EUROPEAN DOCTORAL SCHOOL OF DEMOGRAPHY

CENTER FOR DEMOGRAPHIC STUDIES OF
AUTONOMOUS UNIVERSITY OF BARCELONA, AND
MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH

PROJECT

Forecasting Russian cancer mortality with compositional data analysis

Author
Rustam TURSUN-ZADE

EDSD Supervisor
Timothy RIFFE

CED Supervisor
Albert ESTEVE PALÓS

July 2020

Table of contents:

Introduction:

In this paper we adopt the forecasting method developed by (Kjærgaard et al. 2019), to obtain a coherent forecast of cancer mortality decomposed by groups of cancer localizations for the Russian Federation in compositional framework. Our approach is similar to that used by (Kjærgaard et al. 2019), but we focus on cancer decomposed by groups of localizations, rather than as a single category. We are aiming at making a first attempt to explore the possibility to obtain a valid compositional cancer mortality forecast for Russia that will be based on the full available time-series of mortality. The fluctuations observed in Russian mortality time-series are characterized by different magnitudes when decomposed by sex (Figures 1 and 2). Thus, we are also interested in evaluating if compositional models will capture the erratic mortality patterns in both males and females and provide a better forecast than standard Lee-Carter (LC) model when trained on the longest possible perspective with no additional data smoothing. This will offer a first approximation to the possible limitations of compositional approach imposed by data quality and dynamics. We will start by descriptive analytics of Russian cancer mortality trends, and those obtained by compositional transformations. This will be followed by approximation and interpretation of compositional model parameters and goodness of fit tests and we will end with the twenty year cause-specific compositional cancer mortality forecast.

Background:

An estimated 18.1 million of new cancer cases and 9.6 million cancer deaths were expected to occur worldwide in 2018, with both mortality and incidence numbers growing (Bray et al. 2018). One in five men and one in eight women develop a cancer and one in eight men and one in ten women die from it (Bray et al. 2018). Cancer directly affects the well-being of diseased individuals and their families, and it is one of the costliest diseases to treat (Meropol et al. 2007). The average expenditure made per cancer patient in the United States was almost 4 times higher than per patient with no cancer (Park and Look 2019). A recent study conducted in United Kingdom estimated cancer (specifically pancreatic cancer) as the most costly disease for the National Health Service with 3074 pounds of expenditure per patient (Briggs, Scarborough, Wolstenholme 2018). Evidence suggests that individuals from lower socioeconomic strata also have higher risks of psychological well-being reduction associated with cancer (Simon et al. 2008). The ongoing covid-19 crisis has also increased cancer mortality. The pandemics might cause up to 20% increase in cancer mortality in patients with newly diagnosed cancer due to direct effects of covid-19, late diagnostics or untimely treatment (Wise, 2020). Although, the

major trend breaks and onsets of epidemics are virtually impossible to forecast, it is unlikely that in the near future infections will take precedence over degenerative diseases on an annual horizon. Hence, accurate and coherent forecasts of mortality by cause of death are not only a valuable instrument facilitating efficient public health planning and healthcare actions, but may also help at addressing socioeconomic inequality brought on by cancer. An accurate mortality forecast allows for use of mortality-morbidity correspondence to model disease incidence - a method that is being used by World Health Organization (Girosi, King, 2008).

Several creative mortality forecasting approaches have been developed including proportional rate of change models, Age-Period-Cohort models, regression models, and models applied in a Bayesian framework (Booth and Tickle 2008). In spite of a well-developed tradition and ample toolbox, up until recently the problem of adequate and coherent cause-specific mortality forecasting was not appropriately addressed. The most commonly used method for mortality forecasting is the Lee-Carter model (1992). This approach extrapolates future age-specific mortality rates assuming a constant rate of age-specific mortality improvement over time. This exact feature of the corresponding model results in inconsistent forecasts when considering causes of death. The Lee-Carter model was shown to give more pessimistic results when forecasting mortality by causes of death (Wilmoth 1995; Booth, Tickle 2008). When causes of death are subject to independent forecasts, which in itself is an unlikely simplification as cause-specific survival also depends on overall mortality, the aggregate mortality forecast tends to be dominated by the mortality from causes with smaller rate of mortality improvement in the chosen fitting period (Wilmoth 1995). Another drawback is the assumption of a constant rate of mortality improvement by age. Accounting for interdependence (competing risks) between causes of death imposes an additional complication, since the correspondence between the forecast cause-specific mortality rates and the cumulative density of deaths is not as straightforward as in all-cause-combined mortality model, even if causes are assumed independent (Kjærgaard et al. 2019). Coherency of the corresponding prognosis also tends to be violated since sum of the forecast cumulative densities for individual causes of death should ideally be equal to the corresponding sum of the total forecast density of deaths, a property known to be difficult to obtain in practice (Booth and Tickle 2008). Finally, the best practice model should take into account the covariance between causes, ages and any other chosen factor. These shortcomings motivated the development of new approaches designed specifically for joint forecasts of cause of death series.

