GAUSSIAN PROCESS-BASED MORTALITY MONITORING USING MULTIVARIATE CUMULATIVE SUM PROCEDURES

Karim Barigou, Stéphane Loisel, Yahia Salhi, Rayane Vigneron

REPRINT | 2026 / 32



UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication

Gaussian Process-Based Mortality Monitoring using Multivariate Cumulative Sum Procedures *

Karim Barigou[†1], Stéphane Loisel[2], Yahia Salhi[3], and Rayane Vigneron[3]

[1] *Institute of Statistics, Biostatistics and Actuarial Science (ISBA), Louvain Institute of Data Analysis and Modeling (LIDAM), UCLouvain, Louvain-La-Neuve, Belgium*

[2] *Conservatoire National des Arts et Métiers (Cnam), Efab, Lirsa, 292 rue St Martin, 75003 Paris, France*

[3] *Univ Lyon, Université Claude Bernard Lyon 1, Laboratoire de Sciences Actuarielle et Financière, Institut de Science Financière et d'Assurances (50 Avenue Tony Garnier, F-69007 Lyon, France).*

Version: September 24, 2026

Abstract

This paper proposes an online multivariate cumulative sum (MCUSUM) monitoring procedure for detecting changes in mortality dynamics, with direct applications to mortality and longevity risk management for insurers and pension funds. The method is built on Gaussian process (GP) non-parametric mortality forecasts, and performs surveillance in real time by tracking multivariate forecast errors across ages. We develop MCUSUM schemes targeting two practically relevant forms of change: (i) a change in level, corresponding to an abrupt proportional shift in mortality rates, and (ii) a change in trend, corresponding to a shift in the rate of mortality improvement. In both cases, one-sided monitoring rules allow the practitioner to focus on either adverse mortality shocks or adverse longevity developments. By explicitly exploiting dependence between age groups, the proposed multivariate approach improves detection performance relative to collections of univariate control charts. We evaluate the procedure through simulation experiments and empirical applications to recent mortality data from France, Japan, Canada, and the USA, and we further illustrate its use on a real-world life insurance portfolio.

Keywords: Mortality modeling, Change-point detection, Gaussian processes, longevity risk management.

*Contacts: Karim Barigou (karim.barigou@uclouvain.be), Stéphane Loisel (stephane.loisel@lecnam.net), Yahia Salhi (yahia.salhi@univ-lyon1.fr), Rayane Vigneron (rayane.vigneron@gmail.com)

[†]Corresponding author, mail: karim.barigou@uclouvain.be

1 Introduction

Monitoring mortality in real time is central to the management of both mortality and longevity risk. Life insurers and pension funds rely on forecasting models calibrated on historical data to price products, set reserves, and hedge exposures; when the underlying mortality dynamics shift, these models can become misspecified and generate material basis risk. Recent experience has highlighted that such shifts may occur abruptly (e.g., shocks affecting multiple ages simultaneously) or gradually (e.g., changes in improvement trends), and the financial impact depends critically on how quickly the change is detected and incorporated into pricing and risk management. This motivates prospective surveillance tools that can signal, as soon as possible, departures from the expected mortality evolution while controlling false-alarm rates. A particular challenge is that mortality is inherently multivariate: adjacent ages and cohorts exhibit strong dependence, so effective monitoring should exploit cross-age information rather than treat age-specific series as independent.

Recently, a few researchers have studied methods for detecting changes in mortality dynamics, using multivariate control charts (Díaz-Rojo et al., 2020) or univariate procedures such as the Shiryaev–Roberts procedure (Abgrall et al., 2018). In the latter case, the intensity of the aggregate mortality is obtained by summing up the intensity at the different ages invoking the standard independence assumption. As Pascual and Akhundjanov (2020) pointed out, if different processes are independent, monitoring each process separately can be very effective for detecting any change. However, mortality data are often correlated due to similarities in adjacent ages or cohorts (Xu et al., 2020, Jevtić et al., 2013). In the presence of such correlations, a system of separate univariate monitoring cannot be efficient as they ignore the dependence structure in the data.

For efficient monitoring of mortality rates in a multivariate setting, multivariate control charts should be considered (see Jiang et al. (2011), Rogerson and Yamada (2004), Joner Jr et al. (2008) in a general context). Multivariate methods generate powerful surveillance schemes that are able to capture the dependence structure for which univariate control charts are insensitive. Jiang et al. (2011) using breast cancer incidences and Pascual and Akhundjanov (2020) using hepatitis C counts found that multivariate charts provide faster and more accurate detection of outbreaks than the surveillance methods based on the assumption of independence. For an overview of multivariate statistical process control charts, we refer to Bersimis et al. (2007) and Qiu (2013).

In this paper, we consider the multivariate cusum (MCUSUM) control chart from Healy (1987) for detecting a shift in the mean μ of a multivariate normal process to an out-of-control mean μ_c that the user wants to detect. In an actuarial context, μ_c may correspond to a change of level or a change of trend in the observed death rates. To this end, we use Gaussian processes (Rasmussen and Williams, 2005) as the basis for a Bayesian non-parametric forecasting mortality model and use the difference between the predicted and observed death rates to monitor change in an online fashion. The application of Gaussian process (GP) models for mortality rates was already considered in a single-population case by Ludkovski et al. (2018) and in a multi-population setting by Huynh and Ludkovski (2021). The advantage of using GP as the baseline is that the predicted

log death rates are multivariate normal distributed, allowing a direct application of the MCUSUM algorithm while standard mortality models such as Lee and Carter (1992), Brouhns et al. (2005), Cairns et al. (2006) often impose a relatively low-dimensional dependence structure across ages through shared latent factors. This may be restrictive in a monitoring context when one wishes to capture richer cross-age forecast dependence.

Among early contributions using GP for time series change detection, Chandola and Vatsavai (2011) proposed a GP-based algorithm to identify changes in a univariate Normalized Difference Vegetation Index (NDVI) time series. In contrast, our approach is based on CUSUM rather than Exponentially Weighted Moving Average (EWMA) monitoring, and it is designed for multivariate rather than univariate processes.

More broadly, our work is related to two strands of the change-detection literature. First, several authors have developed GP approaches for change-point detection, including kernel constructions specifically designed to capture structural changes (Saatçi et al., 2010, Caldarelli et al., 2022, Zhao and Pan, 2025). Second, recent work has emphasized the importance of dependence-aware or correlation-aware procedures in sequential detection problems (Deng et al., 2025). Our contribution is complementary to these developments. We do not propose a new GP change-point kernel, nor a new general-purpose dependence-aware detection method. Instead, we use GP regression as a flexible probabilistic forecasting model for age-specific mortality and combine its multivariate Gaussian predictive distribution with a MCUSUM surveillance rule targeted at actuarially meaningful departures from the baseline mortality regime.

To summarize, this paper proposes an online mortality monitoring framework that combines GP forecasting with multivariate CUSUM (MCUSUM) surveillance. We model age-specific log death rates jointly using GP regression, which provides probabilistic forecasts and yields a multivariate Gaussian predictive distribution that is naturally compatible with the MCUSUM procedure. Building on this, we design MCUSUM schemes to detect two actuarially meaningful departures from the baseline dynamics. First, we consider a *change of level* in which mortality shifts from a baseline rate μ to $\rho\,\mu$; this setting allows us to monitor either mortality risk ($\rho \geq 1$, higher-than-expected mortality) or longevity risk ($\rho \leq 1$, lower-than-expected mortality) by selecting the relevant direction of change. Second, we consider a *change of trend* in the mortality improvement rate; this setting allows us to monitor either weaker-than-expected mortality improvements or stronger-than-expected mortality improvements, depending on the direction of change under consideration. The proposed multivariate approach explicitly exploits cross-age dependence through the GP predictive covariance structure, in contrast to collections of univariate charts that ignore this dependence. We evaluate the methodology on national populations (USA, Canada, France and Japan; ages 50–89; 1991–2020) and on an insurer portfolio, and we complement this with a simulation study showing that the multivariate scheme detects these level and trend shifts faster than competing univariate CUSUM procedures, with gains that increase as dependence across age classes strengthens.

The paper is structured as follows. In Section 2, we present the GP-based mortality forecasting model and its multivariate Gaussian predictive distribution. Section 3 introduces the MCUSUM monitoring framework and explains how it can be implemented for two types of actuarially relevant departures from the baseline: (i) level shifts in mortal-

ity and (ii) changes in mortality improvement trends, with one-sided schemes allowing the practitioner to focus specifically on either mortality risk or longevity risk. Section 4 applies the proposed procedures to national populations described above and to a life insurance portfolio. Section 5 reports a simulation study showing that our multivariate monitoring scheme outperforms competing univariate CUSUM charts, with larger gains when dependence across age classes is stronger. Section 6 concludes.

2 Gaussian Process Regression for mortality forecasting

We present briefly mortality forecasting based on GP. For an extended treatment on the parameter choices, we refer to Ludkovski et al. (2018). Consider a (training) set of observations $(X, \boldsymbol{y}) = \left\{\left(\boldsymbol{x^i}, y^i\right) \mid i = 1, \ldots, n\right\}$ where each $\boldsymbol{x^i}$ is an input vector of dimension d and y^i is the corresponding output. In our case, $\boldsymbol{x^i}$ is a bivariate vector $\boldsymbol{x}^i = (x^i_{\text{age}}, x^i_{\text{year}})$ representing the age (or age tranche) and year of the data point, e.g. $\boldsymbol{x} = (65, 2020)$ or $\boldsymbol{x} = ([65, 69], 2020)$, and y^i is the corresponding logarithmic central death rate $y^i = \log(D^i/E^i)$ where D^i and E^i represent the annual deaths and exposures. The goal is to find the relation between inputs and outputs. With a GP, this is done by assuming

$$y^i = f\left(\boldsymbol{x^i}\right) + \varepsilon^i, \tag{2.1}$$

where $f(\boldsymbol{x})$ is a GP and $\varepsilon^i \sim \mathcal{N}(0, \sigma_n^2)$ are i.i.d. random variables representing the noise in the data.

A GP $f(\boldsymbol{x})$ is a, possibly infinite, collection of random variables, any finite subset of it having a joint Gaussian distribution. The process $f(\boldsymbol{x})$ is completely defined by its mean function $m(\boldsymbol{x})$ and a covariance or kernel function $k(\boldsymbol{x}, \boldsymbol{x}')$. For a sample $(X, f) = \left\{\left(\boldsymbol{x^i}, f^i\right) \mid i = 1, \ldots, n\right\}$ generated from $f(\boldsymbol{x})$, we then have

$$f \sim \mathcal{N}(m(\boldsymbol{x}), K(X, X)), \tag{2.2}$$

where $K(X, X)$ is the covariance matrix, defined as

$$K(X, X) = \begin{bmatrix} k(\boldsymbol{x^1}, \boldsymbol{x^1}) & \cdots & k(\boldsymbol{x^1}, \boldsymbol{x^n}) \\ k(\boldsymbol{x^2}, \boldsymbol{x^1}) & \cdots & k(\boldsymbol{x^2}, \boldsymbol{x^n}) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x^n}, \boldsymbol{x^1}) & \cdots & k(\boldsymbol{x^n}, \boldsymbol{x^n}) \end{bmatrix}$$

GP Regression is a Bayesian method as it starts from the prior GP (2.2) to impose initial knowledge about the function f and then combines this with the observed data points (X, y) using the Gaussian assumption (2.1), leading to a posterior distribution of functions.

Consider a matrix X_* consisting of n_* test inputs and denote the corresponding (unknown) vector of function values by f_*. For instance, the new points $\boldsymbol{x}^*$ may correspond

to new age-year inputs for mortality projections. By detrending via $f - m(\boldsymbol{x})$, we may assume without loss of generality that $m \equiv 0$. Then, the GP posterior distribution in the test inputs is multivariate normal (Rasmussen and Williams, 2005) and given by

$$\begin{aligned} f_* \mid X_*, X, \boldsymbol{y} \sim \mathcal{N} \Big(& K\left(X_*, X\right) \left[K(X, X) + \sigma_n^2 I\right]^{-1} \boldsymbol{y}, \\ & K\left(X_*, X_*\right) - K\left(X_*, X\right) \left[K(X, X) + \sigma_n^2 I\right]^{-1} K\left(X, X_*\right) \Big). \end{aligned} \tag{2.3}$$

Point mortality predictions can then be obtained as the mean of the distribution (2.3).

