Latent class models for capturing unobserved heterogeneity in major global causes of mortality

The cases of traffic crashes and COVID-19

Papadimitriou, Eleonora; Afghari, Amir Pooyan; van Gelder, Pieter

DOI
10.1016/j.trip.2024.101147

Publication date
2024

Document Version
Final published version

Published in
Transportation Research Interdisciplinary Perspectives

Citation (APA)
Papadimitriou, E., Afghari, A. P., & van Gelder, P. (2024). Latent class models for capturing unobserved heterogeneity in major global causes of mortality: The cases of traffic crashes and COVID-19. *Transportation Research Interdisciplinary Perspectives*, *26*, Article 101147. https://doi.org/10.1016/j.trip.2024.101147

Transportation Research Interdisciplinary Perspectives 26 (2024) 101147

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.sciencedirect.com/journal/transportation-research-interdisciplinary-perspectives

Latent class models for capturing unobserved heterogeneity in major global causes of mortality: The cases of traffic crashes and COVID-19

Eleonora Papadimitriou *, Amir Pooyan Afghari, Pieter van Gelder

Safety and Security Science Section, Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628BX, Delft, The Netherlands

A R T I C L E I N F O

Keywords:
Traffic fatalities
COVID-19
Cross-sectional data
Latent class models

A B S T R A C T

Existing models for correlating global mortality rates with underlying country-specific factors overlook the variations in the effects of these factors on mortality across different countries. These may arise from social, cultural, and political complexities which are usually not measurable and are therefore referred to as *unobserved heterogeneity* in the statistical literature. Unobserved heterogeneity leads to biased parameter estimates in the models, erroneous inferences about the effects of factors contributing to mortalities, and ultimately inefficient policies. In this paper, latent class modelling is proposed for capturing such unobserved heterogeneity on the cases of traffic mortality and COVID-19 mortality. The 'pyramid' model of safety management is used as a common framework for model formulation. The proposed latent class model is an extension of the Negative Binomial (NB) model used in risk epidemiology. The model is tested with data from 105 countries, retrieved from international databases, including socioeconomic, infrastructure, exposure, transport, and COVID-19 variables. The results suggest that there exist two (different) latent country classes in both causes of mortality. The probability of a country belonging to a certain latent class is a much more efficient metric of country membership than previous deterministic groupings (e.g. income or geographic). Variables such as the elderly population, the GDP per capita or the level of motorization, have different effects in different country classes; these effects are not identifiable by conventional statistical modelling. The impact of ignoring unobserved heterogeneity in country mortality modelling is shown by comparing the results with those of conventional NB models.

Background and objectives

Global mortality rates by different causes have been a critical question receiving continuous attention in the safety and epidemiology disciplines. Several studies have attempted to explore country-specific differences in mortality rates and the role of socioeconomic factors affecting them, in order to identify good practices that can inform policy making at national, regional and global level. Country benchmarking, i.e. a process to identify best practice, the ways in which positive results are achieved in these countries, and the ways in which these can be implemented to other countries, is a useful tool for addressing the burden from a certain cause of mortality (Wegman & Oppe, 2010; Bert et al., 2022). Moreover, the impact of macroscopic country characteristics on mortalities can improve predictions of their evolution and reveal potential areas of intervention (UNECE Sustainable Transport Division, 2018).

One recent example of a critical cause of mortality is the Covid-19 pandemic which was declared by the World Health Organisation (WHO) on 11 March 2020 and within that first year brought 1.96 million fatalities (WHO, 2020; 2022). Countries responded to this outbreak by means of various restrictions aiming to prevent the spread of the virus, as well as to support the health sector. However, there were considerable differences in the type of implemented measures and their compliance / enforcement, resulting in different mortality outcomes. Numerous recent studies aimed to identify these relationships at the international level.

Another long-lasting example is the global "epidemic" of traffic fatalities (more than 1.35 million people annually all over the world) – being the leading cause of death for people aged between 25–55 years (WHO, 2018). Several studies have attempted to model its development in order to transfer knowledge from good performing countries and proactively anticipate future developments; for instance, lessons to be leaned by low-to-middle-income countries in view of their transport development, safety improvements to be anticipated by the penetration of new vehicle technologies etc.

While the two causes of mortality are fundamentally different from

* Corresponding author.
E-mail address: e.papadimitriou@tudelft.nl (E. Papadimitriou).

https://doi.org/10.1016/j.trip.2024.101147
Received 9 November 2023; Received in revised form 11 June 2024; Accepted 11 June 2024

the microscopic perspective, it has been noted in the literature that there are common macroscopic factors affecting the size of the final outcomes (fatalities) in both causes of mortality, ranging from cultural and institutional factors, to specific regulations and policies in place, as well as factors related to the behaviour of individuals (Papadimitriou & Afghari, 2023). However, there are notable differences in countries around the world arising from unobserved country-specific factors (e.g. social structures, cultural attributes, political complexities, etc.) which are often difficult to measure and take into account in macroscopic statistical modelling of mortality rates. Therefore, the existing statistical analysis methods used for correlating mortality rates with their underlying factors are "primitive", in many cases as simple as basic regressions, and they have not considered the complexities associated with the above differences. One of these complexities is varied effects of factors on mortality rates (i.e. most studies have assumed a fixed effect of those factors across all countries whereas this may not be totally accurate). This is referred to as *unobserved heterogeneity* in the statistical literature and, if it is not accounted for, leads to biased and inefficient parameter estimates in the statistical models, results in erroneous inferences about the effects of factors contributing to mortalities, and ultimately leads to inefficient policies for mitigating the mortalities. Therefore, there is a need to use more proper techniques to address this unobserved heterogeneity in country comparisons. Econometric modelling disposes the theoretical and practical tools to model these complexities, and can be used to get more insight on how to model complex relationships between macroscopic factors and major mortality causes.

The objective of this paper is to understand unobserved heterogeneity in the effects of macroscopic factors that are associated with mortality causes, by leveraging our learnings from advanced econometric modelling. For that purpose, a number of latent class models are developed to model mortality rates for two critical causes of mortality: traffic crashes and COVID-19 infections. The main motivation for choosing these two causes of mortality is that recent studies have shown that they have a similar annual burden of casualties (Colonna and Intini, 2020), they share several attributes at global level, and they can be described by analogous macroscopic safety mechanisms (Papadimitriou & Afghari, 2023) – albeit their significant operational differences. In addition, studies have repeatedly shown that COVID-19 pandemic significantly influenced transport systems in general (De Vos, 2020; Peralvo et al., 2022), and road traffic risk in particular (Katrakazas et al., 2020; Saladié et al., 2020). As such, global data from international mortality databases (traffic fatality and COVID-19) are used in this study to apply and compare these models with the classical Poisson-family statistical models that are typically used in ecological studies modelling rare events in a cross-sectional framework.

This paper is structured as follows: Section 2 presents a literature review on the available macroscopic modelling techniques for traffic-related and COVID-19 related fatalities. Section 3 presents our conceptual framework, research hypotheses and model formulation, which expands the classical Negative Binomial model into a Latent Class model that can capture the latent classes underlying the observations; it also presents the data collected to develop the models. Section 4 includes the modelling results for both causes of mortality. Section 5 presents a discussion of the findings, followed by conclusions (section 6).

2. Literature review

A literature review was conducted aiming to identify the main cross-country factors (from macroscopic perspective) associated with traffic and COVID-19 mortalities (epidemiological studies) and the existing methods to study the effects of the above factors on traffic and COVID-19 mortalities. A literature search was carried out to identify the most relevant and representative studies of cross-country macroscopic modelling of the two causes of mortality. The following types of studies were excluded: studies with only descriptive analysis, studies only comparing a few countries (<10), longitudinal COVID-19 studies. In the COVID-19 context, the focus was on studies looking at the 1st year of the pandemic (2020). Moreover, grey literature publications were eventually excluded.

