# BR-EMS 2021 life table for the Brazilian insured population

Mario de Oliveira[*]
Ana Carolina Soares Bertho[**]
Bruno Costa[***]
Flávia Sommerlatte Silva[****]
Mariane Branco Alves[*****]
Milton Ramos Ramirez[******]
Rafael Brandão de Rezende Borges[*******]
Reinaldo Marques[********]
Ricardo Martins da Silva Rosa[*********]
Rodrigo Lima Peregrino[**********]
Viviana das Graças Ribeiro Lobo[**********]
Thais Cristina Oliveira Fonseca[***********]

[*] Instituto de Matemática, Universidade Federal do Rio de Janeiro (IM/UFRJ), Rio de Janeiro-RJ, Brazil (mario@labma.ufrj. br; https://orcid.org/0009-0005-8446-8237).
[**] Escola Nacional de Ciências Estatísticas, Instituto Brasileiro de Geografia e Estatística (Ence/IBGE), Rio de Janeiro-RJ, Brazil (carolina.bertho@ibge.gov.br; https://orcid.org/0000-0003-4822-1948).
[***] Instituto de Matemática, Universidade Federal do Rio de Janeiro (IM/UFRJ), Rio de Janeiro-RJ, Brazil (bcosta@im.ufrj. br; https://orcid.org/0009-0004-1740-3380).
[****] Mercer, Lisboa, Portugal (flavia.sommer@gmail.com; https://orcid.org/0000-0002-3301-6274).
[*****] Instituto de Matemática, Universidade Federal do Rio de Janeiro (IM/UFRJ), Rio de Janeiro-RJ, Brazil (mariane@im.ufrj. br; https://orcid.org/0000-0002-2489-9300).
[******] Instituto de Matemática, Universidade Federal do Rio de Janeiro (IM/UFRJ), Rio de Janeiro-RJ, Brazil (milton@im.ufrj. br; https://orcid.org/0000-0003-0367-5305).
[*******] Universidade Estadual do Rio de Janeiro (UERJ), Rio de Janeiro-RJ, Brazil (rborges@labma.ufrj.br; https://orcid. org/0000-0002-2576-8118).
[********] Universidade Federal de Alfenas (Unifal-MG), Varginha-MG, Brazil (reinaldo.marques@unifal-mg.edu.br; https:// orcid.org/0000-0001-9714-6740).
[*********] Instituto de Matemática, Universidade Federal do Rio de Janeiro (IM/UFRJ), Rio de Janeiro-RJ, Brazil (ricardo@ im.ufrj.br; https://orcid.org/0000-0001-9808-9936).
[**********] Laboratório de Matemática Aplicada, Universidade Federal do Rio de Janeiro (LabMA/UFRJ), Rio de Janeiro-RJ, Brazil (rodrigo@matematica.ufrj.br; https://orcid.org/0009-0008-6834-6493).
[**********] Instituto de Matemática, Universidade Federal do Rio de Janeiro (IM/UFRJ), Rio de Janeiro-RJ, Brazil (viviana@ dme.ufrj.br; https://orcid.org/0000-0002-4076-8327).
[***********] Instituto de Matemática, Universidade Federal do Rio de Janeiro (IM/UFRJ), Rio de Janeiro-RJ, Brazil (thais@ im.ufrj.br; https://orcid.org/0000-0002-4943-3259).

This article presents the Brazilian private insurance market's actuarial life tables, BR- EMS 2021. Using Bayesian inference on the parameters of the Heligman- Pollard law of mortality and data from 23 insurance groups over 15 years, totaling 3.5 billion registers, the data were corrected through a two hidden-layer neural network. The resulting tables show that the insured population exhibits lower mortality rates than the general Brazilian population, even lower than the national populations of well-developed countries such as the USA. Moreover, besides the expected gender gap in mortality rates, there is a clear distance between the death and survivorship insurance coverage groups. Likewise, the insured population characteristics mitigate well-known regional structural discrepancies in the Brazilian population, indicating that being part of the selected population of insured individuals is thus associated with a more effective protection against death than other outstanding factors such as geographic region of residence.

**Keywords**: Actuarial life tables. Death and survivorship coverages. Mortality graduation. Heligman-Pollard model

## Introduction

Over recent years, life expectancy in Brazil has significantly increased, rising from 52.7 years for both sexes in 1960 (UN, 2022) to 76.8 years in 2020 (IBGE, 2021). However, despite this overall improvement, vast differences remain in mortality rates among sub-national groups. For example, in 2010, the difference in life expectancy between southern and northern populations was almost 5 years (IBGE, 2013a). While socioeconomic factors are known to be associated with adult mortality, such data is often inaccurate or restricted. Therefore, researchers often explore other factors that may explain differences in mortality rates among large and heterogeneous populations.

Since 1999, the Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística*, IBGE) has produced and published national life tables on a yearly basis in compliance with a federal decree (IBGE, 2021). These tables are based on mortality projections that reference the year of the last demographic census and do not include subdivisions by educational or income groups. In recent years, researchers have proposed studies to construct life tables for specific strata of the Brazilian population. For example, Siviero, Souza, and Machado (2019) present gender-specific mortality tables for two-time points and various causes of death in the city of São Paulo. Similarly, Ribeiro, Turra, and Pinto (2021) discuss mortality differences based on educational attainment in São Paulo and report a 77% reduction in mortality rates for adults with higher levels of education. Gonzaga *et al.* (2022) analyzed the mortality differentials among beneficiaries from the National Institute of Social Insurance (INSS, in Portuguese). The authors compared their findings with IBGE life tables, though not expecting to find relevant differences, since INSS only excludes public servants and population out of social insurance and without private pension. However, they found lower probabilities of death for all INSS subgroups when compared to IBGE life tables, except Continuous Provision Benefit (BPC, in Portuguese), a

type of benefit granted to elderly people in poverty. Consequently, Gonzaga *et al.* (2022) identified higher life expectancies for INSS beneficiaries compared to those calculated by the IBGE. Di Lego, Turra and Cesar (2017) developed another Brazilian research about mortality differentials that focused on a specific population. Using the Brazilian Air Force (BAF) data, the authors found a higher life expectancy for BAF officers compared to the average Brazilian. Although in this case the analysis is restricted to men, the authors emphasize that the age-specific mortality ratios were similar to those registered in Sweden, Japan and France. Both studies, conducted by Gonzaga *et al.* (2022) and by di Lego, Turra and Cesar (2017), reveal cases in which selected populations have better life conditions than the average population, which can be measured in lower probabilities of death and higher life expectancies.

Similarly, Brazil's insured population has historically shown significantly lower mortality rates than the overall national population. This has prompted several joint public-private initiatives to propose mortality tables for subgroups of this population, which consists of individuals with life insurance policies or annuities. The BR-EMS tables (*Experiência do Mercado Segurador Brasileiro* – the experience of the Brazilian insurance market), published in 2010 (OLIVEIRA *et al.*, 2012), represent the first mortality tables proposed for the Brazilian insurance market. Developed by LabMA/UFRJ and published by the national government regulatory agency for the insurance market (*Superintendência de Seguros Privados*, SUSEP), the BR-EMS 2010 tables were based on data from 23 insurance groups covering the years 2004-2006 and comprising over 39 million individuals. These tables indicated a difference in life expectancy for males with survivorship coverage compared to males in the overall population, with a gap of 11.16 years (81.36 and 70.20 years, respectively). Prior to the development of the BR- EMS tables, the national insurance market relied on foreign tables such as the American Annuity 2000 Mortality Table (AT-2000) (JOHANSEN, 1995; AMERICAN ACADEMY OF ACTUARIES, 2002). In the Brazilian market, the AT-2000 table started to be used as soon as the AT-83's mortality experience came to be considered inadequate in the actuarial practice. The AT-2000 was built by applying the Improvement Scale G to the AT-83 table, from 1983 to 1996, and the resulting table was graduated to form a Basic Table from 1996 (SOCIETY OF ACTUARIES, 1983). After the graduation, a 10% charge was subtracted to meet the Market needs in terms of level of mortality. The Scale G is an improvement scale developed in conjunction with the AT-83 for projecting the mortality probabilities (qx) of a life table. Thus, although the AT-2000 life table was widely applied in the Brazilian life insurance and pension plan market, it did not necessarily reflect the Brazilian experience. It was therefore urgent to construct a life table based on Brazilian empirical data. The book by Oliveira *et al.* (2012) describes the database and methodology used to develop the BR-EMS 2010 tables.

