

Assessing the impact of multiple comorbidities on fatal outcome in young COVID-19

*Paulino Monroy-Castillero¹, *Eitan Friedman², Arturo Revuelta-Herrera³, Arik Yochelis¹

1 Department of Solar Energy and Environmental Physics, Blaustein Institutes for Desert Research (BIDR), Ben-Gurion University of the Negev, Sede Boqer Campus, Midreshet Ben-Gurion 8499000, Israel

2 Chaim Sheba Medical Center, Ramat Gan, 52621 Israel and the Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv, 69978 Israel

3 Health Secretariat of Mexico City, Mexico City 06820, Mexico

* Both authors have equally contributed to this work

Address correspondence to:

Eitan Friedman MD PhD, Chaim Sheba Medical Center, Tel-Hashomer, Israel 52621 and the Sackler School of Medicine, Tel-Aviv University, Tel Aviv 69978 Israel

Tel +972-522891561 FAX +972-3-5357308

emails – eitan.friedman@sheba.gov.il or feitau@tauex.tau.ac.il

Abstract

A Bayesian analysis with the use of a rank-biserial correlation algorithm was applied to identify the impact of multiple comorbid conditions on fatal COVID-19 outcome in young adult cases (40-50 years). The demonstration was conducted for a publicly available database provided by the Mexican authority, in the absence of other alternative free-access repositories with information per patient. The methodology here proposed showed that even in the face of small sample sizes, it is possible to highlight deleterious synergistic comorbid conditions.

Young adult cases with COVID-19 and co-existing diabetes, obesity, hypertension, CRF, or COPD were found more likely to have a fatal outcome compared with having no co-morbidities (X2-6 times). With the methodology proposed, we show that having diabetes or hypertension in addition to CRF increased risk for mortality more than what is expected from independent effect (adverse synergistic effect), whereas in patients with obesity, the additional presence of diabetes or hypertension do not increase markedly the death risk due to COVID-19. Quantitative analysis of having two comorbidities highlights the combinations of morbid conditions that are more likely to be associated with fatal outcomes in younger adults COVID-19 cases in a clinically applicable manner.

The clinical implication of this method needs to be prospectively assessed.

Introduction

The novel SARS-CoV-2 virus first described in China at the end of 2019 is the causative agent of COVID-19 pandemic.¹ Up to December 13, 2020, there were 72,653,057 confirmed COVID-19 cases and 1,618,890 COVID-19 related deaths globally.² While the majority of COVID-19 cases exhibit mild symptoms with a self-limiting clinical course and do not require hospitalization, a subset of cases develop severe disease hallmarked by bilateral pneumonia and acute respiratory distress requiring ventilation, with a minority of COVID-19 cases dying of the disease.^{3,4} By early estimates infection fatality ratio are ~0.5%.⁶ Case fatality rate is hard to assess in real time as multiple confounders limit the ability to compare inter country rates (e.g., testing indications, population genetic makeup, associating a direct causal effect on COVID-19 as a cause of death and timing, associated comorbidities, population median age differences, and gender specific fatalities) to name a few. While age (above 65 years) and male gender adversely affect disease course and are associated with higher mortality rates, other comorbid conditions also play a role. Notably, hypertension, diabetes mellitus, chronic obstructive pulmonary disease (COPD), cardiovascular diseases, obesity, and chronic renal failure (CRF) are all comorbidities that are well established as associated with more adverse COVID-19 outcomes⁷⁻⁹ Specifically in Mexico by November 13, 2020, of 323,671 confirmed cases at the Mexican Institute of Social Security, hypertension (reported in 40.4% of cases), diabetes (33.1%) and obesity (21.7%) were more frequently noted in cases with fatal outcome.¹⁰ Identification and quantitation of risk factors associated with disease severity and/or mortality among COVID-19 patients is urgently needed, ideally using data analysis scheme that is based on actual data without any model driven assumptions.

