COVID-19 Hazard Risk Assessment for Japanese Population

Wirawit Chaochaisit[1], Michel Mommejat[2], Iri Sato Baran[1]

[1] Genesis Institute of Genetic Research, Genesis Healthcare Corporation, Tokyo, Japan.

[2] International Business Division, Genesis Healthcare, Tokyo, Japan.

Genesis Institute of Genetics Research

TABLE OF CONTENTS

INTRODUCTION

It has been known from public health finding that age is a key factor in determining if a person will develop a severe COVID-19 symptom. However, there are still many other factors that contribute to the severity of the disease progression. In a cohort study conducted by a group of researchers from the UK, electronics health record (EHR) of more than 17 million adults from England were pseudonymously linked to 10,926 COVID-19-related deaths from UK Office of National Statistics (ONS) data and then analyzed for relative importance of all possible risk factors, ranging from basic demographic profile like age, gender, BMI, to pre-existing medical conditions, i.e., comorbidities with COVID-19, ethnicity, and socioeconomic [1], hereinafter referred to as base UK study.

The results give a comprehensive insight into a quantifiable degree of effects for each individual factor contributing to the severe outcome of COVID-19 and is being utilized by The Association of Local Authority Medical Advisors (ALAMA), an Occupational Health physicians providing services to Local Authorities forum in UK, to issue a health recommendation for workforce based on their meta-analysis built principally on top of quantified relative risks in the base UK study and related literatures [2].

Research Problem

Though the results from both of the studies can be used as a general reference for relative risk between each factors and to make a general assessment for COVID-19 severity risk in UK, but the studies cannot be applied directly to other populations or countries which may be influenced by many domestic factors like public health policy, population norms of wearing Personal Protective Equipment (PPE) like face mask, and even cultural aspect like the level of exposure via different way of greeting.

Therefore, in order to apply the findings to a different population or country it is important that the study be replicated locally or adjusted to reflect relative differences in the risk level. Unlike the base UK study [1] where NHS England is the host in developing a secured privacy-protected system to aggregate medical history from more than ten million of patient's health records, the situation of accessing anonymized health record for public health research in other countries is still relatively limited which prevents replication of the same approach. Therefore, an alternative method must be explored.

METHODOLOGY

Cox Proportional Hazard Regression

To overcome the described limitation, in our study, we collected England's public death data, including COVID-19-related death, from UK Office for National Statistics (ONS) [3] as well as Japan's National Statistics Center (e-Stat) [4] and Japan's Ministry of Health, Labour and Welfare [5] from the same 90-day study period starting from February to May 2020 applied an identical analysis method, Cox proportional hazards regression, as in [1]. The equation of Cox proportional hazards regression is shown in Figure 1.

$$h(t|x) = b_0(t) \cdot \exp\left(\sum_{i=1}^{n} b_i(x_i - \bar{x}_i)\right)$$

Figure 1. The mathematical model of Cox proportional hazard regression. *h* is the hazard function given *x* as a covariate from a set of range *i* to *n*. The effect of each covariate, i.e. partial hazard, to baseline hazard b_0 is defined as an exponential function of each variable combined whereas its coefficients are estimated by regression method.

By using only a single binary variable, death data in Japan, as the covariate, the model estimates a relative hazard between Japan and UK, which as a whole implies the relative differences in risks between the two countries from the specified periods. Hence, can be used as a delta value for overall adjustment while maintaining relative risk between all other factors being identified in the base UK study. Cox proportional hazard model also taking into account the samples with unknown death period, a.k.a. censor data. In the same settings with UK's study, we regard all the people who still alive in the cohort after the study period and those who died during the study period due to non-COVID-19-related causes as censor data.

Since our goal is to find a relative difference between base UK study and Japan population, we limited the number in UK's population to exact number of samples as in original study to closely mimic the hazard level. However, based on open national statistics data we include all reported COVID-19 related deaths without the limitation of linked patient health records posed in original study. The properties of the cohort are described in Table 1.

	UK	**Japan**
Number of samples (N)	17,278,392	17,229,036
COVID-19-related deaths	36,842	17,920
Non-COVID-19-related deaths	152,313	344,075
Collection Period	Feb 1 – May 6, 2020	Feb 14 – May 19, 2020

Table 1. Summary of population cohort settings for survival analysis.

Key Results #1: Survival Probabilities and Estimate Hazard Ratio Between UK and Japan

After data pre-processing and formatting to above definitions, the analysis results are as shown below. First, as a result of independent Kaplan-Meier fitting, Figure 2 shows a comparison of cumulative hazard where Japan (JA) clearly shows a slower rising trend where UK shows a sharp rise around 60 days of study period.