Due to the sustained increase in life expectancy of the Brazilian population in the past decades, there is a need to keep the BR-EMS tables updated. This has already been done twice, with a previous version published in 2015, based on data spanning 2004-2012

(OLIVEIRA *et al.*, 2016), and the most recent version (BR-EMS 2021), based on data from 2004 to 2017 and published by the Brazilian Institute of Actuary (Instituto Brasileiro de Atuária, IBA). These tables focus on mortality rates of insured sub-national groups in Brazil (male, female, death, and survivorship coverages). From a modeling point of view, the BR-EMS 2021 tables were estimated assuming a Binomial sampling model and Heligman-Pollard law with parameters estimated using the Bayesian paradigm (GAMERMAN; LOPES, 2006). Other studies in Brazil considered this law of mortality table estimation, such as Beltrão and Sugahara (2017) that investigated the mortality among federal civil servants from 1993 to 2014, highlighting differences by educational level assuming a Binomial distribution of deaths and fitting a modified Heligman-Pollard model with the exclusion of the first term, related to infant mortality. Even among public servants who only have high school, the probabilities of death found by Beltrão and Sugahara (2017) were lower than those calculated by IBGE (2013). The authors considered the administrative record of employees (SIAPE) as an identifier and emphasized the importance of using numerator and denominator from the same source, as it reduces sub-enumeration. Likewise, in the case of the data used in the BREMS tables, numerator and denominator have the same source, reducing problems with data quality. Di Lego, Turra and Cesar (2017) used a Poisson regression model to estimate the BAF life tables, examining 10-year age groups, starting with the age group 15-24. Gonzaga *et al.* (2022) investigated mortality inequalities among beneficiaries from INSS in 2015 for individuals older than 65 years old, using three methodologies: Gompertz model, Van der Maen model and Topals regression.

It is also essential to mention that, since its initial version, BR-EMS 2010, the tables have been used throughout Brazil's life insurance and pension market for a so-called liability adequacy requirement, and carried out every calendar year by the insurance companies, as defined by the regulatory agency (SUSEP). The health insurance market has also instituted this requirement since 2020 through its specific regulatory agency (ANS – National Health Insurance Agency). In addition, many products now have the BR-EMS as reference tables for calculating benefits and reserves. Even though the BR-EMS 2021 tables have been available to the public since 2021, the study to construct the tables, the filtering techniques used, the statistical methodology and details about estimation were never published and represent original work. Thus, in this paper, we compare the behavior of the BR-EMS 2021 tables with the national mortality tables issued by IBGE, confirming and updating the significant differences in mortality behavior between the insured and general populations. We also discuss advances in government and market databases, the main reason for this improvement being the universalization of CPF (*Cadastro de Pessoas Físicas)*. This national identification number is now mandatory for every individual, including children of any age. The data was corrected using neural network techniques to construct a coherent database for such a large and complex amount of information. A comparison of mortality rates by region in Brazil reveals that the insured population's tables have a much more homogeneous behavior across regions, indicating that better living conditions of the

insured population mitigate regional structural differences. Furthermore, the mortality rates of Brazilian insured population are comparable to those of well-developed countries such as the United States and Chile.

The paper is organized as follows: section 2 describes the database, the data processing, and filtering techniques. Section 3 details the modeling process based on the Bayesian Heligman-Pollard model used to graduate the BR-EMS 2021 mortality tables. Section 4 presents the BR-EMS 2021 tables and discusses the main features of the resulting estimated mortalities. This section also compares life expectancies calculated via the BR-EMS 2021 and the IBGE 2020 tables. Section 5 compares the insured population tables and the national tables for the five macro-regions of Brazil. Section 6 portrays the BR-EMS 2021 and international tables of well-developed countries. Finally, Section 7 presents the study's conclusions.

## The database

This section provides a description of the data treatment applied to the empirical data to establish a coherent and reliable database, which served as the basis for calculating the observed death rates of BR-EMS 2021. Note that the record files, generated annually by the insurance companies and used in constructing this database, were obtained under a strict confidentiality agreement with the National Private Pension and Life Association (FenaPrevi). The set of files encompass an impressive volume of information, exceeding 3.5 billion records, pertaining to active insurance plans with records containing data on about 94 million individuals from 2004 to 2017. The number of records by year varies from 50 million in 2004 to more than 200 million in 2016, with the death coverage representing around 88% of all data. These records contain valuable data, including the insured individuals' gender, date of birth, and pertinent details regarding their risk exposure, such as the dates of plan initiation and termination due to contract expiration, plan cancellation, or death. Figure 1 illustrates the exposure and number of deaths before and after filtering from 2004 to 2017. With such complex and large amounts of data, it is necessary to build coherence in all this information gathered from several sources. For instance, occasionally, the death of an individual is reported by only one company or a death event is sometimes misreported as the end of a contract (and vice versa). Such inconsistencies, or absence of death information, demanded verification of this information for all insured individuals against the Brazilian government's Social Security information systems. This process was undertaken through a technical cooperation agreement between LabMA/UFRJ and the Ministry of Labor and Social Security. LabMA/UFRJ shared a list of personal identification numbers, retrieved from the database assembled using insured company datasets, with the Data Provider in charge of the Ministry's information system. The Data Provider conducted a systematic search within its information system and returned an identical list augmented with information such as date of birth, gender, and date of death. It is noteworthy that

despite being a government-managed database, it does not encompass data on all Brazilian citizens. The Data Provider affirmed this limitation, specifying that the consulted systems were the CNIS (Cadastro Nacional de Informações Sociais), the national register of social information, and SISOB (Sistema de Controle de Óbitos), the Death Registration System.

Actuarial life tables are built from the number of deaths and exposure to risk at each age. The sex and birth and death dates are extracted from the companies' databases, where all records are associated with a national identification number, the CPF, a unique numeric identifier for all Brazilian individuals, referred to hereinafter as NID. This identifier is used to merge information from different databases. The verification of data is carried out using the list of all NIDs extracted from the companies' databases and concatenated with the social security information of gender, date of birth and, when available, date of death, occupation code and level of education of each individual.

We name the combination of a NID, sex, and date of birth as a triad; although it is a unique number for each individual, prior to 2014, there were records of different insured individuals using the same NID number. For instance, the NID of a family member could be used for children and spouses. We refer to these as wildcard NIDs. These wildcard NIDs represent approximately 18% of the nearly 96 million NIDs present in the database for the years 2004 to 2017. After 2014, there has been a growing momentum in using the NID number as the primary key for identifying data of Brazilian individuals stored in insurance companies' databases as well as government registries. Babies, children, and spouses started to be registered under their own NID numbers. This has resulted in a better quality of the information received by the companies when compared to previous years, particularly regarding the presence of few or no wildcard NIDs.

In previous versions of the BR-EMS table, Wildcard NIDs related to at most four triads were taken into account, where each triad was considered a different person. However, Wildcard NIDs have poor quality information because death registration is accurate only for the "owner" of the NID, resulting in a lower mortality rate.

The new mandatory requirement of unique NIDs even for babies since 2014 allowed for a more precise definition of an individual in the 2021 table. Thus, to decrease the number of Wildcard NIDs, we designed a neural network algorithm, trained on the Government database, to correct the sex information of all triads, allowing the posterior correction of all birth dates as well. The motivation for using neural networks for data correction is their effectiveness in learning complex and non-linear structures from data. They are being used in various applications as a universal method of approximating any function (see Universal Approximation Theorem). The method considered for correction was a two hidden-layer neural network, with the feature being the proportion of times each sex appeared in each company and training was performed by verifying the sex against the sex informed by government's Social Security information systems. The first layer of the network consists of 50 neurons, and the second layer consists of 5 neurons, both with ReLU activation

function. The loss function considered in the training step to penalize poor predictions was the binary cross-entropy given by

$$H(u,p) = -\frac{1}{N}\sum_{i=1}^{N} u_i \, log(p_i) + (1-u_i) \, log(1-p_i) \tag{1}$$

where $u_i$ indicates the label (1 female and 0 male) and $p_i$ is the probability of observation $i$ being labeled $u_i$. The Adam optimization algorithm was employed, and the network was trained for 200 epochs. The implementation of this neural network was done using the Scikit-learn library in the Python language (VAN ROSSUM; DRAKE JR., 1995), ensuring the model's self-sufficiency and independence from proprietary software. These corrections resulted in unique sex and date of birth for 97.7% of all NIDs. The remaining 2.3% have been discarded. In the various tests carried out during the study of BR-EMS tables, the neural network based on two hidden layers reached an accuracy of 95%, with low values of type 1 and type 2 errors, which was considered sufficient for the purposes of life table construction.

Historically, the registered number of deaths in the companies' database always fell short compared to the actual quantity since many deaths of insured individuals are commonly under-reported. Thus, all information on the number of deaths has to be enriched by verification of this data against the Brazilian government's Social Security information systems. After that, individuals were sorted into subpopulations separated by year, company, product, sex, and type of insurance coverage. These subpopulations were selected using four exclusion criteria (OLIVEIRA *et al.*, 2012): (i) information of spurious data provided given by the insurance company themselves, (ii) subpopulations that have a small number of individuals, (iii) subpopulations that fell outside of the established bounds (outliers), inspired by Fence's Tukey (1977), determined by a comparison of the actual and expected number of deaths under two extreme life tables (CSO 2001 as an extreme table for low mortality rates, and the IBGE 2005, which describes the Brazilian population, as the extreme for high mortality rates), and (iv) to having a nonstandard mortality pattern. The total number of deaths after the aggregation with the external database and the filtering processes is 1,189 thousand, with the following composition: 681 thousand (57%) were present only in the Government database; 109 thousand (9%) were informed only by the insurance companies; and 399 thousand of deaths (34%) informed by both. Figure 1 shows the extent of this data filtering process for each year from 2004 to 2017. The exposure after filtering for each year varies from 7,154 (2004) to 15,134 (2017) for males in the death coverage; from 5,653 (2004) to 11,934 (2017) for females in the death coverage; from 1,667 (2004) to 4,457 (2017) for males in the survivorship coverage; from 1,365 (2004) to 3,904 (2017) for females in the survivorship coverage; Regarding the changes in the exposure after filtering, they did not affect substantially the characteristics of the exposed. For instance, the percentual change for males was 10.7% in 2004 and 9.1% in 2017 for the death coverage, with similar results for the other variants.