COVID-19 patients who are younger than 60 years of age (and hence considered at low risk for developing severe disease and fatal outcome).¹¹ Thus, the ability to evaluate the contribution of risk factors associated with severe outcome in this age group is essential for early intervention in order to potentially improve prognosis and for prioritization schemes for vaccine distribution. Most risk factor analyses studies have focused on a single factor and its effect on disease severity or mortality on all age groups, employing assumption-driven models to compensate for and extrapolate over missing data, such as a logistic regression-type approaches,¹²⁻¹⁴ while some others are based on artificial intelligence algorithms and provide a binary outcome (mortality or recovery).¹⁵ Based on correlation with mortality rates, a common theme has emerged that diabetes mellitus, hypertension, and obesity are the most significant comorbidities associated with COVID 19 mortality.^{7,10} However, assessment of an individualized adverse disease course should not rely on retrospective analyses but rather assess these risks as a function of comorbidities in a prospective manner, data that is often incompletely collected and curated.¹⁶ Thus, a key component in identifying the relative contribution of each factor to adverse disease outcome and the possible interaction between factors affecting mortality should be hypothesis-free methodology. Applying a generic Bayesian inference to COVID-19 Mexican real-time data cases, we show that it is possible to identify synergistic effects of comorbid conditions and compare their effects on mortality risk, even in the presence of incomplete data.

Methods

Data Source and Collection

Deidentified, open access, repository of data collected on all individuals who tested positive for SARS-

CoV-2 by reverse transcription polymerase chain reaction (RT PCR) using one of several commercially available kits that were validated and approved for use by the Institute of Epidemiological Diagnosis and Reference (InDRE)¹⁷ in Mexico, were obtained from the web site of the Secretaría de Salud, Mexican Federal Government,¹⁸ (accessed December 10, 2020). This site contains all Mexican subjects tested for SARS CoV-2 in public institutions since February 2020. For each individual, the following data were obtained: age, gender, alive/deceased status, date of symptom onset, and any of the following, self-reported comorbidities: diabetes mellitus, chronic obstructive pulmonary disease (COPD), asthma, immunosuppression, hypertension (HTN), cardiovascular disease (CVD), obesity (BMI>30), chronic renal failure (CRF), and smoking status (never/ever). To avoid bias that may have been introduced by reporting delays of disease outcome (deaths or recovery), we only included individuals with symptom onset date earlier than November 11, 2020.

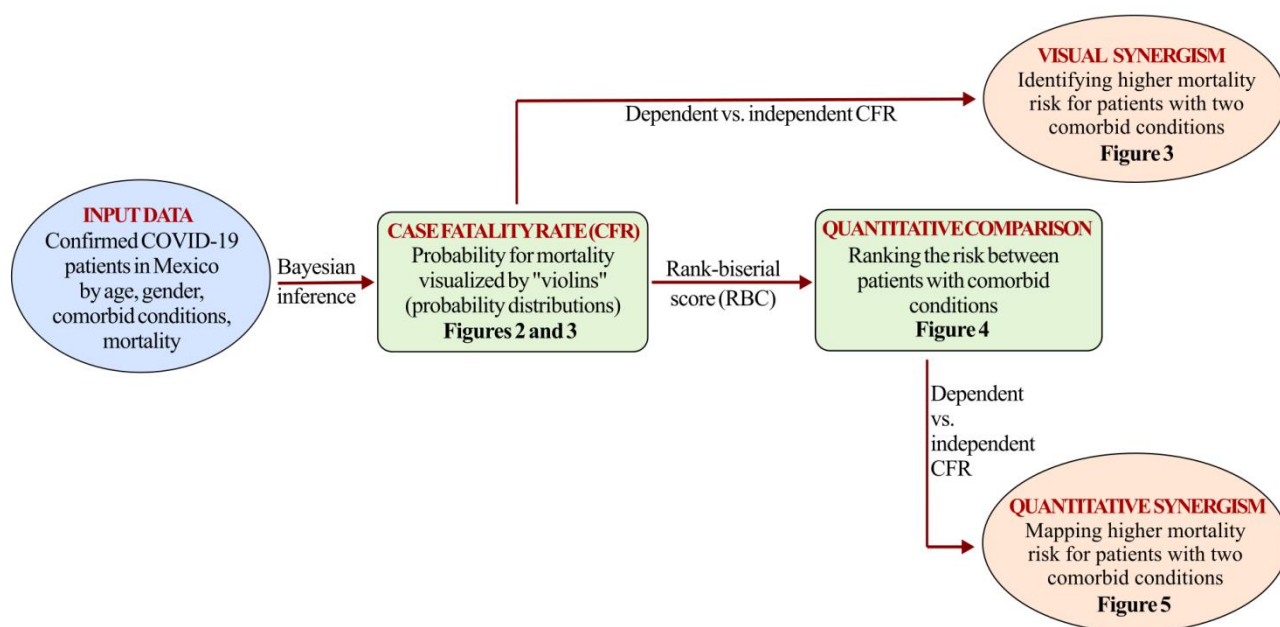


Figure 1: Flowchart of the conducted study.

