**

5 Descriptive data were calculated and are provided as frequencies, percentages, or
6 mean with standard deviation (SD). Characteristics of patients in the model
7 development and validation cohorts were compared through Student's t-test or χ^2 . A
8 Cox regression model was applied to predict the risk of death. The risk of death was
9 assessed through univariate analysis of the following variables: age, sex, current
10 pregnancy, diabetes, COPD, asthma, immunosuppression, hypertension, CVD, obesity,
11 CKD, smoking, and time from symptom onset to medical attention. Risk factors
12 associated with death which were statistically significant ($p < 0.05$) were included to
13 develop the multivariate regression model. All variables with a level of significance
14 $p < 0.1$ were considered in the Cox regression model by using the Enter method.
15 Variables that kept a level of significance $p < 0.05$ were used in the final model, which
16 was evaluated through Harrell's C-statistic to determine its discrimination of death.

17 A scoring system was developed in accordance to the model proposed by Sullivan LM,
18 et al. [14] Each risk factor was organized into categories and the reference value was
19 determined as follows. Age was entered into the Cox regression model as a continuous
20 variable and further categorized into 10 sets of years (<20, 20-29, 30-39, 40-49, 50-59,
21 60-69, 70-79, 80-89, 90-99, and >90); the midpoint between the nine values of each
22 category was set as the reference value (W_{ij}). The reference for age as a risk factor
23 (W_{iREF}) was set in the 20-29 years category since this group had the lowest mortality in

1 a previous study performed in Mexican patients with COVID-19 [3]. The rest of the risk
2 factors were modeled as dummy variables, setting the absence of comorbidities and
3 being woman (sex) as the reference values. Each W_{iREF} was subtracted from W_{ij} and
4 multiplied by the regression coefficient (β_i) of the risk factor to determine units of
5 regression of distancing for every risk factor in the reference category $\beta_i(W_{ij} - W_{iREF})$. A
6 constant B, defined as the constant increase in units of risk for each 5-year increase,
7 was obtained by multiplying $\beta_i \cdot 5$. Values for B were 0.25 in this study. Scores for each
8 category were obtained with the equation $\beta_i(W_{ij} - W_{iREF})/B$. For every point in the risk
9 score, the estimated risk of death (p) was calculated with the Cox proportional hazards
10 regression analysis:

$$p = 1 - S_0(t)^{\exp(\sum_{i=1}^p \beta_i X_i - \sum_{i=1}^p \beta_i i)}$$

11 Where: $S_0(t)$ is the average survival according to mean values of every risk factor;
12 $\sum_{i=1}^p \beta_i X_i$ is substituted with each value of β_i of the risk score, times the B constant, plus the
13 reference age value according to the β_i of age; and $\sum_{i=1}^p \beta_i i$ is the sum of every β_i times
14 the proportion or mean value of every risk factor.

15 To validate the model, we calculated the risk score for every patient and applied the
16 Cox proportional hazards regression analysis. Estimated risks were obtained from both
17 the scoring system and the observed risk in the regression analysis. The values
18 obtained for patients in the validation cohort were distributed in percentiles (1 to 99) to
19 determine the punctuation categories. The estimated and observed risks were
20 compared in each category of punctuation (percentiles 25, 50, 75, and 99).

1 The Kaplan-Meier analysis was performed to determine survival in each category of the
2 scoring system, and a Cox regression analysis, to determine increases in the lethality
3 risk for each category.

4 The association between the risk score and the probability of other adverse events
5 (hospitalization, invasive mechanical ventilation, pneumonia, and admission to ICU)
6 were also studied. The frequencies of each adverse event for every category in the
7 scoring system were quantified and a binomial logistic regression analysis was
8 performed to determine the risk of each adverse event according to the risk score;
9 logarithms of the odds ratios (OR) were graphed to establish the scoring value at which
10 risk for every adverse event is increased.

11 All statistical analyses were performed using the SPSS software v.21 and the R
12 statistical software v.3.6.2; figures were created in GraphPad Prism v.6. A value of
13 $p < 0.05$ was used to establish statistical significance.