Notice that the periods 2010-2011 and 2014-2015 present a substantial difference in the number of deaths before and after filtering. However, such atypical patterns were not consistent with the market experience. This indicates that it is crucial to consider filtering techniques to obtain a consistent database for mortality graduation.

FIGURE 1
Exposure and number of deaths before and after filtering
Brazilian insurance population – 2004-2017
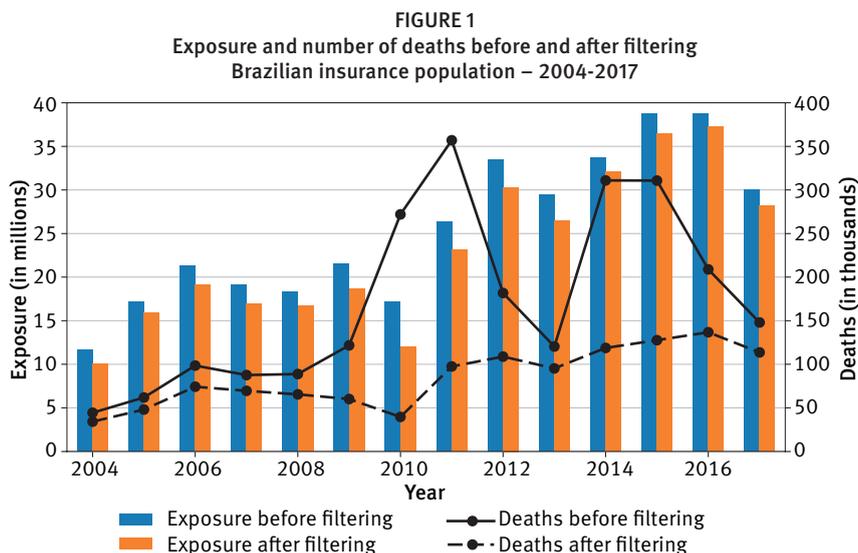


Source: Brazilian insurance companies participating in the analysis, CNIS, and SISOBI.

## Bayesian graduation of the BR-EMS mortality tables

The procedure detailed in Section 2 was used to obtain the yearly number of deaths $O_{c,s,x,t}$ and exposure $E_{c,s,x,t}$ for coverage $c$, sex $s$, age $x$, and year $t$, $t = 2004, \ldots, 2017$. For each coverage $c$, sex $s$ and age $x$, $O_{c,s,x}$ and $E_{c,s,x}$, were obtained using the exponentially weighted average of $O_{c,s,x,t}$ and $E_{c,s,x,t}$, assigning higher weights to recent data and increasingly lower weights to past information, respectively:

$$O_{c,s,x} = \Sigma_t \, \phi_t O_{c,s,x,t} \quad and \; E_{c,s,x} = \Sigma_t \, \phi_t \; E_{c,s,x,t} \tag{2}$$

with weights $\phi_{2017} = \frac{1}{2}, \phi_{2016} = \frac{1}{4}, \ldots, \phi_{2007} = \frac{1}{2048}, \phi_{2006} = \frac{1}{6144}, \phi_{2005} = \frac{1}{6144}, \phi_{2004} = \frac{1}{6144}$ . These weights result from a convex linear combination of the past (weight $\phi$) and the present (weight $1 - \phi$). As for the choice of the parameter $\phi$, if $\phi$ is very small, little weight is given to stability; if $\phi$ is large, little importance is given to innovation. The weight used considers $\phi - \frac{1}{2}$ to the most recent information (2017) and $1 - \phi = \frac{1}{2}$ to the past. More specifically, the weights $\phi$ follow a geometric progression, restricted to $\Sigma_{t=1}^T \phi_t = 1$, T the number of years used for table construction. To respect this restriction, the last two weights are equal. For instance, for a sequence of length 3, we have $\phi_1 = \frac{1}{2}, \phi_2 = \phi_3 = \frac{1}{4}$; and for length 4, we have $\phi_1 = \frac{1}{2}, \phi_2 = \frac{1}{4}, \phi_3 = \phi_4 = \frac{1}{8}$; and so on. Notice that the 2010 table was used as a starting point

assuming a simple average, while an exponentially weighted average was considered for the other years. Thus, for the weighted sum of interest we can write:

$$E_x^{BR-EMS\,2015} = \frac{1}{2}E_x^{2012} + \frac{1}{4}E_x^{2011} + \frac{1}{8}E_x^{2010} + \frac{1}{16}E_x^{2009} + \frac{1}{32}E_x^{2008} + \frac{1}{64}E_x^{2007} + \frac{1}{64}E_x^{BR-EMS\,2010}$$

with $E_x^{BR-EMS\,2010} = (E_x^{2006} + E_x^{2005} + E_x^{2004})/3$. The same rule applied to the 2021 table results in $E_x^{BR-EMS\,2021} = \frac{1}{2}E_x^{2017} + \frac{1}{4}E_x^{2016} + \frac{1}{8}E_x^{2015} + \frac{1}{16}E_x^{2014} + \frac{1}{32}E_x^{2013} + \frac{1}{32}E_x^{BR-EMS\,2015}$.

Putting these two equations together results in

$$E_x^{BR-EMS\,2021} = \frac{1}{2}E_x^{2017} + \frac{1}{4}E_x^{2016} + \frac{1}{8}E_x^{2015} + \frac{1}{16}E_x^{2014} + \frac{1}{32}E_x^{2013} + \frac{1}{64}E_x^{2012} + \frac{1}{128}E_x^{2011}$$

$$+ \frac{1}{256}E_x^{2010} + \frac{1}{512}E_x^{2009} + \frac{1}{1024}E_x^{2008} + \frac{1}{2048}E_x^{2007} + \frac{1}{6144}E_x^{2006}$$

$$+ \frac{1}{6144}E_x^{2005} + \frac{1}{6144}E_x^{2004}$$

The computation of $O_x^{BR-EMS\,2015}$, $O_x^{BR-EMS\,2021}$ follows an analogous rule.

We now describe the statistical methodology used to construct the insured population mortality tables (BR-EMS 2021). For each coverage $c$, each sex $s$ and each age $x$, the probability of death $q_{c,s,x}^{BR2021}$ was estimated assuming a probabilistic model for $O_{c,s,x}$ accounting for uncertainty through a full Bayesian inference approach. Following Oliveira *et al.* (2012), the Binomial smoothing model is formulated as

$$O_{c,s,x}\,q_{c,s,x}^{BR2021}\,Binomial\,(E_{c,s,x}, q_{c,s,x}^{BR2021}), \tag{3}$$

where $q_{c,s,x}^{BR2021}$ is the probability that an individual of age $x$ dies at age $x$ and before age $x + 1$ for coverage $c$ and sex $s$ and $x = 0, 1, 2, \ldots, \omega$, where $\omega$ represents the maximum age. For the younger to intermediary age group $[x_0, x_1]$, the probabilities $q_{c,s,x}^{BR2021}$ were graduated by assuming a nine-parameter version of the Heligman and Pollard model (1980) which is flexible and easily interpretable. This model was selected as the best from other competitors for the insured population table of 2010 (results in OLIVEIRA *et al.*, 2012). We follow this choice so that the methodology is comparable for all insured population tables. Other alternatives that could be considered for mortality graduation are nonparametric smoothing via penalized splines (HYNDMAN; ULLAH, 2007), bidimensional splines (CURRIE *et al.*, 2004) and Generalized additive models (DODD *et al.*, 2019). The first alternative aims to account for heterogeneity in the age dimension, which our model captures via the Binomial assumption, resulting in different variances for each age. The second alternative includes the cohort effect, which is not expected to be significant in our application, as the insured population table is updated every 5 years, and it should keep general characteristics in such a short period. The third option considers mortality estimation based on generalized additive models (GAMs) which lacks the interpretability of HP models. For each coverage $c$ and sex $s$, the death probability $q_x$ at age $x$ is modeled following the HP mortality law as

$$\eta_x = \frac{q_x}{1 - q_x} = A^{(x+B)^C} + De^{-(ln\,x + ln\,F)^2} + \frac{GH^x}{1 + KGH^x} \tag{4}$$

where $A, B, C, D, E, F, G, H$ and $K$ are unknown parameters and the mortality odds $n_x = q_x/(1 - q_x)$ is decomposed into three parts: a child mortality curve, an accident hump, and an adult

mortality curve. Parameters *A, B, C,* and *D* take values within the interval (0, 1), where *A* represents the infant mortality rate and *B* represents the mortality rate for one year old children. Parameter *C* is related to the rate of mortality decline, or the rate at which an individual adapts to his environment. $D \in (0,1)$ indicates the severity of the accident hump. Parameter *E* takes values in $(0,\infty)$ and is associated with the spread, with large values indicating a concentrated accident hump. Parameter *F* is the hump location parameter, with domain on the interval (15, 110). Finally, $G \in (0, 1)$ and $H \in (0,\infty)$ indicate the level and the increase rate of mortality at the adult ages, respectively.