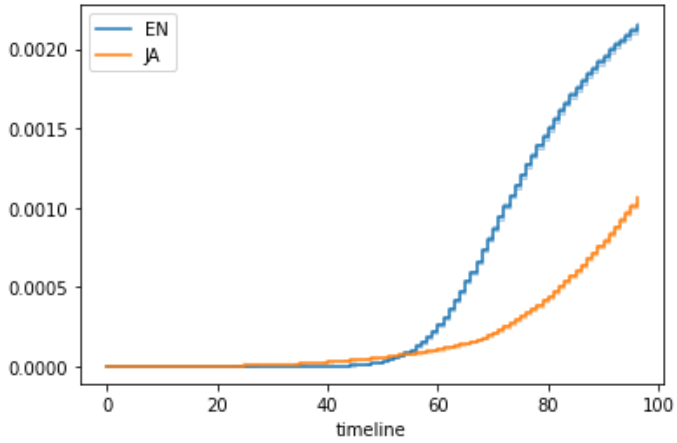


Figure 2. Cumulative Hazard UK vs. Japan populations.

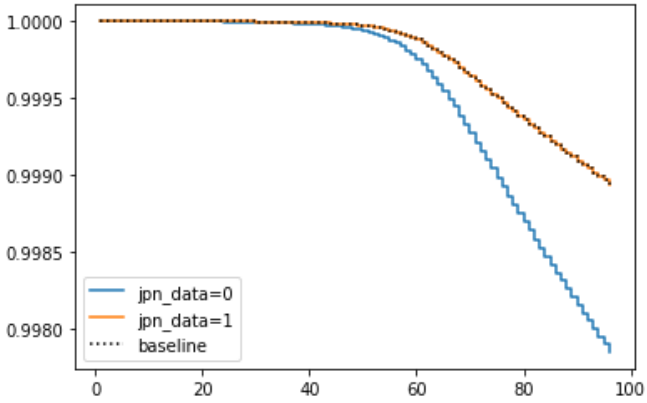
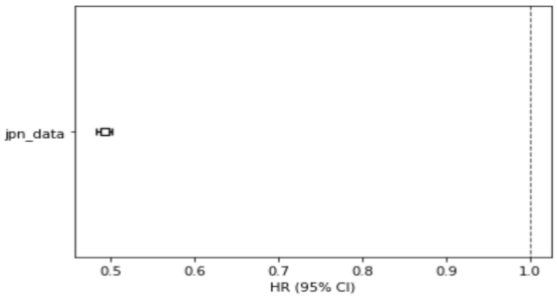


Figure 3. Kaplan-Meier plot of UK vs. Japan populations. It shows cumulative probability of survival from survival function.

Inversely, in Figure 3, the cumulative probability of survival of UK starts to decrease with steeper slope when compared to Japan as a baseline.

model	lifelines.CoxPHFitter
duration col	'duration'
event col	'event'
weights col	'weight'
baseline estimation	breslow
number of observations	3.45074e+07
number of events observed	54762
partial log-likelihood	-946578.13
time fit was run	2021-01-02 15:05:36 UTC



	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	z	p	-log2(p)
jpn_data	-0.71	0.49	0.01	-0.73	-0.69	0.48	0.50	-77.87	<0.005	inf

Concordance	0.42
Partial AIC	1893158.26
log-likelihood ratio test	6458.39 on 1 df
-log2(p) of ll-ratio test	inf

Figure 4. Results from Cox proportional hazard regression between Japan and UK.

Subsequently, we combined the deaths data from both countries and define Japan as a covariate. The results are as shown in Figure 4 The hazard ratio is 0.49 as depicted in column *exp(coef)* of the resulting table with visualized plot on the top right showing narrow confidence interval between 0.48 and 0.50 and P-value of 0.005. The derived hazard ratio has the same meaning as how much risk an individual in Japan's population would have if compared against England's population. In other word, what would be a risk of a person in Japan if it is to be measured by England's hazard baseline.

Individual Risk Calculation and Classification

To make a risk calculation based on hazard ratios in original study and applying adjustment for Japanese population, we adopt the estimated hazard ratios from ALAMA association which summarized relative risks of all medical conditions being reported from different literatures [6]. The association first requested the researcher groups from the base UK study to derive the age-stratified associations for all covariates to analyze possible interaction between age and each covariate [7].