14

15 **Results**

16

17 Out of 264,026 patients in the model development cohort, 84,627 had a positive RT-
18 PCR for SARS-CoV-2, 140,553 were negative, and 38,846 had pending results. After
19 exclusion of patients with incomplete data and duplicated registries, 83,779 patients
20 with a positive test were included to develop the model. Among the 592,160 patients in
21 the model validation cohort, 256,488 patients had a positive result, 253,447 were
22 negative, and 82,225 had unreported results. After excluding duplicated and incomplete
23 registries, and random sampling of positive cases, 100,000 patients were included to

1 validate the model. Descriptive values of demographic characteristics, patient history,
2 and outcomes of patients (survivors and non survivors) in both cohorts are provided in
3 Table 1.

4 Variables included in the Cox regression model are presented in Table 2. Eight risk
5 factors were included in the model; the only quantitative variable was age. Age,
6 diabetes, and CKD were associated with the greatest increases in death. The predictive
7 model has a discrimination of 0.8 (95%CI:0.796-0.804) and an average survival of 0.903
8 with the mean values for every risk factor. The Patient History COVID-19 (PH-Covid19)
9 scoring system assigns a score to every risk factor ultimately included in the predictive
10 model (Table 3); the sum of scores for all risk factors included ranges from -2 to 25
11 points. Predicted probabilities of death in patients with a positive test for SARS-CoV-2
12 for every possible total value in the scoring system range from 0.74% to 99.82% (Table
13 4).

14 Baseline characteristics and outcomes of patients in the validation cohort were
15 statistically significantly different from those in the model development cohort, except for
16 time from symptom onset to medical attention (Table 1). Results of the Cox regression
17 model applied to the validation cohort are provided in Supplementary Table S1.
18 Calculated scores in the validation cohort reflected the following distribution according
19 to percentiles: -2 to 2 points, percentile 1 to 25; 3 to 5 points, percentile 25-50; 6 to 8
20 points, percentile 51 to 75; and 9 to 15 points, percentile 76 to 99. Patients >99
21 percentile were considered as extreme values. Estimated risks and the observed risks
22 of death obtained from the Cox proportional hazards regression analysis, in relation to
23 the scores in each percentile, are presented in Fig. 1. The estimated and observed

1 risks were similar for each group and were strongly correlated ($r=0.98$, $R^2=0.96$,
2 $p<0.0001$).

3 In the survival comparison between categories generated according to percentiles (Fig.
4 2), survival was decreased with increasing risk scores (-2 to 2 points, 99.6%; 3 to 5
5 points, 98.6%; 6 to 8 points; 95.7%, and 9 to 15 points, 84.3%, $p<0.0001$ for all
6 categories). A subsequent Cox regression analysis was performed to determine the risk
7 of death for each individual category compared to the category comprising the lower
8 percentile (-2 to 2 points): 2 to 5 points, HR:3.54 (95%CI:2.85-4.39); 6 to 8 points,
9 HR:10.67 (95%CI:8.78-12.98); 9 to 15 points, HR:41.9 (95%CI:34.7-50.7); and >15
10 points, HR:87.6, (95%CI:70.1-109.5). In accordance with this, risk groups were derived
11 from the scoring system: Low (-2 to 2 points), Medium-Low (3 to 5 points), Medium (6 to
12 8 points), Medium-High (9 to 15 points), and High (>15 points).

13 The prevalence for every adverse event in each category of the scoring system are
14 provided in Fig. 3; results of the logistic regression analysis to determine the risk of
15 adverse events for each risk group in the PH-Covid19 scoring system are given in
16 Supplementary Table S2. Tendencies of risk increasement for each adverse event with
17 augmenting scores (Supplementary Figure S1) reflect that risk for any adverse event
18 starts at a value of 5 points.

19

20 **Discussion**

21

22 We have developed and validated the PH-Covid19 score, a multivariable prediction
23 model of death in Mexican patients with COVID-19, by using different datasets from the

1 Mexican COVID-19 Epidemiologic Surveillance Study. This scoring system has been
2 created to aid clinicians working in resource-strained conditions to early stratify patients
3 with COVID-19 according to their risk of fatal outcomes, without the need to perform
4 laboratory or imaging studies.

5 Similar to other studies, we found that older age is the main risk factor for dying from
6 COVID-19, with every 10-year increase associated with the largest increases in hazard
7 ratio [2,5,15]; in the PH-Covid19 score, 10-year increases added two points, starting
8 from 30 years, and being <20 years subtracted 2 points. Sex (men), which increases
9 two points in our score, was correlated with death in our study and others [2,3,5,6,15].