GP is completely specified by its mean function $m(\boldsymbol{x})$ and its covariance or kernel function $k\left(\boldsymbol{x}, \boldsymbol{x}'\right)$. In this paper, we use the most commonly used covariance function, which is the squared exponential (SE) kernel,

$$\begin{aligned} k\left(\boldsymbol{x}, \boldsymbol{x}'\right) &= \sigma_f^2 \exp\left(\frac{-\left|\boldsymbol{x} - \boldsymbol{x}'\right|^2}{2l^2}\right) \\ &= \sigma_f^2 \exp\left(\frac{-\sum_{k=1}^{2}\left(x_k - x_k'\right)^2}{2l^2}\right) \end{aligned} \tag{2.4}$$

with σ_f^2 the highest possible covariance and l a length-scale parameter that determines the smoothness of the fit. Both parameters are called hyperparameters and are estimated using the training data, e.g. by maximizing the marginal log-likelihood:

$$-\frac{1}{2}\log(\det(K(X, X))) - \frac{1}{2}\boldsymbol{y}^{\mathrm{T}} K(X, X)^{-1}\boldsymbol{y} + \text{ constant}.$$

In several applications, the mean function is often taken equal to zero, but can be modelled using a set of basis functions by defining the function:

$$g(\boldsymbol{x}) = f(\boldsymbol{x}) + h(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\beta},$$

where $f(\boldsymbol{x})$ is a GP with zero mean function $m \equiv 0$ and kernel function $k\left(\boldsymbol{x}, \boldsymbol{x}'\right)$, $h(\boldsymbol{x})$ is a set of fixed basis functions and β are the corresponding coefficients that are estimated from the training data. In this way the model appears as a linear model, where the residuals are modelled via a GP.

This basis-function representation is introduced to emphasize the flexibility of the GP framework. In the remainder of the paper, however, we use the following specific quadratic mean specification, following Ludkovski et al. (2018):

$$m(\boldsymbol{x}) = \beta_0 + \beta_1^{age} x_{age} + \beta_1^{year} x_{year} + \beta_2^{age} x_{age}^2. \tag{2.5}$$

For smoothing the mortality surface, the choice of the mean function has little effect, but for mortality forecasting (that is, extrapolation), it plays an important role. Indeed, if we predict data points far away from the training data points, the predictions will be essentially driven by the prior mean; see Ludkovski et al. (2018) for a detailed discussion. It was shown in Ludkovski et al. (2018) on US mortality data that the quadratic model tends to perform better than the linear model without quadratic term. We use this quadratic specification as a parsimonious baseline mean function, following Ludkovski

et al. (2018), rather than as a universally optimal model for all populations considered here. Figure 1 displays the observed historical and GP-predicted log death rates for representative 5-year age groups 50-54, 65-69, and 85-89 in each country. The transition from the calibration period (1990–2010) to the forecast period (2011–2020) is smooth, with no visible discontinuity at the forecast boundary, and the extrapolated trajectories remain qualitatively consistent with the age-specific historical trends.

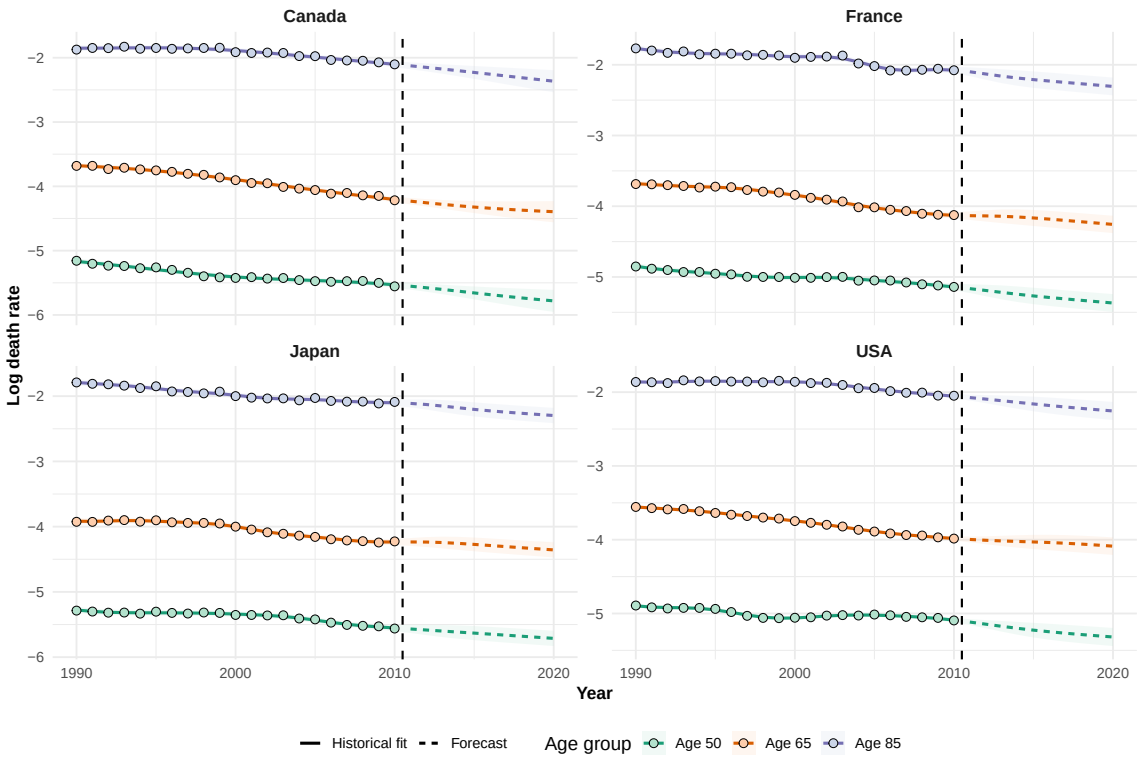


Figure 1: Observed and GP-predicted log death rates for selected ages 50, 65, and 85 in Japan, France, Canada, and the USA. Points correspond to observed historical mortality data over the calibration period 1990–2010. Solid lines represent the fitted Gaussian process mean over the calibration period, while dashed lines represent the forecast mean over 2011–2020. Shaded areas indicate the corresponding 95% predictive intervals. The vertical dashed line marks the transition between the calibration and forecast periods.

Assume that we have mortality data for years $t = 1, \ldots, T$ and we want to forecast N years ahead. From the posterior distribution (2.3), we find that the vector of predicted log death rates is multivariate normal, i.e.

$$\begin{aligned} \boldsymbol{y^t} &:= \log\left(\boldsymbol{\mu_t}\right) \\ &\sim \mathcal{N}(\boldsymbol{m_t}, \boldsymbol{\Sigma}_t) \end{aligned} \tag{2.6}$$

for any prediction year $t = T+1, \ldots, T+N$ where at any time point t, the mean $\boldsymbol{m_t}$ and the covariance matrix $\boldsymbol{\Sigma}_t$ are time-dependent. In this expression, $\log\left(\boldsymbol{\mu_t}\right)$ is the vector of log death rates:

$$\boldsymbol{\mu_t} = (\mu_{z_1,t}, \ldots, \mu_{z_M,t})$$

for different M age-groups, e.g. $z_1 = [50; 55); z_2 = [55; 60); \ldots; z_M = [85; 90)$. Following (2.6), monitoring the central log death rates boils down to monitoring the mean of a multivariate normal process, which is discussed in the next section.

3 Online monitoring via the MCUSUM algorithm

The CUSUM algorithm is a well-known sequential procedure based on likelihood ratios for detecting a change in a process. Assume that we observe a sequence $\boldsymbol{y} = (\boldsymbol{y^t})_{t\geq 1}$ of random variables such that

$$\begin{aligned} \boldsymbol{y^t} &\sim F_1, \quad 1 \leq t \leq \tau, \\ &\sim F_2, \quad \tau + 1 \leq t, \end{aligned}$$

where the distributions F_1 and F_2 are known but the change point τ is unknown and deterministic. The objective is to determine if and when the shift from F_1 to F_2 occurs.

The CUSUM algorithm signals that the change of distribution has happened the first time:

$$S_t = \sum_{i=1}^{t} \log \frac{f_2\left(\boldsymbol{y^i}\right)}{f_1\left(\boldsymbol{y^i}\right)} - \min_{k \leqslant t} \sum_{i=1}^{k} \log \frac{f_2\left(\boldsymbol{y^i}\right)}{f_1\left(\boldsymbol{y^i}\right)} > L, \tag{3.1}$$

where f_1 and f_2 are the density functions corresponding to F_1 and F_2, respectively, and L is a fixed threshold which is based on the rate of false alarms we are ready to accept. More conveniently, the CUSUM can be computed recursively via $S_0 = 0$ and

$$S_t = \max\left(S_{t-1} + \log \frac{f_2\left(\boldsymbol{y^t}\right)}{f_1\left(\boldsymbol{y^t}\right)}, 0\right). \tag{3.2}$$

Therefore, once we know the log ratio of densities, the CUSUM can be easily calculated. Intuitively, the CUSUM increases at time t if the likelihood that $\boldsymbol{y^t}$ is coming from f_2 is greater than the one that $\boldsymbol{y^t}$ is coming from f_1, i.e. $f_2\left(\boldsymbol{y^t}\right) > f_1\left(\boldsymbol{y^t}\right)$.

Assume now $\boldsymbol{y^t}$ may come from a multivariate normal with an *in-control* mean $\boldsymbol{\theta_1}$ or a multivariate normal with an *out-of-control* mean $\boldsymbol{\theta_2}$, i.e. $F_1 = \mathcal{N}(\boldsymbol{\theta_1}, \boldsymbol{\Sigma})$ and $F_2 = \mathcal{N}(\boldsymbol{\theta_2}, \boldsymbol{\Sigma})$ with a known covariance matrix $\boldsymbol{\Sigma}$. In our application, $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ would correspond to best-estimate mortality assumptions and to some mortality shocks, respectively. In that case, the log-likelihood ratio is given by

$$\log \frac{f_2(\boldsymbol{y}^t)}{f_1(\boldsymbol{y}^t)} = (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)'\Sigma^{-1}(\boldsymbol{y}^t - \boldsymbol{\theta}_1) - \frac{1}{2}(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)'\Sigma^{-1}(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1) \tag{3.3}$$

Inserting into Equation (3.2), the CUSUM for multivariate normal distribution, also called MCUSUM, is given by

$$S_t = \max\left(S_{t-1} + (\boldsymbol{\theta_2} - \boldsymbol{\theta_1})' \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{y^t} - \boldsymbol{\theta_1}\right) - \frac{1}{2} (\boldsymbol{\theta_2} - \boldsymbol{\theta_1})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta_2} - \boldsymbol{\theta_1}), 0\right). \tag{3.4}$$

Remark 3.1 *One may wonder why we model mortality by age groups rather than by single ages, and why we monitor the full age range instead of a narrower local window. There are two main reasons.*

First, the computation of the MCUSUM in (3.4) requires inversion of the covariance matrix. For the standard age range $[50, 90)$ at single-year resolution, this leads in our empirical applications to a high-dimensional and nearly singular covariance matrix. Grouping ages reduces dimensionality and yields a more stable multivariate monitoring procedure.

Second, age grouping reflects a trade-off between statistical stability and resolution. Finer groupings preserve more local age-specific information, but they increase sampling noise and estimation error in the covariance structure. Coarser groupings reduce noise and improve numerical stability, but may smooth out localized changes. We use 5-year age groups as a pragmatic compromise that is standard in demographic and actuarial applications and that performed well in our datasets.

This choice is not universal. If the practitioner is mainly concerned with a specific segment of the portfolio or liability structure, the same MCUSUM framework can be applied to a local age window (for example, ages 60–70 for pension-related applications) rather than to the full monitored age range. More generally, the relevant grouping should balance similarity within groups, dependence across groups, and the granularity needed for the decision problem at hand.

3.1 Change of level detection

The best-estimate mortality projections can be monitored by investigating a possible shift from the in-control mean $\boldsymbol{m_t}$ of the log death rates to an alternative out-of-control mean $\overline{\boldsymbol{m}}_t$. The change-point model can be expressed as

$$\mathbb{E}\left[\log(\boldsymbol{\mu_t})\right] \sim \begin{cases} \boldsymbol{m_t} & \text{for } i = 1, \ldots, \tau \\ \overline{\boldsymbol{m_t}} & \text{for } i = \tau + 1, \ldots \end{cases} \tag{3.5}$$

where $\overline{\boldsymbol{m}}_t$ corresponds to a possible change of level of the multivariate mortality process.

When the in-control and out-of-control Gaussian parameters are allowed to vary with time, the corresponding period-by-period Gaussian log-likelihood ratio can be inserted into the usual CUSUM recursion, yielding

$$\begin{aligned} S_t = &\max(S_{t-1} + (\overline{\boldsymbol{m_t}} - \boldsymbol{m_t})' \boldsymbol{\Sigma}_t^{-1} \left(\boldsymbol{y^t} - \boldsymbol{m_t}\right) \\ &- \frac{1}{2} (\overline{\boldsymbol{m_t}} - \boldsymbol{m_t})' \boldsymbol{\Sigma}_t^{-1} (\overline{\boldsymbol{m}}_t - \boldsymbol{m_t}), 0), \end{aligned} \tag{3.6}$$

where $\boldsymbol{y^t}$ is the vector of observed log death rates

$$\boldsymbol{y^t} = (\log(D_{z_1,t}/E_{z_1,t}), \ldots, \log(D_{z_M,t}/E_{z_M,t})), \tag{3.7}$$

for different M age-groups, e.g. $z_1 = [50; 55); z_2 = [55; 60); \ldots; z_M = [85; 90)$. Equation (3.6) is therefore the time-varying analogue of (3.4), obtained by replacing the constant parameters $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ with their time-dependent counterparts $(\boldsymbol{m}_t, \overline{\boldsymbol{m}}_t, \boldsymbol{\Sigma}_t)$ in the Gaussian log-likelihood ratio.

