



Change point detection for COVID-19 excess deaths in Belgium

Han Lin Shang^{1,2} · Ruofan Xu²

Accepted: 18 February 2021 / Published online: 6 March 2021

© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

Abstract

Emerging at the end of 2019, COVID-19 has become a public health threat to people worldwide. Apart from deaths with a positive COVID-19 test, many others have died from causes indirectly related to COVID-19. Therefore, the COVID-19 confirmed deaths underestimate the influence of the pandemic on society; instead, the measure of ‘excess deaths’ is a more objective and comparable way to assess the scale of the epidemic and formulate lessons. One common practical issue in analysing the impact of COVID-19 is to determine the ‘pre-COVID-19’ period and the ‘post-COVID-19’ period. We apply a change point detection method to identify any change points using excess deaths in Belgium.

Keywords Multivariate time series · Structural change · Mortality

Introduction

Emerging at the end of 2019, COVID-19 has become a public health threat to people worldwide. According to statistics compiled by Johns Hopkins University (COVID-19 Dashboard, 2020), up until early June 2020, the COVID-19 pandemic claimed nearly 400,000 lives globally. With a population of just 11.5 million, Belgium had the highest death to population ratio, with 78 deaths per 100,000 people (Coronavirus Resource Center, Johns Hopkins University 2020) at that time. This figure includes only deaths directly linked to COVID-19 and is likely inaccurate due to different reporting practices or people dying of COVID-19 related causes without being tested. The physical, psychological, and social effects of lockdown and economic changes stemming from COVID-19 have indirectly caused deaths, which are

✉ Han Lin Shang
hanlin.shang@mq.edu.au

¹ Department of Actuarial Studies and Business Analytics, Macquarie University, Sydney, NSW 2109, Australia

² Research School of Finance, Actuarial Studies and Statistics, Australian National University, Canberra, Australia

not counted in official figures. Therefore, the deaths directly linked to COVID-19 underestimate the impact of the pandemic on society.

One way to explore the pandemic's actual mortality effect is to look at the number of 'excess deaths'. This measure can be constructed by comparing the observed weekly deaths throughout 2020 to values from the previous non-pandemic period. This provides a more objective and comparable way to assess the scale of the pandemic and formulate lessons. Attempts have been made to continuously track and examine comparative excess mortality data for Europe, the United Kingdom, the United States of America, and other countries. Simple visualizations and discussions can be found in *The Financial Times* (2020), *The Economist* (2020) etc. Researchers have recently developed different approaches to understand the dynamics of the pandemic. In the field of demography, researchers have focused on the impact of COVID-19 on different demographic groups, such as age, gender, and income level. For example, Dowd et al. (2020) examine the role of age structure in deaths, showing that countries with a much older population have a dramatically higher mortality rate than those with a much younger population. Cairns et al. (2020) analysed the impact of COVID-19 on future higher-age mortality. Riley et al. (2020) assessed the potential impact of the COVID-19 pandemic on sexual and reproductive health in low-and middle-income countries. Banerjee et al. (2020) estimated the one-year excess mortality associated with the COVID-19 pandemic. One common practical issue in analysing the impact of COVID-19 is to classify the 'pre-COVID-19' period and the 'post-COVID-19' period. Some researchers choose the date when the first death was reported (see, e.g., Vandenborgh, 2020), while the others choose the date when the number of fatalities began increasing steeply as the change point (see, e.g., Barbero, 2020). The former choice may underestimate the magnitude of any effect on non-COVID-19 deaths, while the latter choice is likely to exaggerate any findings. Hence, it is necessary to propose a statistical method for determining the breakpoint.

As age group highly influences the COVID-19 mortality (Dowd et al., 2020), a natural question to ask is whether the phase changes also exhibit age patterns? If this is the case, practitioners should classify the 'pre-COVID-19' and post-COVID-19' period for each age group independently when analysing COVID-19 impact differences between age groups. In addition, the phase changes differences among age groups may aid in establishing more targeted policy and public guidance towards different age groups in different pandemic periods.