2.1. Traffic mortality

In the traffic safety field, cross-country comparisons have been the focus of analysis by several international organisations, e.g. WHO, the EC, the UN etc. Several researchers have used regional (e.g. European) or global data to develop statistical and econometric models of traffic mortality. Because of the difficulty in obtaining accurate and comparable measures of traffic exposure (e.g. vehicle- and person-kilometres of travel) at global level, the mortality rate is typically used as a dependent variable, while correcting for the level of motorization and / or vulnerable road users' exposure by means of other independent variables (e.g. vehicle ownership, bicycle ownership). The majority of relevant studies aims to i) develop prediction models for the future development of traffic fatalities, or ii) investigate the impact of specific interventions on mortality, with the impacts of economic recessions receiving a lot of attention.

In the first family of models, studies focused on the methodological challenges of modelling macroscopic developments of fatalities and exposure (vehicle-kilometres-based, or population-based) by means of pertinent time series analysis techniques. In Lassarre (2001) structural time series models were developed for road safety developments in Europe, whereas Commandeur et al. (2013) further developed a framework for the application of state-space models in these countries by specifying the pertinent ways to account for unobserved trends. Borsos et al. (2012) proposed a function to model the dynamic impact of motorization on traffic fatalities. In Antoniou et al. (2016), a methodological framework was proposed for modelling temporal panel data of traffic mortalities, taking into account both the panel size and the time horizon size. These studies focus on the temporal dimension of traffic mortality.

In the second family of studies, the correlation between traffic mortality and socioeconomic developments is examined; these are expressed by the gross domestic product (GDP) per capita, the level of motorisation or the unemployment rate (Page, 2001; Kopits & Cropper, 2005; Yannis et al., 2011). A dedicated group of studies take an in-depth look into socioeconomic disruptions e.g. the energy crisis of the decade of 1980 (Hedlund et al., 1982), or the more recent economic recession of 2008, either from a long-term (Kweon, 2015) or from a short-term differences perspective (Yannis et al., 2014).

An attempt to model both short-term temporal developments of mortality and the cross-sectional unobserved factors of the countries panel was presented by (UNECE Sustainable Transport Division, 2018), where a short-term prediction model was developed on global mortality data within the 3-year intervals of WHO data. This research tested a geographical grouping of countries, which accounted for a minor share of the variability in the data.

While numerous studies have investigated unobserved heterogeneity in traffic crash occurrence and severity at microscopic or local level (e.g. Mannering et al., 2016; Yu et al., 2019) and found it to be a significant factor affecting model accuracy, the knowledge from these microscopic studies has not been transferred and applied to macroscopic traffic safety modelling at global level.

2.2. COVID-19 mortality

There are numerous recent studies that use ANOVA and basic correlation tests, or linear / log-linear regression models to compare country mortality rates from COVID-19 (Liu & Eggleston, 2022; Kapitsinis, 2021; Bouba et al., 2021), while seldom taking into account the panel effect or the impact of unobserved factors (e.g. geographical, cultural).

There were a few attempts to take into account latent country characteristics in the modelling, however in a cluster-based or fixed-effects approach. In Marginean & Orastean (2022) the authors clustered the 27 European countries on the basis of their health spending, and correlated the three groups identified with the COVID-19 health outcomes. El Mouhayyar et al. (2022) analysed data from 187 countries, grouped on the basis of the UN regional classification, by means of square-root adjusted multiple regression models; the authors note the potential presence of "ecological fallacy" in their inferences: the aggregate country effects may not fully represent partial country characteristics. Liang et al. (2020) applied linear regressions to a cross-sectional dataset of 169 countries, while checking for the variability of the results with respect to four country classifications (low, moderate, high): on the basis of their per capita incomes, government effectiveness scores, proportions of population aged 65 or older, and numbers of hospital beds. Focusing on the effect of numbers of tests, the study found a negative correlation of Covid-19 mortality with test number, and the correlation varied with country characteristics.

A few studies took a non-parametric or machine learning approach to modelling COVID-19 related mortality. Chen et al. (2023) applied a fuzzy-set qualitative comparative analysis on years of life lost (YLL) for 2021 in 80 countries, and identified several different configurations, and four distinct 'pathways' to the final outcomes. As the authors state, overall *"some countries failed differently, whereas others succeeded differently"*, which clearly indicates the presence of unobserved heterogeneity. Arulanandam et al. (2021) developed a non-parametric regression model for the impact of obesity on COVID-19 mortality in 154 countries. They authors note a low robustness of the effects of their control variables (socioeconomic factors), which is an indication of unobserved heterogeneity between countries.

One of the few studies that accounted for unobserved effects in COVID-19 mortality outcomes is that of Bjørnskov (2021), which compared European countries' lockdown policies with final outcomes. The author proposed a multilevel model with seasonal effects, lagged policy implementation effects, and country fixed effects; two methodological issues are noted: i) the endogeneity between public policy and mortality outcomes, and ii) the lack of knowledge (at the time) about the impact of the virus incubation period (hence the inclusion of lagged effects). The findings suggest that mortality outcomes were independent from the policy measures over the long term, although some significant short-term effects could be identified. Other studies exploiting the features of advanced econometric modelling are those of Antonietti et al. (2023) who included country-level variables and week-level variables, as well as a geographical random effect over 138 countries, and Debajyoti et al. (2023) who developed a dynamic panel regression model to account for daily patterns of COVID-19 mortality together with static socioeconomic indicators and region-specific control variables for 119 countries.

2.3. Research gaps

Cross-country comparison of mortalities around the world can identify macroscopic factors contributing to the global burden of mortalities and benchmark best policy practices to combat them effectively. From the above literature review, it is concluded that several methodological and knowledge gaps exist in the modelling of global causes of mortality:

(i) Statistical tests, ANOVA and simple linear regression models have been widely used for such a comparison. However, these techniques only look at the effects of the macroscopic factors on the mean mortality of the countries and overlook the differences in these effects arising from the social structures, cultural attributes, and political complexities. Traffic and COVID-19 deaths are two examples of major global mortalities that have been studied extensively but without considering these differences.

(ii) At the same time, previous studies in both contexts suggest that there are common macroscopic factors behind the two causes of mortality. As such, there is a need for a methodological framework that can disentangle the effects of observed and unobserved macroscopic factors contributing to global mortalities and capture the resulted heterogeneity across countries.

This study aims to address this research gap by formulating a common conceptual model for macroscopic causes of global mortalities, and then by developing a latent class statistical modelling framework which looks at the different effects of these factors on mortalities across different countries (or group of countries).

3. Methods and data

To the best of the author's knowledge, there are no global studies explicitly taking into account the unobserved factors that express the heterogeneity between countries' mortality rates. A few studies attempt to correct for this bias (both in the traffic safety and COVID-19 contexts), by grouping (or clustering) countries into economic or geographical groups, however this is not sufficient. Such a clustering approach is deterministic separating countries into clusters without any uncertainties (i.e. a country can only belong to one cluster). However the heterogeneity between countries is determined by a number of unobserved cultural and socioeconomic factors, making it difficult to formulate a distinct grouping with certainty. In this paper, it is postulated that these unobserved factors may be optimally represented by latent classes with probabilities (i.e. a country can belong to multiple latent classes with probabilities that sum to unity across all the classes).

A first step for implementing this analysis is the selection of a conceptual framework allowing for the parallel examination of the two causes of mortality in an accordant way. For that purpose, we introduce the pyramid model for traffic safety systems analysis (Koornstra et al., 2002; Wegman et al., 2005), which can assist in variables selection. Subsequently, we propose a poisson-family latent class modelling approach.

