Biometrics 64, 1–17 DOI: 10.1111/j.1541-0420.2005.00454.x
December 2008

Comparing COVID-19 incidences longitudinally per economic sector against the background of preventive measures and vaccination

Florian Stijven[1,2*], Johan Verbeeck[1,], and Geert Molenberghs[1,2,***]**

[1]Data Science Institute (DSI), Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Hasselt University, Hasselt, Belgium

[2]Leuven Biostatistics and Statistical Bioinformatics Centre, University of Leuven, Leuven, Belgium

*email: florian.stijven@kuleuven.be

**email: johan.verbeeck@uhasselt.be

***email: geert.molenberghs@uhasselt.be

Summary:

In the COVID-19 pandemic, workplace transmission plays an important role. For this type of transmission, the longitudinal 14-day incidence curve of SARS-CoV-2 infections per economic sector is a proxy. In Belgium, a census of confirmed 14-day incidences per NACE–BEL sector level three is available from September 2020 until June 2021, encompassing two waves of infections. However, these high-dimensional data, with a relatively small number of NACE-BEL sectors, are challenging to analyze. We propose a non-linear Gaussian-Gaussian model that combines parametric and semi-parametric elements to describe the incidence curves with a small set of meaningful parameters. These parameters are further analyzed with conventional statistical methods, such as canonical correlation analysis and linear models, to provide insight into predictive characteristics of the first wave for the second wave. Those non-linear models classify economic sectors into three groups: sectors with two regular waves of infections, sectors with only a first wave and sectors with a more irregular profile, which may indicate a clear effect of COVID-19 vaccination. The Gaussian-Gaussian model thus allows for analyzing and comparing incidence curves and to bring out key characteristics of such curves. Finally, we consider in which other settings the proposed approach could be applied, together with possible pitfalls.

Key words: COVID-19; Gaussian-Gaussian model; NACE-BEL sectors; Non-linear modelling; Vaccination.

This paper has been submitted for consideration for publication in *Biometrics*

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/biom.13766

1. Introduction

SARS-CoV-2, the virus causing COVID-19, was first detected in December 2019 in Wuhan (Zhu et al., 2020). The virus spread rapidly across the world, causing a pandemic. Initially, only non-pharmaceutical interventions (NPIs) such as social distancing, mask mandates and work sector specific preventive protocols were available as mitigation measures. Marinaccio et al. (2020) showed that during the first wave in Italy, about 20% of the reported COVID-19 cases were work-related, which underscores the importance of workplace related infections. The fraction of workplace related infections decreased to 3–4% in September–October 2020, possibly due in part to effective mitigation measures at the workplace (Marinaccio et al., 2021). Indeed, a recent meta-analysis showed that appropriate NPIs in the workplace effectively reduce workplace transmission. NPIs in the workplace are thus an important component of the interventions to mitigate the overall spread of SARS-CoV-2 (Ingram et al., 2021).

Although NPIs in the workplace have been extensively studied, most studies were conducted in sectors related to health care. Only five of 61 studies, in a meta-analysis of NPIs in the workplace, included non-health care related workplaces (Ingram et al., 2021). This indicates an underexposure of non-health care sectors in research. Moreover, occupational risk for COVID-19 has mostly been studied in subsets of the working population or in workplaces with a SARS-CoV-2 outbreak. Exceptions are the studies by Marinaccio et al., which examined the COVID-19 risk for the entire Italian working population *indirectly* by means of analyzing compensation claim applications for work-related COVID-19 (Marinaccio et al., 2020, 2021). To the authors' knowledge, the COVID-19 risk has not been *directly* studied for a country's entire working population except by Verbeeck et al. (2021) who used a subset of the data considered in this article.

Studying the spread of SARS-CoV-2 across a country's economic sectors can provide

valuable insights into the role of workplace related infections during the pandemic, and the effect of different types of interventions on the virus's spread in the workplace. Although every workplace and sector is unique in its own right, some sectors may behave similarly. We wish to identify such groups of sectors. In addition, we wish to study whether within an economic sector, the past wave's characteristics are predictive for characteristics of the next wave. However, forecasting future incidences per economic sector is outside the scope of the current investigation. To this extent, a census of all reported SARS-CoV-2 infections from September 2020 to June 2021 is analyzed longitudinally by economic sector.

The analysis of these high-dimensional data with multiple waves, including a relatively small number of economic sectors, is challenging. Moreover, grouping sectors based on their similarity in evolution of incidences requires the quantification of similarity between incidence profiles. Finally, in order to study whether the past wave is predictive for the characteristics of a sector's next wave, the sequence of incidence values that makes up a wave should be reduced to some key parameters that describe distinct aspects of this wave. We propose a non-linear Gaussian-Gaussian model to resolve all these difficulties by reducing each sector's incidence profile, and corresponding waves, to a small set of meaningful values. In these non-linear models, waves are described parametrically while periods in between waves are described semi-parametrically.

