

Assess the presence of disruptions in cause-specific mortality series

Carlo G. Camarda* Markéta Pechholdová†

Extended abstract submitted to the
European Population Conference 2014, Budapest, Hungary
November 2013

This paper was written within the following projects:

- “*AXA project on Mortality Divergence and Causes of Death*”
- “*Project ANR-12-FRAL-0003-01 DIMOCHA*”

Abstract

Time series of mortality by cause of deaths may present, in the year of ICD revision, disruptions which are undue to variability in the mortality trend. Assessing the presence of these eventual disruptions is of great help for drawing conclusions on redistribution of death counts among other causes of death and, consequently, for the estimation of continuous mortality series. Our approach aims to detect statistically significant discontinuity in an ICD-revision year by the mean of (back-)forecasting the estimated trend in the revision year. The estimation of the trends is done via smoothing techniques on the standardized deaths within a Poisson framework for accounting the variability of the phenomenon. A simulation study is given to demonstrate the performance of the method. Actual applications on West German cause-specific mortality data illustrate the outcome of the approach.

KEYWORDS: International Classification of Diseases; Mortality; Cause of death; Disruptions; Detection, Poisson, Smoothing, Forecast.

*Corresponding author: Institut National d'Études Démographiques. 133, Bd Davout, 75980 Paris Cédex 20, France; Tel. +33(0)1 5606 2155, email: carlo-giovanni.camarda@ined.fr

†Department of Demography, University of Economics, Prague; email: Marketa.Pechholdova@seznam.cz

1 Introduction

Periodic revisions of the international classification of diseases (ICD) are a known issue in long-term cause of death analyses. The only efficient method to deal with the ruptures developed so far employs a systematic comparison of medical and statistical content across ICD revisions and is known to perform the better the greater is the detail of the ICD grouping (Meslé and Vallin, 1996; Pechholdová, 2009). An integral part of the method consists of a systematic inspection of continuity performed at several steps of the method. These checks are the only way to detect ruptures which cannot be anticipated from any available information source (such as the ICD manuals themselves, bridge coding studies or WHO ICD10 to ICD9 mapping tool).

However, a great level of detail represents a great deal of time series to be checked, especially when the series are broken down by age and sex. Moreover, the issue of small death counts appears in some combinations of cause, age, and sex. Among several methods tested and used, visual inspection has yielded the best results. Given the subjectiveness and the time load represented by the visual inspection, we are proposing a statistically-oriented alternative based on prediction of the series and a cutoff level of decision whether the actual and the reconstructed data form an acceptably coherent series. We believe this method can replace, at least in a large part, the visual inspection, and facilitate the process of bridging the two ICD revisions at its best practice.

2 Method

We formulate the method for a specific cause of death and sex. The presence of eventual disruption is assessed for the whole age range of the population.

Let $\mathbf{D} = (d_{ij})$ be the matrix of deaths at age $i, i = 1, \dots, m$ and year $j, j = 1, \dots, n$ for a specific cause of death. Likewise the matrix $\mathbf{E} = (e_{ij})$ contains the exposures over the same dimensions for the whole population. The rows of \mathbf{D} and \mathbf{E} are labeled by age(-group) of death, $\mathbf{a}' = (a_1, \dots, a_m)$, and the columns by year of death, $\mathbf{y}' = (y_1, \dots, y_n)$.

An additional $m \times n$ matrix of cause-specific death rates can easily be computed: $\mathbf{M} = \left(m_{ij} = \frac{d_{ij}}{e_{ij}}\right)$.

2.1 Standardizing data

We aim to gauge the presence of disruption devoid of changes in age structure of the population. We thus decide to compute the standardized death rates, i.e. death rates of a population adjusted to a standard age structure, $\mathbf{p} = (p_i) : \sum_i p_i = 1$.

In formula the standardized death rates are given by:

$$\mathbf{M}^S = \text{diag}(\mathbf{p}) \cdot \mathbf{M}$$

where $\text{diag}(\mathbf{p})$ is the $m \times m$ diagonal matrix with p_i as entries. In the following we will use a “standard European population” defined by the World Health Organization (WHO). Other options are equally suitable.

In order to obtain the age-standardized crude death rate and the corresponding total expected deaths over years, we can now sum up over ages \mathbf{M}^S and multiply the outcomes by a given factor κ :

$$\begin{aligned} \mathbf{CDR}^S &= \mathbf{1}_m \cdot \mathbf{M}^S \\ \mathbf{D}^S &= \kappa \mathbf{CDR}^S \end{aligned}$$

with $\mathbf{1}_m$ a vector of 1s of length m .