3.1. Conceptual framework: The pyramid of systems safety management

The proposed pyramid, also known as the 'SUNflower' model (Koornstra et al., 2002; Wegman et al., 2005) has been used by several national and international studies as an appropriate framework to describe road safety management systems at a specific 'snapshot' in time. Since its introduction, several variations of its structure and contents have been proposed e.g. by the World Bank road safety management capacity evaluation protocol (Bliss & Breene, 2009), the European Road Safety Observatory country profiles (European Road Safety Observatory, 2021), or the UNECE SafeFITS global road safety model (UNECE Sustainable Transport Division, 2018) (see Fig. 1).

Despite slight context-specific adjustments, the overall framework can be described as follows:

- The first (bottom) layer, Structure and Culture, reflects the institutional, economic, cultural and regulatory characteristics (i.e. policy input) of each country, that are background factors related to road safety performance. Typically, characteristics of the transport system such as exposure, urbanisation, modal split and road network type are included in this layer, although in (UNECE Sustainable Transport Division, 2018) they were in a separate layer.
- Measures and Programmes (policy output), at the second layer, are the result of structural and economical characteristics and include the legislation and measures in place to contain risk factors, as well as their enforcement.
- To link the above layers to the actual traffic safety outcomes in a country, an intermediate layer specifies the operational level of safety in that country, including key performance indicators (KPIs) –

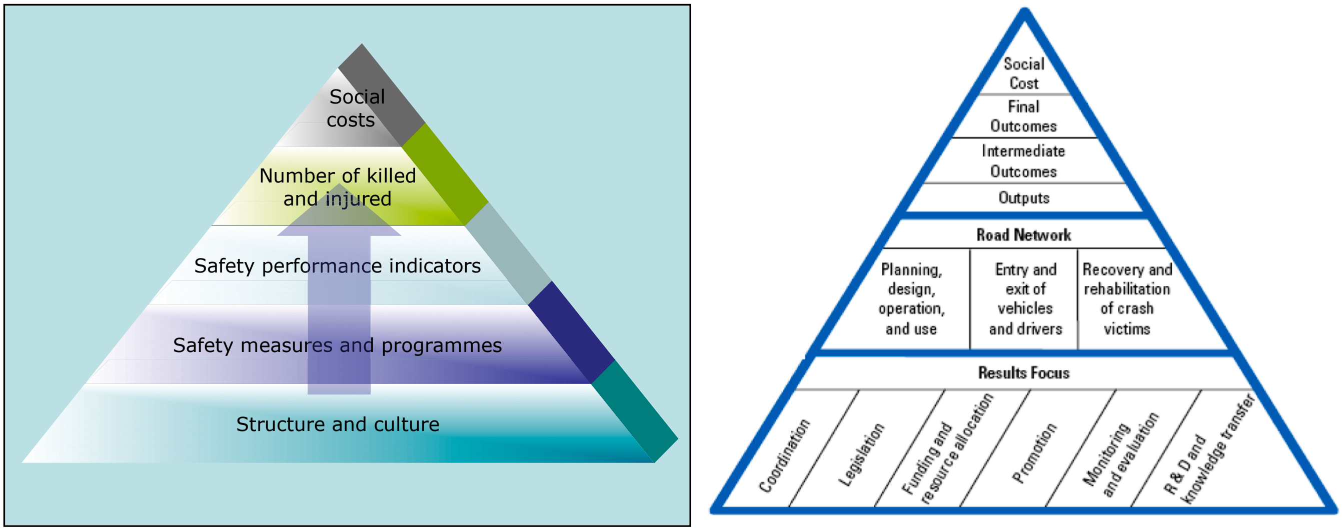

Fig. 1. Generic structure of the pyramid of road safety management systems [sources: adapted from (Koornstra et al., 2002; ERSO, 2021) – left panel, Bliss & Breen, 2009 – right panel].

also referred to as Safety Performance Indicators (SPIs) – on road safety, they include metrics related to user behaviour (e.g. speeding, drinking and driving, use of protective systems etc.), the state of the road network and vehicle fleet etc.

- The risk outcomes in terms of fatalities are then assumed to be the result of this interaction between layers from the bottom up.
- A fifth layer of the pyramid includes the socioeconomic cost of mortality, as the final outcome.

A limited number of studies have attempted to quantitatively model traffic fatalities on the basis of the pyramid; these studies typically estimate composite variables for each layer of the pyramid, including several indicators to calculate a score for each layer – for example, Hermans et al. (2009), Papadimitriou & Yannis (2013) for benchmarking purposes, and the SafeFITS model (UNECE Sustainable Transport Division, 2018) for forecasting road safety developments. In this paper, we argue that the 'pyramid' model could be a suitable conceptual framework to serve as a backbone for modelling different causes of mortality in a consistent way.

3.2. Research hypotheses

The main background hypothesis of this research is that:

(a) Both causes of mortality can be described on the basis of the pyramid model for safety management systems for a given 'snapshot' in time;
(b) There are certain common observed and unobserved attributes that affect the final outcomes of the two causes of mortality, e.g. socioeconomic, geographical, cultural, policy-related, exposure and behaviour-related; biases will occur if unobserved attributes are not accounted for.
(c) There are also distinct factors that are explicitly critical to one cause of mortality or the other, caused by their different mechanisms.

The 'pyramid' framework has been applied to traffic safety in several studies, and it has been established that certain indicators are conceptually related to different layers of the pyramid. An analogous taxonomy regarding the Covid-19 pandemic is proposed in this research and different types of indicators are identified that are relevant to the different layers of the pyramid as shown in Table 1. The purpose of this exercise is to allow developing a coherent and analogous model formulation, introduce explanatory variables in a structured way, and represent all the layers of the examined systems. This is particularly important given that the number and types of variables that have been examined in previous studies (of traffic and COVID-19 mortalities) is extremely diverse and in most cases focuses on a particular set of variables. However, this exercise does not imply that the two causes of mortality are affected by the same variables.

Table 1
Taxonomy of traffic and Covid-19 indicators for different management layers.

	Traffic mortality indicators	Covid-19 mortality indicators
Structural	• GDP per capita • Demographics • Population density, urbanisation, state of road network • Existence / funding of a road safety Lead Agency • Existence of fatality reduction vision, strategy & targets • …	• Life expectancy • Demographics • Human life quality index • Prevalence of diseases • Number of hospital beds • …
Exposure	• Vehicle-kilometres of travel • Level of motorization • Share of vulnerable road users • …	• Number of cases / infections • Share of vulnerable groups • …
Programmes and measures	• Duration and intensity of traffic enforcement • Speed limits • Alcohol limits • Road & vehicle standards • …	• Duration and intensity of enforcement • Mobility restrictions • Activity closures • Social distancing • Vaccination campaigns • Personal protection measures • …
KPIs	• Share of drivers exceeding speed limits • Share of drivers under the influence of alcohol • Share of road users using protective devices • …	• Share of population complying with measures • Reproduction rate (R) • …
Final outcomes	• Number of fatalities	• Number of fatalities

For instance, structural factors in the traffic mortality context include socioeconomic and demographic indicators, together with indicators related to the state of the road and urban environment, as well as the institutional setup and organisation of road safety management in a country. Analogous indicators for the Covid-19 mortality may also include socioeconomic and demographics, as well as indicators related to the quality of the healthcare system and the overall wellbeing of the population.

Regarding exposure, there are well known traffic risk indicators e.g. vehicle-kilometres of travel, and relevant indicators in the Covid-19 context concern the number of infections. Of course, there is a fundamental difference: traffic fatalities are happen 'immediately' after crashes that are random localised events, often involving human error or infrastructural deficiencies. In contrast, Covid-19 fatalities arise from a viral infection that spreads from person to person. Therefore, it is acknowledged that the mechanism of exposure and outcome is fundamentally different between the two causes; however, the role of exposure is key, and it is important to include it in the analysis. In both cases, the share of vulnerable groups of each pandemic among the population also need to be taken into account.