Statistical Analysis

Given the incompleteness of data and lack of established knowledge on specific a priori probability, Bayesian methodology was chosen for inference.^{19–23} Here, we focus on factors that putatively affect the case fatality rate (CFR) by following the path outlined in Fig. 1: (i) using clinical data to substantiate the posterior probability distribution of the CFR, referred to as case fatality rate distribution (CFRD), and their respective relative risks (RR), and (ii) identifying synergistic effects in the presence of two comorbidities in a qualitative and a quantitative way, the latter by using the rank-biserial correlation (RBC).²⁴

Case Fatality Rate (CFR) and Relative Risk (RR)

The compiled data encompassed the total number of patients (n) and the number of deaths (d) by

gender, age, and comorbid conditions. Using Bayesian inference²¹ data-based case fatality rate distribution (CFRD) which outlines the likelihood of the CFR was calculated. From a statistical view point, CFR is the true value, and CFRD is the so-called posterior probability distribution that displays less spreading along the vertical axis as n increases.²¹⁻²³

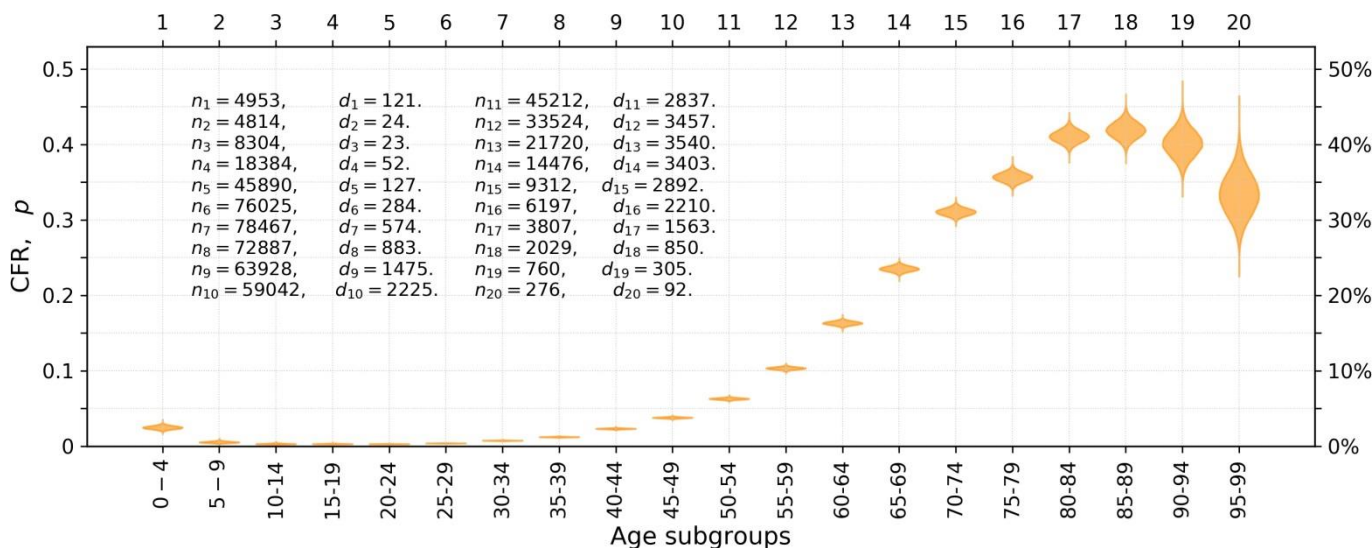


Figure 2: **Age dependent risk for developing adverse disease course in the absence of comorbid conditions.** Case fatality rate (CFR) likelihood outlined by “violins” (posterior beta distributions) of individuals within age intervals. The violins are constructed using a standard Bayesian update procedure^{22,23} by starting from a uniform prior distribution and by incorporating data about the number of patients, n_i , and the number of deaths, d_i , in each subgroup ($i = 1, \dots, 20$). Specifically, we use a large sample of 50,000 random variates (realizations via Monte Carlo method^{25,26}) from the beta distribution, $B(p; \alpha, \beta)$, where $\alpha = d_i + 1$ and $\beta = n_i - d_i + 1$. Violin’s elongation along the vertical axis indicates the precision in the probability of CFR, i.e., the smaller vertical extent is the closer is the result to “true” value.