In demography $\kappa = 10^6$ is commonly used when dealing with cause of death analyses. This option leads to a straightforward interpretation of the standardized crude death rates as number of deaths over a million of individuals. Instead of following this convention, we decided to account for the size of the population in hands: κ will be equal to the average of the total population size from \mathbf{E} . In this manner the number of standardized expected deaths \mathbf{D}^S can be interpreted as the total number of deaths from a specific population and cause of death, only devoid of the changes in the age structure. Along these lines, we will model the trend of \mathbf{D}^S , and its eventual disruption will thus be assessed accounting for stochasticity of the actual phenomenon. The same distance between neighboring values of \mathbf{D}^S for two different populations will be considered differently if the sizes of these populations are different.

As example, figure 1 presents a selected cause of death from the West German series which will be used in following demonstrations: acute pulmonary heart disease (ICD9 code 415). In West Germany, the ICD revision between the 8th and the 9th classification took place between 1978 and 1979. The developments of actual death rates (left panel) does not show any particular disruption in the revision year. On the right panel, the trend of the expected standardized deaths for all ages reveals a clear jump in 1979. Our aim is to assess whether this increment is statistically significant. Once the decision is adopted, different methodologies could be employed for redistributing death counts from acute pulmonary heart disease to related causes, in order to create coherent mortality time series.

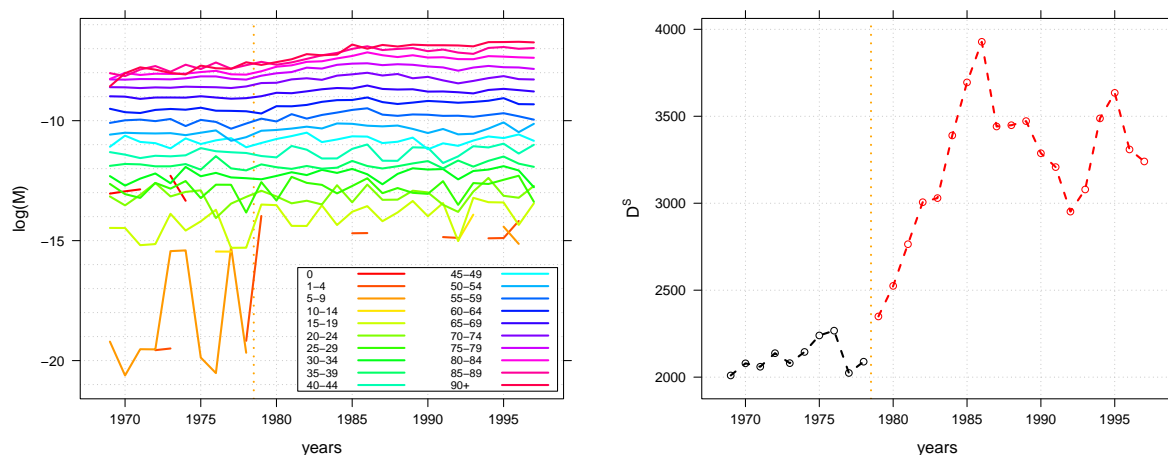


Figure 1: Actual death rates in logarithmic scale (left panel) and standardized deaths (right panel). West Germany. 1969-1997, 20 age-groups, both ages combined, standard European population, $\kappa = 62531434$. ICD-revision in 1979.

2.2 Smoothing and forecasting data

Assessment of discontinuity needs to be done essentially at the change of the classification. Therefore we will separately consider the trends before and after the revision year. Let D_1^S and D_2^S the actual standardized deaths for these two periods: $j = 1, \dots, n_1$ and $j = n_1 + 1, \dots, n$. Instead of comparing directly the standardized deaths, we assume them to be realization of a Poisson distribution with a smooth force of mortality multiplied by the factor κ repeated for the number of associated years:

$$D_1^S \sim \mathcal{P}(\mu_1 \kappa_{n_1}) \quad \text{and} \quad D_2^S \sim \mathcal{P}(\mu_2 \kappa_{n_2}).$$

In this way we have a statistical framework for modeling D_1^S and D_2^S accounting for the size of the population in a suitable manner. Without going into details, a P -splines model is applied on both periods estimating the smooth mortality trends $\hat{\mu}_1$ and $\hat{\mu}_2$ (Camarda, 2012; Eilers and Marx, 1996). In particular, this technique can be easily embedded in a Poisson setting, allowing us to smooth our data accounting for the population size via κ , as offset. The more dealing with Poisson counts, the smoother operates on the logarithmic of the force of mortality: $\ln(\mu) = \eta$.