Programs and measures, in the next layer, may include in each case the specific regulations and measures in place for risk prevention and control; in the traffic mortality case, these range from speed limits and protective devices, to vehicle and road regulations, while in the Covid-19 case they mostly include social distancing, hygiene measures, mobility restrictions and activity area closures. It is noted that the latter type of measures are very rare and in practice not applicable in the traffic context, although it is well known that traffic mortality decreases with decreased exposure. In both cases, the enforcement of measures is a key additional component.

Finally, the 'behavioural' layer of KPIs may be the most challenging to assign specific quantitative indicators to. In the traffic context, the behaviour of drivers regarding key regulations is typically used through a number of indicators (see Table 1). In the case of the Covid-19 pandemic, there are no data available measuring the compliance of the population. However, it can be assumed that the virus reproduction rate (R) is an approximation of the operational level of risk, since it expresses the speed of virus spread among the population and the risk of observing more fatalities as a result.

The selection of indicators in Table 1 is not exhaustive; it serves as an indicative classification of certain commonly used indicators in traffic safety and other epidemiological studies, in order to facilitate models formulation – a full description of the data used in this research is given in the section 3.5.

3.3. Poisson-family modelling

Let Y_i represent the total number of mortalities in the i^{th} country. The epidemiology and traffic safety literature have shown that Y_i follows a Poisson-family distribution, namely a negative binomial (NB) distirbution with mean μ_i and inverse dispersion parameter φ:

$$Y_i\ NB(\mu_i, \varphi)$$

Assuming an exponential function for the mean of the negative binomial distribution, the total predicted fatality count (μ_i) in the i^{th} country can be expressed as a function of exogenous explanatory variables:

$$\mu_i = \exp(\sum \beta X_i + \varepsilon_i) \tag{1}$$

where X_i are other explanatory variables and β are estimated regression parameters (including the intercept) and $\exp(\varepsilon_i)$ is a random error term, which follows a Gamma distribution with mean 1 and variance $1/\varphi$. The probability of the total number of fatalities is then stated as:

$$P(y_i = Y_i) = \frac{\Gamma(y_i + \varphi)}{y_i!\Gamma(\varphi)} \left[\frac{\varphi}{\varphi + \mu_i}\right]^{\varphi} \left[\frac{\mu_i}{\varphi + \mu_i}\right]^{y_i} \tag{2}$$

where $\Gamma(.)$ is the gamma function. The log-likelihood function (LL) of the model is obtained by applying the logarithm transformation and summing it over observations to yield:

$$LL = \sum_{i=1}^{N} \log(P(y_i = Y_i)) \tag{3}$$

This model is referred to as the Negative Binomial (NB) count model for fatalities (e.g. traffic fatalities) and is now extended into the latent class specification to account for unobserved heterogeneity in the data.

3.4. Latent class modelling

Assuming that there are S number of latent classes over the population, the probability of observations belonging to each distinct class, $P(C_s)$, can be computed using a logit model with the following specifications:

$$P(C_s) = \frac{e^{U_s}}{\sum_{s=1}^{S} e^{\beta X_i}} \text{ and } U_s = \Omega_s Z_s \tag{4}$$

where $\boldsymbol{\Omega}$ is the vector of parameters (including an intercept), and $\boldsymbol{Z}$ is the vector of class-specific covariates (including the intercept). Such covariates determine the probabilities of observations being assigned to each specific class.

Within each class, the probability Y_i of conditioned to that class can be computed using the equation (2). Applying the rules of conditional probabilities, the marginal probability of the latent class Negative Binomial regression model is stated as:

$$P(Y_i) = \sum_{s=1}^{S} P(Y_i|C_s) \times P(C_s) \tag{5}$$

where $P(Y_i)$ is the unconditional probability of total number of mortalities, $P(Y_i|C_s)$ is the conditional probability of total number of mortalitiesin class C_s (same as equation (2), and $P(C_s)$ is the probability of class C_s.

The overall log-likelihood function can be determined by the product of equation (5) over the entire observations. The Maximum Likelihood Estimation (MLE) approach is employed for the estimation of the latent class regression model.

In the above formulation of the latent class model, the classes are assumed to be latent across observations, and thus the number of latent classes is not known a priori. Therefore, the model is empirically tested with a different number of classes (S), and the preferred number of classes is selected based on the model with the superior statistical fit. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are employed to compare the statistical fit of the model candidates (Washington et al., 2020).

$$AIC = -2LL + 2P$$

$$BIC = -2LL + PLog(N)$$

where LL is the log-likelihood of the estimated model at convergence, P is the number of estimated parameters, and N is the number of observations or sample size. The model with lower AIC and BIC is regarded as a superior model in terms of statistical fit.

3.5. Data collection

The data used in this research were selected from a number of international databases and sources, in order to obtain the types of indicators shown in Table 2. The traffic fatality data are retrieved from the WHO Global Status Report on Road Safety (2018), and the World Bank and International Road Federation (IRF) databases. The Covid-19 data and health / policy related indicators are retrieved from a publicly

Table 2
Variables descriptive statistics and sources.

Socioeconomic data	Source	Mean	SD	Min	Max
Population (million people)	World Bank	66.82	197.635	2.689	1444.216
Population density (number of people per square kilometres)	World Bank	206.94	777.684	1.98	7915.731
Mean age (years)	various*	31.99	9.323	15.1	48.2
Proportion of population with age above 65 years	Various	9.933	6.648	1.144	27.049
Proportion of population with age above 70 years	Various	6.332	4.562	0.526	18.493
GDP per capita (US dollars)	Various	20.473	18.714	0.661	85.535
Life expectancy at birth (years)	Various	74.205	7.024	53.28	84.63
Human development index** (out of 1)	Various	0.749	0.154	0.394	0.957
Death rate from cardiovascular disease in 2017 (annual number of deaths per 100,000 people)	Various	252.895	130.22	79.37	724.417
Diabetes prevalence (% of population aged 20 to 79) in 2017	Various	7.36	3.421	0.99	17.72
Number of hospital beds per thousand people	Various	3.029	2.607	0.1	13.05
Covid-19 data		**Mean**	**SD**	**Min**	**Max**
COVID-19 mortalities in 2020	CSSE-JHU	24258.4	63182.47	2	515,513
Total confirmed cases of COVID-19	CSSE-JHU	1,054,106	3,205,886	509	28,805,150
Mean reproduction rate (R) across 2020	various*	1.065	0.144	0.036	1.22
Std deviation of reproduction rate across 2020	Various	0.306	0.106	0.084	0.64
Maximum reproduction rate across 2020	Various	2.205	0.723	0.4	5.39
Minimum reproduction rate across 2020	Various	0.608	0.201	0.1	1
Proportion of days in 2020 with R higher than 1	Various	0.672	0.194	0	1.061
Mean stringency index*** (out of 100)	Various	61.961	12.518	14.755	87.918
Standard deviation stringency index	Various	15.504	4.656	3.06	31.031
Maximum stringency index	Various	85.415	12.45	27.31	100
Minimum stringency index	Various	17.041	13.119	0	61.11
Traffic safety data		**Mean**	**SD**	**Min**	**Max**
Road traffic fatalities	WHO	11249.752	38353.642	143	299,091
Proportion of population with no access to public transport	UN & EC DG REGIO	0.454	0.307	0.013	0.926
% of paved roads on the road network (2018 or latest year available)	IRF	60.665	30.109	2.920	100.00
% of motorways on the road network (2018 or latest year available)	IRF	2.422	2.335	0.010	21.410
% of motorcycles in the vehicle fleet (2018 or latest year available)	WHO	22.757	21.986	0.002	93.023
Existence of a road safety lead agency	WHO	0.924	0.267	0	1
The lead agency is funded	WHO	0.743	0.439	0	1
Existence of national road safety strategy	WHO	0.848	0.361	0	1
The strategy is funded	WHO	0.481	0.285	0	1
Existence of fatality reduction target	WHO	0.714	0.454	0	1
Maximum speed limits on urban roads > 50 km/hr	WHO	0.048	0.214	0	1
Maximum speed limits on rural roads > 120 km/hr	WHO	0.305	0.463	0	1
BAC limits less than or equal to 0.05 g/dl	WHO	0.581	0.498	0	1
Effectiveness of seat-belt law enforcement	WHO	5.92	2.315	0	10
Effectiveness of drink-driving law enforcement	WHO	5.98	2.457	0	10
Effectiveness of speed law enforcement	WHO	5.782	2.281	0	10
Effectiveness of helmet law enforcement	WHO	6.109	2.545	0	10
Seat-Belt wearing rate-Front (2016 or latest available year)***	WHO	59.767	28.434	3.5	98
Helmet wearing rate-driver (2016 or latest available year)***	WHO	67.8	28.179	6.2	100