Oeppen (2008) proposed an expression of the LC model framed as a Compositional Data Analysis (CoDA) problem (Aitchison 1986). This setup maintains the compositional form of the

data, forecasting the death densities of the multiple-decrement life-table rather than log-mortality hazards, and circumvents the fundamental problem of independent cause-specific forecasts pointed out by others. Life-table death densities sum to a constant - life table radix. CoDA models circumvent a well-known problem of identification of dependence between causes of death, by capturing and forecasting transitions between causes (Kjærgaard et al. 2019). The real power of CoDA models is that they can potentially consider any arbitrary number of dimensions treating them as sub-compositions guaranteeing the coherency of the forecast and sum constraints. Using CoDA techniques it is possible to forecast future cause-specific death densities, their relative structure (relative importance of causes of death), redistributions occurring between ages and causes and general time trends all following a single unified framework. Oeppen's proposal was thus further updated and used to obtain coherent regional mortality forecasts (Bergeron-Boucher et al. 2017, 2018), and for forecasting mortality by causes of death with a specific focus on cancer (Kjærgaard et al. 2019).

Forecasting Russian mortality is in itself a complex problem, since mortality patterns (especially in males) are characterized by large fluctuations during the early 1990-2000`s (see Figure 1). This is the period on which a forecast must be based, but the LC model is premised on capturing a linear trend, and it performs erratically in periods of intense fluctuation. Hence the direct application of the standard LC model is ill-advised. One option is to extend the LC model to include additional information for better capturing of period effects (Oeppen 2008). When forecasting mortality with LC, it is common to use singular value decomposition (SVD), to simplify the calculations and reduce the information in data. Default procedure is to consider only the first singular vectors, returned by SVD since they dominate the observed dynamics in the data. Inclusion of the second and third singular vectors might allow for better capturing of period effects such as were experienced by Russia during transition to market economy, a property that is also applicable for compositional mortality forecasting models (Oeppen, 2008). Alternatively, the data could be preliminarily smoothed, to exclude the period fluctuations, or the forecast period should be shortened.

Cancer was chosen as it forms the second leading cause of death in Russia (Figure 2) and by its nature (Baili et al. 2015; Doll 1998; IARC 2004) is generally characterized by more uniform dynamics (Figures 2 and 3) than those seen in other main groups of causes of death (Figure 2) which holds true for the Russian mortality pattern in 1990's and early 2000's (Leon et al. 1997, Shkolnikov 1999). There is also evidence that Russian cancer mortality data has the high quality of cause of death registration relative to other main groups of causes of death (Danilova et al. 2016).

Data and methods:

Data:

This project implies use of national mortality data decomposed by causes of death, age and sex alongside with the national data on population at risk. Russian Fertility and Mortality Database commonly referred to as RosBRiS is used as the primary data source (Russian Fertility and Mortality Database, 2018). RosBRiS offers quantitative, time-series, national mortality and population data made freely available for public use. These data are covering the period of 1965-2018 and exactly correspond to the official state statistics. The longest possible time-series 1965-2018 is considered. We fit the forecast using the original data by 5-year age groups, as Kjærgaard et al. (2019) found that graduating the age dimension into single year intervals adds no additional accuracy to the forecast. In contrast to Kjærgaard et al. (2019), we do not consider the similar sample of causes for males and females.

In our analysis the following groups of localizations are considered (ICD 10 code is provided in the brackets):

- For males: Malignant neoplasms of the trachea, bronchus and lungs (C33-C34), malignant neoplasms of the stomach (C16), malignant neoplasms of the prostate (C61);
- For females: Malignant neoplasms of the trachea, bronchus and lungs (C33-C34), malignant neoplasms of the stomach (C16), malignant neoplasms of breast (C50), malignant neoplasms of cervix uteri (C53).