The value of κ is crucial for determining the amount of smoothness which is optimized here by the Bayesian Information Criterion. In a nutshell, the larger the value of κ , the larger are the expected standardized deaths. Since we work in a Poisson setting, the larger the value of counts, the more any statistical model will be prone to trust actual pattern because it will detect small variability compared to the trends. Therefore a large κ (i.e. a large population) will lead to a fitted curves that will follow the standardized deaths more

closely, i.e. given the same trend, a larger κ leads to a rougher fitted curve. Following these arguments, incorporating the actual population size into the computation of \mathbf{D}^S becomes attractive from a modeling point of view as well as from its interpretation.

Once the trends are estimated, we need to evaluate the smooth curves at the revision year. Instead of taking $\hat{\eta}_1$ and $\hat{\eta}_2$ for the last and first year of their period (y_{n_1} and y_{n_1+1} respectively), we (back-)forecast both fitted curves in the mid-year $y_r = [y_{n_1} + 0.5]$. In P -splines models, forecasting is a natural consequence of the smoothing process, and extrapolation for half year does not pose any issue in terms of reliability of the estimates (Currie et al., 2004).

Given the estimates, we can compute the expected disruption in the revision year:

$$\Delta = \hat{\eta}_1[y_r] - \hat{\eta}_2[y_r].$$

Our aim is now to decide whether Δ is statistically different from 0, i.e. we have a clear disruption in the revision year.

2.3 Testing disruption

An additional advantage of the P -splines methodology lays in the capability to easily compute diagnostics such as the standard errors of our estimates. In the conjunction time-point y_r , we can compute the standard errors $SE_{\hat{\eta}_1[y_r]}$ and $SE_{\hat{\eta}_2[y_r]}$.

This convenience is essential for testing the value of Δ . Specifically, for a given confidence level $1 - \alpha$, we can calculate the upper and lower confidence limits for the estimated difference as follows:

$$\begin{aligned} L_{\Delta} &= (\hat{\eta}_1[y_r] - c_{\alpha}^1 \cdot SE_{\hat{\eta}_1[y_r]}) - (\hat{\eta}_2[y_r] + c_{\alpha}^2 \cdot SE_{\hat{\eta}_2[y_r]}) \\ U_{\Delta} &= (\hat{\eta}_1[y_r] + c_{\alpha}^1 \cdot SE_{\hat{\eta}_1[y_r]}) - (\hat{\eta}_2[y_r] - c_{\alpha}^2 \cdot SE_{\hat{\eta}_2[y_r]}) , \end{aligned}$$

where c_{α}^1 and c_{α}^2 denote the $(1 - \frac{\alpha}{2})$ percentile of two Student's t -distributions with n_1 and n_2 degrees of freedom, respectively. We opted for the t -distributions given the relatively few data-points (i.e. years) in the large majority of the cause-specific mortality time series.

Once L_{Δ} and U_{Δ} are computed, we can just check whether this confidence interval contains the value zero. If so, we can reject the hypothesis of disruption and assume the two trends as part of a common pattern. Alternatively, if zero is not between L_{Δ} and U_{Δ} , we can deduce the presence of a disruption (given a confidence level) and consequently search for possible redistribution of death counts among causes of death.

The proposed approach has similarities with the Regression Discontinuity Design which is a quasi-experimental pretest-posttest design that captures the causal effects of interventions by assigning a threshold above or below which an intervention is assigned.

For a fuller treatment of the method, we refer to Imbens and Lemieux (2010).

3 Simulation

Before we apply the model to a demographic dataset, we want to demonstrate the performance of the method in a simulation study.

We simulated 10000 instances for 12 different scenarios in which we varied amount of smoothness of the true trend, size of the population (i.e. size of the death counts), and amplitude of the true disruption. Specifically we have two different sample sizes ($N=100$ and $N=500$ which represent the minimum death counts in the dataset and leads to different final sizes), two different curves (two exponentials with either minor or strong sine signal) and three different scenarios concerning the presence of disruption: no disruption, a disruption equal to two and five times the variance in the data.