As an illustrative actuarial benchmark, one may consider the 20% longevity stress used in the Solvency II Standard Formula to compute the solvency capital requirement for the longevity risk submodule.[1] In our framework, such an externally specified stress can be represented as a proportional downward shift in mortality rates across all age groups:

$$\boldsymbol{\mu}_t^{\boldsymbol{shock}} = 0.8\boldsymbol{\mu_t} \iff \log\Big(\boldsymbol{\mu}_t^{\boldsymbol{shock}}\Big) = \log(\boldsymbol{\mu_t}) + \log(0.8)\mathbf{1}.$$

This example is used here only to illustrate how a pre-specified actuarial stress can be translated into the monitoring framework; it should not be interpreted as a realistic model of a pure level change, nor as a full representation of the Solvency II longevity stress.

Therefore, setting $\overline{\boldsymbol{m_t}} = \boldsymbol{m_t} + \log(\alpha)\mathbf{1}$ in Equation (3.5) with $\alpha = 0.8$ provides a simple illustration of how a uniform proportional mortality shock can be incorporated into the change-detection framework. Such a persistent proportional departure from the baseline is referred to here as a *change of level*; see also Abgrall et al. (2018) for related level-shift monitoring ideas in the Shiryaev–Roberts setting within a Poisson model.

In mortality applications, this terminology should be interpreted with care. Our objective is not to assume a literal discontinuity in mortality rates, but rather to detect the onset of a period during which the previously calibrated forecasting model becomes persistently misspecified relative to subsequent observations. In practice, such departures may correspond to sustained structural changes in mortality dynamics, for example prolonged slowdowns in mortality improvement rates (Djeundje et al., 2022) or structural changes in mortality projections documented in time-series models (Van Berkum et al., 2016). By contrast, shocks such as Covid-19 may be transitory or persistent depending on the context; see, for instance, the discussion in Richards (2024).

If one is only concerned with the change of level in the i-th age-group, we can of course set $\overline{\boldsymbol{m_t}} = \boldsymbol{m_t} + \log(\alpha)e_i$ where e_i is the vector of zeros except 1 in the i-th place.

3.2 Change of trend detection

Apart from detecting a change of level in mortality rates, actuaries are also concerned with the detection of a change of trend in mortality dynamics. In that case, rather than observing a sudden and permanent shift in mortality rates, the change develops gradually over time. To monitor such changes, it is natural to focus on *mortality improvement rates*. We define the yearly mortality improvement by

$$\begin{aligned}\boldsymbol{I}_t &:= \log\left(\frac{\boldsymbol{\mu}_{t-1}}{\boldsymbol{\mu}_t}\right) = -\Big(\log(\boldsymbol{\mu}_t) - \log\Big(\boldsymbol{\mu}_{t-1}\Big)\Big) \\ &= -\,\Delta\log(\boldsymbol{\mu}_t). \end{aligned} \tag{3.8}$$

With this convention, positive values of $\boldsymbol{I}_t$ correspond to improving mortality, and a decrease in $\boldsymbol{I}_t$ corresponds to a deterioration in the trend of mortality improvements.

[1] For further details on the Solvency II technical specifications, we refer to the Directive of European Commission (2009).

Ludkovski et al. (2018) remarked that there are two possibilities to model mortality improvements using GP. A first possibility is to compute historical mortality improvements and then model these directly with a GP. This approach is in the same spirit as Mitchell et al. (2013) who applied Lee-Carter on mortality improvements. The second possiblity is to remark that the *instanteneous* mortality improvement is also a GP, therefore obtaining it as a by-product, see Proposition 1 in Ludkovski et al. (2018). Since $\Delta \log(\boldsymbol{\mu}_t)$ is Gaussian as a difference of two Gaussian vectors, the mortality improvement vector $\boldsymbol{I}_t = -\Delta \log(\boldsymbol{\mu}_t)$ is also Gaussian:

$$\boldsymbol{I}_t \sim \mathcal{N}(\boldsymbol{m}_t^I, \boldsymbol{\Sigma}_t^I)$$

for some mean vector $\boldsymbol{m}_t^I$ and covariance matrix $\boldsymbol{\Sigma}_t^I$.

The change-point model for trend detection can then be expressed as

$$\mathbb{E}[\boldsymbol{I}_t] \sim \begin{cases} \boldsymbol{m}_t^I & \text{for } t = 1, \dots, \tau, \\ \overline{\boldsymbol{m}}_t^I & \text{for } t = \tau + 1, \dots, \end{cases} \tag{3.9}$$

where $\overline{\boldsymbol{m}}_t^I$ corresponds to a possible change in the mortality improvement trend. For instance, if we want to detect a deterioration in mortality improvements of size $\alpha > 0$, we may consider a downward shift in the improvement factor from $\boldsymbol{c}_t$ to $\boldsymbol{c}_t - \alpha$, where

$$\boldsymbol{c}_t := \exp\Big(\boldsymbol{m}_t^I\Big).$$

This yields the out-of-control mean

$$\overline{\boldsymbol{m}}_t^I = \log\left(\exp\Big(\boldsymbol{m}_t^I\Big) - \alpha\right),$$

which corresponds to weaker improvements than expected.

By a direct adaptation of (3.6), the MCUSUM for trend detection is given by

$$\begin{aligned} S_t = \max\Bigg(& S_{t-1} + \left(\overline{\boldsymbol{m}}_t^I - \boldsymbol{m}_t^I\right)' \left(\boldsymbol{\Sigma}_t^I\right)^{-1} \left(\boldsymbol{y}_t^I - \boldsymbol{m}_t^I\right) \\ & - \frac{1}{2}\left(\overline{\boldsymbol{m}}_t^I - \boldsymbol{m}_t^I\right)' \left(\boldsymbol{\Sigma}_t^I\right)^{-1} \left(\overline{\boldsymbol{m}}_t^I - \boldsymbol{m}_t^I\right), 0\Bigg), \end{aligned} \tag{3.10}$$

where $\boldsymbol{y}_t^I$ is the vector of observed mortality improvements, namely

$$\boldsymbol{y}_t^I = \log\Big(\boldsymbol{\mu}_{t-1}\Big) - \log(\boldsymbol{\mu}_t) = -(\boldsymbol{y}_t - \boldsymbol{y}_{t-1}).$$

3.3 Choice of threshold

The choice of the threshold L in (3.1) which determines the moment when we stop the algorithm is closely related to the concept of false alarm. In the literature of sequential control, the standard approach is to compute the *in-control average run length* ARL_0, that is the expected number of sampling periods until we hit the threshold L, in case there is no change (so the name "in-control") (Li et al., 2014, Pascual and Akhundjanov,

2020). For instance, if $ARL_0 = 100$ and the periods are years, this means that we have, on average, a false alarm once every 100 years. Another approach to determine the threshold is to look at the probability of false alarm rather than the expected average time (Abgrall et al., 2018). In this case, the threshold L is determined by solving

$$\mathbb{P}\left[\max_{1\leq i\leq T} S_i \geq L\right] = \alpha \tag{3.11}$$

where $(S_i)_{1\leq i\leq T}$ is the MCUSUM sequence when no change occurs (i.e. in our case, the sequence is following the expected GP-based mortality forecasts). The parameter α needs to be determined by the user based on the desired trade-off between false alarms and detection delay. In particular, a more risk-averse decision-maker may be willing to tolerate a higher false-alarm probability in exchange for faster detection of adverse changes.

4 Monitoring longevity and mortality risks: Applications to real mortality data

In this section we apply the MCUSUM algorithm described in Section 3 to national mortality data from France, Canada, USA and Japan, and on a life insurance portfolio. For the national data, we focus on the male death data and the corresponding exposures, for ages 50-89 by age tranches of 5 years and the years 1991-2020 extracted from the Human Mortality Database (HMD).[2] In our analysis, we split the data into two parts: the first 20 years (1991-2010) are used to calibrate the model using the GP approach described in Section 2 and the last 10 years (2011-2020) are used to detect a possible longevity or mortality risk in the data using the MCUSUM approach. This choice reflects a trade-off. A sufficiently long pre-monitoring period is needed to estimate the baseline mortality surface and, in particular, the improvement trend in a stable manner. At the same time, using a relatively recent calibration window helps keep the baseline representative of contemporary mortality dynamics. To assess the sensitivity of our conclusions to this choice, we also considered an alternative split with an earlier calibration period (1981–2000) and a longer monitoring window (2001–2020); the corresponding results are reported in Appendix A. The main qualitative conclusions remain unchanged. Similar to Ludkovski et al. (2018), the R package **DiceKriging** (Roustant et al., 2012) was used to implement our GP model. The implementation of the MCUSUM for level and trend detection given in Equations (3.6) and (3.10) respectively is straightforward. To determine the threshold (3.11), the probability of false alarm is set to 1%. Since the resulting threshold L may differ across countries and monitoring specifications, we report in the figures the normalized monitoring statistic

$$\widetilde{S}_t = \frac{S_t}{L}.$$

This rescaling is used only for visual comparability across datasets: the alarm threshold then becomes $\widetilde{S}_t = 1$ in all cases, and the stopping rule is equivalently given by $\widetilde{S}_t > 1$.

[2] See www.mortality.org.

4.1 Detecting a change of level

4.1.1 Study on HMD data

First, we apply the MCUSUM to detect a possible increase or decrease of 5% in the level of death rates, i.e. we set

$$\overline{\boldsymbol{m_t}} = \boldsymbol{m_t} + \log(\rho)\mathbf{1}$$

in Equation (3.5) with $\rho = 0.95$ for the longevity level risk and $\rho = 1.05$ for the mortality level risk.

Figure 2 shows the MCUSUM for the mortality and longevity risk over the 10-year detection period 2011-2020 for Japan, France, Canada and USA. We presented also the percentage change in the 40-year period survival probability of a male individual aged 50 to illustrate the change between the observed survival probability and the GP-predicted survival probability. We observe that a change of level is detected at $\rho = 1.05$ for the USA and Canada indicating a mortality risk for both countries, and a change at $\rho = 0.95$ for Japan, which shows evidence of a 5% decrease in Japanese male death rates for all age tranches. This is in agreement with other case studies that observed a mortality deterioration in North America but improvements in Japan (Case and Deaton, 2015, Leon et al., 2019, Raleigh, 2019, Djeundje et al., 2022). A close look at the survival probability indicates a sharp decline in 2020 due to Covid-19 with a 7% percentage loss for the survival probability in the USA.

The case of Japan is particularly interesting because the longevity-risk MCUSUM is triggered (Figure 2, panel **C**) even though the observed and GP-predicted 40-year survival probabilities remain close (Figure 2, panel **A**). This illustrates that the multivariate monitoring procedure can reveal changes that are not visible from a simple aggregate survival indicator. To better understand this signal, Figure 3 reports the percentage difference between observed and GP-predicted death rates by age group and year for Japanese males. A compensation effect appears: mortality worsens at the oldest ages, but this is offset in the aggregate by mortality improvements at younger ages, especially around ages 50–59. This suggests that the MCUSUM signal is driven by a change in the age pattern of mortality deviations rather than by a uniform shift in aggregate mortality.

The previous analysis suggests that, at least for Japan, the relevant deviation from the baseline is not well represented by a uniform level shift across all ages. Instead, the age-specific residuals point to a structured pattern combining lower-than-expected mortality at younger ages and higher-than-expected mortality at the oldest ages.

To illustrate the flexibility of the proposed framework, we next consider a *retrospective age-structured alternative* in which the out-of-control mean varies by age group rather than through a common multiplicative factor. This exercise is not intended as a genuine ex ante detection experiment, since the specification is motivated by the age profile observed in Figure 3. Its purpose is instead to show that, when a practitioner has prior reason to monitor a non-uniform age pattern, an appropriately specified MCUSUM can detect such departures earlier than a uniform-shift detector.