Data for each table variant is typically scarce for ages greater than a certain value $x_1$, for each table variant. Nevertheless, for the male survival variant, data are consistent across all ages, up to 89 years. Thus, parameter *K*, which mainly influences the mortality laws at an older age, has been determined, for all the variants, according to the assumption of joint closure, from the age at which they first met, with the male survival modality curve. This assumption is based on the expected behavior for risk and benefit plans and implicitly reflects the belief that, conditional on having survived to a certain advanced age, deaths of all variants are ruled by the same mortality process.

Estimation of unknown parameters $\theta = (A, B, \ldots, K)'$ via least squares or weighted least squares presents convergence difficulties in many cases, as previously discussed. One of the main disadvantages of nonlinear least square methods is that they can be quite sensitive to the choice of starting values. This is even more important in the case of several interdependent parameters, such as in the HP model. Moreover, it lacks uncertainty quantification, and even in the event of some sort of asymptotic assumption based on Gaussian noise, the interval estimation for the nonlinear function of parameters is troublesome and depends on large samples. To overcome these difficulties of the weighted least square method, we consider the Bayesian estimation approach for the Heligman–Pollard (HP) model, which makes use of the posterior distribution of parameters to obtain point estimators, credibility intervals, and predictive distributions. Let D be the vector of observed deaths and exposures in all ages. If we assume a prior distribution $p_0(\theta)$, $\theta \in \Theta$ for the unknown parameters, then the posterior distribution $p_1(\theta \mid D)$ is obtained as

$$p_1(\theta \mid D) = \frac{f(D \mid \theta) p_0(\theta)}{g(D)} \tag{5}$$

where $f(D \mid \theta)$ denotes the Binomial model with death probabilities induced by (4), $p_0(\theta)$ denotes the prior density for θ, and

$$g(D) = \int_\theta f(D \mid \theta) p_0(\theta) d\theta \tag{6}$$

is the prior predictive distribution for the observables, for which there is no analytical solution, which makes us resort to computational methods for approximation of the posterior distribution via Markov Chain Monte Carlo (MCMC) methods (GAMERMAN; LOPES, 2006). Model specification is completed after assigning independent prior distributions for the vector θ. In particular, we assume priors such that A ~ Beta($a_A$, $b_A$), B ~ Beta($a_B$, $b_B$),

C ~ Beta(a$_C$, b$_C$), D ~ Beta(a$_D$, b$_D$), $E$ ~ Gamma(a$_E$, b$_E$), F ~ N(μ$_F$, σ$^2_F$ ), G ~ Beta(a$_G$, b$_G$) and H ~ Gamma(a$_H$, b$_H$). The hyperparameters a$_A$, b$_A$, a$_B$, b$_B$, . . . were selected to respect the restrictions in the domain for all parameters. The parameter $K$ was chosen after a sensitivity study. Thus, the computation of point estimators and credibility intervals based on equation (5) was obtained via MCMC by sampling θ from $p_1(θ \mid D)$ and using Monte Carlo integration to obtain the required summaries. That is, after performing the MCMC algorithm, we obtain θ$^{(1)}$,..., θ$^{(M)}$ which are independent samples from the posterior distribution of θ. Point estimates for any function $h(\cdot)$ of parameters θ may be obtained by Monte Carlo integration, with the posterior mean being the optimal estimator under quadratic loss, which is approximated by

$$E[h(\theta) \mid D] \approx \frac{1}{M} \sum_{j=1}^{M} h(\theta^{(j)}) \tag{7}$$

with (h(θ$^{(1)}$), …, h(θ$^{(M)}$)) being the function $h(\cdot)$ applied to each sampled value θ$^{(1)}$,..., θ$^{(M)}$.

For instance, the death probabilities are

$$q_x = \frac{e^{\eta_x}}{1 + e^{\eta_x}} \tag{8}$$

with η$_x$ defined in equation (4) being a nonlinear function of θ. The function is computed for each value of θ resulting in a sample q$_x^{MCMC}$ = (q$_x^{(1)}$, …, q$_x^{(M)}$). Point estimates can be obtained as in equation (7).

Credible intervals with 1 − $α$ credibility level for the parameters θ$_j$ or any nonlinear function g(θ$_j$) may be obtained through the marginal posterior distributions p$_{1j}$(θ$_j$ | D), j = 1, . . . , 9. We consider equal-tail credible interval approach with intervals given by CI(1-alpha; θ$_j$) = [$Q_{p1j}(α/2)$, $Q_{p1j}(1-α/2)$], where $Q_{p1j}(α)$ denotes the theoretical quantile $α$ under distribution p$_{1j}$. Another possibility considers the highest posterior density intervals (CHEN *et al.*, 1999). All these summaries are empirically approximated by operations on the Monte Carlo samples. For instance, the credible interval of 95% for the death probabilities is estimated as CI(95%, q$_x$) = [Q(0.025; q$_x^{MCMC}$), Q(0.975; q$_x^{MCMC}$)], with Q(alpha, q$_x^{MCMC}$) being the empirical quantile based on the samples q$_x^{MCMC}$.

The quality of fitting for the proposed estimated tables can be evaluated based on the Standardized Mortality Ratio (SMR) test proposed by Liddel (1984). The SMR is defined as the ratio between observed and predicted counts for the model with SMR greater than 1 indicating underestimation. The test statistic is calculated, assuming that the death counts follow a $Poisson(E_{c,s,x} \times q_{c,s,x}^{BR2021})$ distribution. Under this assumption, the test statistic has an asymptotically normal distribution, allowing to test the hypotheses: H$_0$: there is no underestimation or overestimation of; H$_1$: there is underestimation or overestimation. In this study, we assume a Binomial model for the death counts, which, under the conditions of exposure and observed death rates, approximates the Poisson model, validating the requirements for the test.

**Insured population mortality tables**

This section discusses the main results regarding the mortality tables of the insured population, BR-EMS 2021. Mortality for insured individuals is significantly lower than for the national population. In addition, sex and coverage represent significant differences in mortality for all age groups.

The private market subpopulation has two components: death coverage and survivorship coverage. These two groups show much lower mortality rates than the general population, with the latter showing even lower death rates than the first.

Figure 2 shows the weighted exposure and death counts, following equation (2) and distributed by age, for males and females, under the death and survivorship coverage. Almost 3/4 of this population is between the ages of 20 and 50. This concentration of working-age policyholders occurs because most of the database comprises collective insurance groups, often related to employment contracts. We can also see that, under the death coverage, the maximum number of deaths is achieved at age 62 for males and females. On the other hand, the death counts peak at a much older age for the population of survivorship coverage, around twenty years later.

FIGURE 2
**The weighted exposure and deaths of the death and survivorship coverages for males and females
Brazilian insurance population − 2004-2017**
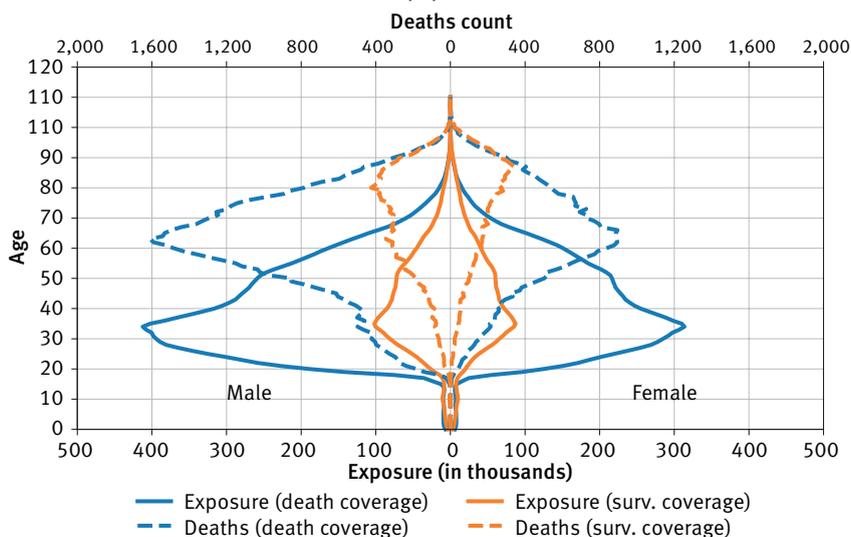


Source: Brazilian insurance companies participating in the analysis, CNIS, and SISOBI.

Table 1 shows the minimum, mean, and maximum exposures observed across ages for death and survivorship coverages, for males and females, of the BR-EMS v.2021 tables, for the years 2015, 2016 and 2017, which correspond to 87.5 % of the weight used in the

definition of $Ox$ and $Ex$. High exposures are observed in all coverages, sexes and ages, assuring a robust estimation of death probabilities.