The idea is to check when and how often the proposed approach fails in finding disruption when in reality there are and vice versa. Outcomes are given in Figure 2. The solid lines represent the true values of the standardized deaths, the dots an instance among the 10000 datasets we simulated and colors depict the different classification periods.

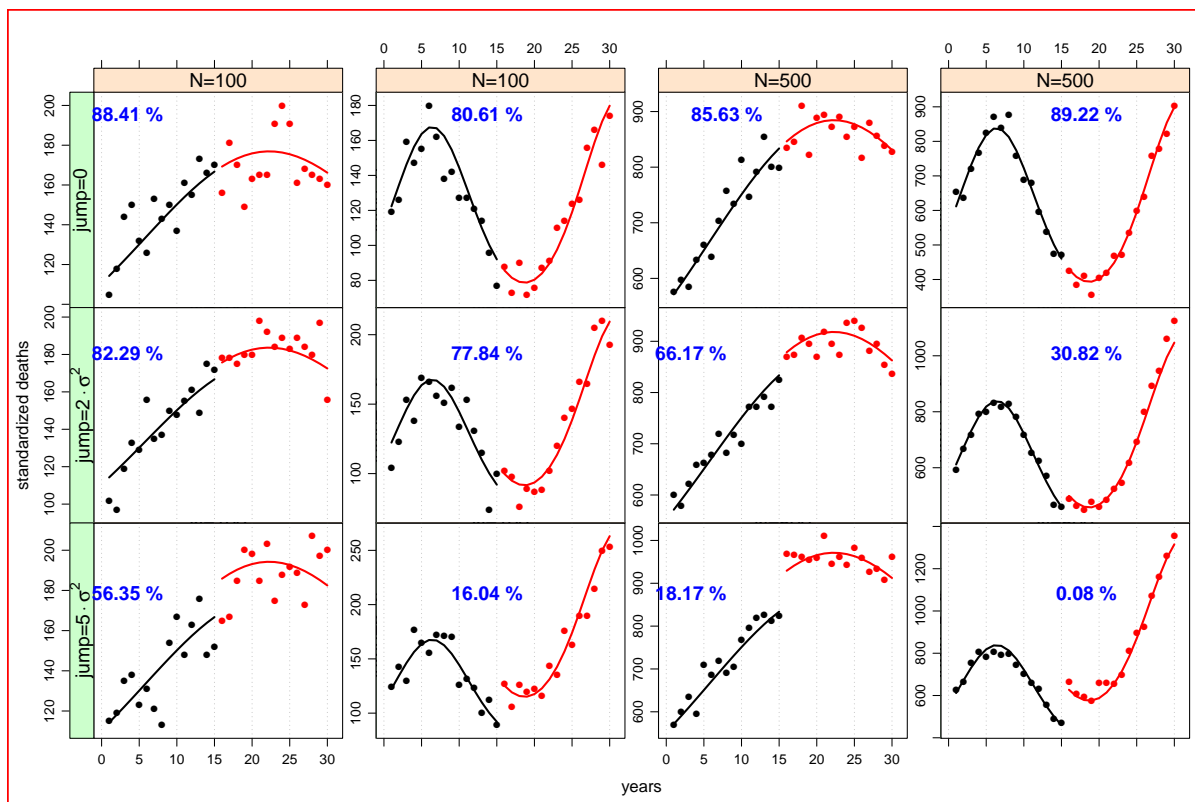


Figure 2: The effect of various sample size, smoothness and disruption level on the results of detection

The percentages printed in each panel of Figure 2 represent the number of times

(among 10000 times) the approach has not found a disruption with 80% confidence level. The confidence level is selected based essentially on experience, i.e. compliance with the results of visual inspection.

A perfect model would have 100% in the first rows and 0% in the others. Of course this is not the case (and it won't be the case with any method). Nevertheless several messages can be read:

- an increasing number of counts leads to better outcomes: see decreasing percentage of “disruption not found” from the first (second) to the third (forth) column in the scenarios in which we assumed the presence of disruptions (second and third row);
- a smoother latent trend tends to hide eventual disruptions, especially with small sample size: see the left-bottom panel where with a relatively big true-disruption, the approach overlooks about 56% of them;
- a larger disruption is better found by the proposed method: for all columns we have a decreasing percentage of “disruption not found”;
- if we have got a large dataset and clear disruption, the approach will detect 99.92% of them within an unsmooth trend and 81.83% of them when the trend is smoother;
- the effect of multiply by 5 the sample size is similar to having a disruption within an unsmoother trend: see the similar percentages in the last two rows between the second and third column.