Initially, CFR from COVID-19 cases by age group (at 5-year intervals) for individuals with no comorbid conditions were calculated (Fig. 2). While the “true” CFR remains unknown, the so-called “violins” in Fig. 2 outline the likelihood and precision of CFR. Next, the RR in a presence of a single comorbid condition for males and females separately, was calculated (Fig. 3). The likelihood of RR is again outlined by a distribution of the CFRD of patients that is normalized by the CFRD of cases with no comorbid condition, as is shown in Fig. 2. Fig. 3 outlines the CFRD of patients having two comorbid conditions. In order to identify any possible deleterious comorbidity synergism, the combined “theoretical” CFRD was plotted as if both comorbidities have no synergistic effect but rather an additive effect (gray color violins in Fig. 3).

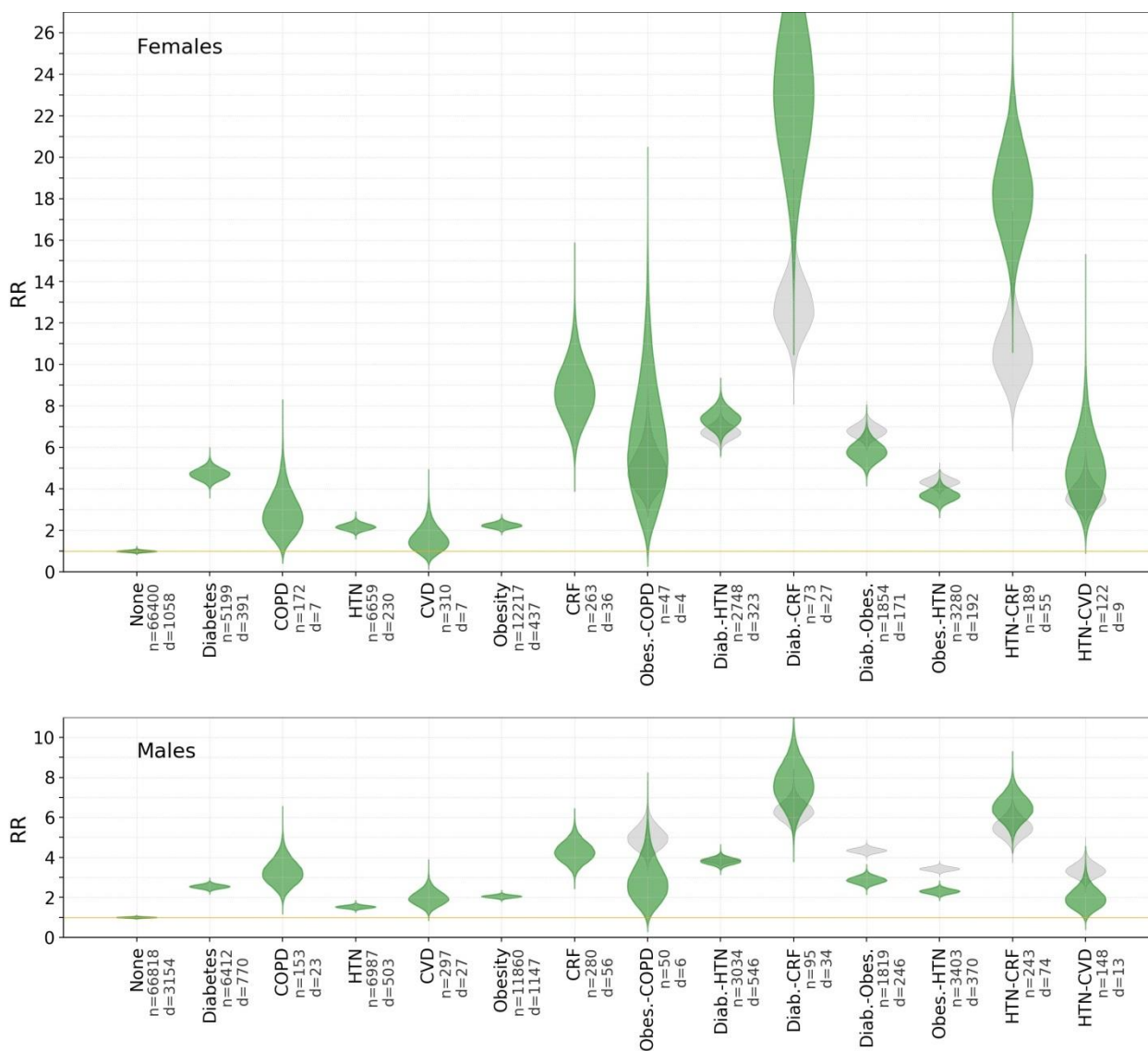


Figure 3: **Relative risks (RR) for comorbid conditions.** Green color violins depict the RR distributions obtained from the data while the gray violins depict a reference of distributions for conditions of two independent comorbid conditions having additive contributions. Specifically, computation of the relative risk distribution (RRD) of each subgroup is obtained by Bayesian update²³ (as in Fig. 2) but for $RR = p/p_{ref}$, where p and p_{ref} are random variates taken from the data for comorbid condition forming $B_{data}(p)$ and a CFRD for the reference subgroup in the absence of comorbid condition $B_{ref}(p)$. For a pair of independent comorbidities, gray shaded violins, we define the distribution with additive property, $CFRD_{add}$, that is generated via the Monte Carlo method that randomly samples the probability for mortality $p_{add} = 1 - (1 - p_{comorbid 1})(1 - p_{comorbid 2})$, where $p_{comorbid 1}$ and $p_{comorbid 2}$ are random variates of each comorbid condition CFRD separately.²⁶