This set of causes was chosen in order to ensure the coincidence when ICD changes occur. Since we forecast the death densities in the multiple-decrement life table, all other cancers and all other causes of death that are not implicitly forecasted were collapsed into two r categories: “Other cancers” and “Other causes” to ensure the death density sum constraint. Since the dynamics of these two groups of causes are practically meaningless and uninterpretable (since it is characterized by many omnidirectional trends and risk factors), the information and results for these 2 groups of causes will be presented in analysis only where appropriate. We use the Russian mortality data as-is, with no additional cleaning or smoothing, but we removed the information on deaths in early ages (under age 30) to avoid fluctuations due to very small numbers. To achieve continuity, zero death counts were replaced by a half death, similar to Kjærgaard et al. (2019).

Methods:

For detailed information on methods and models the reader is referred to the original study by Kjærgaard et al. (2019). Since we apply a similar methodological framework only general description of methodology is given.

Compositional data can be seen as percentages, densities or parts of the whole that sum to an arbitrary chosen constant, and are restricted to a simplex space. In general case, composition can be seen as strictly positive vector of length P $(P \geq 2)$ with a fixed sum that can be transformed from the simplex to the real space by log-ratio or other transformations. Compositional data arise when we have many compositions all having the same sum. This concept was developed by Aitchison (1986) and further improved by Vera Pawlovsky-Glann and Antonella Buccianti (2011). It is independent of time-series used in forecasting and is used to reduce the information in the data.

Kjærgaard et al. (2019) offered three compositional models that complement the original work by Oeppen (2008). These models are:

1) CoDA model with common time trend (CT-CoDA). This is the original model proposed by Oeppen (2008) extended for the cause-specific mortality forecast purposes. This model allows only a single time trend for all causes of death and uses the rank three approximation for singular value decomposition (SVD).
2) Two step CoDA model (2S-CoDA). This model introduces an additional weighting for ages, causes and times based on their relative importance in the overall structure of mortality composition. This model was found to produce the best forecasts in original paper by Kjærgaard et al. (2019) and it ought to produce a particularly good fit if a single category(such as cancer) is decomposed into its integral parts (e.g. cancer localizations). This model uses rank one approximation for SVD, since in 2S-CoDA higher rank singular vectors are modelled by cause-specific terms in generalized SVD (Kjærgaard et al. 2019).
3) Model with cointegrating vector errors (VECM-CoDA). This model allows for estimation of more than one time trends for rank approximations and uses rank three for SVD.

The performance of these models is compared with Lee-Carter model (LC). All three compositional models use the multidimensional compositional matrices of death densities as input. The output is the set of time-series of cross-sectional multiple-decrement life-tables calculated with the closed population assumption. With these three models twenty year out-of-

sample forecasts for males and females were obtained. To check the goodness of fit of these models the ten years point estimate of in sample root mean squared errors is evaluated (RMSE).

In contrast to the Lee-Carter model that models the death rate and forecasts its changes, CoDA forecasts the life-table death density dynamics consisting in the redistribution of deaths from one cause and age to another overtime. This property of CoDA models is called the life-saving process (Bergeron-Boucher et al. 2018; Vaupel, Yashin 1987).

Results:

This section starts with descriptive analysis of Russian cancer mortality trends (SMR) and dynamics in transformed compositional data, followed by the inspection of the compositional and LC model parameters, goodness of fit tests and resulting death density forecasts.

Descriptive analysis of SMR and CoDA:

Among males, the age-standardized death rate (SDR) (Figure 4) for all cancers combined was gradually increasing to the historical maximum observed in 1993 after which the trend has been declining until the present. Among females cancer mortality was quite stable before 1981, slightly increased in 1982-1993, and has been declining since 1994. Cancer mortality rates for Russian males are about twice as high as for females. SDR from cancer of the trachea, bronchus and lungs was increasing in males to a maximum in 1993, when it began to decline back to the levels similar to those observed in 1965 reached by 2018. Among females the patterns are similar to males, but at lower levels. Stomach cancer mortality was gradually declining through the period of observations for both males and females. The opposite holds true for cancer of the prostate, which has more than tripled since 1965. SDR from female breast cancer rose until 2004, followed by a decline accelerating after 2009. The SDR of cancer of the cervix uteri was declining with some short-term fluctuations.