To this end, a change point detection method for time series may provide a statistically reliable result for phase change detection of excess deaths during the COVID-19 pandemic. In the literature, change point analysis either deals with the mean change (Bai & Perron, 2003), variance change (Hawkins & Zamba, 2005), or the distributional change (Matteson & James, 2014) within time-ordered observations. The number of change points is usually assumed to be single or a known number. However, as the COVID-19 pandemic may change the mean or variance of the underlying data structure, we seek to detect distributional changes. Using the hierarchical divisive estimation approach proposed by Matteson and James (2014), any distributional changes within an independent time-ordered sequence can be identified simultaneously without assuming the exact number of change points.

This paper's contribution is to provide a statistical change point detection method to identify the change points in the counts of COVID-19 caused excess death. The rest of the paper is organized as follows. In Sect. 2.1, we introduce the dataset and provide a brief analysis of the excess deaths in Belgium. The change point detection method is provided in Sect. 2.2, followed by a discussion of the change point detection results for the excess deaths in Belgium in Sect. 3. We examine if the phase changes exhibit age differences —finally, we summarize our research findings in Sect. 4.

Methodology

Dataset overview

To examine the excess death counts in Belgium, we extract the data from the Human Mortality Database (2020). The original data contains the weekly death counts in Belgium from January 1st, 2010 to May 16th, 2020. To obtain a stationary time series, we only use the past five-year data. In summary, we use the weekly data from week 1 in 2015 to week 20 in 2020, numbering 279 weeks in total. The data are recorded by gender (male and female) and age group (ages 0–14, 15–64, 65–74, 75–84, 85 years and above).

Figure 1 presents the weekly total death counts from 2015 to the present plotted against the week of the year. There is a clear annual seasonal pattern due to temperature and other weather-related effects (see also Healy, 2003). The graph shows the COVID-19 pandemic effect in 2020, and it also indicates increased mortality rates at the start of 2015 and 2018 and the first half of March 2018. This increase is probably due to the flu epidemic.

To minimize the impact of large random fluctuation, we use the weekly median death counts in the past five years to capture the pattern of 'pre-COVID-19' death counts. The excess death counts are computed as the difference between the actual weekly death counts and the weekly median death counts (Roser et al., 2020). Then, to adjust for the subpopulation size, the excess death rate is computed as the excess death counts divided by the median death counts from 2015 to 2019.