TABLE 1
**Exposure summaries for the BR-EMS v.2021 tables**
**Brazilian insurance population – 2015-2017**

| Year | BR-EMSmt v.2021-male | | | BR-EMSmt v.2021-female | | |
|------|------|------|------|------|------|------|
|      | Min | Mean | Max | Min | Mean | Max |
| 2015 | 2035.29 | 189695.50 | 544926.20 | 1993.75 | 142566.80 | 400529.70 |
| 2016 | 2210.50 | 181340.60 | 520049.30 | 2562.67 | 147251.50 | 409318.10 |
| 2017 | 1336.75 | 126912.10 | 343786.00 | 2020.17 | 100768.70 | 261670.90 |
|      | BR-EMSsb v.2021-male | | | BR-EMSsb v.2021-female | | |
| 2015 | 1632.08 | 34354.48 | 91353.46 | 1862.54 | 29564.65 | 78466.75 |
| 2016 | 2217.50 | 41207.53 | 109483.58 | 1921.83 | 35243.33 | 92532.33 |
| 2017 | 3042.38 | 42410.10 | 111209.33 | 3237.63 | 36952.97 | 96416.92 |

Source: Brazilian insurance companies participating in the analysis, CNIS, and SISOBI.

Assuming the model proposed in section 3, and after running a Markov Chain Monte Carlo algorithm, the chains obtained for all parameters reached convergence, and results for both coverages and sexes indicated a good fit of the HP model to the data.

It is worth noting that the BR-EMS v.2010 life tables were based on HP models but other models were compared (details in OLIVEIRA *et al.*, 2012) and the HP model had a better performance. For the insured population, we expect smooth changes over a short range of years as the insurance market has experienced a robust behavior in the last decade. Thus, it is reasonable to assume the same HP model as in the 2010 table. Moreover, the dynamic linear model (details in WEST; HARRISON, 1997) was also fitted to the 2021 data. This class of models provides a flexible alternative similar to splines and when fitted to the 2021 data revealed adherence in the goodness of fit tests (CHARPENTIER, 2015, cap. 9) that assess the overall fit of two proposed tables. In particular, the performed tests accepted the hypothesis that the fitted tables are the same using both methods with p-values near 1 for the BR-EMSmt-v.2021-f, BR-EMSmt-v.2021-m, BR-EMSsb-v.2021-f, BR-EMSsb-v.2021-m tables. For details on the tests performed for model selection purposes, see supplementary material (https://github.com/LabMA-UFRJ/brems2021REBEP).

To check the adequate fit of the HP model to the data, the SMR test described in section 3 was considered and for all variants, p-values above 0.68 were observed, an indication that we do not reject $H_0$, concluding that there is no underestimation or overestimation of death probabilities based on the proposed models. We then obtained the points estimates for death probabilities using Monte Carlo integration as proposed in section 3. The following summaries for the HP model parameters are exhibited in Table 2.

TABLE 2
Estimates of model parameters and 95% credible intervals for both sexes (BREMS tables
Brazilian insurance population − 2004-2017

| Parameter | Death coverage estimates | | | | | |
|---|---|---|---|---|---|---|
| | Lower 95% CI | Male | Upper 95% CI | Lower 95% CI | Female | Upper 95% CI |
| A | 0.00018 | 0.00021 | 0.00025 | 0.00021 | 0.00021 | 0.00022 |
| B | 0.16092 | 0.20063 | 0.23635 | 0.19596 | 0,199628 | 0,203843 |
| C | 0.02739 | 0.03393 | 0.03903 | 0.03319 | 0.03380 | 0.03448 |
| D | 0.00064 | 0,000651 | 0.00066 | 0.00020 | 0.00020 | 0.00020 |
| E | 5.68846 | 5.79439 | 5.91055 | 5.67515 | 5.77632 | 5.90385 |
| F | 22.31090 | 22.67817 | 23.05258 | 18.74745 | 19.14469 | 19.51159 |
| G | 0.00003 | 0.00004 | 0.00004 | 0.00002 | 0.00002 | 0.00002 |
| H | 1.09550 | 1.09621 | 1.09711 | 1.09597 | 1.09683 | 1.09773 |
| K | -0.25000 | -0.25000 | -0.25000 | -0.80000 | -0.80000 | -0.80000 |
| | Survivorship coverage estimates | | | | | |
| A | 0.00017 | 0.00021 | 0.00025 | 0.00019 | 0.00023 | 0.00028 |
| B | 0.16150 | 0.20023 | 0.24079 | 0.55695 | 0.70264 | 0.83865 |
| C | 0.02639 | 0.03311 | 0.03874 | 0.05933 | 0.06059 | 0.06171 |
| D | 0.00050 | 0.00062 | 0.00071 | 0.00020 | 0.00020 | 0.00020 |
| E | 3.37384 | 7.12956 | 9.58086 | 3.38031 | 3.44430 | 3.51650 |
| F | 20.36749 | 22.47901 | 23.64581 | 18.66215 | 19.04279 | 19.39363 |
| G | 0.00002 | 0.00002 | 0.00002 | 0.00001 | 0.00001 | 0.00001 |
| H | 1.09826 | 1.09934 | 1.10141 | 1.09857 | 1.09989 | 1.10165 |
| K | -0.72000 | -0.72000 | -0.72000 | -1.25000 | -1.25000 | -1.25000 |

Source: Source: Brazilian insurance companies participating in the analysis, CNIS, and SISOBI.

Figure 3 displays all four BR-EMS 2021 probabilities of death obtained from mortality tables fitted using the HP model, under their official names registered in the insurance supervisory authority: BR-EMSmt-v.2021-f, BR-EMSmt-v.2021-m (death coverage for females and males, respectively), BR-EMSsb-v.2021-f, and BR-EMSsb-v.2021-m (survivorship coverage for females and males, respectively), along with the IBGE-2020 curves, separated by coverage. It is evident that, for all subgroups of sex and coverage, the insured population has lower mortality levels. Moreover, the death probabilities of the different risk groups indicate higher values for males than those for females.

The complete life tables considering $q_x$, 95% credible intervals, the expected number of survivors at age $x$ ($l_x$), and life expectancy at age $x$ ($e_x$) are publicly available at https://www.gov.br/susep/pt-br/assuntos/informacoes-ao-mercado/informacoes-tecnicas-e-planos-padroes/tabuas-biometricas-br-ems and can also be accessed at https://github.com/LabMA-UFRJ/brems2021REBEP.

FIGURE 3
The BR-EMS 2021: mortality tables for the insured population by coverage and sex and the IBGE 2020
tables. The probability of death for the age group 80+ equals 1 for the IBGE tables
Brazilian insurance population – 2004-2017; Brazilian population – 2020
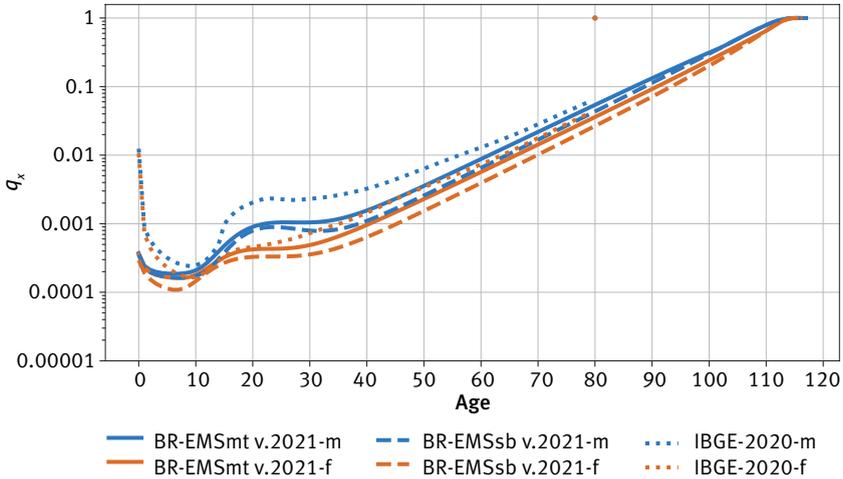


Source: Brazilian insurance companies participating in the analysis, CNIS, and SISOBI; IBGE (2021).

A comparison of life expectancy for different coverages indicates a difference of 3.3 years at age 0 for females, with life expectancy 83.2 years (BR-EMSmt-v.2021-f) and 86.5 years (BR-EMSsb-v.2021-f). This difference reaches 6.2 years for females when comparing BR-EMSsb-v.2021-f and IBGE-2020 (80.3 years). Expectancy at 60 years is estimated as 24.6 (IBGE-2020), 25.8 (BR-EMSmt-v.2021-f) and 28.4 (BR-EMSsb-v.2021-f) revealing a large difference for the general population when compared to the insured population. In the context of males, a similar pattern is observed. For instance, life expectancy at age 0 can differ in 7.8 years (73.3 years for IBGE and 81.1 years for BR-EMSsb-v.2021-m), indicating a significant reduction in mortality for the insured group when compared to the national population. Expectancy at 60 years for males is estimated as 20.8 (IBGE-2020), 22.1 (BR-EMSmt-v.2021-m) and 24.1 (BR-EMSsb-v.2021-m). Again, there is a certain selectivity in beneficiaries from the insurance market. Although individual insurance products are not included in the BR-EMS analysis, there are two types of beneficiaries in group life insurance: workers employed in formal job positions; and people who bought some product and have life insurance during the payment period. Therefore, the economic conditions of the beneficiaries compounding the denominator of the BR-EMS Life Tables are better than the Brazilian population as a whole.