The distribution of all cancer (C00-C97) death densities (Figure 5) has shifted towards older ages and it became more compressed in both males and females. The geometric mean of death distributions in males is close to the distribution of 1980, and in females to that of 2000. Centered log-transformed multiple-decrement life-table cancer death densities (Figure 6) have only positive values in both males and females. This means that during the period the deaths were transferred from all other causes towards cancer (all cancers C00-C97). The centered deaths distribution in 2018 is sharply increasing with age, meaning that there is more cancer death that occurs in the older ages.

Parameter approximation:

The time pattern of mortality development and its twenty year forecast is captured by the *k(t)* parameter (Figure 7). For compositional models, an increasing *k(t)* pattern implies that mortality levels decline with time, while for the Lee-Carter model the interpretation of *k(t)* dynamics is opposite to that of the CoDA models. For CT-CoDA model the general development of mortality in time is shown, as the model does not allow the cause-specific *k(t)* decomposition. CT-CoDA captures an increasing mortality pattern for both males and females that is forecasted to sustain this dynamics for the next twenty years in a linear way. 2S-CoDA model captures no clear time trend of mortality development. In males, for the fitting period model suggests that mortality from cancer of the trachea, bronchus and lungs and stomach cancer were declining, while mortality levels for other causes considered were increasing. Cause-specific development of mortality is forecast to reach zero equilibrium in twenty years. In females, the time pattern captured by the 2S-CoDA model is more stable and distinct. The model captures the mortality decline from all causes considered except for the cancer of the trachea, bronchus and lungs and other cancers before 2005, when mortality has started to increase again. The forecast suggests that in females mortality will further increase in a slow manner, and the mortality from cancer of female breast, for which a more significant increase is expected. Contrast to 2S-CoDA model, the VECM-CoDA captures the declining mortality pattern for all causes considered except for the other diseases, that are expected to be stable in future and other cancers, that are expected to increase in males. In females, the mortality from all causes declines and is expected to do so in the future. Prostate cancer mortality is expected to decline most in males, while in females the stomach cancer will have the primacy in terms of mortality decline. For all other causes (and especially for other non-cancer causes that also showed a decline in 1985) an increasing pattern is observed in 1990`s, followed by a decline that began around 2005. Similar to VECM-CoDA model, LC model suggests a gradual decline in mortality from stomach cancer and cancer of the trachea, bronchus and lungs in males with time, but does not capture the decline in prostate cancer mortality and increase in all other cancers, while LC model forecasts these two groups of cancers to stay relatively stable in future. For females LC model also captures and forecasts the decline in stomach cancer mortality and most of the other causes considered, but contrast to VECM-CoDA, does not forecasts the further decline in female breast cancer and other cancers.

The *b(x)* parameter (Figure 8) for compositional models profiles the redistribution of death from ages and causes where its values are negative to those where its values are positive given the *k(t)* parameter is positive. For LC the positive values of the *b(x)* imply that the mortality declines in these ages, but no cause redistribution can be definitively claimed, since this model uses the

death rates rather than death densities. According to the CT-CoDA model prostate cancer causes fewer deaths in younger ages while redistribution primarily goes from stomach cancer and trachea, bronchus and lungs cancer to other cancers other diseases and prostate cancer in older ages. In females, stomach cancer deaths are redistributed across all ages. In younger ages the primacy goes to the cancer of cervix uteri and other diseases, in mid ages deaths are being redistributed from cancer of cervix uteri towards the female breast cancer that remains being important role in older ages. In males shifts of mortality to older ages is clearly observed for prostate cancer while in females the opposite is expected for cancer of cervix uteri. 2S-CoDA model output is similar to that of CT-CoDA in males and females, although in females the age-pattern of cause-specific mortality redistribution starts at older ages. It is important to remember that redistribution is only occurring if $k(t)$ parameter is positive. VECM-CoDA model captures different pattern compared to two previous models. In males the redistribution of deaths also goes to prostate cancer in mid and older ages but cancers of trachea, bronchus and lungs have primacy over the latter in older ages. For females, the parameter estimated by VECM-CoDA model are similar to those of two previous models, except for the trachea, bronchus and lungs and other cancers that are estimated to be of a greater importance in older ages, and cancer of cervix uteri, which has a pattern similar to that captured by two previous models, albeit with younger age schedule. LC model agrees with the output of compositional models on the dynamics of parameter for prostate cancer in males and cancer of cervix uteri in females. The estimated parameter for all other causes considered does not follow a clear redistribution pattern.