Fig. 1 The weekly total death counts in Belgium

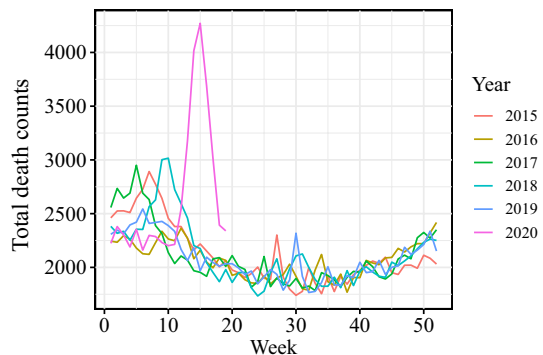


Fig. 2 The weekly total excess death rate in Belgium

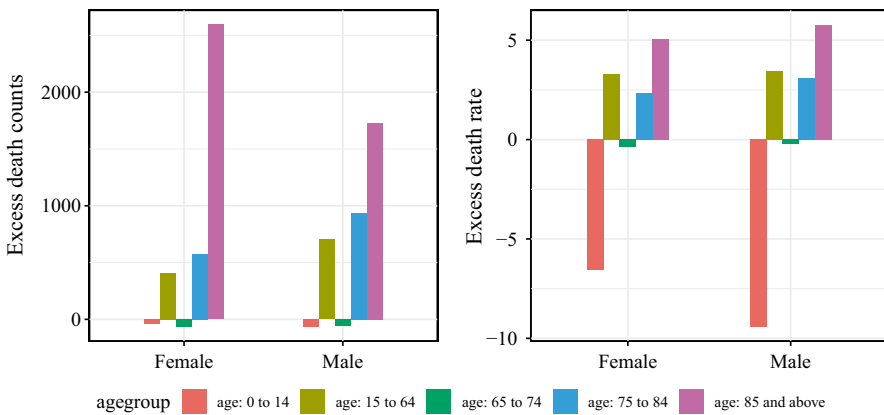
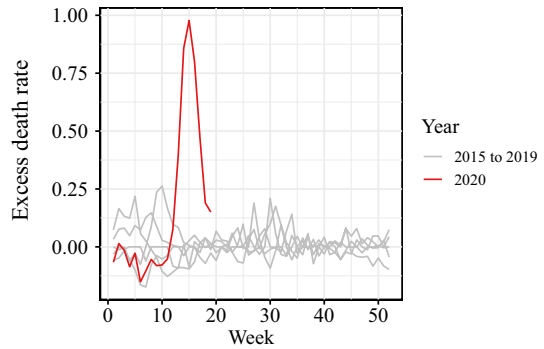


Fig. 3 The total excess death counts and rate in 2020, grouped by age, for female and male subpopulations in Belgium

Compared with the excess death counts, the rate is more suitable for comparing the impact across groups. Therefore, in the following analysis, we work with the excess death rate labelled by gender and age groups from 2015 to the present. The weekly total excess death rates are shown in Fig. 2, with the rates in 2020 highlighted in red. In early 2020 (before week 11), the weekly excess death rate is negative. We might not draw any conclusions based on this information as the deaths generally vary much at the beginning of a year. However, from week 13 (end of March) in 2020, the weekly excess death rate increased dramatically, and it continued to increase until week 15. Based on the graphical interpretation, we have gained insight into the critical time points in the COVID-19 pandemic. In the following sections, we utilise statistical methods to identify the change points that mark the different phases in the COVID-19 pandemic.

The bar charts in Fig. 3 show a different level of the negative impact of COVID-19 for different subpopulations in Belgium. Both genders show similar excess death patterns.

Surprisingly, the total excess death rate in 2020 for ages between 65 and 74 is negative. However, when examining excess deaths by week in Sect. 3, we notice that the excess deaths for those aged between 65 and 74 after week 13 are positive, which suggests that the elderly (age above 64) are more susceptible to COVID-19. In particular, those aged above 85 are more vulnerable to COVID-19 than other age groups. Meanwhile, COVID-19 has little impact on excess deaths amongst children and teenagers, as they travel and go out less. By contrast, COVID-19 is a much greater threat to older people, as they generally suffer from the physiological changes associated with ageing, decreased immune function, and multimorbidity. Additionally, the gender difference in excess death counts has different patterns among age groups. For age groups below 85, males generally have a higher excess death counts, while there are more excess deaths for females than males for those aged over 85. This is because the female subpopulation aged above 85 is larger than that of males. The median death count for females above 85 is also higher than males. Therefore, after we control for median death counts, males have higher death rates than females for all age groups. This phenomenon is consistent with the findings in many existing empirical studies that males tend to have more severe infection outcomes (e.g., higher mortality rate) in all age groups (see, e.g., Haitao et al., 2020; Bhopal & Bhopal, 2020).

Change point detection

We consider a hierarchical divisive estimation approach (Szekely & Rizzo, 2005) to detect single or multiple change points in the excess death counts. We adopt this approach because it can detect a distribution change without assuming any specific distribution structure. Therefore, it is more flexible than Bai and Perron (2003) and Hawkins and Zamba (2005), where only the mean and variance change can be detected.