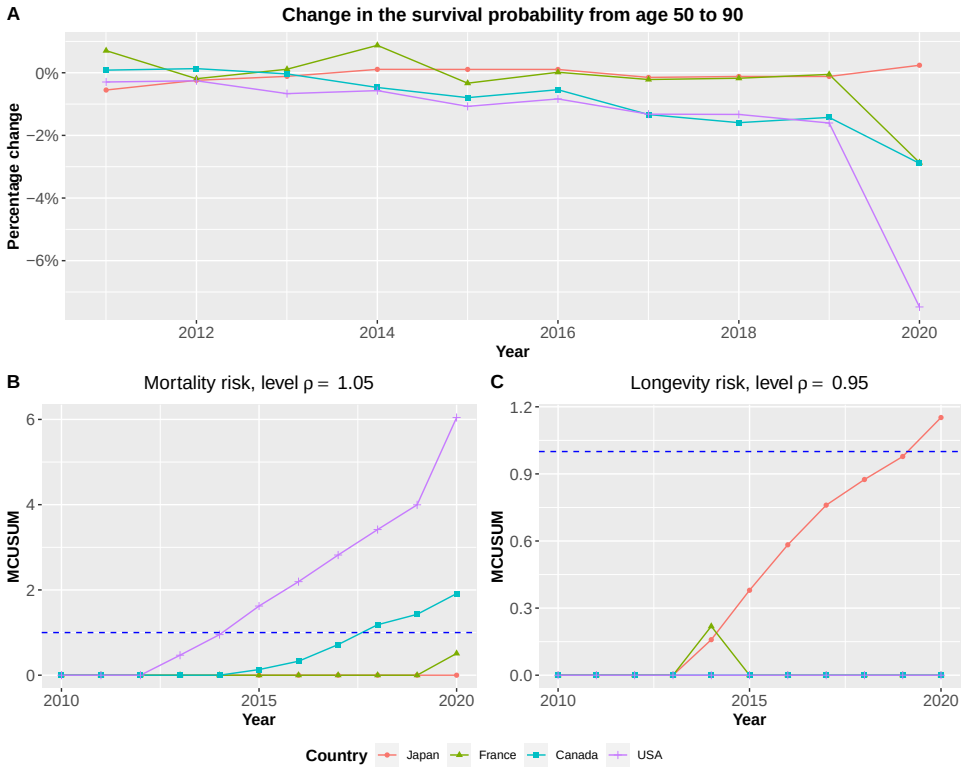


Figure 2: **A:** Percentage change difference of the 40-year survival probability of an individual aged 50 between the observed survival probability and the GP-predicted survival probability. **B:** MCUSUM for the detection of a change at the level $\rho = 1.05$ (mortality risk). **C:** MCUSUM for the detection of a change at the level $\rho = 0.95$ (longevity risk). Dotted lines represent the detection threshold.

To illustrate this idea, Figure 4 reports two stylized age-structured alternatives applied to all four populations based on HMD data for ages 50-99 by 5-year age groups. We first consider a stylized age-structured alternative, loosely motivated by the U.S. pattern: mortality rates below age 70 are multiplied by 0.95, rates between ages 70 and 80 are unchanged, and rates above age 80 are multiplied by 1.05 (left panel of Figure 4). Relative to the uniform-shift specification of Figure 2, this alternative illustrates how the MCUSUM can be adapted to monitor simultaneous longevity improvements at younger ages and mortality deterioration at older ages.

We then consider a second stylized age-structured alternative motivated by the Japanese age profile: mortality rates below age 80 are multiplied by 0.95, rates between ages 80 and 90 are unchanged, and rates above age 90 are multiplied by 1.05 (right panel of Figure 4). Under this retrospective specification, the monitoring statistic signals earlier than under the uniform level-shift alternative. This does not mean that the age pattern was identified prospectively in 2014; rather, it shows that non-uniform alternatives can be substantially more informative when the underlying departure from the baseline operates through a changing age structure.

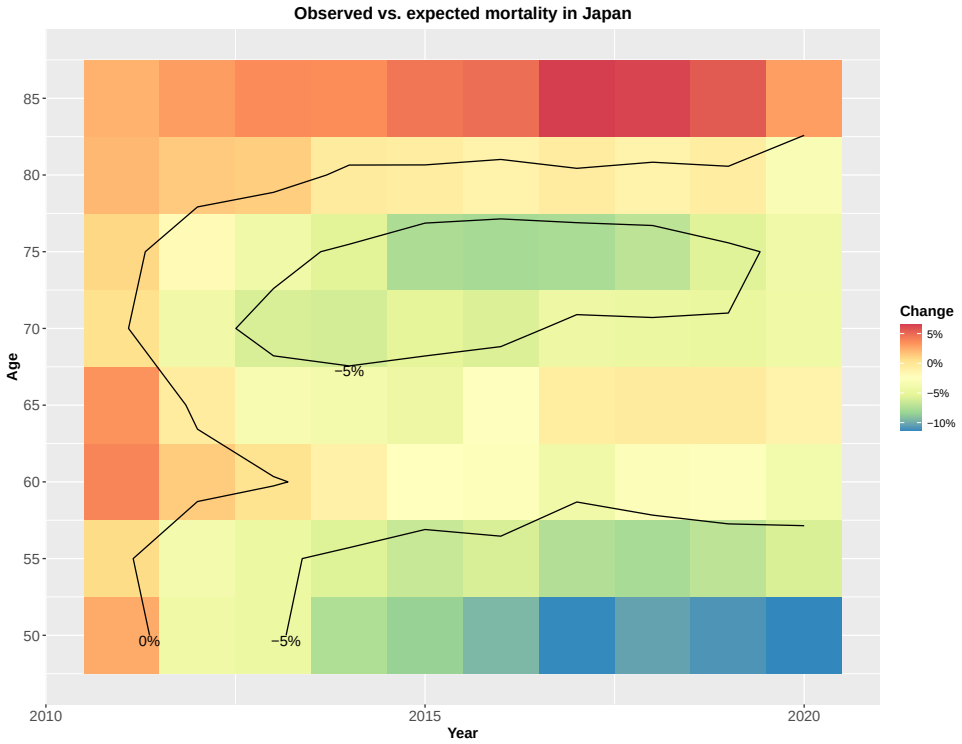


Figure 3: Percentage change between observed and GP-predicted death rates by age tranches for Japanese males.

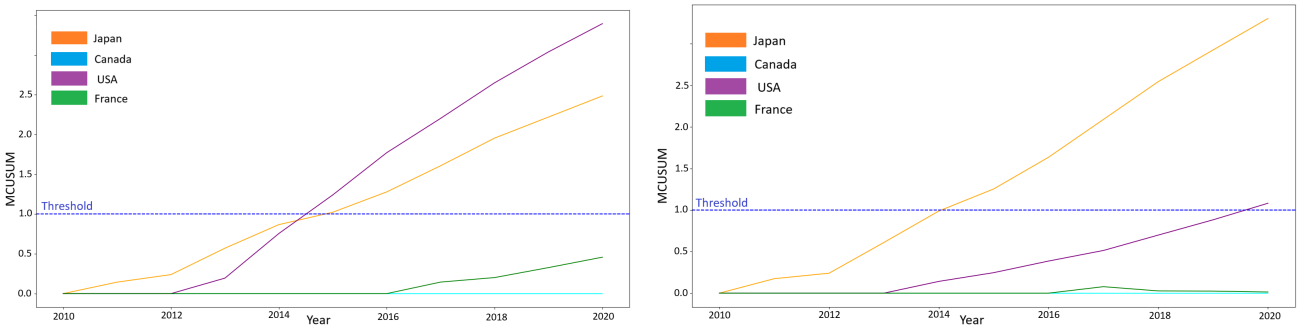


Figure 4: **Left:** MCUSUM for the HMD national-population data with a changing detection level:$\rho = 0.95$ before 70 years, $\rho = 1$ between 70 and 80 years, $\rho = 1.05$ after 80 years. **Right:** MCUSUM for for the HMD national-population data with a changing detection level: $\rho = 0.95$ before 80 years, $\rho = 1$ between 80 and 90 years, $\rho = 1.05$ after 90 years.

4.1.2 Study on real-world insurance data

To complement the study based on HMD data, we now present an illustration using a real-world French life insurance portfolio. The underlying dataset covers the period 1970–2024 and contains approximately 179,000 insured individuals, around 70% of whom are male. In the present analysis, we focus on the richest male quartile of the portfolio, corresponding

to roughly 31,000 policyholders. The age range considered in this illustration is 65–99 years.

After a preliminary mortality analysis based on standardized mortality ratios (SMRs), comparing the insurer population mortality with the French national insured population prospective mortality table for males (TGH05), we observe that for this wealthy male subpopulation the SMRs are below 1 at younger old ages and above 1 at older ages (see Figure 5). This pattern is consistent with a possible rectangularization phenomenon (Fries, 1980, Carter et al., 2001) and further supports the relevance of refining classical detection techniques in order to allow for different behaviours across age groups.

We perform a detection analysis similar to the one conducted for HMD data on this portfolio and motivated by the observed mortality rate switch from Figure 5. These results (see Figure 6) confirm the hypotheses of mortality level changes across different age groups. Indeed, we do not detect anything in the first three graphs, which correspond to the cases where:

1. we have $\rho = 1.05$ constant for all age groups, indicating an overall increase in mortality across the entire portfolio;
2. we leave ρ unchanged below 80 years old and $\rho = 1.05$ after, which corresponds to an increase in mortality after 80 years;
3. we have $\rho = 0.95$ below 80 years old and unchanged afterward, indicating an increase in longevity before 80 years.

Finally, we detect a regime change in the last case, when we monitor both a 5% increase in longevity for age groups under 80 ($\rho = 0.95$) and a 5% increase in mortality for those over 80 ($\rho = 1.05$).

These results seem to be in line with the rectangularization of mortality curves, characterized by a compression of mortality rates towards older ages and predicted by demographers like Olshansky et al. (2005). Gavrilova and Gavrilov (2015) believe that there is a biological age limit or a *Gompertzialization* of mortality (which increases exponentially with age). This means that even if mortality rates among younger age groups continue to decline, overall life expectancy will not increase significantly, even in well-developed countries, due to the increase in mortality rates among older age groups. This trend can be explained by the *biological limits to longevity* (see Olshansky et al. (2005)). While the human body has inherent biological constraints that make it more susceptible to age-related illnesses, such as cancer, cardiovascular diseases, and neurodegenerative disorders. These illnesses tend to be more severe in old age and are less easily recoverable. Others factors (global warming, pollution, health system saturation) can negatively affect older populations, increasing vulnerability to health risks.

We now modify our real-world dataset (a group of male policyholders from a portfolio of a French life insurer) by a uniform and constant factor that does not affect the interpretation of the results. The mortality rates calculated from this dataset represent average mortality rates over the five-year period from 2017 to 2022. These rates have

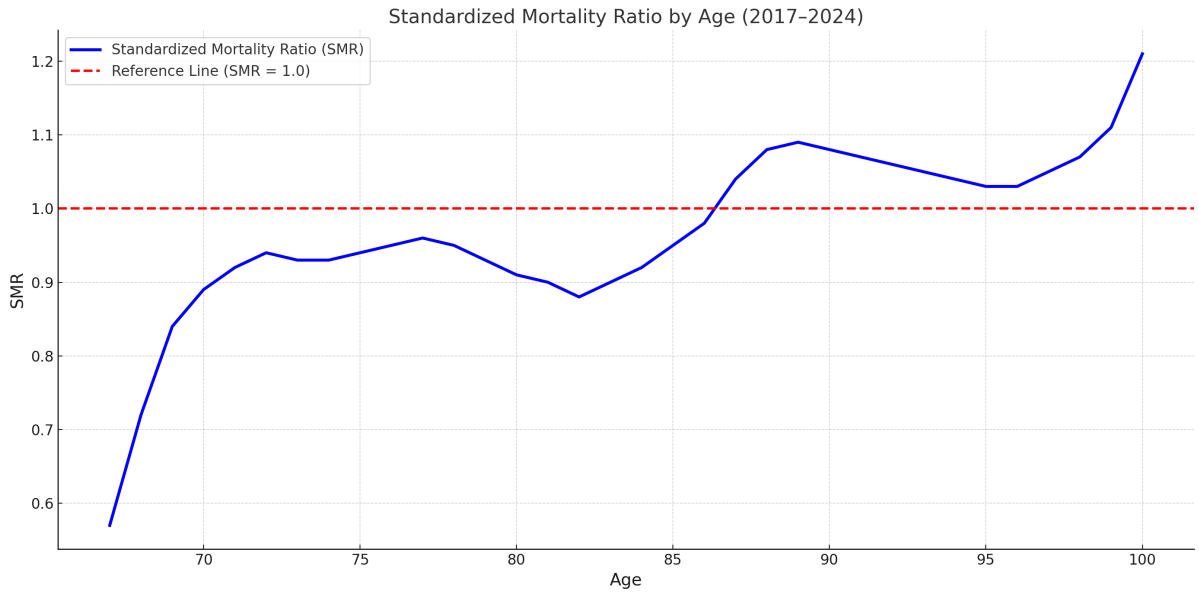


Figure 5: Mean Standardized Mortality Ratio by single age within the richest quartile of the insurer's portfolio and the TGH05 data from 2017 to 2024.

been smoothed using a Savitzky–Golay (SG) filter (see Schafer (2011)), which is a signal-processing technique designed to preserve broad trends while reducing noise. Our filter applies a fifth-order localized polynomial regression to obtain a smoother representation of mortality rates and make the age pattern easier to visualize. We emphasize, however, that this smoothing step is used here for descriptive purposes only, in order to make the rectangularization pattern more visible. Smoothing may affect the behavior of monitoring procedures such as MCUSUM: while it reduces high-frequency noise, it can also attenuate abrupt local deviations and thus alter the timing or magnitude of detection signals. The smoothed curves should therefore be interpreted as illustrative rather than as the basis for a formal real-time monitoring exercise. Additionally, we use again the prospective mortality table TGH05 (2006) (still in force for male insured population in France). It provides age-specific mortality rates for male insured individuals and is commonly used in French actuarial calculations to estimate life expectancy and assess longevity risk since 2007.