The article is organised as follows: the data and statistical methods are described in Section 2, the results are presented in Section 3. Finally, we end with a discussion of the results and statistical methods in Section 4.

2. Methods

2.1 *Data*

Daily confirmed SARS-CoV-2 infections registered by the Belgian health institute, Sciensano, are linked to their employer's main activity and aggregated by the National Social Security Office to weekly incidences (number of cases per 100,000) by NACE-BEL code, an economic sector classification system. The weekly incidences from 8 September 2020 to 5 July 2021 are mapped to 14-day incidences, as is common in epidemiology, by joining two consecutive weeks. During this period, two SARS-CoV-2 waves of infections occurred in Belgium, a large first wave from the end of September 2020 to the end of November and a somewhat smaller wave from March to May 2021. Although, some sectors did not experience the second wave. The NACE-BEL classification consists of five levels (STATBEL, 2017). Only sectors at the third level are analyzed, since this level contains the largest number of sectors with more than 10,000 employees, which is deemed to be the minimum requirement for sufficiently stable profiles.

2.2 *Gaussian-Gaussian Models*

Each sector's longitudinal profile of 14-day incidences, spanning 42 weeks, is modeled with a non-linear model. Because a sector's longitudinal profile constitutes a census of all reported infections, the model fitting can be seen as a data transformation. The sector's 42-dimensional response vector is transformed to a set of meaningful values, i.e., the non-linear model's parameter estimates. Two models are considered depending on whether the profile contains one or two waves. Longitudinal profiles with two regular waves are modeled with a Gaussian-Gaussian curve for each wave (Verbeeck et al., 2021) and with natural cubic splines with five equidistant knots between the waves, as shown in Equation (1). This model is further referred to as the *two-peak model* and contains 1 variance and 13 mean parameters. The modeled incidence for time point t in sector s is denoted by μ_{st}. The parameters of the Gaussian-

Gaussian curves have an appealing interpretation in terms of the wave's characteristics, as shown in Table 1 and Figure 1. The natural cubic splines part is denoted by $\widehat{f}_s(t)$. The number of knots is chosen such that the spline part is sufficiently flexible to describe the period between the waves accurately. However, this period is as such not of interest. Restrictions are imposed on the spline parameters to ensure a continuous and differentiable modeled profile (further details in Web Appendix A). Note that the second and third half Gaussian are assumed to have the same asymptotic height, δ_{s2}. The connection between the spline and Gaussian part is assumed to be at 10% of the peak height for the first connection, $\nu_{s1}+\sigma_{s2}\cdot\sqrt{\log(10)}$, and at 10% of the peak height minus three weeks, $\nu_{s2}-\sigma_{s3}\cdot\sqrt{\log(10)}-3$, for the second connection. These time points were based on a visual inspection of the model fit for varying choices. A sensitivity analysis (Web Appendix C) shows that different choices with respect to this connection and the number of knots may result in a slightly higher convergence rate, but those models provide a lesser-quality fit to the observed profiles. A prototypical longitudinal profile of the two-peak model is depicted in Figure 1.

$$\mu_{st}=\begin{cases}\delta_{s1}+\alpha_{s1}\cdot\exp\left\{-\frac{(t-\nu_{s1})^2}{\sigma_{s1}^2}\right\} & \text{if } t<\nu_{s1}\\ \delta_{s2}+\alpha_{s2}\cdot\exp\left\{-\frac{(t-\nu_{s1})^2}{\sigma_{s2}^2}\right\} & \text{if } \nu_{s1}\leqslant t<\nu_{s1}+\sigma_{s2}\cdot\sqrt{\log(10)}\\ \widehat{f}_s(t) & \text{if } \nu_{s1}+\sigma_{s2}\cdot\sqrt{\log(10)}\leqslant t<\nu_{s2}-\sigma_{s3}\cdot\sqrt{\log(10)}-3\\ \delta_{s2}+\alpha_{s3}\cdot\exp\left\{-\frac{(t-\nu_{s2})^2}{\sigma_{s3}^2}\right\} & \text{if } \nu_{s2}-\sigma_{s3}\cdot\sqrt{\log(10)}-3\leqslant t<\nu_{s2}\\ \alpha_{s4}\cdot\exp\left\{-\frac{(t-\nu_{s2})^2}{\sigma_{s4}^2}\right\} & \text{if } \nu_{s2}\leqslant t\end{cases} \quad (1)$$

The *one-peak model* is fitted for profiles with only one regular wave and contains 1 variance and 12 mean parameters. This model describes one wave with a Gaussian-Gaussian curve followed by natural cubic splines with six equidistant knots. The period modeled by splines in the one-peak model is longer than in the two-peak model, therefore more knots are used to model this longer period. As in the two-peak model, restrictions are imposed on the parameters to ensure a continuous and differentiable modeled profile. A more detailed description of both models can be found in the Web Appendix A.