Let $\{Z_t, t = 1, \dots, T\}$ be an independent sequence of time-ordered observations (either univariate or multivariate). Given a change point t_* , the observed data is partitioned into two clusters $A_{t_*} = \{Z_t, t = 1, \dots, t_*\}$ and $B_{T-t_*} = \{Z_t, t = t_* + 1, \dots, T\}$. Both A_{t_*} and B_{T-t_*} are independent and identically distributed (i.i.d.) samples from the distribution $A, B \in \mathbb{R}^d$ respectively, such that $\mathbb{E}|A|^\alpha, \mathbb{E}|B|^\alpha$ for some $\alpha \in (0, 2)$. We compute the generalized energy distance between A and B as follows:

$$\hat{e}(A_{t_*}, B_{T-t_*}; \alpha) = \frac{2}{t_*(T-t_*)} \sum_{i=1}^{t_*} \sum_{j=t_*+1}^T |Z_i - Z_j|^\alpha - \frac{1}{\binom{t_*}{2}} \sum_{i=2}^{t_*} \sum_{j=1}^{i-1} |Z_i - Z_j|^\alpha - \frac{1}{\binom{T-t_*}{2}} \sum_{i=t_*+2}^T \sum_{j=t_*+1}^{i-1} |Z_i - Z_j|^\alpha,$$

where $|\cdot|$ is the Euclidean norm. The first term on the right-hand side of (1) corresponds to the between-distance measure of A_{t_*} and B_{T-t_*} . The second and third terms on the right-hand side of (1) correspond to the within-distance of A_{t_*} and B_{T-t_*} , respectively. To incorporate different sample sizes, Szekely and Rizzo (2005) proposed a Q statistic defined as

$$\hat{Q}(A_{t^*}, B_{T-t^*}; \alpha) = \frac{t^*(T-t^*)}{T} \hat{\varepsilon}(A_{t^*}, B_{T-t^*}; \alpha).$$

A change point location t^* is determined as the value that maximizes $\hat{Q}(A_{t^*}, B_{T-t^*}; \alpha)$. For detecting a difference in mean, α is set to 2.

To estimate multiple breakpoints, we iteratively apply the above techniques as follows.

Suppose that $k-1$ change points have been estimated at time $0 = \hat{t}_0 < \hat{t}_1 < \dots < \hat{t}_k = T$. These change points can partition the observations into k clusters $(\hat{C}_1, \dots, \hat{C}_k)$ such that $\hat{C}_i = \{Z_{\hat{t}_{i-1}+1}, \dots, Z_{\hat{t}_i}\}$. Given these clusters, we then apply the procedure for finding a single change point to the observations within each of the k clusters, so that we obtain k candidate change points. Then we choose the one resulting in the largest Q statistics as the estimated k th change points. Then, we check if the selected k th change point is significant using a permutation hypothesis test. Details of the hypothesis test can be found in Szekely and Rizzo (2005). The iteration terminates if the k th estimated change point tested is insignificant. The implementation of this distance-based approach is included in the *ecp* package (James & Matteson, 2014) in R (R Core Team, 2020).

Results and discussion

Since our interest lies in the change point detection caused by COVID-19, we perform the change point detection on data from the latter half of 2019 until the most recent observation in 2020. Let $\{X_t, t = 1, \dots, T\}$ be the observed excess death rate starting from the 27th week of Year 2019, so that $T = 46$ and $X_t = (X_{t,i}^{(f)}, X_{t,i}^{(m)}, i, j = 1, \dots, 5)'$. The superscript 'm' and 'f' are used to distinguish the gender 'male' and 'female', and the subscript t denotes the week index and i, j denotes the five age groups (age 0—14, 15—64, 65—74, 75—84 and above (and including) 85). We first apply the change point detection to the matrix-valued time series $\{X_t, t = 1, \dots, T\}$.

In this way, we obtain identical change points for all age groups and genders. There is no change point detected in the latter half of 2019. The change points detected in 2020 are in week 13 and week 18 (corresponds to $\hat{t}^* = 39$ and 44). Then, the excess death rate in 2020 fall into three clusters (see Fig. 4f). Week 12 is the end of the 'pre-COVID-19' period, and starting from week 13, the surge of weekly excess rates reveals the outbreak of COVID-19 among the population. From week 18, the weekly excess rate drops to one-fourth of the amount in the peak week (week 15). However, it is still too early to tell if the excess death curve is about to be flat since the collected data at the time of this study are only available up to week 19 of 2020.