The toy model adopted in Figure 7 is as follows: starting from the TGH05 mortality table, 20% of individuals who were originally expected to die before the age of 87 are assumed to avoid *death* due to advancements in medical science, their access to new treatments, improved healthcare facilities, better preventive measures, and enhanced disease management strategies: they continue their life and we assume in this model that they do not have any additional mortality risk penalty (compared to other ones) after this avoided death instant. Secondly, we increase the mortality rates projected initially by this table in 2006 by 10% due to factors such as the increasing prevalence of chronic diseases, environmental changes affecting public health, an aging population with higher frailty, and the potential long-term effects of emerging health crises. The impact of this modification to the TGH05 mortality table is illustrated in Figure 7. We observe that our toy model

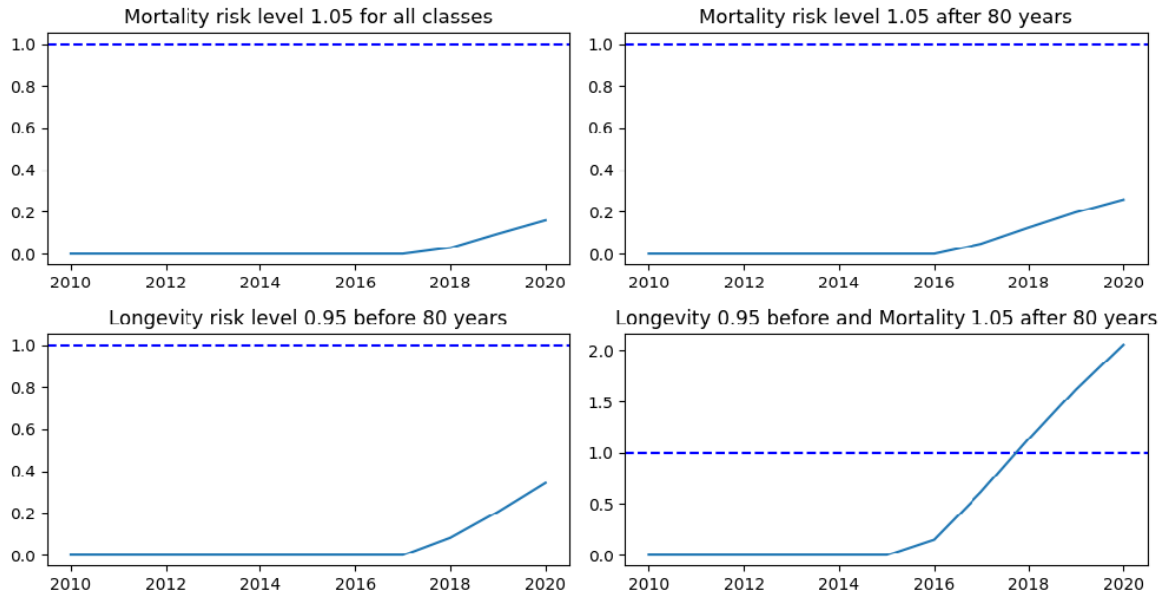


Figure 6: Multivariate CUSUM for insurer's richest quartile portfolio.

closely aligns with the filtered and adjusted mortality rates derived from the insurer's population data. The fact that we can reproduce real-world mortality patterns with a very simple model shows that it is likely that some future actuarial assumptions may include a rectangularisation component. It is therefore desirable to anticipate this evolution and to opt for monitoring techniques that can accomodate such refinement, which is the case for MCUSUM. As mortality rates become more rectangular, with a greater concentration of deaths occurring at older ages, the detection of changes in mortality patterns becomes even more critical. The MCUSUM, with its ability to capture joint behavior and correlations among age tranches, offers a valuable tool for monitoring these changes and assessing longevity and mortality risks. The comparison of MCUSUM performance with other techniques to monitor rectangularisation is left for future research.

4.2 Detecting a change of trend

As pointed out by Hunt and Villegas (2022), actuaries in practice are often interested in the mortality improvement rates rather than the level of death rates. Indeed, practitioners usually make a level assumption to move from the national population to the specific insurance portfolio or pension scheme. Then, improvements rates, often estimated at the national level, are applied to the sub-populations even though there might be a basis risk (Villegas et al., 2017). The concept of mortality improvement rates is widely adopted e.g. in the United Kingdom with the CMI Mortality Projection Model (see CMI (2009)) and the Scale AA improvement rates by the Society of Actuaries in the United States (see SOA (2020)).

Following Section 3.2, we consider the detection of a change in mortality improvement

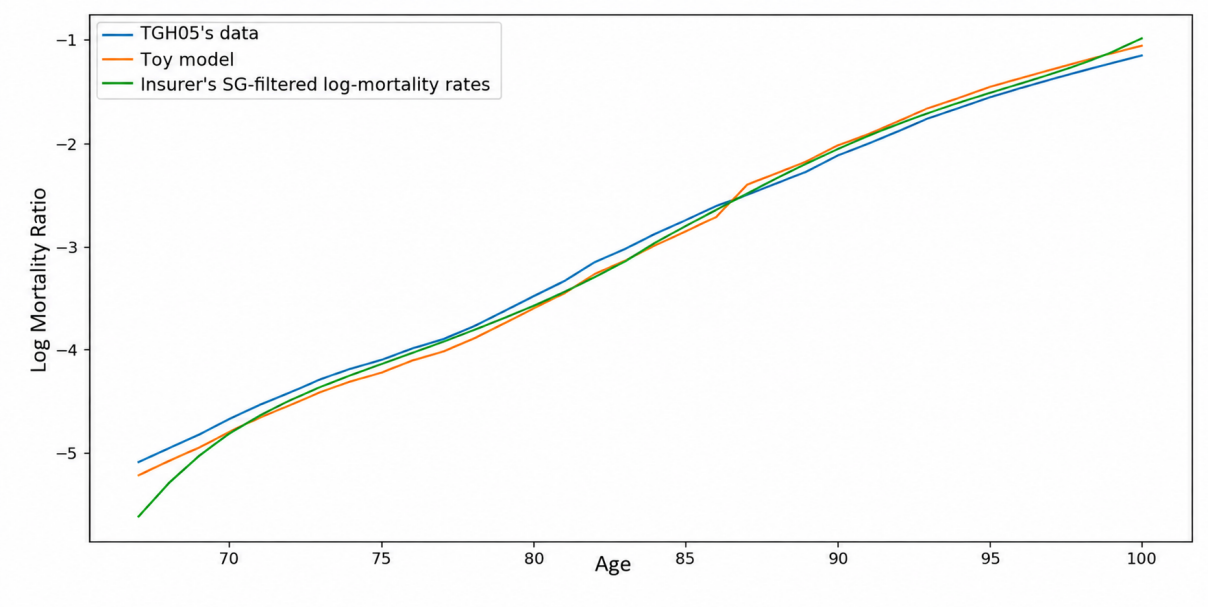


Figure 7: TGH05's, SG-filtered insurer's mortality rates by and toy model mortality rate from 2017 to 2022.

trends. More specifically, we consider an increase or decrease of 1% in the mortality improvement factor by setting

$$\overline{\boldsymbol{m}}_t^I = \log\Big(\exp\Big(\boldsymbol{m}_t^I\Big) - \alpha\Big)$$

with $\alpha = 0.01$ for weaker-than-expected improvements and

$$\overline{\boldsymbol{m}}_t^I = \log\Big(\exp\Big(\boldsymbol{m}_t^I\Big) + \alpha\Big)$$

with $\alpha = 0.01$ for stronger-than-expected improvements.

Figure 8 shows the MCUSUM for the mortality and longevity trend risk over the 10-year detection period 2011-2020 for Japan, France, Canada and USA. The figure also represents for each year the observed difference in mortality improvements averaged over age tranches:

$$\text{Difference}_t = \frac{1}{M}\sum_{i=1}^{M}\Big(\text{MI}_{z,t}^{obs} - \text{MI}_{z,t}^{GP}\Big) \tag{4.1}$$

where M is the number of age classes, $\text{MI}_{z,t}^{obs}$ and $\text{MI}_{z,t}^{GP}$ are the observed mortality improvements and the GP-predicted mortality improvements for the age-tranche z and year t.

Contrary to the MCUSUM for the level detection, the MCUSUM for the trend detection is more unstable due to the volatility of the mortality improvements themselves. Indeed, it is well documented that observed mortality improvements are extremely noisy as being the ratio of two small quantities (see Hunt and Villegas (2022) and the references therein). In particular, for a change of 1% with a probability of false alarm of 1%, the MCUSUM for the longevity and mortality trend risk hits the threshold for almost

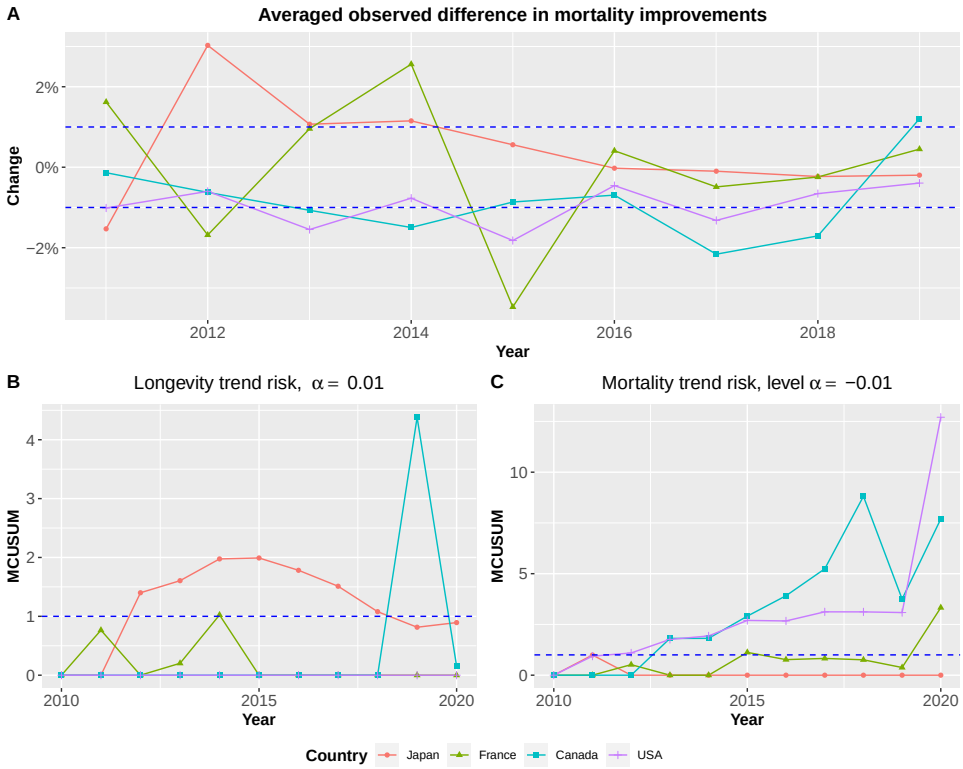


Figure 8: **A:** Difference averaged by age tranches between the observed improvement rates and the GP-predicted improvement rates. Dotted lines represent the levels 1% and −1%. **B:** MCUSUM for the detection of a change in the trend at the level $\alpha = 1\%$ (longevity trend risk). **C:** MCUSUM for the detection of a change in the trend at the level $\alpha = -1\%$ (mortality trend risk). Dotted lines in **B** and **C** represent the detection threshold ($L = 1$).

all countries (except the USA for the longevity trend risk). These results are consistent with Figure 8 (Graph **A**) where we notice that the year of detection of the MCUSUM corresponds to the year in which the average observed difference (4.1) exits the range $[-1\%, 1\%]$. Given the volatility of mortality improvements, this analysis suggests to increase the parameter $|\alpha| > 0.01$ or decrease the probability of false alarm to e.g. 0.1 % to focus on large long-term changes of the trend.

To complete our analysis, we provide in Figure 9 the difference between the observed and GP-predicted mortality improvements in Japan, complementary figure to mortality rates in Figure 3. We remark that the MCUSUM for the longevity trend risk was essentially driven by the mortality improvements in the younger age classes (50-70) during the period 2011-2015 but since then, there is a stagnation of mortality improvements. Such stagnation can also be observed by the decrease of the MCUSUM for longevity trend risk for Japan on Figure 8 (Graph **B**).