In both the one and two-peak model, the first peak is modeled identically. Parameters describing the first peak can therefore be compared across these models. The period after the first peak is described differently, parameters describing the period after the first peak can thus not be compared across the two models.

[Table 1 about here.]

The model parameters are estimated by maximum likelihood via the Broyden–Fletcher–Goldfarb–Shanno (BFGS) minimisation algorithm (Nash, 1990) with the `optim` R-function. The deviation from the working model is assumed to be identically and independently normally distributed. The resulting minus log-likelihood is thus minimised. The choice for the starting values was partly data-driven. Note that the maximum likelihood framework is used here for curve fitting, and not for statistical inference. In essence, the likelihood is used as a measure of how well the working model fits the observed profile without claiming that the likelihood is correctly specified. Further details on the parameter estimation and the choices for the starting values can be found in Web Appendix B.

[Figure 1 about here.]

Each sector is classified into the two-peak model group or the one-peak model group based on two criteria. First, if the optimisation for the two-peak model does not converge properly, then the one-peak model is fitted. Lack of convergence is interpreted as the observed profile deviating from the "regular two-waves profile". Second, the complement of the coefficient of determination, $1 - R^2$, is computed for each converged model. The 95% percentile of these measures for the converged two-peak models is considered as the cut-off for a sufficient fit. For converged two-peak models with a larger value for $1 - R^2$ than this 95% percentile, the one-peak model is fitted. Finally, profiles for which both models do not converge and/or the $1 - R^2$ is larger than the cut-off, are classified into a third group: *other sectors*. These are profiles that do not adhere to the regular one or two peak pattern. Note that this does not

mean that the profile does not contain two (or one) wave(s), it merely indicates that the profile does not contain two (or one) *regular* waves(s).

An alternative method for the classification of sectors is multi-dimensional scaling (MDS) in which the profiles are treated as high-dimensional response vectors. The similarity between sector i and j is defined by the Pearson correlation between the response vectors: ρ_{ij}. Because ρ_{ij} is invariant to linear transformations, a high correlation does not necessarily imply similar absolute incidences. The distance between sector i and j is defined as $d_{ij} = 1 - \rho_{ij}$. The distances between all pairs of profiles constitute a distance matrix, from which a two-dimensional MDS-plot is constructed. In this plot, the distance between two points reflects how similar the corresponding profiles are. This method does not rely on a parametric model, as in the classification method based on the Gaussian-Gaussian models described above.

2.3 *Comparison of two waves*

Characteristics of the two waves are analyzed within the two-peak sectors with a canonical correlation analysis (CCA) and a linear model. The CCA studies the relationship between the first and second wave within a sector. It takes two separate linear combinations, one from each set of variables corresponding to the Gaussian-Gaussian curves of the first and second wave, respectively. The CCA then explores the correlation between these two linear combinations. Large values of the canonical correlation indicate that there is an important correlation between the first and second set of variables. A likelihood ratio test, with an α level of 0.05, investigates the null hypothesis of the canonical correlation between two sets of variables being equal to zero. Two CCA analyses are performed. The first CCA uses only parameters of the waves (widths, height and timing). The second CCA uses the mean age and the mean income of employees in the sector, in addition to the first wave parameters, in the first set of variables. The relation between the first and second wave is further explored by a linear model, where it is evaluated whether the height of the second peak can be predicted

from the shape parameters of the first wave and the mean age and income. The variables are selected by a forward selection algorithm with the selection probability equal to 0.05. Note that the prediction serves here as a means to identify sectors at risk, but not to forecast future waves.

3. Results

3.1 *Sector Grouping*

Of the 272 level three economic sectors, 94 have more than 10,000 employees, representing 4,110,028 of the total of 4,585,970 (90%) registered employees. Of these 94 sectors, 79 are classified as two-peak sectors, representing around 3,570,000 employees. Five sectors are classified as one-peak sectors. All five are healthcare (related) sectors, representing around 386,000 employees. The remaining 10 sectors are classified as other sectors, representing around 154,000 employees. The sectors in each of the three groups are listed in Web Tables 1-3 and the individual profiles for the other and one-peak sectors are plotted in Web Figures 1 and 2. Figures of the fitted model superimposed on the observed profile per sector are available in the Supplementary Information.