This resulted in hazard ratio breakdown of each covariate into each age group to which the association then estimated the hazard ratio for each single year of age. This age-specific hazard ratio is serving as the main reference for the study in the subsequent meta-analysis and manual adjustment of relative risk level.

$$HR = 1.1084^{age}$$

$$\ln(HR) = age \cdot \ln(1.1084)$$

$$\frac{\ln(HR)}{\ln(1.1084)} = age\_hr(HR) = age \quad (1)$$

In their study, the estimated hazard ratio for each comorbidity will be converted to age using a formula which describes a relationship between hazard ratio (*HR*) and age as a continuous covariate as shown in (1). The equation is presumably estimated from exponential fitting between *HR* and *age*, resulting in the exponential base of 1.1084. The sum of converted age from all applicable factors with the person's true age is defined as Covid-age, or the equivalent level of vulnerability to a healthy white man in UK of the same age. Infectious Fatality Rates (IFR) as per Covid-age group are estimated by many studies based on the death statistics data. The vulnerability level classification has also been defined [8] based on the age-group guideline from Public Health England (PHE) which then translated into Covid-age by looking at its distribution and follow the same ratio of percentile as implied by PHE guideline.

The calculation of an individual hazard score in our study is based on estimated hazard ratios from [6] where hazard ratio from applicable factors are multiplied to created cumulative hazard ratio according to Cox hazard regression model where partial hazard is an exponential function of all covariates added sum. The cumulative hazard is then adjusted by relative hazard ratio between UK and Japan from our Cox regression analysis using the same approach as other covariates. In contrast with [2], we have decided not to use converted Covid-age as a unit for calculation because it represents an age of healthy white male population in UK which might mislead the interpretation in our context. Also, in their process of converting a hazard ratio to age, the age's decimal digits are rounded to an integer, making a slightly decrease in resolution when adding multiple risk factors. Instead, we converted Covid-age from their risk level classification table to hazard ratio and then apply natural logarithm to normalize and define it as COVID-19 Health Hazard Score, in short hazard score.

$$Covid19\ Health\ Hazard\ Score = \ln(Cumulative\ Hazard \times 0.49 \times True\ Age\ Hazard) \quad (2)$$

To calculate a person's hazard score, the person true age is converted into hazard ratio and multiplied with cumulative hazard ratios from all risk factors and 0.49 constant for Japan's population adjustment as described in the equation (2). As a result, the risk level classification table from [8] has been updated as shown in Table 2.

Risk Level	Health Hazard Score	Equivalent Covid-age
Very High	8.75 and above	85 and above
High	[7.2,8.75]	70 - 85
Moderate	[5.15,7.2)	50 - 70
Low	[0,5.15)	Below 50

Table 2. Risk level classification based on Health Hazard Score

Key Results #2: COVID-19 Hazard Visualization and Reference Critical Hazard

As part of our company's commitment in contribution to public health and the society, we have developed and provide an online risk assessment tool with no cost. It is designed to be easy to understand yet having solid scientific research evidences and transparent risk calculation. As one of the key features, we have normalized accumulated hazard risk in different categories into a scale comprehensible in a form of radar chart (Figure 5) where we provide a reference critical hazard based on the criteria defined in Table 3.

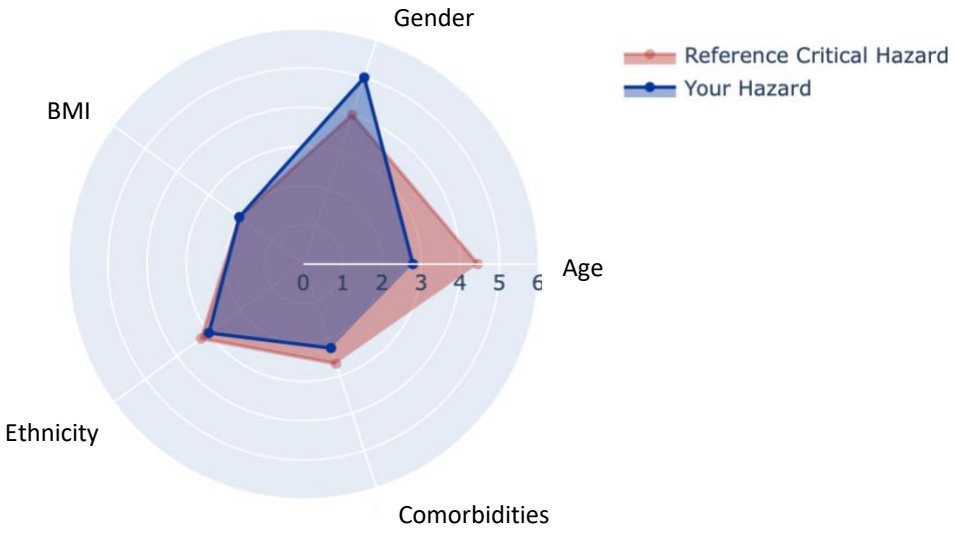


Figure 5. Radar chart depicting relative risk in each category with reference critical hazard dynamically changed according to user's age.