4 Actual Application

Time series of West German cause-specific mortality are analyzed in this section. Figure 3 presents actual and fitted values for acute pulmonary heart disease. Moreover we add the estimated points at the conjunction time-point between two classifications (1978.5) along with the associated 80% confidence intervals.

The estimated value for Δ is equal -0.1594. This is the expected difference in 1978.5 between the estimated log standardized crude death rates from the fitted curves. The associated confidence interval is $[-0.2262; -0.0926]$. This interval does not include zero, hence we could reject the hypothesis that the two trends belong to a common pattern, i.e. we have a statistically significant disruption in the trend of mortality from acute pulmonary heart disease at the moment of the change between the 8th and 9th ICD revisions.

Furthermore routines for building each part of the model were implemented in R (R Development Core Team, 2013) and, together with user-friendly examples, they are available from the first author upon request. In our implemented R functions several arguments

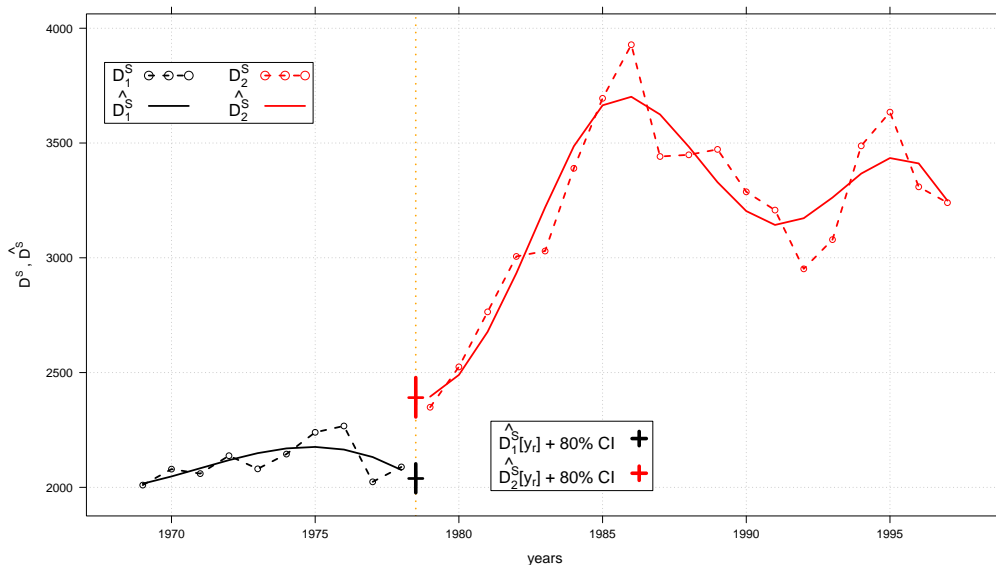


Figure 3: Actual and fitted standardized deaths by P -splines. Extrapolated trends in 1978.5 along with 80% confidence intervals. West Germany. 1969-1997, 20 age-groups, both ages combined, standard European population, $\kappa = 62531434$. ICD-revision in 1979.

could be called and many outcomes could be extracted (see appendix). Particularly useful for facilitating the process of bridging the two ICD revision is the automatic plotting of the actual and fitted data along with a flagged legend which immediately informs about the presence of disruption.

Any researcher could easily employ our R-function with his/her datasets and quickly check which causes of death show significant ruptures and thereupon adjust them with suitable reassignment of deaths to other causes. Figure 4 presents four different outcomes from the same dataset as given by the mentioned R-function. We modified the confidence level (80% and 95%) as well as the multiplicative factor κ (the proposed one with the average population and the conventional value of 10^6). Whereas remarkable differences are due to varying the level of κ , changing the confidence level does not lead to different decision with respect to the presence of disruption, at least in this dataset: see the boxes on the top-right corner of the panels which have automatically a red background in case of statistically significant disruption.