Goodness of model fit:

To evaluate the quality of fit, a 10 year in-sample forecast (Figure 9) was done. Only selected causes with meaningful dynamics are shown here. For females all models produce a reasonably good fit. In males, the situation is opposite, all models fail to capture the dynamics of prostate cancer and only the LC model produces forecast that are somewhat close to observed trachea, bronchus and lungs cancer dynamics. All models underestimate the cancer mortality decline in males. To evaluate the goodness of fit of all the causes shown in Figure 9, they were aggregated into a single quantity and are shown in Figure 10. We can see, that in males the LC model results are the closest to the observed dynamics followed by VECM-CoDA, 2S-CoDA and CT-CoDA, while in females 2S-CoDA produces the best results, followed by VECM-CoDA, LC and CT-CoDA. For numerical evaluation of the goodness of fit, the point 10 year estimate of RMSE (residual mean squared error) is shown in Table 1.

Out of sample mortality forecast:

The final goal of the paper was to forecast the life table cancer death densities. The 20 year forecast is presented in Figure 11. It is expected that for trachea, bronchus and lungs, all models except the VECM-CoDA forecasts the decline in densities of death, for stomach cancer all models show a continuous secular decline, while for prostate cancer the increase in life-table death is expected to occur. For prostate cancer VECM-CoDA and CT-CoDA show a more rapid increase than the other 2 models. In females, all models show similar patterns for trachea, bronchus and lungs, stomach and cervical cancer, but for female breast cancer, the 2S-CoDA shows a small decline in contrast to other models.

Discussion:

In this paper we tried to make a forecast of the rather erratic Russian cancer mortality series using the CoDA framework. In terms of general dynamics of cancer in Russia, prostate and female breast cancers are becoming more important while deaths are being redistributed from stomach and trachea, bronchus and lungs (except for the old ages) cancers. In females the application of CoDA framework substantially improved the forecast outcome. The 2S-CoDA provides best fit for the data at hand for females. This result is similar to that obtained by Kjærgaard et al. (2019) for Denmark and France, but in contrast to the original study, the VECM-CoDA has provided a second best fit followed by LC and CT-CoDA models. Kjærgaard et al. (2019) says that assuming the constant time trend in cause-specific mortality is a reasonable simplification; however, it seems that given the unusual development of Russian mortality, assuming the common cause-specific time-trend is not appropriate. 2S-CoDA improves the forecast accuracy by introducing the cause, age and time weights allowing for better capturing of most recent trends. VECM-CoDA was able to improve the forecast accuracy by allowing for different time trends for cause-specific mortality. CT-CoDA which assumes the constant time trend failed at providing a better forecast than the general LC method. In fact this model forecasts that the future mortality levels will increase as a result of a dynamics in the long-term time-series.

We know that the mortality in Russian Federation began to decline and LE began to increase in 2004 (Shkolnikov et. al. 2013), and this increase is stable in time. The Russian male mortality series forecast offers more interesting results. None of the models were able to provide an appropriate fit. Among the four models, LC model does the best job. We argue nonetheless, that this has happened by a mere coincidence. All models underestimate the mortality in Russian

males, but the exact quality of LC mentioned in introduction, that it tends to overestimate the mortality when cause-specific forecasts are being aggregated, results in a better overall RMSE score. Fluctuations in Russian male mortality does not allow for an appropriate forecast be it in compositional or standard framework. Kjærgaard et al. (2019) warns for using of any the extrapolative models in presence of many breaks in time-series trend. In Russia the secular decline in mortality started in 2004 (Shkolnikov et. al. 2013), and it seems that previous fluctuations still not allow for good capturing of the appropriate mortality time-trend. Among the causes considered in males, CoDA models primarily failed to forecast the development of mortality from the cancer of the trachea, bronchus and lungs that in males have the most erratic pattern of mortality development and that of the prostate cancer that has unexpectedly increased. This increase in prostate cancer mortality may be a result of over diagnostics (Bray, Parkin 2009; Taitt 2018) that followed the introduction of Prostate-Specific Antigen testing in Russia. A similar prostate cancer mortality trend was observed in many western countries (Dickinson et al. 2016; Houston et al. 2018). Prostate cancer mortality increase may thus be real and maybe (at least partly) caused by better diagnostics resulting in an increase of registration of this cancer site as the underlying cause of death. Male mortality in Russia is being shaped by a very unique combination of factors that makes it impossible to produce a robust forecast using a compositional framework in males if the time-series of mortality is takes as-is.