The results represent data for males and females at age interval 40-50 years.

Rank-biserial correlation (RBC)

While it is possible to visualize the synergistic effects in patients with two comorbid conditions (Fig. 3), the calculated CFRD can also be used to get a quantitative comparative ranking. To that end, the RBC can be used to compare the risk of two patients with distinct comorbid conditions. Formally, the RBC is a correlation effect size for the Mann-Whitney U test.²⁴ Then, we set a threshold $|RBC| \geq 0.9$ which provides the region for which it is possible to identify with a significance of 10%, who is at higher

risk. Several examples are shown in Fig. 4. Finally, we also use RBC to detect some synergetic effect of two comorbidities, as shown in Fig. 5.

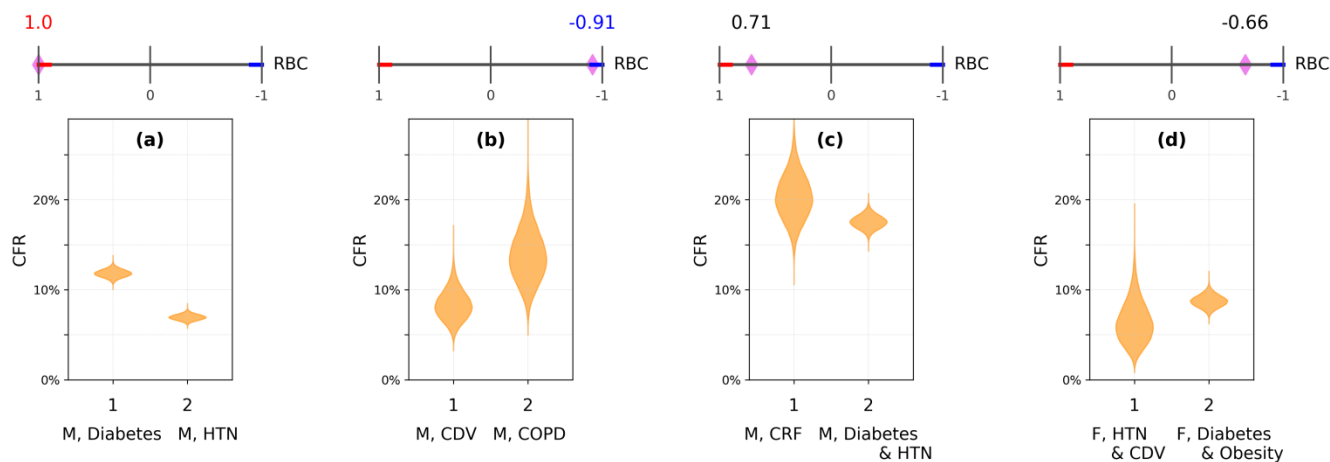


Figure 4: Risk ranking of patients with comorbid conditions. Selected cases of patients (labeled as 1 and 2) with distinct comorbid conditions, shown by the CFRD. The results taken for males (M) and females (F) at age interval 40-50 years. The top bar indicates the rank-biserial correlation (RBC) performed from a Monte Carlo sampling of each CFRD, where the supra-thresholds $RBC \geq 0.9$ and $RBC \leq -0.9$ are marked by red and blue colors, respectively.²⁴ The diamond symbol marks the RBC value according to: $RBC(B_1, B_2) = P(p_1 > p_2) - P(p_2 > p_1)$, where $P(p_1 > p_2)$ is the probability that the CFR for patient 1 (p_1) is higher than CFR for patient 2 (p_2) based on their respective CFRD, which are represented by the yellow violins. $P(p_2 > p_1) = 1 - P(p_1 > p_2)$.