Although a decrease in mortality is expected over time, the 2021 tables are approximately on the same level as the tables published in 2015. However, the two tables are constructed based on data with different qualities, with the 2021 table resulting from a substantial improvement in the data, given the requirement of unique NIDs from 2014. It is expected that future data quality remains high and tables constructed over the years will be comparable.

### Regional tables for the Brazilian private insured market

The Brazilian territory is divided into five macro-regions: North, Northeast, South, Southeast, and Midwest. The structural differences among these regions lead to heterogeneity in the mortality patterns for the general population as evidenced by the regional mortality tables elaborated by IBGE (IBGE, 2013a). These mortality curves are shown in Figure 4. In particular, notice the higher mortality associated with the North and Northeast regions. It should also be noted that the regional and global IBGE tables (2013a) exhibit higher death probabilities than those associated with BR-EMSmt v.2021-m, which is the highest table for the private market insured population (BR-EMS v.2021). Aiming to assess whether such regional differences also occur in the private insurance market, Figure 4 also compares the regional mortality profiles of the male insured population (death coverage) with those presented by the IBGE tables.

The regional tables for the private insurance market were built using the same database and methodology adopted to build the BR-EMS v.2021 tables. It contains information on 88,474.027 insured individuals and shows the following exposure distribution among regions: 7.3% in the Midwest, 3.6% in the North, 15.5% in the Northeast, 18.9% in the South, and 54.7% in the Southeast. Male exposures accounted for around 54% in the Southeast and Midwest regions, and approximately 57% in the other regions. However, the reduced information on the private insurance market for minors (under 18), already observed at the national level, worsens after stratification by macro-region. Thus, the regional tables of the private insurance market considered only ages 18 and older. Tables A.1 to A.4, available in the supplementary material (https://github.com/LabMA-UFRJ/brems2021REBEP), display descriptive statistics on exposures and deaths for each region/ sex/ coverage stratum of the population for the years 2015 to 2017. They show a notable discrepancy between the maximum and minimum volumes of exposure, recorded respectively in the Southeast and North regions. In addition to the lower exposures in some strata, the summary statistics make it clear that, except for death coverage for males, there is a shortage of deaths when considering single ages in tables adjusted individually for each variant and region, which could compromise the precision of mortality probability estimates. This situation is further exacerbated when considering the survivorship coverage for females. As a result, we restricted our estimates to regional tables adjusted for death coverage for males, exhibited in Figure 4, which show that for this stratum of the private insurance market, regional tables are very similar to the respective probabilities of death obtained from the BR-EMSmt v.2021-m.

Notice that the likely better living conditions of the insured population mitigate the regional structural differences observed in the national population. As a result, for all the estimated regional mortality tables, the probabilities of death for the insured population are significantly lower than those for the general population, a characteristic that persists even if compared with the 2030 IBGE's projected curves (IBGE, 2013b) (results not displayed

here) – suggesting that even in 2030, the general population will not have reached the mortality levels observed for the insured populations nowadays, with the exception of the general populations of the South and Southeast regions. Other studies reveal regional disparities in mortality patterns worldwide. Woolf and Schoomaker (2019) analyzed US life expectancy and mortality trends from 1959 to 2017, finding significant heterogeneity in geographic mortality patterns. The authors show that disparities in life expectancy between states were observed in the eastern US, particularly in the Ohio Valley, Appalachia, and upper New England, whereas Pacific states were less affected. Similar findings indicating significant spatial variability were reported in Germany by Kibele *et al.* (2017). In Brazil, Queiroz *et al.* (2020) examined mortality patterns from 1980 to 2020, highlighting persistent regional disparities with mortality increase in the north-Northeast and decrease in the south-Southeast regions. These findings align with those of França *et al.* (2017), who analyzed causes of death in Brazil, in a contribution to the Global Burden of Disease Study (2015). The authors observed variations among states, with less pronounced declines in mortality rates in the Northeast and North regions.

**FIGURE 4**
**Comparison between probabilities of death from regional tables for the general population of males (IBGE, 2013a), regional tables for the private insured market (death coverage of males), and BR-EMSmt v.2021-m**
Brazilian insurance population – 2004-2017; Brazil – 2013
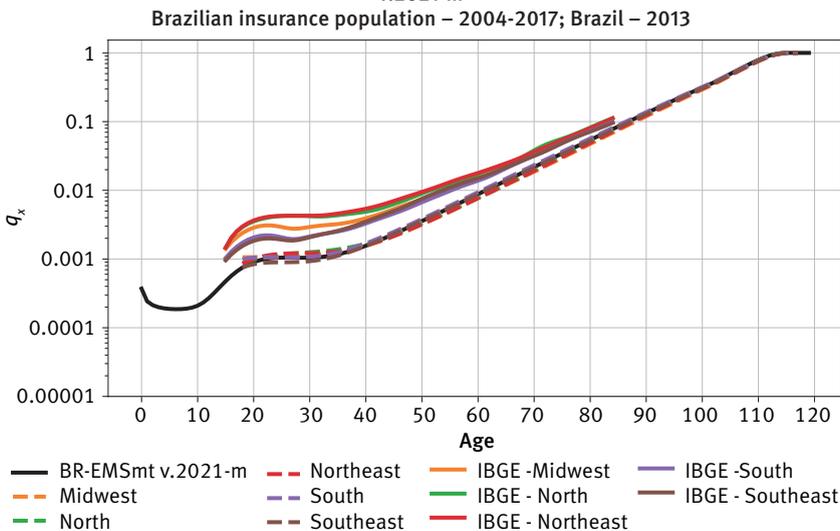


Source: Brazilian insurance companies participating in the analysis, CNIS, and SISOBI and IBGE (2013b).

## Comparison with international tables

In this section, the comparison between the probabilities of death obtained from the BR- EMS v.2021 tables and some international mortality tables (available on Human Mortality Database – HMD) confirms the existence of a national sub-population with low

mortality rates. However, it is also remarkable that mortality levels are comparable to those of developed countries, as seen in Figure 11, where both coverages of insured males are shown along with national male mortality curves of Chile, the US, and Japan for 2017. The Japanese table was included in the comparison, as it is well known that Japan observes one of the lowest mortality rates in the globe and representing a lower bound for the level of mortality we could expect our insured population to achieve. Chile is a representative of Latin America, which has consistent data publicly available over several years. Note that the Chilean rates are closer to the Brazilian ones, and that both insured populations have even lower mortality rates than the US population. In addition, the survivorship coverage has similar rates to the Japanese in the 40-80 year range. Future work will show a more extensive comparison with other nations for both sexes and coverages.

**FIGURE 5**
**Comparison of BR-EMS V.2021 probabilities of death for males, for both coverages, with international male mortality tables (HMD, 2022)**
**Brazilian insurance population – 2004-2017; Chile, Japan and United States of America – 2017**
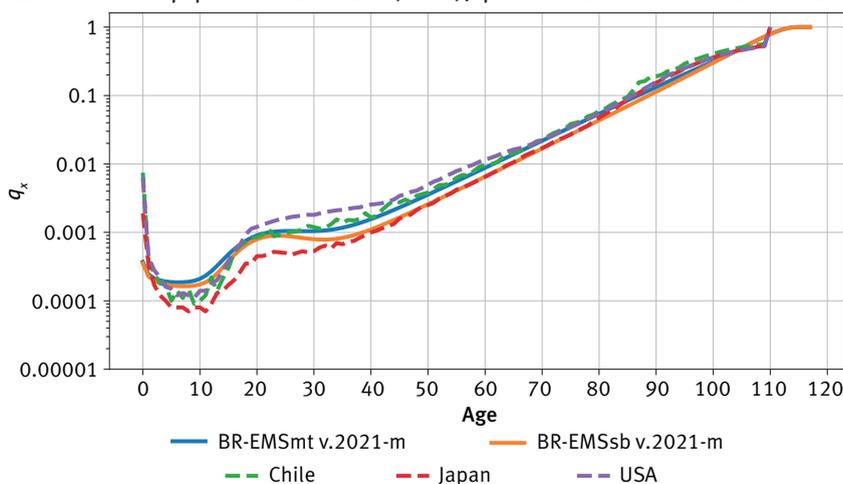


Source: Brazilian insurance companies participating in the analysis, CNIS, and SISOBI; Human Mortality Database (2022).

## Conclusions

This paper presented the 2021 version of the Brazilian insured population mortality tables, BR- EMS 2021, along with the database and procedures used to build the four tables, classified under sex and insurance coverage types: death and survivorship. In addition, a Bayesian estimation procedure was proposed, based on the well-known Heligman-Pollard mortality model, relying on compiled information of around 94 million insured individuals provided by insurance companies. The resulting insured population tables present lower mortality levels compared to national tables, showing this study's importance in avoiding additional underwriting risk for the Brazilian Insurance Market. Before 2010, tables from

other countries were used for pricing risk and life insurance products in Brazil. Thus, the publication of BR-EMS 2010, BR-EMS 2015, and its newest version, BR-EMS 2021, has represented a significant achievement for the Brazilian insurance market.