5 Concluding remarks

The proposed method of rupture detection in mortality time series based on P -splines smoothing and taking into account the actual size of population has yielded satisfactory results on the sample of tested causes of death. As next step, the method will be applied on a large dataset of time series of causes of death in France at the most detailed ICD

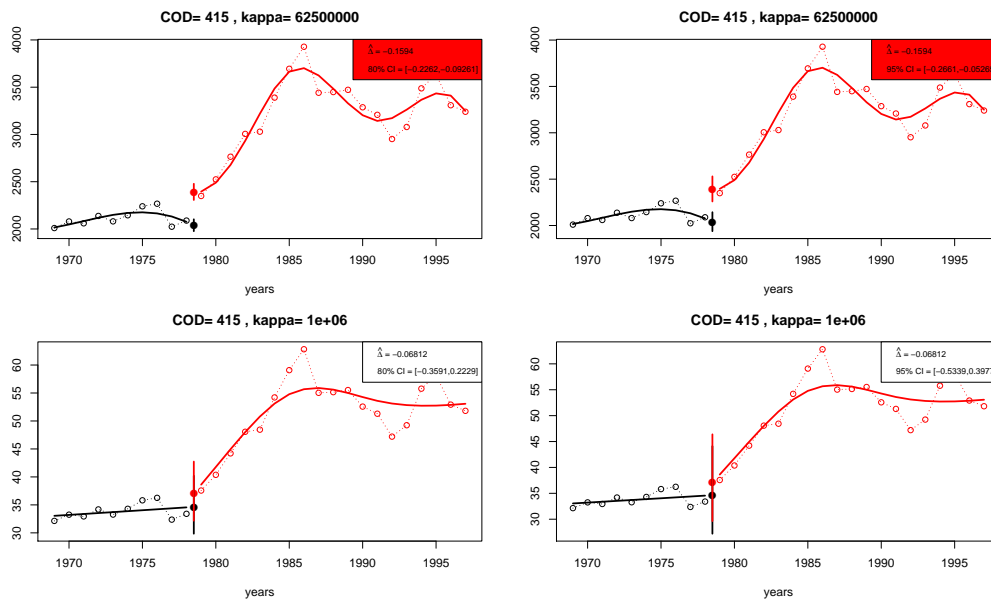


Figure 4: Actual and fitted standardized deaths by P -splines. Extrapolated trends in 1978.5 along with 80% and 95% confidence intervals. Two different κ employed. West Germany. 1969-1997, 20 age-groups, both ages combined, standard European population. ICD-revision in 1979.

level. The method has ambitions to become an important part of automatized procedures designed to facilitate the meticulous task of bridging two successive ICD revisions.

References

- Camarda, C. G. (2012). MortalitySmooth: An R Package for Smoothing Poisson Counts with P -Splines. *Journal of Statistical Software* 50, 1–24. Available on www.jstatsoft.org/v50/i01.
- Currie, I. D., M. Durbán, and P. H. C. Eilers (2004). Smoothing and Forecasting Mortality Rates. *Statistical Modelling* 4, 279–298.
- Eilers, P. H. C. and B. D. Marx (1996). Flexible Smoothing with B -splines and Penalties (with discussion). *Statistical Science* 11, 89–102.
- Imbens, G. and T. Lemieux (2010). Regression Discontinuity Designs: A Guide to Practice. *Journal of Economic Literature* 48, 281–355.
- Meslé, F. and J. Vallin (1996). Reconstructing long-term series of causes of death: The case of France. *Historical Methods* 29, 72–87.
- Pechholdová, M. (2009). Results and observations from the reconstruction of continuous time series of mortality by cause of death: Case of West Germany, 1968-1997. *Demographic Research* 21, 535–568.
- R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Appendix: Software considerations

In this appendix we briefly explain how to use the proposed method by an R-function specially designed to carry out the analysis and available upon request.

We denote with D and E the input matrices of deaths and exposures over age-groups and years. The argument p is also needed for providing the vector of standard population. Finally the last required input concerns the length of the first ICD revision ($n1$).

Given these 4 arguments, the function is able to reproduce the proposed methodology and return several important objects. The following code-snippet present the mentioned function and all its arguments:

```
CheckDisruption <- function(D, E, p, n1,
                             y, cod, kappa, level=80,
                             plot=TRUE, Type=c("SD", "logSDR")){
  [...]
}
```

The user can also modify the other arguments:

- y : a vector with the actual years
- cod : a string with the name of the cause of death
- $kappa$: the multiplicative factor. By default the average population from E
- $level$: the confidence level. By default 80%
- $plot$: whether a plot should be produced. By default the function will give you a plot
- $Type$: whether the user prefer a plot with standardized deaths or standardized crude death rates in a log scale

The arguments y , cod , $plot$ and $Type$ do not interfere with the final results. They are only required to enhance the final plot according to the user's wishes.

The function returns several objects such as the fitted standardized deaths, its standard errors, the estimated Δ along with its confidence interval, etc.