* Various official sources compiled by Our World in Data (https://github.com/owid/covid-19-data/blob/master/public/data/README.md).
**A composite index measuring average achievement in three basic dimensions of human development—a long and healthy life, knowledge and a decent standard of living. Values for 2019, imported from https://hdr.undp.org/en/indicators/137506.
**Government Response Stringency Index: composite measure based on 9 response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest response).
*** Missing values of seat belt and helmet wearing rates of drivers were replaced by the mean value per country income category (with low income countries defined by GDP<\$1,036, low-middle income countries \$1,036 ≤ GDP≤\$4,045, high-middle income countries \$4,046 ≤ GDP≤\$12,035 and high income countries > 12,535, as per the World Bank categorization of year 2020).

available dataset including indicators from official sources for year 2020, compiled by Our World in Data (Ritchie et al., 2020), including Covid-19 cases and mortality statistics from the Data Repository of the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU).

The descriptive statistics of the examined variables are shown in Table 1. Countries with population < 250,000 inhabitants were excluded due to very low number of fatalities, and a number of other countries due to lack of key indicators for either one of the mortality causes, as well as outliers (e.g. very high Covid-19 mortality rate for Peru). This resulted in a usable dataset of 105 countries; their geographical distribution is shown in Fig. 2.

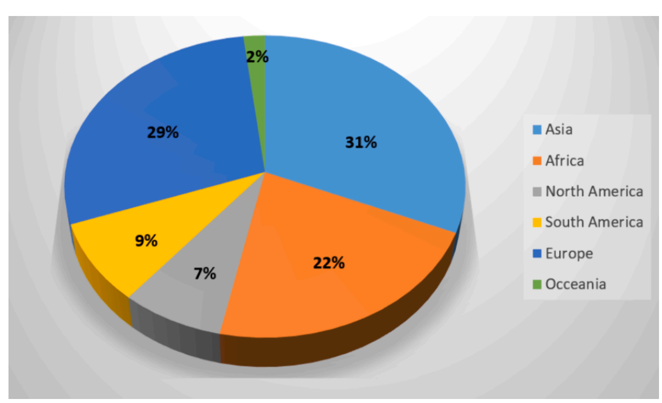

Fig. 2. Distribution of studied countries per continent in the sample.

4. Results

In a recent study (Papadimitriou & Afghari, 2023), it was indicated that both causes of mortality, and despite their differences, can be modelled by taking into account structural and economic indicators, policy and measures related indicators (and their enforcement), as well as indicators reflecting the operational level of safety (performance indicators) at the given "snapshot" in time. The authors developed basic regression models to confirm this hypothesis on the basis of the taxonomy of indicators of Table 1, associated with each cause of mortality.

A similar variable selection approach is implemented by taking indicators from all layers of the 'pyramid', in order to account for all the policy and operational factors that affect the mortality outcomes, and a more advanced econometric modelling approach is applied. The following sections focus on the methodological contribution and its implications in terms of interpretation of the modelling results.

4.1. Traffic mortality

Table 3 shows the parameter estimates and goodness-of-fit of a "classical" Negative Binomial model of traffic mortality. It is observed that the larger share of the variability in the data is captured by the constant term, the dispersion parameter, the population (offset variable) and the GDP per capita. The latter has a negative parameter, which indicates that a higher GDP is associated with lower traffic mortality, confirming previous findings especially from industrialised countries.

Socioeconomic and traffic policy related indicators are found to be of very low magnitude, and not statistically significant. These include the percentage of paved roads, the percentage of motorcycles in the fleet, the percentage of population without convenient access to public transport (i.e. a proxy of urbanization) and the quality of seat-belt use enforcement in the country. Although previous studies have found all these variables to have significant effects on traffic safety outcomes in different cross-sectional settings, our finding is not surprising. It is possible that the heterogeneity on the effect of these variables in different countries cancels out certain opposite effects (e.g. positive effect of a variable in some countries, but negative in others).

In Table 4, the respective results of the Latent Class model of traffic mortality are presented. It is found that two latent classes yield a robust model: class 1 with probability of 23.8 % and class 2 with probability of 76.2 %. No variables were found to significantly determine the splitting of the data (class membership) and only the constant terms of the classes play a role in this respect. The parameter estimates per class can be interpreted as follows:

- GDP per capita is not significant in class 1, but it is significant in class 2, with a negative coefficient, which is intuitive, as in the simple Negative Binomial model.
- The percentage of paved roads is significant in class 1, with a negative coefficient which indicates that a higher share of paved roads in a country is associated with a lower mortality rate; however it is not significant in class 2.
- The effect of the percentage of people with insufficient access to public transport is now statistically significant and positive in both classes. It is noted that the magnitude of the effect is 9 times higher in class 1 compared to class 2.
- The share of motorcycles in the fleet is statistically significant and positive in both classes, with the magnitude of the effect being double in class 1 compared to class 2.
- The quality of seat-belt use enforcement is statistically significant in class 1, with a positive coefficient, suggesting that a higher quality of enforcement is associated with higher traffic mortality. This is counter-intuitive, and may be indicated as a low effectiveness of enforcement in class 1 countries; it should be also taken into account that this is a self-reported variable on the basis of the assessment of a panel of country experts, and therefore may include some bias. This variable is not significant in class 2.

These findings may imply that LMIC countries are more likely to belong to class 1 whereas industrialised countries are more likely to belong to class 2 (also looking at the constant terms of the models). The counter-intuitive parameter of seatbelt use enforcement in class 1 (if it is more likely to include LMIC) might then indicate that the perceived quality of enforcement is not sufficient to address the traffic safety situation in those countries. None of the tested KPI layer indicators was found to be statistically significant.

The statistical fit measures (AIC and BIC) show that the latent class model is substantially superior to the regular NB model in terms of fit. It is also interesting to note that the dispersion parameter of the Negative Binomial distribution is only significant in Class 2, which may reflect that there is more remaining variability among industrialised countries than LMICs – this is also indicated by the fact that there is well known large variability of many of the identified fixed effects within LMICs, e.g. GDP, share of paved roads, seatbelt enforcement).

4.2. COVID-19 mortality

In this section, the same exercise is repeated with COVID-19 mortality data. The results of a Negative Binomial regression model of COVID-19 mortality (not shown here for the economy of space) indicated that none of the examined variables are statistically significant, and only the constant term and the dispersion parameter capture the variability in the observations. In a previous exploratory study (Papadimitriou & Afghari, 2023), as well as in other cross-sectional studies (see section 2.2), these variables were found to be significant in an Ordinary Least Squares log-linear model, but it appears that their identified effects might be artefacts of the overdispersion in the data. The presence of unobserved heterogeneity will be examined next.