To conclude this section, both level risk and trend risk are important for effective mortality risk management. Informally, level risk concerns whether mortality is immediately above or below the baseline forecast, whereas trend risk concerns whether the *rate of mortality improvement* has changed. Monitoring trend risk is more challenging because a

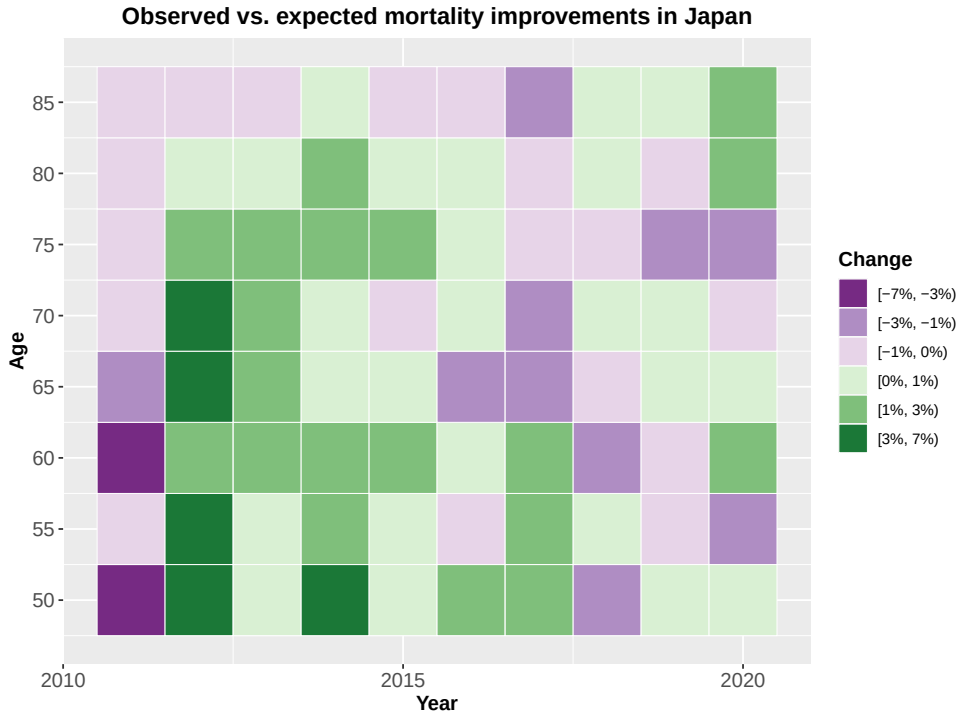


Figure 9: Difference between observed and GP-predicted mortality improvement rates by age tranches for Japanese males.

change in trend affects the slope of the mortality trajectory rather than its current level, so the resulting deviation from the baseline accumulates only gradually over time. Consequently, more post-change observations are typically needed to detect a trend change reliably than a level shift. For this reason, in our real-world dataset we only report results for level-risk monitoring.

5 Simulation study: Comparison of the MCUSUM and CUSUM charts

In this section, we evaluate the performance of the proposed multivariate control chart based on GP regression in monitoring changes in the distribution of death rates. Specifically, we compare the performance of the multivariate cumulative sum (MCUSUM) chart with two univariate counterparts, the Min-MCUSUM and the comonotonic CUSUM defined hereafter which do not account for the dependence between age classes. We perform a simulation study to investigate the effect of the dependence structure between age groups on the performance of the control charts. In particular, we show that the MCUSUM outperforms the Min-MCUSUM and the comonotonic CUSUM and such effect is increasing with the correlation between age classes. Such result is line with other studies, e.g. Jiang et al. (2011) who studied spatial correlation between regions and Pascual and Akhundjanov (2020) who considered correlated Poisson processes.

For the simulation study, we assume that we want to detect a proportional change of level that affects all age tranches, that is

$$\mathbb{E}\left[\log(\boldsymbol{\mu_t})\right] \sim \begin{cases} \boldsymbol{m_t} & \text{for } t = 1, \dots, \tau \\ \overline{\boldsymbol{m_t}} = \boldsymbol{m_t} + \log(\alpha)\mathbf{1} & \text{for } t = \tau + 1, \dots \end{cases} \tag{5.1}$$

as introduced in Section 3.1.

Hereafter, we compare the performance of the MCUSUM (3.6) with two CUSUM alternatives that do not account for the dependence between age classes that are introduced hereafter.

5.1 Univariate CUSUM alternatives

5.1.1 Min-MCUSUM

One naive way to detect a change that affects all age tranches is to apply a univariate CUSUM for each age tranche $i = 1, \dots, M$ separately and stop the detection algorithm once all M univariate CUSUMs are above the threshold.

The univariate CUSUM for the age tranche i is given by

$$S_{i,t} = \max(S_{i,t-1} + (\overline{m}_{i,t} - m_{i,t})\frac{(y_{i,t} - m_{i,t})}{\sigma_{i,t}} - \frac{1}{2}\frac{(\overline{m}_{i,t} - m_{i,t})^2}{\sigma_{i,t}}, 0), \quad \text{for } i = 1, \dots, M,$$

which is the univariate special case of the MCUSUM (3.6) with $m_{i,t}$ and $\sigma_{i,t}$, the mean and standard deviations of the i-th component of the log death rates vector $\boldsymbol{y_t} = (y_{1,t}, \dots, y_{M,t})$, and $\overline{m}_{i,t} = m_{i,t} + \log(\alpha)$ in line with (5.1).

The run time until detection (also called run length) for the min-MCUSUM is defined as

$$N = \min\left\{t \geq 1 : \min_{1 \leq i \leq M} S_{i,t} \geq H\right\}. \tag{5.2}$$

Hence, the detection algorithm is stopped the first time all M CUSUM processes are above the threshold. We remark that in this case, all age classes are monitored independently using the univariate CUSUM and the process min-MCUSUM is considered out of control when all age classes are out of control.

5.1.2 The comonotonic CUSUM

Another interesting case is to compare the MCUSUM with the CUSUM obtained under the assumption of perfect dependence which is motivated by common factor models such as the Lee-Carter. Indeed, in the Lee-Carter model, log death rates are driven by the common period effect κ_t, and therefore changes in log death rates over time are perfectly correlated across age. Such perfect correlation is unrealistic and undesirable.

As pointed out by Cairns et al. (2008), there are two problems that arise from using a model that assumes perfect correlation. First, the model may overstate the aggregate

levels of uncertainty at the portfolio level because it assumes that there are no diversification benefits across ages. Second, perfect correlation would suggest that derivative instruments linked to mortality at one age could be used to hedge perfectly mortality improvements at a different age, which is not accurate. If this problem is ignored, it may result in a hidden basis risk if hedging strategies are adopted that rely on perfect correlation.

Consider for instance the log death rate for age x and time t in the Lee-Carter model:

$$\log(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t,$$

with $\kappa_t \sim \mathcal{N}(\mu, \sigma)$. Then, the sum across age is given by

$$\sum_x \log(\mu_{x,t}) \sim \mathcal{N}(\mu^c, \sigma^c),$$

with $\mu^c = \sum_x \alpha_x + \sum_x \beta_x \mu$ and $\sigma^c = \sum_x \beta_x \sigma$. Therefore, the sum of log death rates across age is a comonotonic sum completely driven by the period effect κ_t, and the standard deviation of the sum is the sum of the standard deviation (Dhaene et al., 2002*a*,*b*).

More generally, if we consider a change of level α in the death rates, this would lead to

$$\begin{aligned}\sum_{x=1}^{M} \log(\alpha \mu_{x,t}) &= \sum_{x=1}^{M} \log(\mu_{x,t}) + M \log(\alpha) \\ &\sim \mathcal{N}(\mu^c + M \log(\alpha), \sigma^c),\end{aligned}$$

This motivates the following comonotonic CUSUM, denoted C-CUSUM, defined as

$$S_t^c = \max\left(S_{t-1}^c + (\bar{m}_t^c - m_t^c) \frac{(s_t^c - m_t^c)}{\sigma_t^c} - \frac{1}{2} \frac{(\bar{m}_t^c - m_t^c)^2}{\sigma_t^c}, 0 \right).$$

with

$$\begin{aligned} m_t^c &= \sum_{i=1}^{M} m_{i,t}, \\ \bar{m}_t^c &= \sum_{i=1}^{M} m_{i,t} + M \log(\alpha) \\ \sigma_t^c &= \sum_{i=1}^{M} \sigma_{i,t} \\ s_t^c &= \sum_{i=1}^{M} y_{i,t} \end{aligned}$$

where we recall that $m_{i,t}$ and $\sigma_{i,t}$, are the mean and standard deviations of the i-th component of the log death rates vector $\boldsymbol{y_t} = (y_{1,t}, \ldots, y_{M,t})$.

The C-CUSUM corresponds to the univariate CUSUM of the comonotonic sum s_t obtained by summing the log death rates across age.

The run time until detection for the C-CUSUM is defined as

$$N = \min\left\{ t \geq 1 : S_t^c \geq L \right\}. \tag{5.3}$$

5.2 The simulation setup

For a given monitoring scheme, the *run length* N is defined as the number of sampling periods until the procedure signals, i.e. the first time the detection statistic crosses the decision threshold:

$$N = \inf\{i \geq 1 : S_i \geq L\},$$

see Equations (3.1), (5.2), (5.3) for the precise definitions of the charting statistics used in the MCUSUM, Min-MCUSUM and C-CUSUM procedures, respectively.

The expected run length, $ARL = \mathbb{E}(N)$, is called the *average run length* and is the standard performance metric for control charts and sequential monitoring procedures (Montgomery, 2007, Qiu, 2013, Bersimis et al., 2007). When no change occurs, ARL quantifies false-alarm behavior (the *in-control* average run length, often denoted ARL_0). When a change occurs at some time point, ARL measures detection speed (the *out-of-control* average run length, ARL_1), with smaller values corresponding to faster detection.

The thresholds for the MCUSUM, Min-MCUSUM and C-CUSUM charts are selected to control the probability of at least one false alarm over a fixed monitoring horizon T:

$$\mathbb{P}\left[\max_{1\leq i\leq T} S_i \geq L\right] = \alpha, \tag{5.4}$$

where $(S_i)_{1\leq i\leq T}$ denotes the detection process under the in-control model (i.e. when no change occurs). In the simulation study, we set $\alpha = 5\%$ and $T = 100$. Since one sampling period corresponds to one year in our applications, this means that under the in-control model there is a 5% probability that the chart signals at least once over the next 100 years. For a fixed false-alarm specification, we compare procedures using their out-of-control ARL_1; the most efficient procedure is the one with the lowest ARL_1, since it detects the change fastest on average.

To align with the empirical analysis, we use Japanese male mortality data from 1990 to 2010 for ages 50 to 89 grouped into 5-year age tranches. Mortality forecasts are estimated using GP regression as outlined in Section 2. Then, for each detection procedure (MCUSUM, Min-MCUSUM, C-CUSUM), we proceed in two steps:

- **Determine the threshold L in (5.4).** We simulate 25,000 in-control trajectories of future death rates from the fitted GP model and compute the empirical $(1-\alpha)$-quantile of $\max_{1\leq i\leq T} S_i$.
- **Estimate the out-of-control average run length.** Once L is fixed, we simulate 25,000 out-of-control trajectories and estimate ARL_1 by the sample mean of the corresponding run lengths, considering a proportional level shift affecting all age tranches from time 0.

For the dependence between age classes, we consider the stylized correlation matrix given in Table 1. In this specification, adjacent age classes have correlation coefficient ρ, whereas non-adjacent age classes are assumed to be uncorrelated. The parameter ρ varies from 0 to 0.5 to represent independent, weakly correlated, and moderately correlated

settings. This parsimonious structure is used only for the simulation study and is intended to reproduce the broad short-range dependence pattern suggested by the GP-implied predictive correlation matrix in Figure 10, in which dependence is strongest between neighboring age groups and rapidly decreases with age distance.



Figure 10: GP-implied predictive correlation matrix for Japanese male log death rates in forecast year 2011.

	[50,55)	[55,60)	[60,65)	[65,70)	...	[85,90)
[50; 55)	1	ρ	0	0	0	0
[55; 60)	ρ	1	ρ	0	0	0
[60,65)	0	ρ	1	ρ	0	0
[65,70)	0	0	ρ	1	ρ	0
...	0	0	0	ρ	1	ρ
[85; 90).	0	0	0	0	ρ	1

Table 1: Correlation matrix between age tranches used for the simulation study.

5.3 Simulation results

5.3.1 Comparing MCUSUM and Min-MCUSUM

Figure 11 reports the ratio

$$\mathrm{ARL}_1(\text{Min-MCUSUM})/\mathrm{ARL}_1(\text{MCUSUM})$$

of the out-of-control ARLs of the Min-MCUSUM to those of the MCUSUM for different change levels $\alpha = 0.9, 0.95, 1.05, 1.08, 1.1$ and correlation parameters $\rho = 0, 0.1, 0.2, 0.3, 0.4, 0.5$ in the correlation matrix of Table 1. Ratios above one indicate superior performance of the MCUSUM, which is observed throughout. The gain is especially pronounced when the cross-age correlation is stronger and when the change of level is smaller. This reflects the fact that the MCUSUM uses the full covariance structure and can therefore exploit joint information across age groups, whereas the Min-MCUSUM is based on separate univariate monitoring.