Limitations and future directions:

We did not explicitly considered the difference in compositional model output resulting from the choice of fitting period. In case of Russia, one possible way to obtain a more valid forecast that will be better in capturing the recent mortality dynamics is to reduce the length of the time-series, by ignoring the period of high mortality fluctuation and prior years, essentially starting the forecast at around 2004, when stable pattern of mortality decline was established in Russia. This would allow for better capturing of the recent dynamics of the $k(t)$ parameter and to avoid a problem of negative dynamics of $k(t)$ such as that portrayed by CT-CoDA. The other option for obtaining a better forecast is the smoothing the data. Fitting smooth curves at the period of intense mortality fluctuations or simply changing the dynamics to the linearly approximated pattern can potentially improve the forecast accuracy.

Conclusion:

In this paper we obtained the coherent mortality forecasts of Russian cancer mortality decomposed by groups of cancer localizations using the compositional framework. Thus, we had

a first glimpse at the possible limitations of compositional models by applying them in practice to the mortality series with many trend breaks and unclear development pattern. Compositional framework improves cancer mortality forecasts for Russian females, but fails to do so for Russian males. It seems that using the full time-series of mortality data for forecasting Russian male mortality is a false approach and there is no good method of forecasting it as such. We were determined to use the mortality series as-is, however, reducing the number of years on which the forecast is based (for example censoring the data prior to 2004), interpolation or smoothing of mortality time-series for years where significant fluctuations occurred may have resulted in better model fit. Compositional forecasting models are new to demography. The flexibility of these models allows robust and coherent time-series forecasting that goes way beyond the cause-specific or regional mortality forecasts suggesting that the full potential of CoDA forecasting models is yet to be discovered.

Bibliography:

1) Aitchison, J. (1986). The statistical analysis of compositional data. London: *Chapman and Hall.*

2) Baili, P., Di Salvo, F., Marcos-Gragera, R., Siesling, S., Mallone, S., Santaquilani, M., Otter, R. (2015). Age and case mix-standardized survival for all cancer patients in Europe 1999–2007: Results of EUROCARE-5, a population-based study. *European Journal of Cancer*, 51(15), 2120–2129. https://doi.org/10.1016/j.ejca.2015.07.025.

3) Bergeron-Boucher, Marie-Pier, Vladimir Canudas-Romo, Jim Oeppen, and James W. Vaupel. 2017. "Coherent Forecasts of Mortality with Compositional Data Analysis." *Demographic Research* 37 (August): 527–66. https://doi.org/10.4054/DemRes.2017.37.17.

4) Bergeron-Boucher, Marie-Pier, Violetta Simonacci, Jim Oeppen, and Michele Gallo. "Coherent Modeling and Forecasting of Mortality Patterns for Subpopulations Using Multiway Analysis of Compositions: An Application to Canadian Provinces and Territories." *North American Actuarial Journal 22, no. 1* (January 2, 2018): 92–118. https://doi.org/10.1080/10920277.2017.1377620.

5) Booth, H., and L. Tickle. 2008. "Mortality Modelling and Forecasting: A Review of Methods." *Annals of Actuarial Science* 3 (1): 3–43. https://doi.org/10.1017/S1748499500000440.

6) Bray, F., & Parkin, D. M. (2009). Evaluation of data quality in the cancer registry: Principles and methods. Part I: Comparability, validity and timeliness. *European Journal of Cancer*, *45*(5), 747–755. https://doi.org/10.1016/j.ejca.2008.11.032.

7) Briggs, Adam D. M., Peter Scarborough, and Jane Wolstenholme. 2018. “Estimating Comparable English Healthcare Costs for Multiple Diseases and Unrelated Future Costs for Use in Health and Public Health Economic Modelling.” *PLoS ONE* 13 (5). https://doi.org/10.1371/journal.pone.0197257.