Results

Cohort Description

By December 10, 2020, 1,174,784 individuals (578,791 females and 595,993 males) who tested positive for SARS-CoV-2 in Mexico were available for analysis from the Secretaría de Salud, Mexican Federal Government database.¹⁸ Eligible individuals (with symptom onset date earlier than November 11, 2020) were subdivided into three groups: 570,094 individuals with no known or reported comorbidities (as enumerated in the methods section), 264,703 individuals with a single morbid condition, and 121,774 individuals with two comorbidities. Noteworthy, the current analysis included only cases with up to two comorbidities whereas individuals with more than 2 comorbid conditions were excluded. Given the major effect of age on mortality and the paucity of data on the effect of comorbidities on mortality in younger cases, we focused on the seemingly less susceptible age group of 40 to 50 years, that encompassed 224,616 patients (11,219 deaths) in total. For that subgroup, of 110,561 females 11,219 deaths were reported, and 7,866 of 114,055 males died.

Relative risk by comorbidity status

Effect of Single Comorbid Condition on Relative Risk (RR) - As shown in Fig. 3, diabetes mellitus, CRF, and obesity were associated with a mean mortality RR of X2-3 in males, and higher means in females. Furthermore, while the RR distributions (RRD) for diabetes mellitus and obesity showed restricted distributions, the RRD for CRF and COPD exhibit smeared RRD (i.e., elongated along the y-axis violins).

This implies that there is more uncertainty as to the RR conferred by the CRF compared with diabetes and obesity. In addition, COPD and hypertension were also associated with an increased RR in females that is more pronounced compared with the effects of these factors in males. Overall, RRD in females are more elongated (hence more uncertain) and have a mean of higher values compared with males.

Effect of Two Comorbid Conditions on RR- As shown in Fig. 3, having more than one comorbid condition in both genders may affect the likelihood of having higher RR. For example, having both CRF and diabetes mellitus increases the RR and the RRD spreads, implying an adverse effect, along with increased uncertainty due to the limited number of cases displaying these two conditions. Hypothetically, RRD in cases reporting two comorbidities could have independent additive effects, indicated by the gray color “violins” in Fig. 3. If the actual RRD (green violins in Fig. 3) are above those outlined by the grey violins, with no or minor overlap then it may indicate a synergistic deleterious impact on disease outcome. Conversely, comorbid conditions may have a seemingly “protective effect” on mortality in COVID-19 cases, namely the RRD is lower than the independent combination. Specific examples of both adverse and seemingly protective effects are shown in Fig. 3.

While the impact of having two comorbidities can be visually deduced from Fig. 3, the analysis may also enable quantitative differences, as shown by pair-wise combination in Fig. 5. Here, the information corresponds only to a dichotomous depiction of the possibility of synergy while disregarding relative risk values (see Methods section for details). Specifically, the results show that only for a few combinations of comorbidities it is possible to determine a clear-cut synergistic effect (e.g., diabetes or HTN with CRF), information that quantifies the visual comparison in Fig. 3. Notably, the combinations of obesity with either asthma, diabetes or HTN, show a non-synergistic and even a seemingly protective effect.