10 Diabetes and CKD resulted in two-point increases in the score. Other studies had
11 similar findings to ours [3,4,6,15]; one study found an uncertain increased risk
12 attributable to diabetes (HR:1.14, 95%CI:0.91-1.43) in critically-ill patients [5]; in another
13 study, uncontrolled diabetes further increased the risk (HR:1.95, 95%CI:1.83–2.08) with
14 respect to controlled diabetes (HR:1.31, 95%CI:1.24–1.37), while CKD stages 4-5
15 caused a greater increase (HR:2.52, 95%CI:2.33–2.72) than CKD stages 3a-3b
16 (HR:2.52, 95%CI:2.33–2.72) [2]. Obesity had the third strongest association with death,
17 resulting in a 1-point increase in our scoring system; other studies have had similar
18 findings [3-15]. Adjusted risk of death by obesity occurred with gradual increases in
19 higher obesity classes ² and no evident or clear risk was identified in another study for
20 class I and II obesity, whereas class III obesity was significantly correlated with death
21 [5]. This last study, however, included only patients admitted to ICUs who had a high
22 prevalence of comorbidities, which are independent risk factors for ICU admission [3].

1 Hypertension was a risk factor for death similar to other studies [3,4], and resulted in a
2 1-point increase in the scoring system. In one large study, adjusted risk of death for
3 hypertension adjusted by all covariates apparently reduced the risk (HR:0.89,
4 95%CI:0.85–0.93), likely reflecting an artificial reduction of risk driven by diabetes and
5 obesity since age and sex adjusted risk increased the risk (HR:1.09, 95%CI:1.05–1.14)
6 [2]. COPD also increased 1 point in our model, being also associated to death in other
7 studies [2-4,15].

8 Results from our validation did not provide different patient history predictors which
9 could enhance the performance of our model. Thus, model updating was not required.

10 Some of our findings raise concerns regarding quality and access to healthcare in
11 Mexico. 10.3-13.5% of patients who died in our cohorts did not receive in-hospital care
12 at any moment of the disease, only 15.5- 18.3% of patients were admitted to an ICU
13 before dying, and only 21.3-28.1% of patients who died were intubated; conversely,
14 intubation in survivors was unusually low (0.8-1.3%). In other studies of hospitalized-
15 only patients, 53-72% of non-survivors were admitted to an ICU and 51-59% received
16 invasive mechanical ventilation [6,16,17]. Observed mortality in COVID-19 patients
17 under 60 years is lower when access to healthcare is not a limitation [18]; non-survivors
18 in our cohorts were younger (mean age 60-61.4 years) than those in other studies (67-
19 80 years) [5,6,15,19].

20 Patients diagnosed with pneumonia in both cohorts were 12.8-21% among survivors
21 and 77.4-87.5% of non-survivors. These numbers are low compared to the prevalence
22 of chest CT-scan abnormalities which occur in 67.3-70.8% of asymptomatic/pre-
23 symptomatic patients [20,21], 95.5% of patients with mild COVID-19 [22], and 98% of all

1 COVID-19 patients included in a meta-analysis [23]. Chest X-ray, on the other hand,
2 may be normal in up to 63% of patients with early COVID-19 pneumonia [24].
3 Nonetheless, the low proportion of pneumonia in non-survivors suggests non-optimal
4 diagnosis of pneumonia may be occurring in Mexico. The lack of an operational
5 definition may have contributed since clinicians could have defined pneumonia
6 differently based on clinical and/or radiographical findings. Other possibilities should be
7 explored, including knowledge of Mexican clinicians on how to diagnose pneumonia and
8 access to radiological studies during the pandemic in low-resource settings.

9 Three prognostic COVID-19 models have been developed in Mexican patients. The
10 LOW-HARM model is a 100-point scoring system calculated by inputting patient history
11 and laboratory values, setting 65 points as the cut-off value of greater discrimination of
12 risk of death [25]. Prediction of death (0.81, 95%CI:0.77-0.84) is similar to the PH-
13 Covid19 scoring system (0.80, CI95%:0.796-0.804), which advantageously only requires
14 patient history predictors. Another scoring system uses age (cut-off 65 years),
15 comorbidities, and pneumonia to predict death [26]. This model was accurate at
16 predicting death and other adverse events but has the limitation of not accounting for
17 the large increases in risk for every 10-year category or similar. Also, the model by
18 Bello-Chavoya, et al. was developed using one dataset of the Mexican Epidemiologic
19 Surveillance Study, which unfortunately had no operational definition for pneumonia as
20 discussed earlier. A third model was developed and validated to predict the risk of
21 admission to ICU; the ABC-GOALS was developed in three versions: clinical, clinical +
22 laboratory, and clinical + laboratory + imaging predictors [27].