The mean incidence curves for the one and two-peak sectors are shown in Figure 2. Both groups differ on more aspects than merely the presence of one or two peaks. The height of the first peak is almost twice as large for the one-peak sectors than for the two-peak sectors. About five weeks after the first peak, the incidence reaches a plateau in the two-peak sectors. There is no plateau in the one-peak sectors. The incidence decreases more slowly after the first peak in the one-peak sectors, but continues to decrease to almost zero in June 2021. The incidence in the one-peak sectors is also consistently higher than in the two-peak sectors up to and including the week of March 9, 2021, and consistently lower afterwards. Using the Mann-Whitney U test, the difference in height of the first wave between the one and two-

peak sectors is 1776 cases per 100,000 employees (95% CI: $[1030, 2557]$). The widths of the increasing and decreasing phases do not differ between the one and two-peak sectors, -0.042 (95% CI: $[-0.192, 0.091]$) and 0.198 (95% CI: $[-0.097, 1.022]$), respectively.

[Figure 2 about here.]

The classification of sectors in the three groups is confirmed by the MDS-plot (Figure 3). The two MDS dimensions, rather than the full data, explain 61.4% of the similarity between the sectors. There is clear separation between the one and two-peak sectors. Two sectors from the group of other sectors are distant from all other sectors: 101 (Processing and preserving of meat and production of meat products) and 011 (Growing of non-perennial crops). These sectors are further discussed in detail in Section 3.3. The remaining other sectors are not well separated from the two-peak sectors.

[Figure 3 about here.]

3.2 *Comparing the two waves in the two-peak sectors*

3.2.1 *Canonical Correlation Analysis.* When contemplating only the widths, height, and timing of the peak of the first and second wave within a sector, only the first two canonical correlations are important: 0.669 (p <0.001) and 0.549 (p <0.001). The correlations between each set of variables and the first two corresponding canonical variates are listed in Table 2. The first canonical variate is strongly correlated with the height of the second peak and therefore has an attractive interpretation. Note that the width of the decreasing phase of the second wave is also moderately correlated with the corresponding first canonical variate, suggesting that the proposed latent variable interpretation may be too much a simplification. There is no obvious interpretation for the second canonical correlation.

[Table 2 about here.]

The CCA is repeated by adding mean income and mean age to the first set of variables.

Again, only the first two canonical variates are important, 0.707 (p <0.001) and 0.645 (p <0.001), respectively. The correlations between each set of variables and the first two corresponding canonical variates are listed in Table 3. The height of the second peak is the only variable that is correlated with the corresponding first canonical variate, suggesting a strong relationship between the first set of parameters and the second peak height. There is again no clear interpretation for the second canonical correlation.

[Table 3 about here.]

3.2.2 *Predicting Height of Second Peak.* The forward selection procedure identified the width of the decreasing phase, the height and timing of the first peak, and the mean income as predictors for the height of the second peak. An increase of 1000 cases per 100,000 employees in the peak incidence during the first wave results in an average increase of 57.0 (s.e. 23.4) cases per 100,000 employees in the height of the second peak. An increase of 1 in the width parameter of the decreasing phase of the first peak or an increase of 1 day in the time to the first peak results in an average increase of 222.6 (s.e. 49.9) respectively 38.1 (s.e. 9.8) cases per 100,000 employees in the height of the second peak. Finally, an increase of 10 euros in the mean day income results in a decrease of 9.4 (s.e. 3.2) cases per 100,000 employees in the height of the second peak. Only 39.57% of the variability in the height of the second peak can be explained by this linear model.

3.3 *Outlying Sectors*

The two sectors that displayed outlying behaviour in Figure 3 are examined in more detail here. The longitudinal profiles of these sectors are plotted in Figure 4 together with the average of all employees. Sector 011, *Growing of non-perennial crops*, contains about 13,500 employees and has multiple distinct features. First, the absolute incidence is very small during the entire period. Second, there are three distinct minor waves. The timing of the

first and third wave correspond to the timing of the other sectors' waves. The second wave occurred after the end of the year's holiday season. Sector 101, *Processing and preserving of meat and production of meat products*, contains about 18,000 employees. This sector has two waves, as most other sectors, but the second wave is larger in comparison with the first wave. Moreover, the second wave is also wider than the first wave.

[Figure 4 about here.]

The additional eight sectors in the "other sectors" group are generally conformable to the two waves, and are therefore mapped close to the two-peak sectors in the MDS-plot (Figure 3). However, they have a distorted second wave; those second waves are more sharply peaked or multi-modal (see Web Figure 1). This cannot be properly modeled by the proposed one and two-peak models. In fact, one could alternatively refer to these eight sectors as "irregular two-peak sectors".