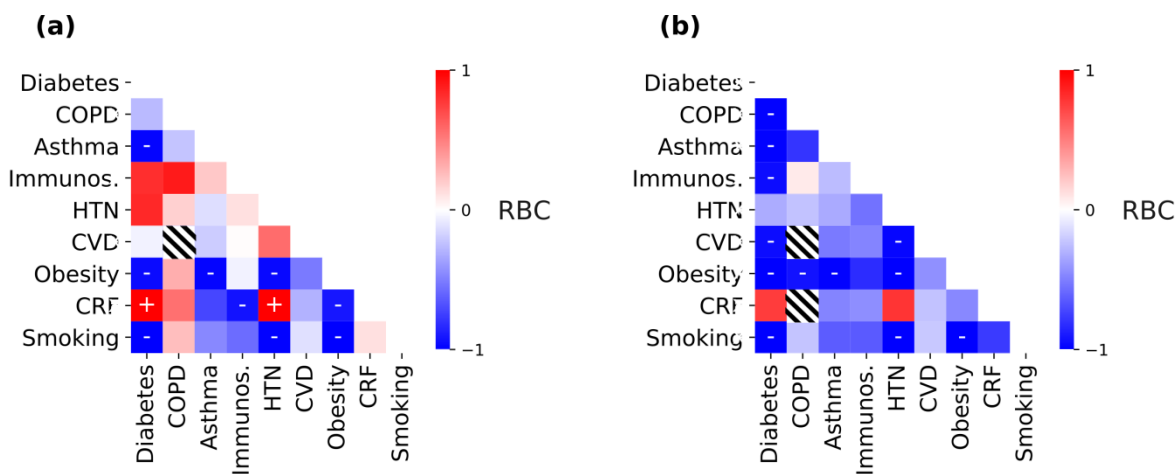


Figure 5: **Identification of synergism between two comorbid conditions.** Mapping of $RBC(B_1, B_2)$, where B_1 is related to a pair of concurrent comorbid conditions and $B_2 = B_{add}$, is the additive CFRD of the respective pair, see also Fig. 4. The RBC and the respective supra-threshold values (together with '+' and '-', respectively), are computed and marked as in Fig. 4; we do not reflect and in fact ignore the comparisons that yield sub-threshold score. Half-filled squares indicate the absence of sufficient data (i.e., subgroups with less than 3 patients). (a) Females. (b) Males. Both at age interval 40-50 years.

Discussion

The analysis used herein is assumption-free and relies only on statistical inference of the available data. Specifically, it enables to quantitatively compare between two risk profiles (Fig. 4), a seemingly useful aid in clinical decision making. Although the statistical sampling used herein is still not big enough to establish conclusive determination whether some of the comorbidities pairs are associated or not with synergistic deleterious effects, is indeed enough to assert that there are combinations that deserve special attention: CRF with diabetes mellitus or HTN significantly increase RR and RRD for mortality. Conversely, the combinations of obesity with hypertension or diabetes mellitus seem not to have a synergistic effect on mortality RR. On the other hand, diabetes with HTN seems to present an independent combination effect; although is not synergistic according to our definition, the increase of risk is quite substantial to be highlighted for medical purposes. The mechanisms that underlie the observed synergistic or lack of such effect of combined comorbidities, remain unknown. A major limitation of this study that can be easily improved by worldwide information sharing is the incompleteness of data collection. Another limitation may be inherent to the fact that the data were collected retrospectively and pertain to a single country with distinct genetic makeup. Lastly, there are a limited number of cases in some categories (e.g., COPD females n=172, d=7, and males n=153, d=23 or CRF females n=263, d=36, and males n=280, d=56).

Unlike other available predictive risk methods that rely on estimations of parametric models from the whole dataset to predict risk profiles,¹²⁻¹⁴ the method applied and described herein used solely data of subgroups that are of interest, without further assumptions, and provides an order relation based on statistical comparison of risks. In the meta-analysis of observational studies by Parohan and coworkers³ risk factors (e.g., hypertension, diabetes mellitus, obesity, cerebrovascular stroke (CVS), COPD) individually were associated with an odds ratio (OR) of 1.5-4.59 for fatal COVID-19 outcome. Additional analyses refined these risks by individual risk factors (some of which are shown in Supplementary Table 1). Specifically in Mexico, a study encompassing 51,633 subjects with SARS-CoV-2 and 5,332 deaths, fatal outcome risk factors include early-onset diabetes mellitus, obesity, COPD, advanced age, HTN, immunosuppression, and CRF.¹⁴ Yet, it had been difficult to clinically apply the known risk factors in order to prioritize care for COVID-19 patients, due to the inaccuracy of assessing the relative contribution of each risk factor and their possible interactions as determinants of deleterious disease outcome.

This methodology can be employed clinically with a real-time updating to stratify cases of different mortality risks in the emergence of epidemics and to guide allocation of medical resources or prioritizing vaccine distribution, if the ability of this tool can be ascertained and validated prospectively.

The accuracy of the tool developed reported herein should be validated in other countries and using more data with more cases in each of the morbid conditions categories. In addition, the accuracy of this tool should be tested in a prospective study design and its added value as a clinical support decision tool should be evaluated.