Risk Factor Category	Normalization Formula, Rationale, and Critical Hazard Definition
Age	$$Normalized\ Age\ HR = \ln\left(1 + \left(\lfloor e^6 \rfloor - 1\right) \cdot \left(\frac{(age\_hr - AGE\_HR\_MIN)}{(AGE\_HR\_MAX - AGE\_HR\_MIN)}\right)\right)$$ where: $$AGE\_HR\_MIN = 7.833039962098921$$ $$AGE\_HR\_MAX = 2250.2915183152363$$ $$AGE\_HR\_MIN \leq age\_hr \leq AGE\_HR\_MAX$$ Critical hazard value is defined to be equivalent to people at age 60. Since hazard ratio for age increase exponentially, we take natural log after scaling to derive the range of [1,6]. Minimum and maximum hazard ratio is converted from the equivalent age of 20 and 75 respectively using Equation (1). $$Critical\ Age\ HR = age\_hr(60) = 480$$

Gender	$$Normalized\ Gender\ HR\ =\ 3\ +\ 2\ \cdot\ \frac{(gender\_hr\ -\ GENDER\_HR\_MIN)}{(GENDER\_HR\_MAX\ -\ GENDER\_HR\_MIN)}$$ where: $$GENDER\_HR\_MIN\ =\ 0.6$$ $$GENDER\_HR\_MAX\ =\ 1$$ $$GENDER\_HR\_MIN\ \leq\ gender\_hr\ \leq\ GENDER\_HR\_MAX$$ The relative risk difference between gender is 0.6 when compare female to male as the reference (*HR*=1), we normalize this relationship to a scale between [3, 5]. The critical hazard value is derived from average gender hazard ratio. $$Critical\ Gender\ HR\ =\ (0.6\ +\ 1.0)\ \div\ 2.0\ =\ 0.8$$
BMI	$$Normalized\ BMI\ HR\ =\ 1\ +\ 5\ \cdot\ \frac{\ln(bmi\_hr)\ -\ BMI\_HR\_MIN}{\ln(BMI\_HR\_MAX)\ -\ \ln(BMI\_HR\_MIN)}$$ where: $$BMI\_HR\_MIN\ =\ 1$$ $$BMI\_HR\_MAX\ =\ 13$$ $$BMI\_HR\_MIN\ \leq\ bmi\_hr\ \leq\ BMI\_HR\_MAX$$ Since the hazard ratio for BMI is stratified into different level (BMI=[30,35), [35,40), >=40) rather than continuous value, we first take natural logarithm to emphasize the effect hazard increase in lower range then scale it to [1,6]. The critical hazard value is derived from hazard ratio for obesity class I, BMI=[30-35) Kg/m2, for each age. $$Critical\ BMI\ HR\ =\ bmi\_hr(age,\ 'obesity\ class\ I')$$
Ethnicity	$$Normalized\ Ethnic\ HR\ =\ 1\ +\ 4\ \cdot\ \left(\frac{ethnic\_hr\ -\ ETHNIC\_HR\_MIN}{ETHNIC\_HR\_MAX\ -\ ETHNIC\_HR\_MIN}\right)$$ where: $$ETHNIC\_HR\_MIN\ =\ 1.0$$ $$ETHNIC\_HR\_MAX\ =\ 2.0$$ $$ETHNIC\_HR\_MIN\ \leq\ ethnic\_hr\ \leq\ ETHNIC\_HR\_MAX$$ Ethnicity hazard ratio is scaled to range between [1,5] and the critical hazard is defined as an average of all ethnicity's hazard ratio. $$Critical\ Ethnic\ HR = ave(ethnicity\_hrs) = \frac{\sum_{i=1}^{N} ethnicity\_hr_i}{N}\ =\ 1.56$$
Comorbidities	$$Normalized\ Comorbid\ HR\ =\ \ln(comorbid\_hr)$$

	where: $$0 \leq comorbid\_hr \leq \lfloor e^6 \rfloor$$ Since there is no theoretically upper bound to cumulative hazard ratio for comorbidities, i.e. the more conditions you one has the higher the value, we decide to limit the maximum value to that equivalent of Euler's constant to the power of 6 to fix the range after natural logarithm to [0,6]. The critical hazard value is defined as a ratio between hazard risk equivalent to a person of age 70 divided by their true age's hazard ratio: $$Critical\ Comorbid\ HR = \frac{age\_hr(70)}{age\_hr(true\_age)}$$

Table 3. Definition, rationale, and formula of normalization and reference critical hazard value for each risk factor category

The reference critical hazard value for BMI and comorbidities categories change dynamically according to age since it's based on the same hazard ratios table used in risk score calculation which incorporates age-interaction into the model. The critical hazard value for factors interacting with age like BMI and comorbidities should be interpreted as a relative measure to tell which factor is more important to be the cause for COVID-19 death when compared with age. For an example, a person at age 60 has comorbidities' critical hazard value of 1.03 compared to 4.12 for a person at age 30. This doesn't mean that comorbidities in people at age 60 is any less significant but their age is relatively more importance as a determining factor for COVID-19 mortality.