4. Discussion and Conclusions

The data of interest consist of high-dimensional longitudinal responses (i.e., the longitudinal profiles). Analyzing these data is challenging because the sample size is relatively small (94 sectors). Even with a large sample size, deriving meaningful results from a direct analysis of the response vectors would be difficult. These issues are tackled by the Gaussian-Gaussian models. Each sector's longitudinal profile, which is *completely* observed, is described by these models with a small set of meaningful parameters. These parameters have a direct interpretation in terms of the height and timing of the peak, and the width of the increasing and decreasing phase of the wave. Let g_i and θ_i respectively be the longitudinal profile and Gaussian-Gaussian model parameter estimates for sector i. Because g_i is completely observed, there is no sampling variability in the parameter estimates, θ_i. We can thus see θ_i as a transformation of the original data g_i, and analyze θ_i as such with conventional methods

that do not require very large sample sizes. Given that the observed and fitted longitudinal profiles match closely, there is not much loss of information with this transformation.

The results revealed interesting patterns. Using the Gaussian-Gaussian models, economic sectors were classified into three groups based on the sectors' longitudinal profiles: one-peak, two-peak and other sectors. All one-peak sectors encompass residential care centres and healthcare staff in hospitals (NACE-BEL code: 861, 869, 871, 872, and 873), which were prioritised for vaccination from January 2021 onward. The height of the first peak is much larger in those sectors than in the two-peak sectors. This is expected as the exposure to SARS-CoV-2 is large and personnel is in close proximity with others. After the large first wave, however, no second wave occurred. Since February 2021, directly after the residential care and hospital staff, personnel from first line care and non-residential collective care institutions were vaccinated (Belgian Government, 2021). The corresponding sectors are classified as two-peak sectors (NACE-BEL code: 862, 881 and 879). Although these sectors are two-peak sectors, the second peaks are smaller than what would be expected from the linear model (in terms of the residuals). These findings could be attributed to the prioritised vaccinations of the personnel in the mentioned sectors. It is not possible to *causally* attribute these findings to the early vaccinations in the respective sectors. However, there is no other explanation that can reasonably explain (the size of) the observed trend in the one-peak sectors. Moreover, there is a trend related to how early each sector was vaccinated. The sectors that were vaccinated first showed no second peak. The sectors that were vaccinated next, showed a second wave, but that wave is smaller than expected.

The canonical correlation analysis indicates that characteristics of the first and second wave are related. More specifically, a strong correlation is found between a linear combination of the first wave's characteristics and the height of the second wave's peak. Evaluating this correlation further in a linear model, the width of the decreasing phase, the height, and the

timing of the first peak are predictive for the second peak's height. A more severe first wave (larger peak, slower decrease in cases, and later peak) is thus associated with a higher second peak's height. Important for policy makers, this result suggests that extra effort should go to sectors with previously higher reported incidences as those sectors are at increased risk during subsequent waves. Also, the mean income of a sector is predictive for the height of the second peak. A sector's mean income might be related to COVID-19 risk factors, such as the possibility of telework. As the linear model explains only a moderate proportion (39.57%) of the variability, inclusion of other variables may be relevant to explain the second peak's height, such as sector specific measures, the vaccination grade, a mask mandate and increased ventilation. However, some of those variables are hard to quantify or not readily available. Furthermore, the effectiveness of (combinations of) measures might also depend on the workplace context (Ingram et al., 2021).

Sector 011, *Growing of non-perennial crops*, and 101, *Processing and preserving of meat and production of meat products*, are outlying in the MDS-plot. The outlying behaviour of the former could be explained by the return of foreign workers from abroad after the year's holiday season. The latter sector is known to be distinct because this sector's environment facilitates viral spread which may lead to large outbreaks (Waltenburg, 2020; Middleton, Reintjes and Lopes, 2020).

The non-linear model was constructed pragmatically such that it describes the data well *and* that the parameters have a meaningful interpretation. The specifics of the non-linear model can be adapted as long as the former two conditions are satisfied. For example, loess can replace the splines, and other (bell-shaped) curves can replace the Gaussian-Gaussian curves if different parameters are needed for interpretation. This is useful towards application in other settings. The proposed Gaussian-Gaussian model is generally applicable in settings where the goal is to compare aspects of asymmetrical curves such as height and timing of

the peak, plateau phases, and speed of increase and decrease. Besides incidence, hospital admission and mortality due to COVID-19 (Molenberghs et al., 2021; Nishimoto and Inoue, 2020), the Gaussian-Gaussian models may be applicable in the seasonal comparison of asymmetric growth-curves of components of plants (Werker and Jaggard, 1997), quantification of tidal asymmetry (Guo et al., 2019), price and volume volatility (Liu and Chen, 2020), or studying hormonal homeostasis of circadian rhythms (Gnocchi and Bruscalupi, 2017). The data in each of those settings consist of (repeated) longitudinal asymmetric profiles with a similar general pattern. Our approach can be applied to reduce these profiles to a small set of meaningful parameters, which can be analyzed with conventional multivariate methods. This would provide the researchers with relatively easy to interpret results. The appropriateness of this approach depends to a large extent on the appropriateness of the non-linear model. Our approach is expected to perform poorly if the parameters have an ambiguous interpretation and/or the non-linear model fails to describe the profiles accurately. The latter may happen if there is no general pattern underlying all profiles.