In Table 5, the respective results of the Latent Class model of COVID-19 mortality are presented. It is found that two latent classes yield a robust model: class 1 with probability of 82.2 % and class 2 with probability of 17.8 %. In addition to the constant term, GDP per capita was found to significantly determine country membership; countries

Table 3
Parameter estimates and goodness-of-fit of the Negative Binomial model of traffic mortality.

Variable	Estimate	St. E.	t-Stat	p-Value	95 % CI	
Constant	8.465	0.172	49.340	0.000*	8.129	8.802
Population	0.007	0.000	16.620	0.000*	0.007	0.008
GDP per capita	−0.030	0.006	−5.010	0.000*	−0.042	−0.018
% of paved roads	0.000	0.000	1.070	0.285	0.000	0.001
% of population without access to public transport	0.000	0.000	−0.580	0.560	−0.001	0.000
% motorcycle	0.000	0.000	−1.800	0.071	−0.001	0.000
Seatbelt enforcement	0.000	0.001	−0.930	0.355	−0.002	0.001
Dispersion parameter φ	0.817	0.153	5.340	0.000*	0.517	1.117
LL	−978.300					
N	105					
P	8					
AIC	1972.6					
BIC	1993.831683					

* indicates a statistically significant effect at 95 % confidence level.

Table 4
Parameter estimates and goodness-of-fit of the Latent Class model of traffic mortality.

Variable	Estimate	St. E.	t-Stat	p-Value	95 % CI	
Class 1 (class probability = 0.238)						
Constant	9.305	0.516	18.030	0.000*	8.294	10.316
Population	0.002	0.000	6.440	0.000*	0.002	0.003
GDP per capita	−0.007	0.007	−1.090	0.278	−0.021	0.006
%of paved roads	−0.027	0.003	−8.770	0.000*	−0.032	−0.021
% of population without access to public transport	−0.021	0.007	−3.020	0.003*	−0.035	−0.007
% motorcycle	0.015	0.007	2.270	0.023*	0.002	0.028
Seatbelt enforcement	0.375	0.078	4.820	0.000*	0.222	0.527
Dispersion parameter	36.121	32.586	1.110	0.268	−27.748	99.989
Class 2 (class probability = 0.762)						
Constant	7.566	0.288	26.310	0.000*	7.002	8.130
Population	0.042	0.004	10.930	0.000*	0.034	0.049
GDP per capita	−0.024	0.006	−3.750	0.000*	−0.037	−0.012
% of paved roads	−0.003	0.002	−1.080	0.279	−0.007	0.002
% of population without access to public transport	−0.006	0.003	−2.000	0.045*	−0.011	0.000
% motorcycle	0.006	0.003	2.040	0.041*	0.000	0.012
Seatbelt enforcement	−0.024	0.050	−0.480	0.634	−0.123	0.075
Dispersion parameter	6.999	2.382	2.940	0.003*	2.331	11.667
LL	−485.700					
N	105.000					
P	17					
AIC	1005					
BIC	1050.517326					

* indicates a significant effect; highlighted rows show fixed variables effects that differentiate between classes.

Table 5
Parameter estimates and goodness-of-fit of the Latent Class model of COVID-19 mortality.

	Estimate	SE	t Stat	p Value	95 % CI	
Class 1 (class probability = 0.822)						
Constant	4.642	0.736	6.310	0.000*	3.200	6.084
Population	0.023	0.003	7.500	0.000*	0.017	0.030
Cases per population	0.052	0.010	5.380	0.000*	0.033	0.070
% of population older than 65	0.037	0.033	1.130	0.256	−0.027	0.101
Min Stringency index	0.010	0.013	0.740	0.459	−0.016	0.035
St Dev of Stringency index	0.042	0.032	1.310	0.191	−0.021	0.104
Diabetes prevalence	0.104	0.053	1.970	0.049*	0.000	0.207
Dispersion parameter	0.858	0.143	5.990	0.000*	0.577	1.139
Class 2 (class probability = 0.178)						
Constant	16.236	2.116	7.670	0.000*	12.089	20.383
Population	0.004	0.000	8.500	0.000*	0.003	0.005
Cases per population	0.043	0.007	6.220	0.000*	0.029	0.056
% of population older than 65	−0.122	0.054	−2.280	0.023*	−0.228	−0.017
Min Stringency index	−0.159	0.023	−6.820	0.000*	−0.204	−0.113
St Dev of Stringency index	−0.251	0.075	−3.370	0.001*	−0.398	−0.105
Diabetes prevalence	−0.246	0.098	−2.500	0.012*	−0.439	−0.053
Dispersion parameter	2.962	2.203	1.340	0.179	−1.355	7.279
Class membership						
Class 1						
Constant	2.531	0.738	3.430	0.001	1.084	3.977
GDP Per Capita	−0.049	0.025	−1.970	0.049	−0.097	0.000
Class 2						
Constant	Reference					
GDP Per Capita						
LL	−1038.400					
N	105.000					
P	18					
AIC	2112					
BIC	2160.5713					

* indicates a significant effect; highlighted rows show fixed variables effects that differentiate between classes.

with higher GDP are less likely to belong in Class 1. The parameter estimates per class can be interpreted as follows:

- The number of COVID-19 cases per population is an effective measure of risk exposure, as it is found to be positively associated with mortality rates in both classes.
- The share of population older than 65 years is non significant in class 1, but significant in class 2. A negative coefficient in class 2 suggests that a higher share of elderly in the population is associated with a lower mortality rate. Although there have been studies suggesting the opposite, it should be noted that, at the macroscopic level, a higher share of elderly corresponds to higher GDP and overall wellbeing of the population.
- The impacts of the stringency index of COVID-related restrictions is non significant in class 1, but significant in class 2. More specifically, a higher minimum stringency index (indicating stricter baseline

restrictions to contain the spread of the virus) and a higher variation of the stringency index (indicating a larger range of restrictions implemented) are associated with lower mortality rates in class 2 countries.

- The prevalence of diabetes in the population has opposite effects in the two classes. In Class 1, a higher prevalence of diabetes is associated with higher COVID-19 mortality (although the effect is marginally significant at 95 % confidence level), while in class 2 it is associated with a lower COVID-19 mortality.

The results suggest that LMICs are more likely to belong to class 1, whereas industrialised countries are more likely to belong to class 2. From this perspective, the non-significant effect of stringency index in class 1 can be attributed to the overall low stringency of COVID-19 restriction policies during the year 2020 (a year in which mostly industrialised countries were heavily affected by the outburst of the pandemic); the low size and variability of the stringency index in LMICs may then result in this non-significance. Accordingly, the not statistically significant effect of ageing population may be considered reasonable given the smaller shares of elderly people in many disadvantaged countries. At the same time, prevalence of diabetes and ageing population can be considered "luxuries" of the highly industrialised countries, who had more resources and institutional capacity to apply policies for successfully limiting mortality from COVID-19.

In this case too, KPIs expressing the operational level of safety (e.g. the reproduction rate) were not found to be significant, while structural, exposure and policy indicators are found to play a more important role. It is possible that these KPIs, conceptually related to behavioural characteristics that may result from local attitudes, safety culture and compliance, can not be clearly associated with fixed country effects and are instead captured by the latent factors included in the model.

4.3. Commonalities between the two mortalities

The modelling results of the two causes of mortality confirm the first hypothesis in this study in that there are common factors contributing to the two. Population is a structural measure in both causes of mortality which is associated with the first layer of the pyramid. Similarly, the GDP per capita withing the traffic mortality, and the prevalence of diabetes within the COVID-19 mortality are also structural measures. The share of paved roads and population without access to public transport and the share of motorcycles within he traffic mortality and the share of population older than 65 years old within the COVID-19 mortality are both exposure metrics. Seatbelt enforcement influencing the traffic mortality and stringency index influencing the COVID-19 mortality are factors associated with the second layer of the pyramid, safety measures and programs.