References

1. WHO announces COVID-19 outbreak a pandemic. 2020. <http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic>.

2. Worldometer. <https://www.worldometers.info/coronavirus/> (accessed 19 June 2020).
3. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 2020;382:1708–20.
4. Kakodkar P, Kaka N, Baig M. A Comprehensive Literature Review on the Clinical Presentation , and Management of the Pandemic Coronavirus Disease 2019 (COVID-19). *Cureus* 2020;12(4);e7560.
5. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet*. 2020;395:470–3.
6. Roques L, Klein E, Papaix J, Sar A, Soubeyrand S. Using Early Data to Estimate the Actual Infection Fatality Ratio from COVID-19 in France. *Biology (Basel)* 2020;9:1–10.
7. Parohan M, Yaghoubi S, Seraji A, Hassan M, Sarraf P, Djalali M. Risk factors for mortality in patients with Coronavirus disease 2019 (COVID-19) infection: a systematic review and meta-analysis of observational studies. *Aging Male* 2020; published online 8 June. doi.org/10.1080/13685538.2020.1774748.
8. Klang E, Kassim G, Soffer S, Freeman R, Levin MA, Reich DL. Morbid Obesity as an Independent Risk Factor for COVID-19 Mortality in Hospitalized Patients Younger than 50. *Obesity* 2020; published online May 23. doi.org/10.1002/oby.22913.
9. Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int* 2020;97(5):829–38.
10. Escobedo-de la Peña J, Rascón-Pacheco RA, Ascencio-Montiel IJ, González-Figueroa E, et al. Clinical characteristics and risk factors for mortality of patients with COVID-19 in a large data set from Mexico. *Annals of Epidemiology* 2020;52:93-98.e2.
11. Kang YJ. Mortality rate of infection with COVID-19 in Korea from the perspective of underlying disease. *Disaster Med. Public Health Prep* 2020; published online 31 March. doi.org/10.1017/dmp.2020.60.
12. Jehi L, Ji X, Milinovich A, Erzurum S, Rubin B, Gordon S, et al. Individualizing Risk Prediction for Positive COVID-19 Testing Results from 11 , 672 Patients. *Chest* 2020;S0012-3692:31654–8.
13. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, et al. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. *JAMA Intern Med* 2020;180(8):1081–1089.
14. Bello-Chavolla OY, Bahena-lópez JP, Antonio- NE. Predicting mortality due to SARS-CoV-2: A mechanistic score relating obesity and diabetes to COVID-19 outcomes in Mexico. *J Clin Endocrinol Metab* 2020;105:1–10.
15. An C, Lim H, Kim DW, Chang JH, Choi YJ, Kim SW. Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study. *Sci Rep* 2020;10:18716.
16. Chong WF, Ding YY, Heng BH. A comparison of comorbidities obtained from hospital administrative data and medical charts in older patients with pneumonia. *BMC Health Serv Res* 2011;11:105.
17. Secretaría de Salud, Federal Government of Mexico. Listado de pruebas moleculares útiles para el diagnóstico de SARS-CoV-2 durante la contingencia de COVID-19 en Mexico. 2020. https://www.gob.mx/cms/uploads/attachment/file/569610/Listado_de_estuches_comerciales_uitiles_para_el_diagnostico_de_SARS-CoV-2.pdf (accessed 14 Aug 2020).
18. Data web open acces. Secretaría de Salud, Federal Government of Mexico. 2020. <https://www.gob.mx/salud/documentos/datos-abiertos-bases-historicas-direccion-general-de-epidemiologia> (accessed 19 June 2020).
19. Spiegelhalter D., Myles J., Jones D., Abrams K. An introduction to Bayesian methods in health technology assessment. *Br Med J* 1999;319:508–12.
20. Ibrahim JG, Chen MH, Chu H. Bayesian methods in clinical trials: A Bayesian analysis of ECOG trials

- E1684 and E1690. *BMC Med Res Methodol* 2012;12:1–12.
21. Kruschke JK. Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd ed. Academic Press. Amsterdam: Elsevier, 2014.
 22. Held L, Sabanés Bové D. Applied statistical inference: Likelihood and bayes. Applied Statistical Inference: Likelihood and Bayes. Berlin, GER: Springer-Verlag, 2014.
 23. Lu S. Visualizing Beta Distribution and Bayesian Updating. *Towards Data Science*. 2019; published online 1 April. <https://towardsdatascience.com/visualizing-beta-distribution-7391c18031f1> (accessed 19 June 2020).
 24. Kerby DS. The simple difference formula: an approach to teaching nonparametric correlation. *Innovative Teaching* 2014;3,1.
 25. Kroese DP, Brereton T, Taimre T, Botev ZI. Why the Monte Carlo method is so important today. *WIREs Comput Stat* 2014;6:386–92.
 26. Koch K. Monte Carlo methods. *GEM - Int J Geomathematics*. 2018;9:117–43.