1 In one systematic review of existing prediction models for diagnosis and prognosis of
2 COVID-19, the use of any of the reviewed models was discouraged since, out of 91
3 diagnostic and 50 prognostic models, all were at high risk of bias due to methodological
4 constraints and poor reporting [8]. Predictive models of death (eight) often excluded
5 patients that had not developed the outcome of interest, did not account for censorship,
6 inadequately reported discrimination and calibration of the model, and had a high risk of
7 bias according to PROBAST evaluation, despite authors claiming good global
8 performances of their models. We have addressed these concerns in our development
9 and validation of the PH-Covid19 scoring system.

10 Another strength of our study is that we were able to perform a type 3 analysis
11 according to TRIPOD by using individual datasets to develop and validate our model;
12 this design allows for external validation of the performance of a model [28]. It is worth
13 highlighting that sample sizes in both cohorts include thousands of patients, which adds
14 robustness to our model.

15 One limitation of our study is that certain diseases (cancer, hematological malignancies,
16 and neurologic diseases) and specific states of a disease (obesity class, former or
17 current smoker, and control of diabetes, hypertension, and asthma) which increase the
18 risk of dying from COVID-19 [2] were not studied since they are not provided in the
19 datasets. However, our model accounts for the main risk factors associated with death
20 in patients with COVID-19, and not requiring inputting specific disease states makes it
21 easier to be used by clinicians while minimizing the risk of not having enough data to
22 use the score with precision.

1 Another limitation is that the epidemiologic surveillance strategy in Mexico allows testing
2 of only 10% of ambulatory patients. Furthermore, the operational definition of suspected
3 COVID-19 case used in Mexico until August 24, 2020 had a low sensitivity (58.2%), but
4 a high specificity (63.7%) compared to that used by the CDC (85.8% and 25.8%,
5 respectively) [29]. Altogether, this means that our cohorts may include very few patients
6 with asymptomatic COVID-19 and less patients with mild COVID-19 compared to other
7 national datasets with higher testing rates. Since it was not possible to determine the
8 exact number of patients with mild disease in our cohorts, we can only indirectly
9 suggest that mild-disease patients could comprise around 71.5 and 84.4% of patients
10 according to the fraction of non-hospitalized patients who survived, a number high
11 enough to permit the use of this score in patients with mild-to-severe COVID-19.

12 The PH-Covid19 score was created to be used in limited-resource settings where
13 access to laboratory and imaging studies may be restricted. In places where this is not a
14 limitation and for patients who are likely to have already been admitted to hospital
15 (critical patients), other prognostic models may have a better performance than ours.
16 However, clinicians should consider that most models have not been validated before
17 deciding to use any COVID-19 diagnostic or prognostic model.

18 The PH-Covid19 score uses patient history predictors which are frequently known at the
19 first contact with a patient, or can be interrogated rapidly, to predict death in patients
20 with COVID-19. This score was developed and validated in Mexican patients to be used
21 in low-resource settings where obtaining laboratory and radiographic studies may not be
22 immediately possible. This score will aid clinicians to stratify patients with COVID-19 at
23 risk of fatal outcomes to use healthcare resources more efficiently.

1

2 **Ethical disclosures:** This article is a retrospective study of an open-source dataset of
3 patients in Mexico. The Ministry of Health of Mexico approved the recollection of
4 information.

5

6 **Declarations of interest:** none

7

8 **Funding:** This research did not receive any specific grant from funding agencies in the
9 public, commercial, or not-for-profit sectors.

10

11 **Acknowledgements**

12

13 J M-G would like to thank “Dirección General de Calidad y Educación en Salud” for
14 supporting his participation in “Programa Nacional de Servicio Social en Investigación
15 en Salud”.

16

17

18 **References**

19

- 20 1. **Gates B.** Responding to Covid-19 — A Once-in-a-Century Pandemic? *New*
21 *England Journal of Medicine* 2020; **382**: 1677–1679.
- 22 2. **Williamson EJ, et al.** Factors associated with COVID-19-related death using
23 OpenSAFELY. *Nature* Springer US, 2020; **584**: 430–436.