For the purposes of this investigation, other modelling approaches are less appropriate. Mixed-effects models could be formulated, but this would require, in our case, 13 random effects in its most general form, which is not feasible in a (maximum) likelihood framework. In addition, a mixed model would "flatten out" possibly important differences between sectors. Another possible modelling approach is a differential equation-based model, as in Wang et al. (2020). Although such a model could provide an accurate description of our data, the parameters would have an interpretation at a different level: the level of disease spread dynamics. Contrary to this, the parameters of our models are merely descriptive and therefore do not rely on assumptions regarding the underlying disease spread.

In conclusion, we have analyzed sector-specific COVID-19 incidence profiles. Each sector's longitudinal profile was reduced to a small number of meaningful parameters by means of

fitting a carefully parameterised non-linear model. These parameters were then analyzed using conventional methods such as a canonical correlation analysis and a linear model. These methods revealed interesting patterns.

Acknowledgements

We thank the Belgian National Social Security Office and Sciensano for providing the data.

Data Availability Statement

The data that support the findings of this paper are available on request from the corresponding author. The data are not publicly available because the data are owned by a third party.

References

Belgian Government (2021). Coronavirus COVID-19 - Vaccination. Available at: `https://www.info-coronavirus.be/en/vaccination/#when-will-i-get-my-vaccine` Accessed on 8 November 2021

Carethers, J. M. (2021). Insights into disparities observed with COVID-19. *Journal of internal medicine* **289,** 463–473.

Gnocchi, D. and Bruscalupi, G. (2021). Circadian rhythms and hormonal homeostasis: pathophysiological implications. *Biology* **6,** 10.

Guo, L., Wang, Z.B., Townend, I., He, Q. (2019). Quantification of tidal asymmetry and its nonstationary variations. *Journal of Geophysical Research: Oceans* **124,** 773–787.

Ingram, C., Downey, V., Roe, M., Chen, Y., Archibald, M. Kallas, K. A., et al. (2021). COVID-19 prevention and control measures in workplace settings: a rapid review and meta-analysis. *International journal of environmental research and public health* **18,** 7847.

Liu, R., Chen, Y. (2020). Analysis of Stock Price Motion Asymmetry via Visibility-Graph Algorithm. *Frontiers in Physics* **8,** 539521.

Marinaccio, A., Boccuni, F., Rondinone, B. M., Brusco, A., D'Amario, S., Iavicoli, S., (2020). Occupational factors in the COVID-19 pandemic in Italy: compensation claims applications support establishing an occupational surveillance system. *Occupational and Environmental Medicine* **77,** 818–821.

Marinaccio, A., Brusco, A., Bucciarelli, A., D'amario, S. Iavicoli, S. (2021). Temporal trend in the compensation claim applications for work-related COVID-19 in Italy. *La Medicina del lavoro* **112,** 219.

Middleton, J., Reintjes, R., Lopes, H. (2020). Meat plants—a new front line in the covid-19 pandemic. *The BMJ* **370,** m2716.

Murti, M., Achonu, C., Smith, B. T., Brown, K. A., Kim, J. H., Johnson, J., et al. (2021). COVID-19 Workplace Outbreaks by Industry Sector and Their Associated Household Transmission, Ontario, Canada, January to June, 2020. *Journal of Occupational and Environmental Medicine* **63,** 574.

Mutambudzi, M., Niedwiedz, C., Macdonald, E. B., Leyland, A., Mair, F., Anderson, J., et al. (2021). Occupation and risk of severe COVID-19: prospective cohort study of 120.075 UK Biobank participants. *Occupational and Environmental Medicine* **78,** 307–314.

Molenberghs, G., Faes, C., Verbeeck, J., Deboosere, P., Abrams, S. Willem, L., et al. (2022) COVID-19 mortality, excess mortality, deaths per million and infection fatality ratio, Belgium, 9 March 2020 to 28 June 2020. *Eurosurveillance* **27,** 2002060.

Nash, J. C. (1990). Compact numerical methods for computers: linear algebra and function minimisation. *IOP Publishing Ltd*

Nishimoto, Y. and Inoue, K. (2020). Curve-fitting approach for COVID-19 data and its physical background. *medRxiv* (preprint), https://doi.org/10.1101/2020.07.02.20144899.