8) Danilova, Inna, Vladimir M. Shkolnikov, Dmitri A. Jdanov, France Meslé, and Jacques Vallin. “Identifying Potential Differences in Cause-of-Death Coding Practices across

9) Russian Regions.” *Population Health Metrics* 14, no. 1 (December 2016). https://doi.org/10.1186/s12963-016-0078-0.

10) Doll, R. (1998). Uncovering the effects of smoking: Historical perspective. *Statistical Methods in Medical Research*, 7(2), 87–117. https://doi.org/10.1177/096228029800700202.

11) Girosi, Federico, and Gary King. *Demographic Forecasting.* Princeton: Princeton University Press, 2008.

12) Human Cause-of-Death Database. French Institute for Demographic Studies (France) and Max Planck Institute for Demographic Research (Germany). Available at http://www.causeofdeath.org (data downloaded on [18.01.2020]).

13) Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at at www.mortality.org or http://www.humanmortality.de (data downloaded on [18.01.2020]).

14) International Agency for Research on Cancer, ed. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Volume 83, Tobacco Smoke and Involuntary Smoking: This Publication Represents the Views and Expert Opinions of an IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, Which Met in Lyon, 11 - 18 June 2002*. Lyon: IARC, 2004.

15) Kjærgaard, Søren, Yunus Emre Ergemen, Malene Kallestrup-Lamb, Jim Oeppen, and Rune Lindahl-Jacobsen. 2019. “Forecasting Causes of Death by Using Compositional Data Analysis: The Case of Cancer Deaths.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 68 (5): 1351–70. https://doi.org/10.1111/rssc.12357.

16) Lee, Ronald D., and Lawrence R. Carter. 1992. “Modeling and Forecasting U.S. Mortality.” *Journal of the American Statistical Association* 87 (419): 659–71. https://doi.org/10.1080/01621459.1992.10475265.

17) Leon, D. A., L. Chenet, V. M. Shkolnikov, S. Zakharov, J. Shapiro, G. Rakhmanova, S. Vassin, and M. McKee. “Huge Variation in Russian Mortality Rates 1984-94: Artefact, Alcohol, or What?” *Lancet (London, England)* 350, no. 9075 (August 9, 1997): 383–88. https://doi.org/10.1016/S0140-6736(97)03360-6.

18) Meropol, Neal J., and Kevin A. Schulman. “Cost of Cancer Care: Issues and Implications.” *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 25, no. 2 (January 10, 2007): 180–86. https://doi.org/10.1200/JCO.2006.09.6081.

19) Oeppen, Jim. 2008. “Coherent Forecasting of Multiple-Decrement Life Tables: A Test Using Japanese Cause of Death Data.” 23.

20) Park, Joohyun, and Kevin A. Look. 2019. “Health Care Expenditure Burden of Cancer Care in the United States.” *Inquiry: A Journal of Medical Care Organization, Provision and Financing* 56 (October). https://doi.org/10.1177/0046958019880696.

21) Pawlowsky-Glahn, Vera, and Antonella Buccianti. *Compositional Data Analysis: Theory and Applications.* John Wiley & Sons, 2011.

22) *Russian Fertility and Mortality Database*. Center for Demographic Research, Moscow (Russia). Available at http://demogr.nes.ru/index.php/ru/demogr_indicat/data (data downloaded on [18.03.2020]).

23) Shkolnikov, V. “Cancer Mortality in Russia and Ukraine: Validity, Competing Risks and Cohort Effects.” *International Journal of Epidemiology* 28, no. 1 (February 1, 1999): 19–29. https://doi.org/10.1093/ije/28.1.19.

24) Shkolnikov, Vladimir M., Evgeny M. Andreev, Martin McKee, and David A. Leon. “Components and Possible Determinants of Decrease in Russian Mortality in 2004-2010.” *Demographic Research* 28 (April 24, 2013): 917–50. https://doi.org/10.4054/DemRes.2013.28.32.

25) Simon, Alice E., and Jane Wardle. “Socioeconomic Disparities in Psychosocial Wellbeing in Cancer Patients.” *European Journal of Cancer (Oxford, England: 1990)* 44, no. 4 (March 2008): 572–78. https://doi.org/10.1016/j.ejca.2007.12.013.

26) Taitt, H. E. (2018). Global Trends and Prostate Cancer: A Review of Incidence, Detection, and Mortality as Influenced by Race, Ethnicity, and Geographic Location. *American Journal of Men’s Health*, *12*(6), 1807. https://doi.org/10.1177/1557988318798279.