Nevertheless, it is worth noting that the likely better living conditions of the insured population are sufficient to mitigate the regional structural differences observed in the national population of Brazil. Likewise, the insured population mortality levels are comparable to those of developed countries, being lower than the United States and similar to the Chilean population. It is noteworthy that the male population in the survival coverage, between the ages of 40 and 80, has mortality levels comparable to those of the Japanese male population. This contributes to the argument that the insured Brazilian population has better life quality and overall health.

Finally, it is important to observe that the present paper does not reflect the effects of COVID-19 in Brazilian population, since the data are from 2004 to 2017. Castro *et al.* (2021) estimated a reduction of 1.3 years in Brazilian life expectancy from 2019 to 2020 due to COVID-19.

## Acknowledgments

## References

AMERICAN ACADEMY OF ACTUARIES. **Final report of the American Academy of Actuaries' Commissioners Standard Ordinary Task Force**. Philadelphia, PA, June 2002. Available in: https://www.actuary.org/sites/default/files/files/publications/CSO_taskforce_report_june2002.pdf . Access in: 10 Dec. 2022.

ARMSTRONG, J. S.; COLLOPY, F. Error measures for generalizing about forecasting methods: Empirical comparisons. **International Journal of Forecasting**, v. 8, n. 1, p. 69-80, 1992.

BELTRÃO, K. I.; SUGAHARA, S. Mortalidade dos funcionários públicos civis do Executivo por sexo e escolaridade – 1993/2014. **Revista Contabilidade & Finanças**, v. 28, n. 75, p. 445-464, set. /dez. 2017. Available in: https://www.revistas.usp.br/rcf/article/view/138289. Access in: 8 Jun. 2023.

BORGES, G. M. Health transition in Brazil: regional variations and divergence/convergence in mortality. **Cadernos de Saúde Pública**, v. 33, n. 8, 2017.

CASTRO, M. C.; GURZENDA, S.; TURRA, C. M.; KIM, S.; ANDRASFAY, T.; GOLDMAN, N. Reduction in life expectancy in Brazil after COVID-19. **Nature Medicine**, n. 27, p. 1629-1635, 2021. Available in: https://doi.org/10.1038/s41591-021-01437-z. Access in: 8 Jun. 2023.

CHARPENTIER, A. (Ed.). **Computational Actuarial Sciences with R**. Chapman & Hall, 2015.

CHEN, M.-H.; QI-MAN, S. Monte Carlo estimation of bayesian credible and HPD intervals. **Journal of Computational and Graphical Statistics**, v. 8, n. 1, p. 69-92, 1999. https://doi.org/10.2307/1390921.

CONGDON, P. Statistical graduation in local demographic analysis and projection. **Journal of the Royal Statistical Society. Series A (Statistics in Society)**, v. 156, n. 2, p. 237-270, 1993.

CURRIE, I. D.; DURBAN, M.; EILERS, P. H. Smoothing and forecasting mortality rates. **Statistical Modelling**, v. 4, n. 4, p. 279-298, 2004. https://doi.org/10.1191/1471082X04st080oa.

DI LEGO, V.; TURRA, C. M.; CESAR, C. Mortality selection among adults in Brazil: the survival advantage of Air Force officers. **Demographic Research**, v. 37, article 41, p. 1339-1350, 2017. https://doi.org/10.4054/DemRes.2017.37.41.

DOBSON, A. J.; BARNETT, A. G. **An introduction to generalized linear models**. 4. ed. Boca Raton, London, New York: CRC press, 2018.

FRANÇA, E. B. *et al.* Cause-specific mortality for 249 causes in Brazil and states during 1990–2015: a systematic analysis for the global burden of disease study 2015. **Population Health Metrics**, v. 15, n. 1, 2017.

GAMERMAN, D.; LOPES, H. F. **Markov Chain Monte Carlo**: stochastic simulation for bayesian inference. 2. ed. London: Chapman & Hall/CRC, 2006.

GLOBAL BURDEN OF DISEASE COLLABORATIVE NETWORK. **Global Burden of Disease Study 2015 (GBD 2015)**. Life expectancy, all-cause and cause-specific mortality 1980-2015 (2015). Technical report.

GONZAGA, M. R.; LIMA, E. E. C.; QUEIROZ, B. L.; ANSILIERO, G.; FREIRE, F. H. M. A. Diferenciais de mortalidade, beneficiários do Instituto Nacional do Seguro Social do Brasil em 2015. **Revista Contabilidade & Finanças**, v. 33, n. 90, e1556, 2022.

HELIGMAN, L.; POLLARD, J. H. The age pattern of mortality. **Journal of the Institute of Actuaries**, v. 107, n. 1, p. 49-80, 1980.

HILTON, J.; DODD, E.; FORSTER J. J.; SMITH, P. W. Projecting UK mortality by using Bayesian generalized additive models. **Journal of the Royal Statistical Society, Series C (Applied Statistics)**, v. 68, n. 1, p. 29-49, 2019.

HMD. **Human Mortality Database**. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available in: www.mortality.org. Access in: 10 Dec. 2022.

HYNDMAN, R. J.; SHAHID, U. M. Robust forecasting of mortality and fertility rates: a functional data approach. **Computational Statistics & Data Analysis**, v. 51, n. 10, p. 4942-4956, 2007. https://doi.org/10.1016/j.csda.2006.07.028.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Tábuas completas de mortalidade** – edição 2020, 2021. Available in: https://www.ibge.gov.br/estatisticas/sociais/populacao/9126-tabuas-completas-de-mortalidade.html?edicao=32297&t=resultados. Access in: 10 Dec. 2022.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Tábuas abreviadas de mortalidade por sexo e idade**: Brasil, grandes regiões e unidades da federação: 2010. Rio de Janeiro: IBGE, 2013a. Available in: https://biblioteca.ibge.gov.br/visualizacao/livros/liv65137.pdf. Access in: 10 Dec. 2022.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Projeções de população por sexo e idade** – Brasil 2000-2060 e unidades da federação 2000-2030. Rio de Janeiro: IBGE, 2013b. Available in: https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html?edicao=9116. Access in: 10 Dec. 2022.

JOHANSEN, R. Review of adequacy of 1983 individual annuity mortality table. **Transactions of Society of Actuaries**, v. 47, p. 211-249, 1995.

KIBELE, E. U. B.; KLÜSENER, S.; SCHOLZ, R. D. Regional mortality disparities in Germany: long-term dynamics and possible determinants. **Kolner Zeitschrift Fur Soziologie Und Sozialpsychologie**, v. 67, suppl. 1, p. 241-270, 2015.

LEE, R. The Lee-Carter method for forecasting mortality, with various extensions and applications. **North American Actuarial Journal**, v. 4, n. 1, p. 80-91, 2000.

LIDDELL, F. Simple exact analysis of the standardised mortality ratio. **Journal of Epidemiology & Community Health**, v. 38, n. 1, p. 85-88, 1984.

OLIVEIRA, M.; FRISCHTAK, R.; RAMIREZ, M.; BELTRÃO, K.; PINHEIRO, S. **Brazilian mortality and survivorship life tables**: insurance market experience – 2010. Rio de Janeiro: Fundação Escola Nacional de Seguros – Funenseg, 2012.

OLIVEIRA, M. M. C.; RAMIREZ, M. R.; FRISCHTAK, R. M.; BORGES, R. B. R.; COSTA, B.; PEDROSO, R. C. Tábuas de mortalidade para a população brasileira de segurados. **Revista Brasileira de Estudos de População**, v. 33, n. 3, p. 653-677, 2016.

OLIVIERI, A. Uncertainty in mortality projections: an actuarial perspective. **Insurance: Mathematics and Economics**, v. 29, n. 2, p. 231-245, 2001.

QUEIROZ, B. L.; LIMA, E. E. C.; FREIRE, F. H. M. A.; GONZAGA, M. R. Temporal and spatial trends of adult mortality in small areas of Brazil, 1980- 2010. **Genus**, v. 76, n. 36, 2020.

QUEIROZ, B. L.; GONZAGA, M. R.; VASCONCELOS, A. M. N.; LOPES, B. T.; ABREU, D. M. X. Comparative analysis of completeness of death registration, adult mortality and life expectancy at birth in Brazil at the subnational level. **Population Health Metrics**, v. 18, suppl. 1, 2020.

RIBEIRO, M. M.; TURRA, C. M.; PINTO, C. C. X. Mortalidade adulta por nível de escolaridade em São Paulo: análise comparativa a partir de diferentes estratégias metodológicas. **Revista Brasileira de Estudos de População**, v. 38, 2021. Available in: dx.doi.org/10.20947/S0102-3098a0139. Access in: 10 Dec. 2022.

SOCIETY OF ACTUARIES. Report of the Committee to Recommend a New Mortality Table Basis for Individual Annuity Valuation (Derivation of the 1983 Table a). **Transactions of the Society of Actuaries**, vol. 33. (1981) Table 22. Available in: http://www.soa.org/Library/Research/Transactions-Of-Society-Of-Actuaries/1981/January/tsa81v3325.pdf. Access in: Apr. 2013.