- 1 3. **Kammar-García A, et al.** Impact of Comorbidities in Mexican SARS-CoV-2-
2 Positive Patients: A Retrospective Analysis in a National Cohort. *Revista de*
3 *investigacion Clinica* 2020; **72**: 151–8.
- 4 4. **Guan W, et al.** Comorbidity and its impact on 1590 patients with COVID-19 in
5 China: a nationwide analysis. *European Respiratory Journal* 2020; **55**: 2000547.
- 6 5. **Gupta S, et al.** Factors Associated With Death in Critically Ill Patients With
7 Coronavirus Disease 2019 in the US. *JAMA Internal Medicine* 2020; **02115**: 1–11.
- 8 6. **Zhou F, et al.** Clinical course and risk factors for mortality of adult inpatients with
9 COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* 2020; **395**:
10 1054–1062.
- 11 7. **Roder-DeWan S.** Health system quality in the time of COVID-19. *The Lancet*
12 *Global Health* 2020; **8**: e738–e739.
- 13 8. **Wynants L, et al.** Prediction models for diagnosis and prognosis of covid-19:
14 Systematic review and critical appraisal. *The BMJ* 2020; **369**: m1328.
- 15 9. **Collins GS, van Smeden M, Riley RD.** COVID-19 prediction models should
16 adhere to methodological and reporting standards. *European Respiratory Journal*.
17 Published online: 23 July 2020. doi: 10.1183/13993003.02643-2020.
- 18 10. **Hooli S, King C.** Generalizability of Coronavirus Disease 2019 (COVID-19)
19 Clinical Prediction Models. *Clinical Infectious Diseases* 2020; **71**: 897–897.
- 20 11. **Directorate General of Epidemiology MEX.** Standardized Guideline for
21 Epidemiologic and Laboratory Surveillance of viral respiratory diseases, May

- 1 2020 [Spanish]. Mexican Secretariat of Health. Mexico City, 2020, p. 1–80
2 (https://www.gob.mx/cms/uploads/attachment/file/552972/Lineamiento_VE_y_Lab
3 _Enf_Viral_20.05.20.pdf). Accessed 17 August 2020.
- 4 12. **Health Ministry MEX.** Information regarding COVID-19 cases in Mexico
5 [Spanish]. Datos Abiertos 2020 ([https://datos.gob.mx/busca/dataset/informacion-](https://datos.gob.mx/busca/dataset/informacion-referente-a-casos-covid-19-en-mexico)
6 referente-a-casos-covid-19-en-mexico). Accessed 17 August 2020.
- 7 13. **Directorate General of Epidemiology MEX.** Historical COVID-19 Datasets
8 [Spanish]. Datos Abiertos - Bases Históricas.
9 2020([https://www.gob.mx/salud/documentos/datos-abiertos-bases-historicas-](https://www.gob.mx/salud/documentos/datos-abiertos-bases-historicas-direccion-general-de-epidemiologia)
10 direccion-general-de-epidemiologia). Accessed 17 August 2020.
- 11 14. **Sullivan LM, Massaro JM, D’Agostino RB.** Presentation of multivariate data for
12 clinical use: The Framingham Study risk score functions. *Statistics in Medicine*
13 2004; **23**: 1631–1660.
- 14 15. **Docherty AB, et al.** Features of 20 133 UK patients in hospital with covid-19
15 using the ISARIC WHO Clinical Characterisation Protocol: prospective
16 observational cohort study. *BMJ* 2020; **369**: m1985.
- 17 16. **Richardson S, et al.** Presenting Characteristics, Comorbidities, and Outcomes
18 Among 5700 Patients Hospitalized With COVID-19 in the New York City Area.
19 *JAMA* 2020; **323**: 2052.
- 20 17. **Yang X, et al.** Clinical course and outcomes of critically ill patients with SARS-
21 CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective,
22 observational study. *The Lancet Respiratory Medicine* 2020; **8**: 475–481.