STATBEL (2017). NACE-BEL 2008. Available at: https://statbel.fgov.be/nl/over-statbel/methodologie/classificaties/nace-bel-200. Accessed on 8 November 2021.

Vasireddy, D., Vanaparthy, R., Mohan, G., Malayala, S. V., Atluri, P. (2021). Review of COVID-19 Variants and COVID-19 Vaccine Efficacy: What the Clinician Should Know? *Journal of Clinical Medicine Research* **13,** 317.

Verbeeck, J., Vandersmissen, G., Peeters, J., Klamer, S., Hancart, S., Lernout, T., et al. (2021). Confirmed COVID-19 cases per economic activity during Autumn wave in Belgium. *International Journal of Environmental Research and Public Health* **18** 12489.

Waltenburg, M. A., Victoroff, T., Rose, C. E., Butterfield, M., Jervis, R. H., Fedak, K. M., et al. (2020). Update: COVID-19 among workers in meat and poultry processing facilities—United States, April–May 2020. *Morbidity and Mortality Weekly Report* **69,** 887.

Wang, L., Zhou, Y., He, J., Zhu, B., Wang, F., Tang, L., et al. (2020). An epidemiological forecast model and software assessing interventions on the COVID-19 epidemic in China. *Journal of Data Science* **18,** 409–432.

Werker, A. R. and Jaggard, K. W. (1997). Modelling asymmetrical growth curves that rise and then fall: Applications to foliage dynamics of sugar beet (Beta Vulgaris L.). *Annals of Botany* **79,** 657–665.

Zhu, N., Zhang, D., Wang, W., Li, X. Yang, B., Song, J., et al. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine* **382,** 727–733.

Supporting Information

Web Appendices, Tables, and Figures referenced in Section 2-3 are available with this paper at the Biometrics website on Wiley Online Library. The R-code used to fit the described models is given in two R-files, one for each model. Plots of the fitted against the observed

profiles to visually assess goodness of fit are added to a PDF file in the Supplementary Information.

Received October 2007. *Revised February* 2008. *Accepted March* 2008.

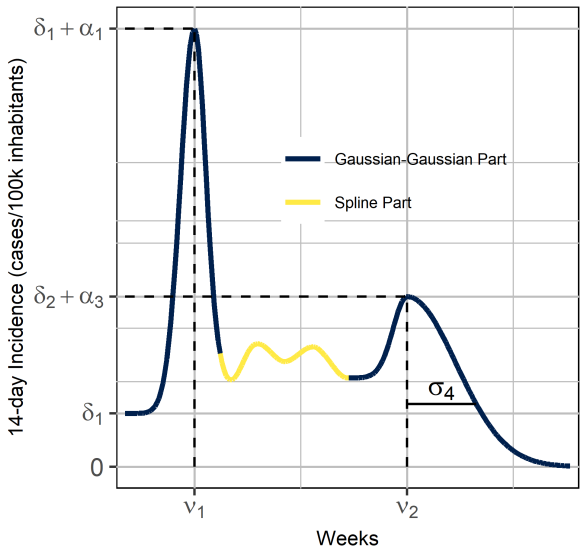

Figure 1. Two-peak Model. The peak incidences of the first and second wave are given by $\delta_1 + \alpha_1$ and $\delta_2 + \alpha_3$, respectively. The timing of the corresponding peak incidences are respectively given by ν_1 and ν_2. The interpretation of the width parameters is illustrated by σ_4; this parameter determines the width of the half-Gaussian curve at 37% of its peak height. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

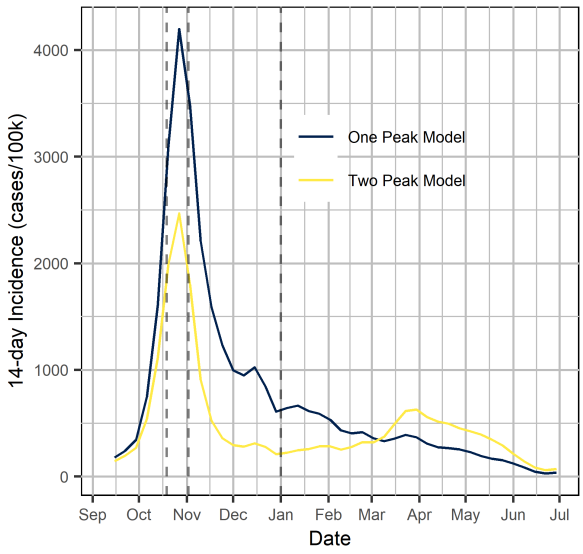

Figure 2. Comparison of the mean 14-day incidences in the group of sectors with one and two peaks. The first and second vertical dashed line are positioned at October 19 and November 2, respectively. At these two dates, stricter mitigation measures were introduced by the Belgian Government. The third dashed vertical line is positioned at January 1; in January the early/prioritised vaccinations started in Belgium. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