SIVIERO, P. C. L.; SOUZA, L. G.; MACHADO, C. J. Diferenciais de mortalidade por sexo no município de São Paulo em 2005 e 2016: contribuição dos grupos etários e das principais causas de óbito. **Revista Brasileira de Estudos de População**, v. 36, 2019. Available in: https://doi.org/10.20947/S0102-3098a0099. Access in: 10 Dec. 2022.

TUKEY, J. W. **Exploratory data analysis**. Reading: Addison-Wesley Publishing Company, 1977.

UN – United Nations. Department of Economic and Social Affairs, Population Division. **World population prospects**: the 2022 revision, custom data acquired via website. 2022. Available in: https://population.un.org/wpp/. Access in: 10 Dec. 2022.

VAN ROSSUM, G.; DRAKE JR., F. L. Python tutorial. **Technical Report CS-R9526**. Amsterdam: Centrum voor Wiskunde en Informatica (CWI), May 1995.

WEST, M.; HARRISON, J. **Bayesian forecasting and dynamic models**. 2. ed. Springer, 1997.

## About the authors

*Mario de Oliveira* holds a PhD in Statistics from the Warwick University, is a professor at the Department of Applied Mathematics (UFRJ) and coordinator of the Laboratory of Applied Mathematics at the Institute of Mathematics, Federal University of Rio de Janeiro.

*Ana Carolina Soares Bertho* has doctorate degree in Demography from Universidade Estadual de Campinas (Unicamp). Researcher and professor at Escola Nacional de Ciências Estatísticas (ENCE/IBGE).

*Bruno Costa* has a doctorate degree in Applied Mathematics from Indiana Univeristy. Professor of the Applied Mathematics Department of Universidade Federal do Rio de Janeiro (UFRJ).

*Flávia Sommerlatte Silva* is Actuary and has a Masters degree in Demography from Universidade Federal de Minas Gerais (UFMG). Actuarial Consultant at Mercer Portugal.

*Mariane Branco Alves* holds a PhD in Statistics from the Federal University of Rio de Janeiro (UFRJ), is a professor at the Department of Statistical Methods and a researcher at the Laboratory of Applied Mathematics at the Institute of Mathematics, Federal University of Rio de Janeiro.

*Milton Ramos Ramirez* has doctorate degree in Systems and Computational Engineer from Universidade Federal do Rio de Janeiro (UFRJ). Professor of the Applied Mathematics Department of Universidade Federal do Rio de Janeiro (UFRJ) and vice-head of the Applied Mathematics Laboratory of UFRJ.

*Rafael Brandão de Rezende Borges* is Assistant Professor at Departamento de Análise Matemática, Universidade Estadual do Rio de Janeiro (UERJ) and researcher at LabMA/UFRJ.

*Reinaldo Marques* holds a PhD in Statistics from the University of Oslo, is a professor at the Department of Actuarial Science and a researcher at the Actuarial Risk Lab, Federal University of Alfenas.

*Ricardo Martins da Silva Rosa* has a doctorate degree in Applied Mathematics from Indiana Univeristy - USA. Professor of the Applied Mathematics Department of Universidade Federal do Rio de Janeiro (UFRJ) and researcher at LabMA/UFRJ.

*Rodrigo Lima Peregrino* is an Applied Mathematician and a researcher at the Laboratory of Applied Mathematics at the Institute of Mathematics, Federal University of Rio de Janeiro.

*Viviana das Graças Ribeiro Lobo* holds a Ph.D. degree in Statistics from Universidade Federal do Rio de Janeiro (UFRJ). She is a tenured Assistant Professor at the Institute of Mathematics at UFRJ and a researcher at the Applied Mathematics Laboratory at UFRJ.

*Thais Cristina Oliveira Fonseca* holds a PhD in Statistics from the Warwick University, is a professor at the Department of Statistical Methods and a researcher at the Laboratory of Applied Mathematics at the Institute of Mathematics, Federal University of Rio de Janeiro.

## Contact address

*Mário de Oliveira*
Universidade Federal do Rio de Janeiro, Centro de Tecnologia
Av. Athos da Silveira Ramos, 149, Bloco A sala ABC110
21941-611 – Rio de Janeiro-RJ, Brazil

*Ana Carolina Soares Bertho*
R. André Cavalcanti, 106, Bairro de Fátima
20231-050 – Rio de Janeiro-RJ, Brazil

*Bruno Costa*
Universidade Federal do Rio de Janeiro, Centro de Tecnologia
Av. Athos da Silveira Ramos, 149, Bloco A sala ABC110
21941-611 – Rio de Janeiro-RJ, Brazil

*Flávia Sommerlatte Silva*
Avenida 25 de Abril de 1974, 51, apartamento 4º direito, Almada
2800-299 – Setúbal, Portugal

*Mariane Branco Alves*
Universidade Federal do Rio de Janeiro, Centro de Tecnologia
Av. Athos da Silveira Ramos, 149, Bloco C sala C125-D
21941-611 – Rio de Janeiro-RJ, Brazil

*Milton Ramos Ramirez*
Universidade Federal do Rio de Janeiro, Instituto de Matemática
Av. Athos da Silveira Ramos, 149, Bloco C
21941-909 – Rio de Janeiro-RJ, Brazil

*Rafael Brandão de Rezende Borges*
Universidade Federal do Rio de Janeiro, Centro de Tecnologia
Av. Athos da Silveira Ramos, 149, Bloco A sala ABC110
21941-611 – Rio de Janeiro-RJ, Brazil

*Reinaldo Marques*
Universidade Federal de Alfenas
Av. Celina Ferreira Ottoni, 4000, sala A118, Padre Vitor
37048-395 – Varginha-MG, Brazil

*Ricardo Martins da Silva Rosa*
Universidade Federal do Rio de Janeiro, Centro de Tecnologia
Av. Athos da Silveira Ramos, 149, Bloco A sala ABC110
21941-611 – Rio de Janeiro-RJ, Brazil

*Rodrigo Lima Peregrino*
Avenida Santo Amaro, 220, Vila Nova Conceição
04506-000 – São Paulo-SP, Brazil

*Viviana das Graças Ribeiro Lobo*
Universidade Federal do Rio de Janeiro, Centro de Tecnologia
Av. Athos da Silveira Ramos, 149, Bloco C sala C109-A
21941-611 – Rio de Janeiro-RJ, Brazil

*Thais Cristina Oliveira Fonseca*
Universidade Federal do Rio de Janeiro, Centro de Tecnologia
Av. Athos da Silveira Ramos, 149, Bloco C sala C114-D
21941-611 – Rio de Janeiro-RJ, Brazil

## Resumo

*Tábuas de mortalidade BR-EMS 2021 do mercado segurador brasileiro*

Este artigo apresenta as tábuas de vida do mercado de seguros privados brasileiro, BR-EMS 2021. Os dados obtidos de 23 grupos seguradores ao longo de 15 anos, totalizando 3,5 bilhões de registros, foram corrigidos por meio de rede neural com duas camadas ocultas. Usando a inferência bayesiana para estimar os parâmetros sob a lei de mortalidade Heligman-Pollard, as tábuas obtidas mostram que a população segurada apresenta probabilidades de morte

mais baixas do que a população brasileira em geral e até mesmo em relação a populações nacionais de países desenvolvidos, como os EUA. Além da esperada diferença de gênero nas taxas de mortalidade, há uma clara distância entre as probabilidades de morte dos grupos de cobertura de risco e cobertura de sobrevivência. Da mesma forma, é demonstrado que as tábuas regionais da população segurada não apresentam as discrepâncias regionais conhecidas no Brasil, indicando que fazer parte da população selecionada de segurados está associado a um fator de proteção mais eficaz do que outros fatores, como a região geográfica de residência.

**Palavras-chave**: Tábuas de vida atuariais. Coberturas de morte e sobrevivência. Graduação de mortalidade. Modelo de Heligman-Pollard.

## Resumen

*Tabla de vida BR-EMS 2021 para la población asegurada brasileña*

Este artículo presenta las tablas de vida del mercado de seguros privados brasileño, BR-EMS 2021. Los datos, obtenidos de 23 grupos de seguros a lo largo de 15 años, totalizando 3,5 mil millones de registros, fueron corregidos usando una red neuronal con dos capas ocultas. Mediante la inferencia bayesiana para estimar los parámetros bajo la ley de mortalidad de Heligman-Pollard, las tablas obtenidas muestran que la población asegurada tiene tasas de mortalidad más bajas que la población general brasileña e incluso más bajas que las poblaciones nacionales de países desarrollados, como los Estados Unidos de Norteamérica. Además de la diferencia de género esperada en las tasas de mortalidad, hay una clara distinción entre las tablas de grupos de cobertura de riesgo y cobertura de sobrevivientes. Asimismo, se demuestra que las tablas regionales de población asegurada no presentan las conocidas discrepancias estructurales regionales en Brasil, lo que indica que participar de la población de asegurados está asociado con una protección contra la muerte más efectiva que otros factores como la región geográfica de residencia.

**Palabras clave**: Tablas actuariales de vida. Cobertura de muerte y supervivencia. Graduación de la mortalidad. Modelo de Heligman-Pollard.