As different age groups have different risk exposures to COVID-19, the change points might also vary across age groups. Therefore, we perform the detection for each age group separately, and the clusters are partitioned accordingly (see Fig. 4a–e).

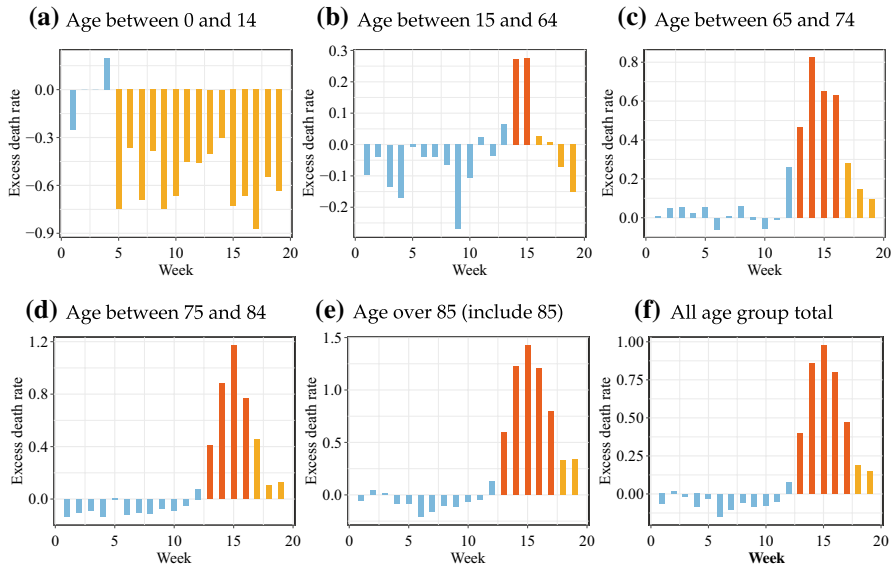


Fig. 4 The weekly excess death rate in different age groups in 2020, with clusters labelled by colours

We can see clear distinctions in the excess death distributions and the phase change points among the young population (age below 14), middle-aged population (age between 15 and 64), and the old population (age above 65). The young population's excess death rates are small, with negative numbers from week 5 (when the first COVID-19 case appeared in Belgium). One possible reason is that children have engaged in fewer outdoor activities and undertaken less international travel, making them less likely to contract the virus. Moreover, lock-down, and social distancing policies mean they are less likely to be involved in accidents (see also Lee et al. 2020). For people aged between 15 and 64, there is a steep increase in the excess death rate in weeks 14 and 15; soon after that, the excess deaths drop to around zero. However, for older people, the COVID-19 deaths surge started one week earlier and lasted longer than the middle-aged population. This might be because older people are more vulnerable to COVID-19 due to their weak immunity and prior/on-going complications/diseases. Moreover, the older the age group is, the more excess death counts appeared, and the longer the “COVID-19 outbreak” period lasted. Additionally, as over 60% of the excess deaths are amongst those over age 85, the detected change points in the whole population are the same as those detected using the population above (and including) 85 only.

Conclusion

We detect the COVID-19 pandemic phase in Belgium using a statistical change point detection technique. According to our analysis, the phase change shows age differentials. A surge in deaths in the elderly population appeared earlier than for the

middle-aged population and also lasted longer. Policymakers should consider this evidence and adjust their policies and focus guidance towards specific age groups. For example, lock-down policies may be first relaxed for younger age groups, while older people should remain extremely cautious.