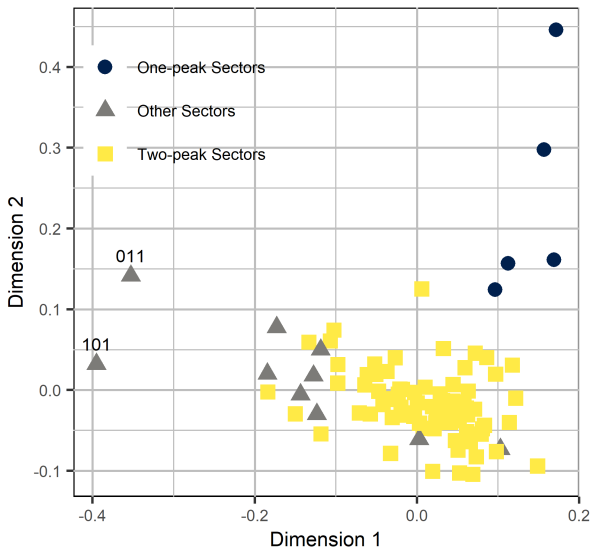

Figure 3. MDS-plot based on the first two MDS dimensions. The two outlying sectors are labelled by their NACE code. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

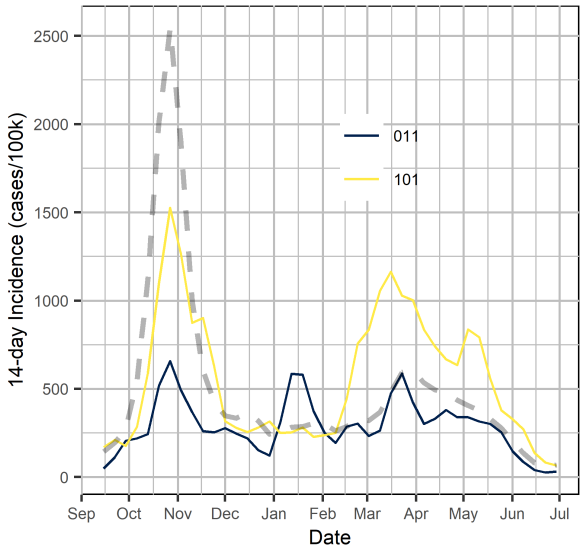

Figure 4. Longitudinal profiles of the sectors of special interest. The gray dashed line is the mean longitudinal profile for all employees. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

Table 1

Interpretation of the parameters of the Gaussian-Gaussian curve that models the first wave in the one- and two-peak models. The parameters of the second Gaussian-Gaussian curve in the two-peak model have a similar interpretation. These parameters are illustrated in Figure 1.

Parameter	Interpretation
δ_1	asymptotic incidence before first wave
$\delta_1 + \alpha_1$	first wave peak height
σ_1	width of increasing phase of first wave
σ_2	width of decreasing phase of first wave
ν_1	timing of first peak

Table 2

Canonical correlation analysis of the sets of variables describing the first and second wave, respectively. The likelihood ratio test (LRT) and the correlation between the wave parameters and the first two corresponding canonical variates are also shown.

Wave	Parameter	Canonical variate 1	Canonical variate 2
	Canonical correlation (LRT)	0.669 (< 0.001)	0.549 (< 0.001)
First Wave	Peak Height	-0.106	0.685
	Timing of Peak	-0.587	-0.472
	Width of Increasing Phase	-0.119	0.267
	Width of Decreasing Phase	-0.540	0.289
Second Wave	Peak Height	-0.813	0.322
	Timing of Peak	-0.257	-0.254
	Width of Increasing Phase	-0.070	0.059
	Width of Decreasing Phase	0.577	0.424

Table 3

Canonical correlation of the sets of variables describing the first and second wave, with the sector's mean income and mean age added to the first set of variables. The LRT and the correlation between the wave parameters, mean income and mean age on the one hand and the first two corresponding canonical variates on the other hand are also shown.

Wave	Parameter	Canonical Variate 1	Canonical Variate 2
	Canonical correlation (LRT)	0.707 (< 0.001)	0.645 (< 0.001)
First Wave	Peak Height	-0.430	-0.328
	Timing of Peak	-0.140	0.715
	Width of Increasing Phase	-0.227	-0.077
	Width of Decreasing Phase	-0.527	0.208
	Mean Income	0.707	0.273
	Mean Age	0.261	0.574
Second Wave	Peak Height	-0.860	0.29
	Timing of Peak	0.069	0.410
	Width of Increasing Phase	0.017	0.083
	Width of Decreasing Phase	0.048	-0.754