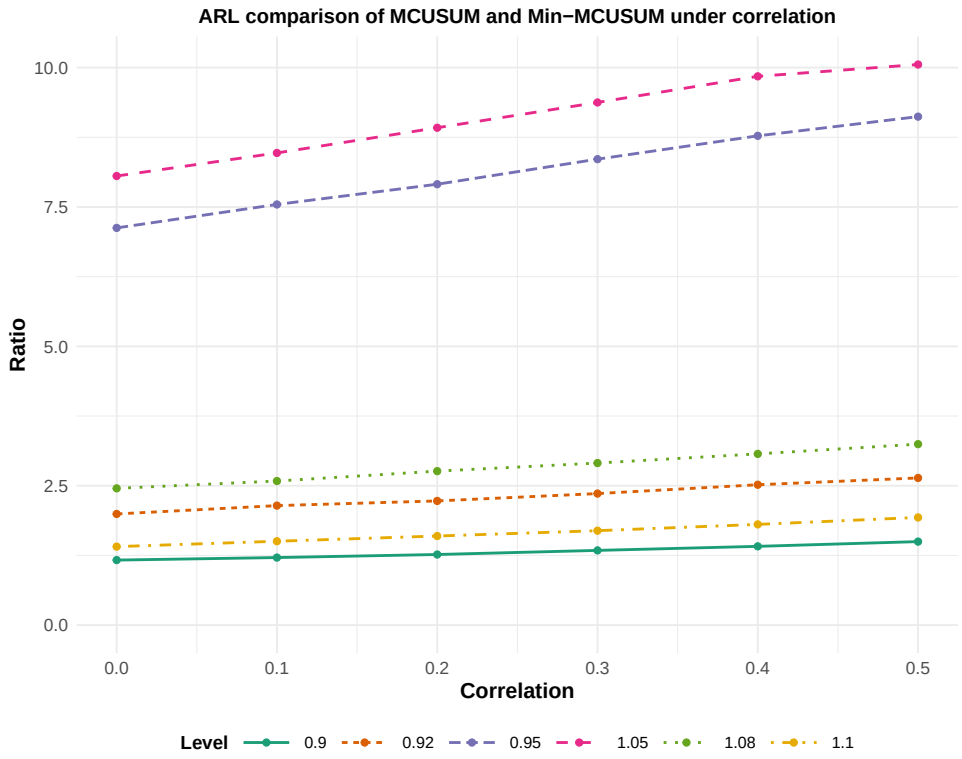


Figure 11: Ratio ARL_1(Min-MCUSUM)/ARL_1(MCUSUM) under the simulated dependence structure of Table 1. Values above one favor the MCUSUM.

5.3.2 Comparing MCUSUM and C-CUSUM

Figure 12 reports the ratio

$$\mathrm{ARL}_1(\text{C-CUSUM})/\mathrm{ARL}_1(\text{MCUSUM})$$

for the same set of change levels and correlation parameters. Again, the MCUSUM dominates across the scenarios considered, with larger gains for stronger dependence and smaller shifts. This indicates that accounting for the full dependence structure is preferable to the comonotonic approximation in the class of common level changes studied here.

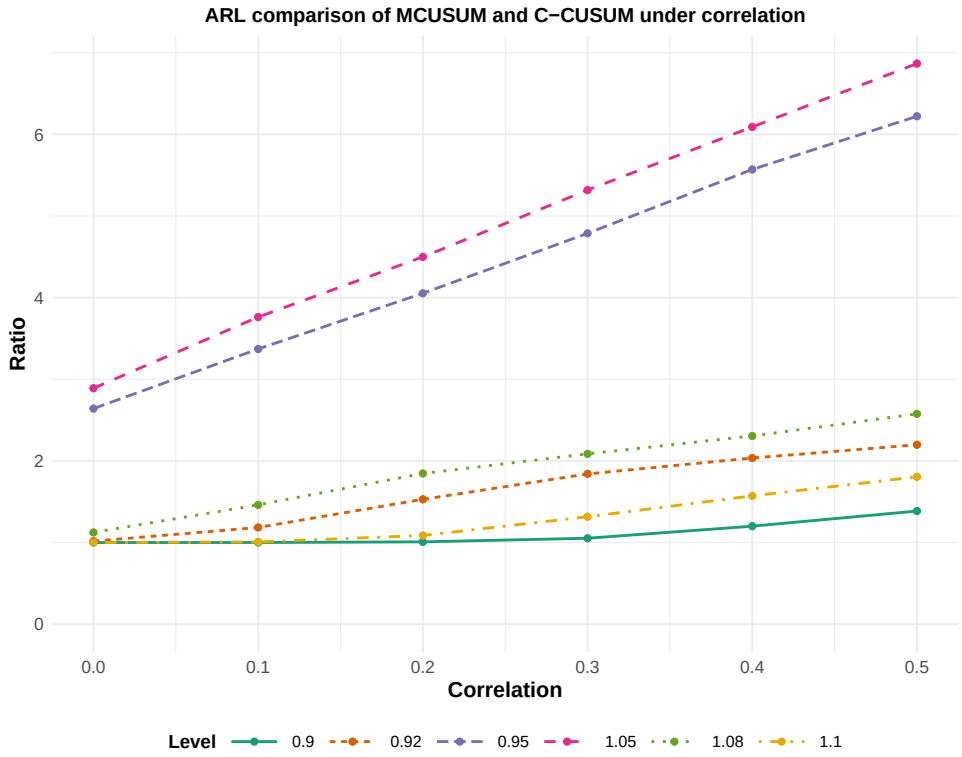


Figure 12: Ratio ARL_1(C-CUSUM)/ARL_1(MCUSUM) under the simulated dependence structure of Table 1. Values above one favor the MCUSUM.

Overall, the simulation study shows that, for the multivariate level-shift alternatives considered here, the MCUSUM outperforms both benchmark procedures in terms of out-of-control average run length. The advantage becomes more pronounced as dependence across age groups increases and as the change becomes more difficult to detect. These results support the use of the MCUSUM when the monitoring objective is to detect structured multivariate departures from the baseline, such as common level changes affecting several age groups simultaneously. We emphasize, however, that this does not make the MCUSUM universally optimal for every surveillance objective. In applications where the priority is to flag any deviation arising in a single age class as early as possible, age-specific charts or max-type procedures may be more appropriate. These local rules should therefore be viewed as complementary to the multivariate approach developed in this paper.

6 Conclusion

This paper presents a multivariate cumulative sum (MCUSUM) procedure to detect changes in mortality intensity, which plays a critical role in risk management for insurance companies and pension funds. Unlike univariate methods, the MCUSUM accounts for interdependence between age-groups and provides a more comprehensive analysis of mortality trends. We have studied the efficacy of the MCUSUM method thanks to a comparison to univariate control charts and a case study of recent mortality data in France,

Japan, Canada, and the USA, as well as an illustration with a real-world life insurance dataset. We have seen that MCUSUM is particularly relevant in the case of rectangularization effect or in presence of inter-age correlations or death postponing mechanisms.

The proposed framework is not restricted to the specific GP mean function used in this paper. More generally, it can be combined with any baseline mortality model that provides best-estimate forecasts together with a suitable description of forecast uncertainty across age groups. In particular, external actuarial assumptions or APC-type models can be incorporated by applying the monitoring procedure to the corresponding forecast errors, provided that their cross-age dependence structure is modeled appropriately. In this sense, the GP specification used here should be viewed as a flexible and convenient baseline rather than as a prerequisite for the methodology. At the same time, the choice of predictive mean and covariance model remains important. Richer GP kernels or hybrid specifications combining actuarial structure in the mean with GP modeling of residual dependence may offer a useful extension. For instance, Ludkovski and Risk (2023) explore compositional GP kernels for mortality modeling, allowing more flexible covariance structures and a closer representation of complex mortality dynamics.

The proposed methodology may be useful to enhance risk management in insurance companies and pension funds, not only in terms of modeling and monitoring longevity and mortality risks, but also in terms of design of index-based hedging solutions, paving the way towards potential basis risk reduction. Further research is of course needed to determine the full scope of potential applications of MCUSUM techniques in life insurance and pension industries. In addition, future work could investigate techniques—such as Andrews Curves (Andrews, 1972) or the Mason-Tracy-Young (MTY) decomposition (Tracy et al., 1992) to help identify the origin and driving factors behind the multivariate signals detected by the MCUSUM procedure.

Acknowledgements

This work is partially supported by the research initiatives « Sustainable actuarial science and climate risks » and « Artificial Intelligence for Actuarial Science and Insurance » sponsored by Milliman Paris, the BNP Paribas Cardif Chair “ACTIONS”, the ANR Research Project DREAMeS (ANR-21-CE46-0002) and the Joint Research Initiative on “Mortality Modeling and Surveillance” funded by AXA Research Fund. The views expressed in this document are the authors owns and do not reflect those endorsed by BNP Paribas Cardif. The authors thank the two reviewers for helpful and constructive comments that led to a much improved manuscript. We are grateful for discussions and comments from Dominique Abgrall, Claire Colinas and Marine Habart.

References

Abgrall, D., Habart, M., Rainer, C. and Sow, A. (2018), 'Exploring the longevity risk using statistical tools derived from the Shiryaev–Roberts procedure', *European Actuarial Journal* **8**(1), 27–51.

Andrews, D. F. (1972), 'Plots of high-dimensional data', *Biometrics* pp. 125–136.

Bersimis, S., Psarakis, S. and Panaretos, J. (2007), 'Multivariate statistical process control charts: An overview', *Quality and Reliability Engineering International* **23**(5), 517–543.

Brouhns, N., Denuit, M. and Van Keilegom, I. (2005), 'Bootstrapping the Poisson log-bilinear model for mortality forecasting', *Scandinavian Actuarial Journal* **2005**(3), 212–224.

Cairns, A. J., Blake, D. and Dowd, K. (2006), 'A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration', *Journal of Risk and Insurance* **73**(4), 687–718.

Cairns, A. J., Blake, D. and Dowd, K. (2008), 'Modelling and management of mortality risk: a review', *Scandinavian Actuarial Journal* **2008**(2-3), 79–113.

Caldarelli, E., Wenk, P., Bauer, S. and Krause, A. (2022), Adaptive Gaussian process change point detection, *in* 'International conference on machine learning', PMLR, pp. 2542–2571.

Carter, L. R., Prskawetz, A. et al. (2001), 'Examining structural shifts in mortality using the Lee–Carter method', *Methoden und Ziele* **39**.

Case, A. and Deaton, A. (2015), 'Rising morbidity and mortality in midlife among White non-Hispanic Americans in the 21st century', *Proceedings of the National Academy of Sciences* **112**(49), 15078–15083.

Chandola, V. and Vatsavai, R. R. (2011), A Gaussian process based online change detection algorithm for monitoring periodic time series, *in* 'Proceedings of the 2011 SIAM international conference on data mining', SIAM, pp. 95–106.

Committee, C. M. I. (2009), 'Continuous mortality investigation: a prototype mortality projections model: part two – detailed analysis', *Institute and Faculty of Actuaries* **Working paper 39**.

Deng, C., Chen, Z., Zhao, X., Wang, H., Wang, J., Gao, J. and Chen, H. (2025), Correlation-aware online change point detection, *in* 'Proceedings of the 34th ACM International Conference on Information and Knowledge Management', pp. 520–530.

Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R. and Vyncke, D. (2002*a*), 'The concept of comonotonicity in actuarial science and finance: applications', *Insurance: Mathematics and Economics* **31**(2), 133–161.

Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R. and Vyncke, D. (2002*b*), 'The concept of comonotonicity in actuarial science and finance: theory', *Insurance: Mathematics and Economics* **31**(1), 3–33.

Díaz-Rojo, G., Debón, A. and Mosquera, J. (2020), 'Multivariate control chart and Lee–Carter models to study mortality changes', *Mathematics* **8**(11), 2093.

Djeundje, V. B., Haberman, S., Bajekal, M. and Lu, J. (2022), 'The slowdown in mortality improvement rates 2011–2017: a multi-country analysis: Vb djeundje et al.', *European Actuarial Journal* **12**(2), 839–878.

European Commission (2009), 'Directive 2009/138/EC of the European Parliament and of the Council of 25 november 2009 on the taking-up and pursuit of the business of insurance and reinsurance (Solvency II)'.

Fries, J. F. (1980), 'Aging, natural death, and the compression of morbidity', *The New England journal of medicine* **303**(3), 130–5.

Gavrilova, N. and Gavrilov, L. (2015), 'Biodemography of old-age mortality in humans and rodents.', *The journals of gerontology. Series A, Biological sciences and medical sciences* **70 1**, 1–9.