- 1 18. **Karagiannidis C, et al.** Case characteristics, resource use, and outcomes of
2 10□021 patients with COVID-19 admitted to 920 German hospitals: an
3 observational study. *The Lancet Respiratory Medicine* 2020; **2600**: 1–10.
- 4 19. **Wu C, et al.** Risk Factors Associated With Acute Respiratory Distress Syndrome
5 and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan,
6 China. *JAMA Internal Medicine* 2020; **180**: 934.
- 7 20. **Hu Z, et al.** Clinical characteristics of 24 asymptomatic infections with COVID-19
8 screened among close contacts in Nanjing, China. *Science China Life Sciences*
9 2020; **63**: 706–711.
- 10 21. **Wang Y, et al.** Clinical Outcomes in 55 Patients With Severe Acute Respiratory
11 Syndrome Coronavirus 2 Who Were Asymptomatic at Hospital Admission in
12 Shenzhen, China. *The Journal of Infectious Diseases* 2020; **221**: 1770–1774.
- 13 22. **Liang T, et al.** Evolution of CT findings in patients with mild COVID-19
14 pneumonia. *European Radiology* European Radiology, 2020; **30**: 4865–4873.
- 15 23. **Awulachew E, et al.** Computed Tomography (CT) Imaging Features of Patients
16 with COVID-19: Systematic Review and Meta-Analysis. *Radiology Research and*
17 *Practice* 2020; **2020**: 1023506.
- 18 24. **Cleverley J, Piper J, Jones MM.** The role of chest radiography in confirming
19 covid-19 pneumonia. *The BMJ* 2020; **370**: 1–9.
- 20 25. **Soto-Mota A, et al.** The LOW-HARM Score for Predicting Mortality in Patients
21 Diagnosed with COVID-19: A Multicentric Validation Study. *medRxiv*. Published

- 1 online: 5 June 2020. doi: 10.1101/2020.05.26.20111120.
- 2 26. **Bello-Chavolla OY, et al.** Predicting Mortality Due to SARS-CoV-2: A
3 Mechanistic Score Relating Obesity and Diabetes to COVID-19 Outcomes in
4 Mexico. *The Journal of Clinical Endocrinology & Metabolism* 2020; **105**: 1–10.
- 5 27. **Mejia-Vilet JM, et al.** A Risk Score to Predict Admission to Intensive Care Unit in
6 Patients With COVID-19: The ABC-GOALS Score. *medRxiv*. Published online: 30
7 May 2020. doi: 10.1101/2020.05.12.20099416.
- 8 28. **Moons KGM, et al.** Transparent Reporting of a multivariable prediction model for
9 Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration.
10 *Annals of Internal Medicine* 2015; **162**: W1–W73.
- 11 29. **Instituto Mexicano del Seguro Social IMSS.** *Operational definition of suspected*
12 *case of respiratory viral disease, including COVID-19 [Spanish]*. Mexico City:
13 Instituto Mexicano del Seguro Social (IMSS), 2020, p. 6.
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22

1

2

3

4

5

6

7

Table 1. Demographic characteristics, patient history data, and outcomes in the model development and validation cohorts.