References

- Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1), 1–22.
- Banerjee, A., Pasea, L., Harris, S., Gonzalez-Izquierdo, A., Torralbo, A., Shallcross, L., Noursadeghi, M., Pillay, D., Sebire, N., Holmes, C., et al. (2020). Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: A population-based cohort study. *The Lancet*, 395, 1715–1725.
- Barbero, C. A. (2020). A statistical forecast of LOW mortality and morbidity due to COVID-19, in Argentina and other Southern Hemisphere countries, Technical report, medRxiv. URL: <https://www.medrxiv.org/content/https://doi.org/10.1101/2020.04.20.20072488v1>
- Bhopal, S. S., & Bhopal, R. (2020). 'Sex differential in covid-19 mortality varies markedly by age', *Lancet*. England.
- Cairns, A. J., Blake, D., Kessler, A. R. & Kessler, M. (2020). The impact of COVID-19 on future higher-age mortality, Technical report, Pensions Institute. URL: <https://openaccess.city.ac.uk/id/eprint/24218>
- Coronavirus Resource Center, Johns Hopkins University (2020), Mortality Analyses. Accessed on June 5, 2020. URL: <https://coronavirus.jhu.edu/data/mortality>
- COVID-19 Dashboard (2020), The center for systems science and Engineering at Johns Hopkins University. Accessed on June 5, 2020. URL: <https://coronavirus.jhu.edu/map.html>
- Dowd, J. B., Andriano, L., Brazel, D. M., Rotondi, V., Block, P., Ding, X., Liu, Y., & Mills, M. C. (2020). Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proceedings of the National Academy of Sciences*, 117(18), 9696–9698.
- Haitao, T., Vermunt, J., Abeykoon, J., Ghamrawi, R., Gunaratne, M., Jayachandran, M., Narang, K., Parashuram, S., Suvakov, S., & Garovic, V. (2020). *Covid-19 and sex differences: Mechanisms and biomarkers*, in 'Mayo Clinic Proceedings.' Elsevier.
- Hawkins, D. M., & Zamba, K. (2005). A change-point model for a shift in variance. *Journal of Quality Technology*, 37(1), 21–31.
- Healy, J. D. (2003). Excess winter mortality in Europe: A cross country analysis identifying key risk factors. *Journal of Epidemiology & Community Health*, 57(10), 784–789.
- Human Mortality Database (2020), University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Retrieved May 27, 2020, from URL: <http://www.mortality.org>
- James, N. A. & Matteson, D. S. (2014). 'ecp: An R package for nonparametric multiple change point analysis of multivariate data', *Journal of Statistical Software* 62(7).
- Lee, P.-I., Hu, Y.-L., Chen, P.-Y., Huang, Y.-C., & Hsueh, P.-R. (2020). Are children less susceptible to COVID-19? *Journal of Microbiology, Immunology, and Infection*, 53(3), 371–372.
- Matteson, D. S., & James, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association: Theory and Methods*, 109(505), 334–345.
- R Core Team (2020), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>
- Riley, T., Sully, E., Ahmed, Z., & Biddlecom, A. (2020). Estimates of the potential impact of the COVID-19 pandemic on sexual and reproductive health in low-and middle-income countries. *International Perspectives on Sexual and Reproductive Health*, 46, 73–76.
- Roser, M., Hannah Ritchie, E. O. & Hasell, J. (2020), 'Coronavirus Pandemic (COVID-19)', Our World in Data. URL: <https://ourworldindata.org/excess-mortality-covid>
- Szekely, G. J., & Rizzo, M. L. (2005). Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method. *Journal of Classification*, 22(2), 151–173.

- The Economist (2020), Tracking covid-19 excess deaths across countries. Retrieved June 5, 2020, from **URL:** <https://www.economist.com/graphic-detail/2020/04/16/tracking-covid-19-excess-deaths-across-countries>
- The Financial Times (2020), Coronavirus tracked: the latest figures as countries fight to contain the pandemic. Retrieved June 5, 2020, from **URL:** <https://www.ft.com/content/a26bf7e-48f8-11ea-aeb3-955839e06441>
- Vandoros, S. (2020). 'Excess mortality during the Covid-19 pandemic: Early evidence from England and Wales.' *Social Science and Medicine*, 258, 113101.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.