Healy, J. D. (1987), 'A note on multivariate CUSUM procedures', *Technometrics* **29**(4), 409–412.

Hunt, A. and Villegas, A. M. (2022), 'Mortality improvement rates: Modeling, parameter uncertainty, and robustness', *North American Actuarial Journal* **0**(0), 1–27.
URL: *https://doi.org/10.1080/10920277.2021.2006068*

Huynh, N. and Ludkovski, M. (2021), 'Multi-output Gaussian processes for multi-population longevity modelling', *Annals of Actuarial Science* **15**(2), 318–345.

Jevtić, P., Luciano, E. and Vigna, E. (2013), 'Mortality surface by means of continuous time cohort models', *Insurance: Mathematics and Economics* **53**(1), 122–133.

Jiang, W., Han, S. W., Tsui, K.-L. and Woodall, W. H. (2011), 'Spatiotemporal surveillance methods in the presence of spatial correlation', *Statistics in Medicine* **30**(5), 569–583.

Joner Jr, M. D., Woodall, W. H., Reynolds Jr, M. R. and Fricker Jr, R. D. (2008), 'A one-sided MEWMA chart for health surveillance', *Quality and Reliability Engineering International* **24**(5), 503–518.

Lee, R. D. and Carter, L. R. (1992), 'Modeling and forecasting US mortality', *Journal of the American statistical association* **87**(419), 659–671.

Leon, D. A., Jdanov, D. A. and Shkolnikov, V. M. (2019), 'Trends in life expectancy and age-specific mortality in england and wales, 1970–2016, in comparison with a set of 22 high-income countries: an analysis of vital statistics data', *The Lancet Public Health* **4**(11), e575–e582.

Li, Z., Zou, C., Gong, Z. and Wang, Z. (2014), 'The computation of average run length and average time to signal: an overview', *Journal of Statistical Computation and Simulation* **84**(8), 1779–1802.

Ludkovski, M. and Risk, J. (2023), 'Expressive mortality models through Gaussian process kernels', *arXiv preprint arXiv:2305.01728* .

Ludkovski, M., Risk, J. and Zail, H. (2018), 'Gaussian process models for mortality rates and improvement factors', *ASTIN Bulletin: The Journal of the IAA* **48**(3), 1307–1347.

Mitchell, D., Brockett, P., Mendoza-Arriaga, R. and Muthuraman, K. (2013), 'Modeling and forecasting mortality rates', *Insurance: Mathematics and economics* **52**(2), 275–285.

Montgomery, D. C. (2007), *Introduction to Statistical Quality Control*, 6 edn, John Wiley & Sons.

of the Society of Actuaries (RPEC), R. P. E. C. (2020), 'Mortality improvement scale MP-2020 report', **Technical report**.
URL: *https://www.soa.org/resources/experience-studies/2020/mortality-improvement-scale-mp-2020/*

Olshansky, S. J., Passaro, D. J., Hershow, R. C., Layden, J., Carnes, B. A., Brody, J., Hayflick, L., Butler, R. N., Allison, D. B. and Ludwig, D. S. (2005), 'A potential decline in life expectancy in the United States in the 21st century', *New England Journal of Medicine* **352**(11), 1138–1145.

Pascual, F. G. and Akhundjanov, S. B. (2020), 'Copula-based control charts for monitoring multivariate Poisson processes with application to hepatitis c counts', *Journal of Quality Technology* **52**(2), 128–144.

Qiu, P. (2013), *Introduction to Statistical Process Control*, Chapman and Hall/CRC.

Raleigh, V. S. (2019), Trends in life expectancy in eu and other oecd countries: Why are improvements slowing?, Technical Report 108, OECD Publishing, Paris.

Rasmussen, C. E. and Williams, C. K. I. (2005), *Gaussian Processes for Machine Learning*, The MIT Press.
URL: *https://doi.org/10.7551/mitpress/3206.001.0001*

Richards, S. J. (2024), 'Robust mortality forecasting in the presence of outliers', *British Actuarial Journal* **29**, e19.

Rogerson, P. A. and Yamada, I. (2004), 'Monitoring change in spatial patterns of disease: comparing univariate and multivariate cumulative sum approaches', *Statistics in Medicine* **23**(14), 2195–2214.

Roustant, O., Ginsbourger, D. and Deville, Y. (2012), 'Dicekriging, diceoptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization', *Journal of statistical software* **51**, 1–55.

Saatçi, Y., Turner, R. D. and Rasmussen, C. E. (2010), Gaussian process change point models, *in* 'Proceedings of the 27th International Conference on Machine Learning (ICML-10)', pp. 927–934.

Schafer, R. (2011), 'What is a Savitzky-Golay filter? [lecture notes]', *IEEE Signal Processing Magazine* **28**, 111–117.

TGH05 (2006), 'TGH05', https://www.moneyvox.fr/r/rente_viagere_2007.pdf.

Tracy, N. D., Young, J. C. and Mason, R. L. (1992), 'Multivariate control charts for individual observations', *Journal of quality technology* **24**(2), 88–95.

Van Berkum, F., Antonio, K. and Vellekoop, M. (2016), 'The impact of multiple structural changes on mortality predictions', *Scandinavian Actuarial Journal* **2016**(7), 581–603.

Villegas, A. M., Haberman, S., Kaishev, V. K. and Millossovich, P. (2017), 'A comparative study of two-population models for the assessment of basis risk in longevity hedges', *ASTIN Bulletin: The Journal of the IAA* **47**(3), 631–679.

Xu, Y., Sherris, M. and Ziveyi, J. (2020), 'Continuous-time multi-cohort mortality modelling with affine processes', *Scandinavian Actuarial Journal* **2020**(6), 526–552.

Zhao, H. and Pan, R. (2025), 'Gaussian derivative change-point detection for early warnings of industrial system failures', *Reliability Engineering & System Safety* **256**, 110681.

A Robustness to the choice of calibration and monitoring periods

To assess the sensitivity of the empirical results to the choice of calibration and monitoring windows, we repeated the real-data analysis using an earlier calibration period and a longer monitoring window. More precisely, we calibrated the Gaussian process model on the years 1981–2000 and then applied the monitoring procedures over the period 2001–2020. Figures 13 and 14 report the corresponding results for level-change and trend-change detection, respectively.

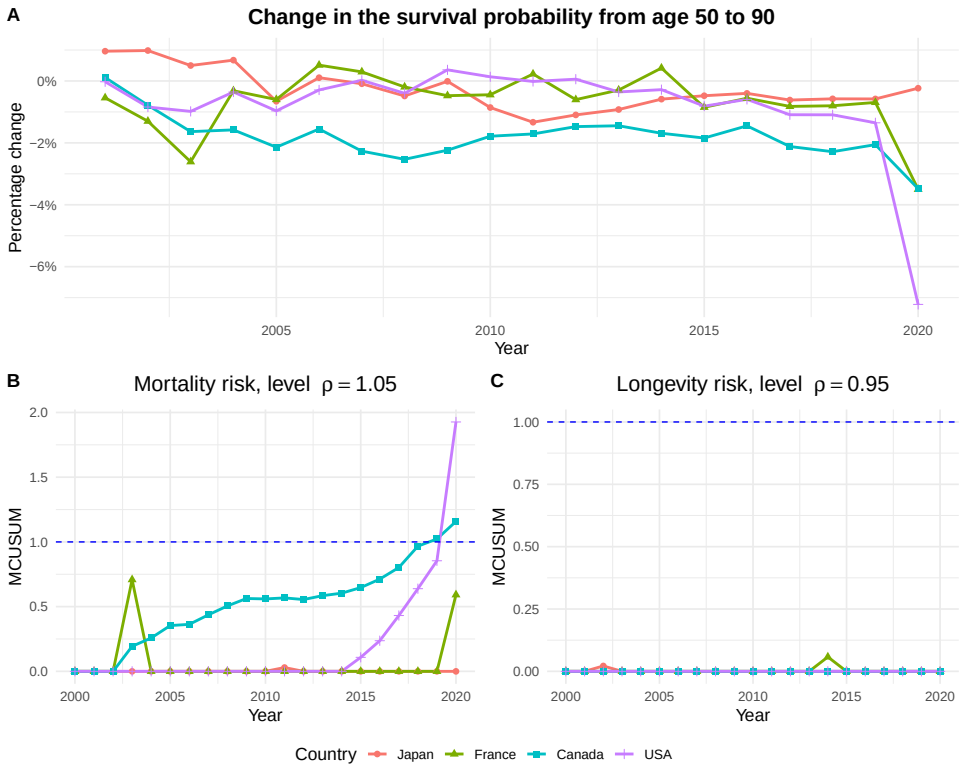


Figure 13: **A:** Percentage change difference of the 40-year survival probability of an individual aged 50 between the observed survival probability and the GP-predicted survival probability. **B:** MCUSUM for the detection of a change at the level $\rho = 1.05$ (mortality risk). **C:** MCUSUM for the detection of a change at the level $\rho = 0.95$ (longevity risk). Dotted lines represent the detection threshold.

For level-change detection, Figure 13 reveals some sensitivity to the choice of the calibration period. In contrast to the main analysis, Japan no longer crosses the longevity-risk threshold associated with $\rho = 0.95$, indicating that the favorable mortality deviations are not sufficiently large and persistent relative to the reference trajectory estimated over 1981–2000. The United States still crosses the mortality-risk threshold associated with $\rho = 1.05$, but the detection occurs in 2019 rather than in 2014. These differences arise because changing the calibration period modifies the estimated reference mortality

trajectory, its predictive uncertainty, and the sequence of deviations accumulated by the MCUSUM.

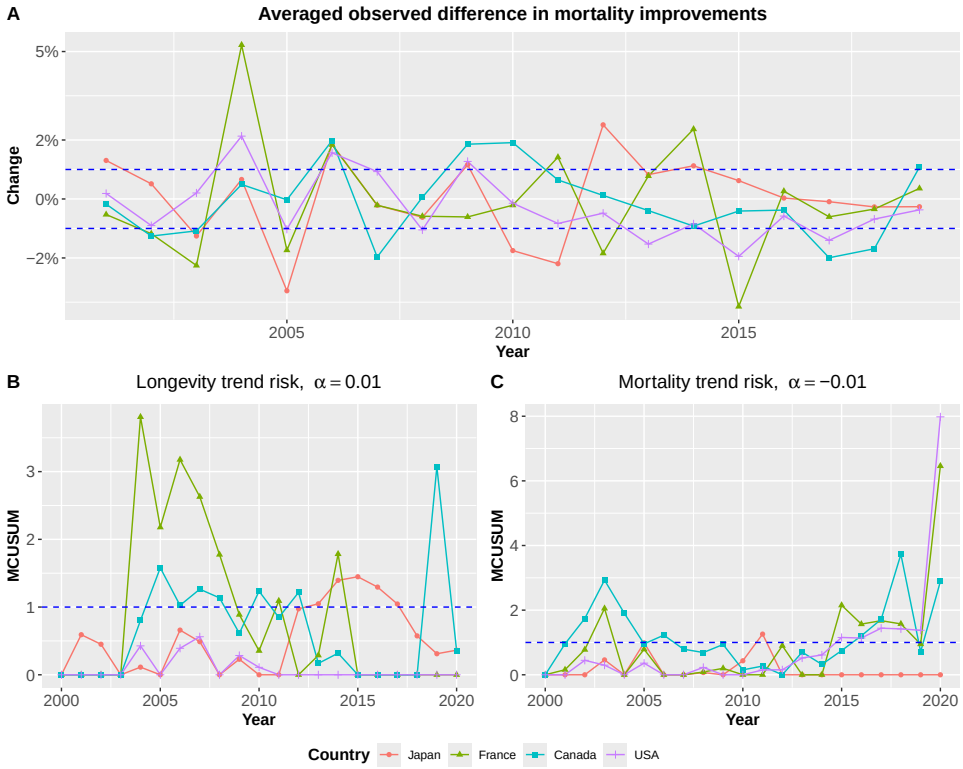


Figure 14: **A:** Difference averaged by age tranches between the observed improvement rates and the GP-predicted improvement rates. Dotted lines represented the levels 1% and −1%. **B:** MCUSUM for the detection of a change in the trend at the level $\alpha = 1\%$ (longevity trend risk). **C:** MCUSUM for the detection of a change in the trend at the level $\alpha = -1\%$ (mortality trend risk). Dotted lines in **B** and **C** represent the detection threshold ($L = 1$).

For trend-change detection, Figure 14 shows that France and Canada cross the longevity trend-risk threshold during the first decade of the monitoring period, whereas these signals are not detected under the main calibration. Japan continues to exhibit a longevity trend-risk signal, although the threshold is crossed later, in 2013. Thus, the Japanese trend-change result remains qualitatively robust, while its detection date is sensitive to the calibration period. The additional signals for France and Canada reflect deviations from the trend estimated over 1981–2000 and illustrate that the trend-change MCUSUM is also sensitive to the reference trend and to the longer monitoring sequence.