These results indicate that while the two causes of mortality are fundamentally different in their mechanisms, from the macroscopic perspective, the same underlying factors contribute to both of them. This is intuitive, as structural factors and policies would play a key role in all epidemiological analyses. Moreover, the two examined causes of mortality have been found to be inter-related, at least during the examined year; traffic fatalities were reduced in most countries, due to the mobility restrictions, while covid-19 fatalities were increasing. It has been found that many travel-related activities might increase the spread of viruses such as the coronavirus responsible for Covid-19. Therefore, the two causes of mortality have been correlated, and it is likely that such correlations may be observed in the future.

5. Discussion

The results of our testing latent class modelling approach on the traffic and COVID-19 mortality show the added value of using advanced statistical models to capture this type of unobserved heterogeneity in macroscopic country mortality modelling. It is found that neither GDP nor geographical classification can sufficiently capture this effect. Without any consideration of latent country characteristics, fixed linear effects of variables may be found, however the overall variability accounted for is low. Many of the variables that were statistically significant in the linear model of the previous paper (Papadimitriou & Afghari, 2023) are not statistically significant once the Poisson assumptions are employed – they are capturing the over-dispersion in the data and once this was accounted for they were found not statistically significant. On the other hand, by clearly separating countries in clusters or other groups, there may be too little or too much within-group variability that leads to difficulty in identifying country-specific effects.

The latent class modelling proved to be efficient in setting up a latent background for country membership, in which countries are more likely to belong, but do not do so in a deterministic way. This allowed to identify mixed effects of certain key variables in each case. For COVID-19 mortality, the stringency index of policies is found to be a meaningful variable only in certain countries (and most probably the industrialised ones). For traffic mortality, the transport infrastructure, demand and supply variables such as the share of motorcycling, paved roads and the quality of public transport provisions were found to explain the fatalities only in certain countries (and most probably LMICs). Moreover, variables such as GDP and aged population were found to have significantly different magnitude and / or sign in different classes of countries. This appears to confirm that simple linear modelling may easily result in partial effects cancelling each other out at the macroscopic level.

The results confirm that the proposed taxonomy of indicators is useful for the models development. Although the two causes of mortality have different mechanisms, and it is not fully understood whether a 'direct' comparison would be meaningful (and therefore this is not the scope of this paper), a conceptual framework for structuring the proposed models in a consistent way was needed. It was found that structural, socioeconomic and welfare characteristics, as well as policy-related indicators are the main determinants of the mortality rates when global level modelling is performed. The lack of behavioural KPIs in the models is somewhat counter-intuitive. However, it is also likely that these effects are better captured by the latent class structure tested here than fixed effects of the KPI variables available in the dataset, since such behavioural KPIs are difficult to measure reliably. Indeed, the available data include very few relevant indicators.

All available variables shown in Table 2 were considered for the models development. Some of them were rejected and not tested due to very small variability between countries (e.g. the existence of a traffic safety Lead Agency, visions and targets for road safety). Others were rejected due to strong correlations with other variables (e.g. the human development index); in each case, the variables selected as optimal ones to be included were those with a significant and robust effect (i.e. effect that was not changing in different model formulations), which was also in line with the existing literature on macroscopic risk modelling in general.

Our study has certain limitations. First of all, our results are dependent on the quality of the data available in international databases. Although we used formal international statistics of both causes of mortality, these are known to suffer from under-reporting of fatalities or inaccurate reporting of independent variables. In this study, we used the WHO-estimated number of traffic fatalities, which is estimated on the basis of multivariate statistical models that take into account the degree of under-reporting in those countries that do not have valid data from death certificates. Recent studies have shown that the WHO-estimated fatalities are a very good estimate of the real number of fatalities, as later corrected by countries implementing cross-sectoral studies (e.g. Thailand, Tunisia) (Papadimitriou et al., 2019). Nevertheless it is important to be aware of these limitations and update such analyses once more reliable data become available. In its current form, the scope of our study was to demonstrate the added value of the methodology and not to conclude on the effects of all indicators on mortality. Recent opinion articles have highlighted the potential lessons to be learned by the Covid-19 pandemic for traffic safety (Jobs, 2020; Yannis, 2020); our study can be useful in a preliminary attempt to jointly assess the two causes of mortality in terms of certain common underlying attributes. A more in-depth study, taking into account the full existing knowledge in the traffic and COVID-19 fatalities' determinants, may lead to more conclusive models on their full macroscopic mortality mechanism.

This paper assumes that the unobserved heterogeneity on mortality effects is due to cultural factors. However, cultural variables were not explicitly included in the dataset. Future research should test the cultural dimensions of Hofstede (2001), who defined country culture on the basis of 6 factors: the power distance index (PDI), individualism vs. collectivism (IDV), masculinity versus femininity (MAS), uncertainty avoidance index (UAI), long term orientation versus short term normative orientation (LTO), and indulgence versus restraint (IVR).

Finally, our results for the two examined causes of mortality may not be representative of other causes of mortality.

6. Conclusions

The modelling results presented above confirm our hypothesis that there are latent country characteristics that affect the impact of socioeconomic factors on different mortalities. The latent class structure associates a probability of country membership to a class, suggesting that there is not a one-to-one country membership, but each country is more likely to belong to a certain class on the basis of its characteristics, although the same country may include features that would be more prevalent in the other class. This is deemed to be a far more accurate representation of underlying country characteristics. For example, there are some European countries that are far less economically robust as others, and may resemble certain LMICs in some (but not all) areas, e.g. Eastern European countries. On the other hand, there may be LMICs that are considered as such from a geographical perspective (e.g. belonging to a less industrialised continent), but may exhibit good performance in certain socioeconomic indicators. This ambiguity in the country classification criteria can be captured by the latent class structure.

The results of our study can be useful to both epidemiology and traffic safety experts working on the topics of mortality developments and country benchmarking. For instance, they can be useful to international organisations and researchers / research groups dealing with heath, transport and other causes of mortality; by developing advanced econometric models, new knowledge and improved understanding of global patterns can emerge, leading to the identification of new good practices. They can also be useful to policy makers in the health, transport or other mortalities sectors, seeking to benchmark their country or region in a reliable and informative way. Moreover, transport planners and system operators can use these models to adjust their transport systems according to international good practice by drawing from a more targeted pool of countries and settings with which they share common latent attributes, making thus the systems safer and more resilient.

It is recommended that more in-depth research is carried out within each specific domain / mortality cause, in order to develop the sophisticated and dedicated models needed for a full picture, on the basis of the methodological considerations proposed in this paper. The monitoring of factors that influence various sources of mortality can help anticipate future 'outbursts' and better understand the ways in which epidemiology, transport and safety policies can jointly affect them, while taking into account the contextual factors of different countries.

CRediT authorship contribution statement

Eleonora Papadimitriou: Conceptualization, Data curation, Investigation, Methodology, Writing – original draft. **Amir Pooyan Afghari:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. **Pieter van Gelder:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Antonietti, R., Falbo, P., Fontini, F., 2023. The wealth of nations and the first wave of COVID-19 diffusion. Ital. Econ. J. 9 (2023), 61–83.

Antoniou, C., Yannis, G., Papadimitriou, E., Lassarre, S., 2016. Relating traffic fatalities to GDP in Europe on the long term. Accid. Anal. Prev. 92, 89–96.

Arulanandam, B., Beladi, H., Chakrabarti, A., 2021. COVID-19 mortality and the overweight: cross-country evidence. Public Health in Practice 2, 100179.

Bert, G., Verschuere, B., Wayenberg, E., Bishoy, L.Z., 2022. A guide to benchmarking COVID-19 performance data. Public Adm. Rev. 80 (4), 696–700.