	Development cohort, positive cases			Validation cohort, positive cases		
	Total n=83,779	Survivors n=74,551	Non-survivors n=9,228	Total n=100,000	Survivors n=94,722	Non-survivors n=5,278
Sex						
Women, n (%)	36,393 (43.4)	33,370 (44.8)	3,023 (32.8) ^a	48,129 (48.1)	46,304 (48.9)	1,825 (34.6) ^a
Men, n (%)	47,386 (56.6)	41,181 (55.2)	6,205 (67.2) ^a	51,871 (51.9)	48,418 (51.1)	3,453 (65.4) ^a
Age, years	46.3 (15.9)	44.6 (15.4)	60 (14.2) ^a	43.9 (16.5)	42.9 (16.1)	61.3 (14.5) ^a
Smoking, n (%)	6,966 (8.3)	6,092 (8.2)	874 (9.5) ^a	6,935 (6.9)	6,538 (6.9)	397 (7.5)
Pregnancy, n (%)	537 (0.6)	523 (0.7)	14 (0.2) ^a	729 (0.7)	724 (0.8)	5 (0.1) ^a
Contact with confirmed case, n (%)	26,408 (31.3)	25,127 (33.7)	1,100 (11.9) ^a	52,047 (52)	50,563 (53.4)	1,484 (28.1) ^a
Comorbidities, n (%)						
Diabetes	14,735 (17.6)	11,235 (15.1)	3,500 (37.9) ^a	14,036 (14)	12,010 (12.7)	2,026 (38.4) ^a
COPD	1,695 (2)	1,183 (1.6)	512 (5.5) ^a	1,244 (1.2)	1,020 (1.1)	224 (4.2) ^a
Asthma	2,466 (2.9)	2,258 (3)	208 (2.3) ^a	2,543 (2.5)	2,427 (2.6)	116 (2.2)
Immunosuppression	1,300 (1.6)	1,020 (1.4)	280 (3) ^a	973 (1)	848 (0.9)	125 (2.4) ^a
Hypertension	17,609 (21)	13,696 (18.4)	3,913 (42.4) ^a	17,162 (17.2)	14,970 (15.8)	2,192 (41.5) ^a
CVD	2,178 (2.6)	1,640 (2.2)	538 (5.8) ^a	1,875 (1.9)	1,601 (1.7)	274 (5.2) ^a
Obesity	17,244 (20.6)	14,752 (19.8)	2,492 (27) ^a	18,161 (18.2)	16,698 (17.6)	1,463 (27.7) ^a
CKD	1,944 (2.3)	1,293 (1.7)	651 (7.1) ^a	1,401 (1.4)	1,131 (1.2)	270 (5.1) ^a
Time from symptom onset to medical attention, days	4.3 (3.2)	4.3 (3.2)	4.5 (3.1) ^a	4.3 (3.1)	4.2 (3.1)	5.1 (3.2) ^a
Hospitalization, n (%)	29,535 (35.3)	21,257 (28.5)	8,278 (89.7) ^a	19,316 (19.3)	14,753 (15.6)	4,563 (86.5) ^a
Intubation, n (%)	2,923 (3.5)	953 (1.3)	1,970 (21.3) ^a	2,260 (2.3)	778 (0.8)	1,482 (28.1) ^a
Pneumonia, n (%)	22,785 (27.2)	15,644 (21)	7,141 (77.4) ^a	16,742 (16.7)	12,122 (12.8)	4,620 (87.5) ^a
ICU admission, n (%)	2,791 (3.3)	1,359 (1.8)	1,432 (15.5) ^a	2,383 (2.4)	1,419 (1.5)	964 (18.3) ^a
Case-Fatality rate (%)	11.02%	-	-	5.3%	-	-

Data are presented as mean values (SD), unless otherwise specified.

a: Statistical significance regarding Survivors

Table 2. Risk factors associated with death in Mexican patients with a positive diagnostic test for SARS-CoV-2 (model development cohort)

	Regression coefficient	Standard error	HR (95%CI)	p value	Mean or proportion
Age	0.050	0.001	1.05 (1.05-1.05)	<0.0001	46.27
Sex (men)	0.473	0.022	1.6 (1.54-1.68)	<0.0001	0.57
Diabetes	0.437	0.023	1.55 (1.48-1.62)	<0.0001	0.18
COPD	0.136	0.047	1.15 (1.05-1.26)	0.004	0.02
Immunosuppression	0.300	0.062	1.35 (1.19-1.52)	<0.0001	0.02
Hypertension	0.206	0.024	1.23 (1.17-1.29)	<0.0001	0.21
Obesity	0.337	0.024	1.4 (1.34-1.47)	<0.0001	0.21
CKD	0.620	0.042	1.86 (1.71-2.02)	<0.0001	0.02

HR: Hazard ratio, 95%CI: 95% Confidence Interval.

COPD: Chronic obstructive pulmonary disease, CKD: Chronic kidney disease

Table 3. The PH-Covid19 risk score to predict death in patients with COVID-19

Risk Factor	Score
Age	
<20	-2
20-29	0
30-39	2
40-49	4
50-59	6
60-69	8
70-79	10
80-89	12
90-99	14
>99	15
Sex	
Male	2
Comorbidities	
Diabetes	2
COPD	1
Immunosuppression	1
Hypertension	1
Obesity	1
CKD	2

Range of values: -2 a 25

COPD: Chronic obstructive pulmonary disease, CKD: Chronic kidney disease.

Table 4. Estimated risk of death according to every possible score in the PH-Covid19 score, in Mexican patients with a positive test for SARS-CoV-2

Total value	Estimated Risk	Total value	Estimated Risk
-2	0.74%	12	21.72%
-1	0.94%	13	26.98%
0	1.21%	14	33.22%
1	1.55%	15	40.45%
2	1.99%	16	48.60%
3	2.55%	17	57.46%
4	3.26%	18	66.63%
5	4.17%	19	75.56%
6	5.32%	20	83.62%
7	6.77%	21	90.20%
8	8.61%	22	94.94%
9	10.92%	23	97.83%
10	13.80%	24	99.27%
11	17.36%	25	99.82%