Discussion

With constraint of individual health data access, the risk assessment based on the model can only tell the relative risk of a person when compared against England's population assuming that no other variables will affect the outcome of certain factors in receiving medical treatments. This may not be the case if there are significant discrepancies in the standard of public health between countries, e.g., an under-resourced country may have a higher rate of death from a certain pre-existing condition due to lack of medical equipment or treatment required. In which case, the hazard ratio of such factor will be deviated from UK's study if the lacking factor is not included in the analysis model for adjustment. Not to mention the physiological differences in different ethnicities which is underrepresented in the base study. Also, the resolution of the death data is not without flaws. Except for daily COVID-19 deaths, average allocation of daily deaths number has to be made for Japan due to lack of daily breakdown in data from national statistics.

Nevertheless, the minorly pre-processed data shouldn't invalidate the overall risk information in this analysis, which aim at understanding relative risk differences between population over the study period. In addition, we believe that the risk for the majority of the medical conditions is generally applicable to all races and ethnicities, if given an equal standard of healthcare access, and is sufficient to be used to assess relative risk level among people from the same population with different demographics like age, gender, and comorbidities.

CONCLUSION

Based on the concrete research evidences in epidemiology modeled upon real patients and clinical data, we developed COVID-19 Health Hazard Score localized for Japan by using Japan's and UK's population death statistics of COVID-19 and other causes to calculate hazard ratio between the two populations for risk level adjustment. As part of our effort to contribute to the public health, we have implemented and provided an easy-to-use tool accessible by anyone to assess their risk of mortality should they become infected with COVID-19. The result includes visualization of both the hazard score and normalized hazard level for different factors with critical hazard level for an individual to make a reference. The explanation of the risk level come with general and workplace related advices considering Japanese society context which may be used as a starting point to consult with an employer or a clinician for a personal preventive measure. The COVID-19 Health Hazard Score tool is available at: https://hazardscore.genesis-healthcare.jp/covid19

REFERENCES

[1] E. J. Williamson *et al.*, "Factors associated with COVID-19-related death using OpenSAFELY," *Nature*, vol. 584, no. 7821, pp. 430–436, Aug. 2020, doi: 10.1038/s41586-020-2521-4.

[2] D. Coggon, P. Croft, P. Cullinan, and A. Williams, "Assessment of workers' personal vulnerability to covid-19 using 'covid-age,'" 2020, Accessed: Mar. 26, 2021. [Online]. Available: https://academic.oup.com/occmed/article-abstract/70/7/461/5881715.

[3] "Coronavirus (COVID-19) - Office for National Statistics." https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases (accessed Mar. 26, 2021).

[4] "政府統計の総合窓口." https://www.e-stat.go.jp/ (accessed Mar. 26, 2021).

[5] "新型コロナウイルス感染症について｜厚生労働省." https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000164708_00001.html (accessed Mar. 26, 2021).

[6] D. Coggon, P. Croft, P. Cullinan, and A. Williams, "ASSESSMENT OF PERSONAL VULNERABILITY TO COVID-19 Sources of evidence and methods leading to adopted risk estimates UPDATED 11 DECEMBER 2020." Accessed: Mar. 26, 2021. [Online]. Available: https://alama.org.uk/wp-content/uploads/2020/12/Methods-at-201211.pdf.

[7] "Post-publications supplement toWilliamson, E.J., Walker, A.J., Bhaskaran, K. etal. OpenSAFELY: factors associated with COVID-19 death in 17 millionpatients. - Google Search." https://www.google.com/search?client=firefox-b-d&q=Post-publications+supplement+toWilliamson%2C+E.J.%2C+Walker%2C+A.J.%2C+Bhaskaran%2C+K.+etal.+OpenSAFELY%3A+factors+associated+with+COVID-19+death+in+17+millionpatients. (accessed Mar. 26, 2021).

[8] "Covid-19 Medical Risk Assessment – Alama." https://alama.org.uk/covid-19-medical-risk-assessment/ (accessed Mar. 26, 2021).