Bjørnskov, C., 2021. Did lockdown work? an economist's cross-country comparison (2021). CESifo Econ. Stud. 67 (3), 318–331.

Bliss, A., Breen, J., 2009. Country guidelines for the conduct of road safety capacity reviews and the specification of lead agency reforms, investment strategies and safe system projects. World Bank Global Road Safety Facility, Washington, D.C.

Borsos, A., Koren, C., Ivan, N.J., Ravishanker, N., 2012. Long-term safety trends as a function of vehicle ownership in 26 countries. Transportation Research Record: Journal of the Transportation Research Board 2280, 154–161.

Bouba, Y., Tsinda, E.K., Fonkou, M.D.M., Mmbando, G.S., Bragazzi, N.L., Kong, J.D., 2021. The determinants of the low COVID-19 transmission and mortality rates in Africa: a cross-country analysis. Front. Public Health 9, 51197.

Chen, B., Liu, Y., Yan, B., Wu, L., Zhang, X., 2023. Why were some countries more successful than others in curbing early COVID-19 mortality impact? a cross-country configurational analysis. Plos ONE 18 (3), e0282617.

Colonna, P., Intini, P., 2020. Compensation effect between deaths from Covid-19 and crashes: the Italian case. Transp. Res. Interdiscip. Perspect. 6, 100170.

Commandeur, J.F.C., Bijleveld, F., Bergel, R., Antoniou, C., Yannis, G., Papadimitriou, E., 2013. On statistical inference in time series analysis of the evolution of road safety. Accid. Anal. Prev. 60, 424–433.

De Vos, J., 2020. The effect of COVID-19 and subsequent social distancing on travel behavior. Transp. Res. Interdiscip. Perspect. 5, 100121.

Debajyoti, C., Bhanu, B., Maneka, J., Low, D., Relative deprivation, inequality and the Covid-19 pandemic (2023). Social Science & Medicine 324, 115858.

El Mouhayyar, C., Jaber, L.T., Bergmann, M., Tighiouart, H., Jaber, B.L., 2022. Country-level determinants of COVID-19 case rates and death rates: an ecological study. Transbound. Emerg. Dis. 69, e906–e915.

European Road Safety Observatory (2021). *Country profiles*, 2021. Available on-line at: https://road- safety.transport.ec.europa.eu/statistics-and-analysis/data-and-analysis/country-profiles_en . Accessed July 2022.

Hedlund, J., Arnold, R., Cerrelli, E., Partyka, S., Hoxie, P., Skinner, D. An assessment of the 1982 traffic fatality decrease (1984) Accident Analysis and Prevention 16(4), pp. 217–261.

Hermans, E., Brijs, T., Wets, G., Vanhoof, K., 2009. Benchmarking road safety: lessons to learn from a data envelopment analysis. Accid. Anal. Prev. 41, 174–182.

Hofstede, G., 2001. Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations, 2nd ed. SAGE, London, England.

Jobs, S. (2020) Can COVID-19 teach us something for the road safety epidemic? Published in Transport for Development, a World Bank blog. Available on-line at: https://blogs.worldbank.org/transport/can-Covid-19-teach- us-something-road-safety-epidemic. Accessed July 2022.

Kapitsinis, N., 2021. The underlying factors of excess mortality in 2020: a cross-country analysis of pre-pandemic healthcare conditions and strategies to cope with Covid-19. BMC Health Serv. Res. 21, 1197.

Katrakazas, C., Michelaraki, E., Sekadakis, M., Yannis, G., 2020. A descriptive analysis of the effect of the COVID-19 pandemic on driving behavior and road safety. Transp. Res. Interdiscip. Perspect. 7, 100186.

Koornstra, M., Lynam, D., Nilsson, G., Noordzij, P., Pettersson, H.E., Wegman, F., Wouters, P., 2002. (2002) SUNflower - A comparative study of the development of road safety in Sweden, the United Kingdom, and the Netherlands. Final report. SWOV Institute for Road Safety Research, Leidshcendam.

Kopits, E., Cropper, M., 2005. Traffic fatalities and economic growth. Accid. Anal. Prev. 37, 169–178.

Kweon, Y.-J., 2015. What affects annual changes in traffic safety? A macroscopic perspective in Virginia. J. Safety Res. 53, 17–21.

Lassarre, S., 2001. Analysis of progress in road safety in ten European countries. Accid. Anal. Prev. 33, 743–751.

Liang, L.L., Tseng, C.H., Ho, H.J., et al., 2020. Covid-19 mortality is negatively associated with test number and government effectiveness. Sci. Rep. 10, 12567.

Liu, J., Eggleston, K., 2022. The association between health workforce and health outcomes: a cross-country econometric study. Soc. Indic. Res. 163, 609–632.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research 11, 1–16.

Marginean, S., Orastean, R., 2022. Health spending patterns and COVID-19 crisis in European union: a cross-country analysis. Systems 10, 238.

Page, Y., 2001. A statistical model to compare road mortality in OECD countries. Accid. Anal. Prev. 33, 371–385.

Papadimitriou, E., Afghari, A.P., (2023) Exploring common attributes between the traffic safety pandemic and the Covid-19 pandemic: a global macroscopic analysis, Presented at the Annual Meeting of the Transportation Research Board, Washington DC, 8-13 January 2023.

Papadimitriou, E., Iaych, K., Adamantiadis, M., 2019. Understanding and bridging the difference between country-reported and WHO-estimated road traffic fatality data. Joint Report of the EuroMed Transport Support Project and the World Health Organization. WHO, Geneva.

Papadimitriou, E., Yannis, G., 2013. Is road safety management linked to road safety performance? Accid. Anal. Prev. 59, 593–603.

Peralvo, F.C., Vanegas, P.C., Avila-Ordóñez, E., 2022. A systematic review of COVID-19 transport policies and mitigation strategies around the globe. Transp. Res. Interdiscip. Perspect. 15, 100653.

Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C, Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D. Roser, M., (2020) Coronavirus Pandemic (COVID-19), 2020. Published online at OurWorldInData.org. Retrieved from: https://ourworldindata.org/coronavirus' [Online Resource].

Saladié, Ò., Bustamante, E., Gutiérrez, A., 2020. COVID-19 lockdown and reduction of traffic accidents in Tarragona province Spain. Transp. Res. Interdiscip. Perspect. 8, 100218.

UNECE Sustainable Transport Division, 2018. SafeFITS – Safe Future Inland Transport Systems, 2018. UNECE, Geneva.

Washington, S., et al., 2020. Statistical and econometric methods for transportation data analysis. CRC Press, 2020.

Wegman, F., Oppe, S., 2010. Benchmarking road safety performances of countries. Saf. Sci. 48 (9), 1203–1211.

WHO, 2018. Global Status Report on Road Safety. WHO, Geneva, p. 2018.

WHO (2020) Director-General's opening remarks at the media briefing on COVID19 -March 2020.

WHO (2022) Coronavirus Covid-19 Dashboard https://covid19.who.int/ (Accessed July 2022).

Yannis, G., Antoniou, C., Papadimitriou, E., Katsochis, D., 2011. When may road fatalities start to decrease? J. Saf. Res. 42 (1), 17–25.

Yannis, G., Papadimitriou, E., Folla, K., 2014. Effect of GDP changes on road traffic fatalities. Saf. Sci. 63, 42–49.

Yannis, G. (2020). Road Traffic Accidents Pandemic versus Covid-19 Pandemic, 2020. Available on-line at: https://www.georgeyannis.com/road-traffic-accidents-pandemic-versus-Covid-19-pandemic/. Accessed July 2022.

Yu, H., Li, Z., Zhang, G., Liu, P., 2019. A latent class approach for driver injury severity analysis in highway single vehicle crash considering unobserved heterogeneity and temporal influence. Analytic Methods in Accident Research 24, 100110.