27) Vaupel, J., and A. Yashin. 1987. Repeated Resuscitation: How Lifesaving Alters Life Tables. *Demography* 24(1), 123–135.

28) Wilmoth, John R. 1995. “Are Mortality Projections Always More Pessimistic When Disaggregated by Cause of Death?” *Mathematical Population Studies* 5 (4): 293–319. https://doi.org/10.1080/08898489509525409.

29) Wise, Jacqui. “Covid-19: Cancer Mortality Could Rise at Least 20% Because of Pandemic, Study Finds.” *BMJ* 369 (April 29, 2020). https://doi.org/10.1136/bmj.m1735.

Appendix:

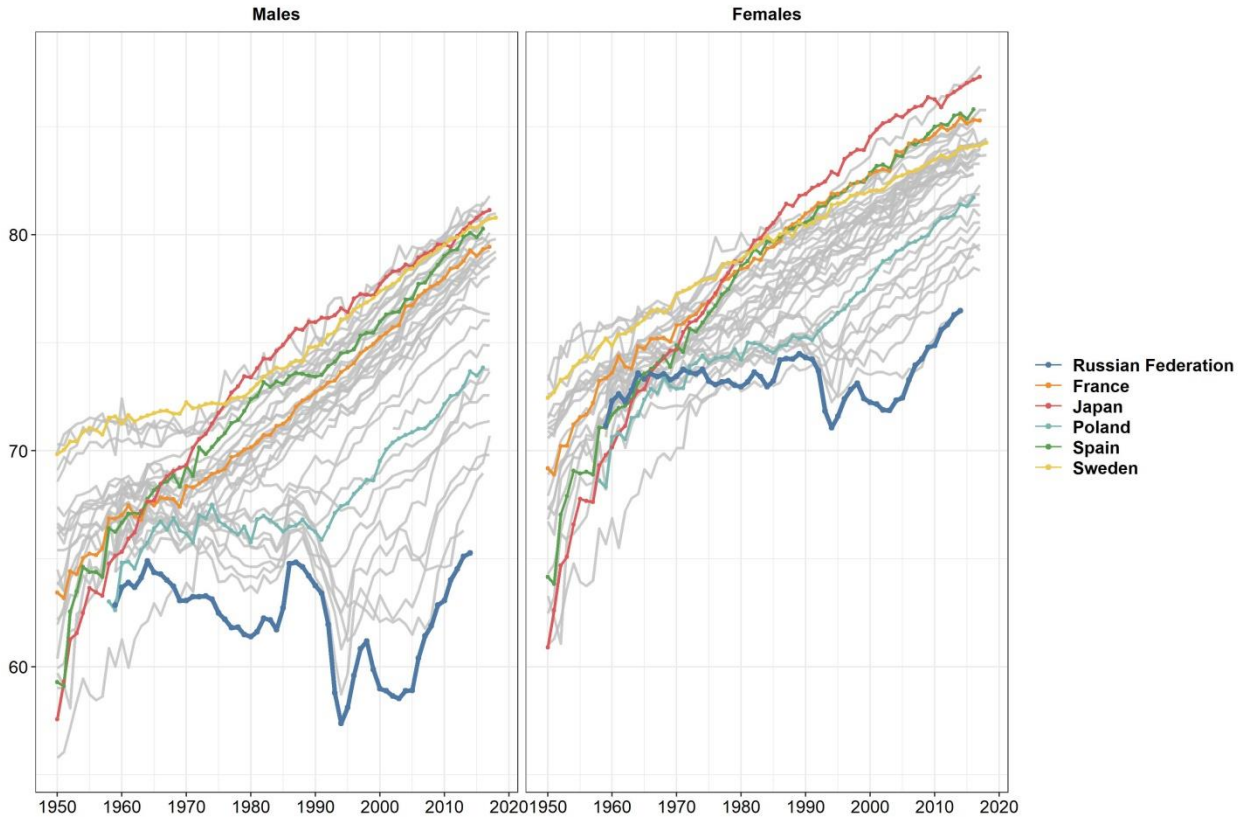

Figure 1 Russian life expectancy at birth compared to other HMD countries. The Russian time series is the most erratic in this set of populations. **Source:** *Human Mortality database*.

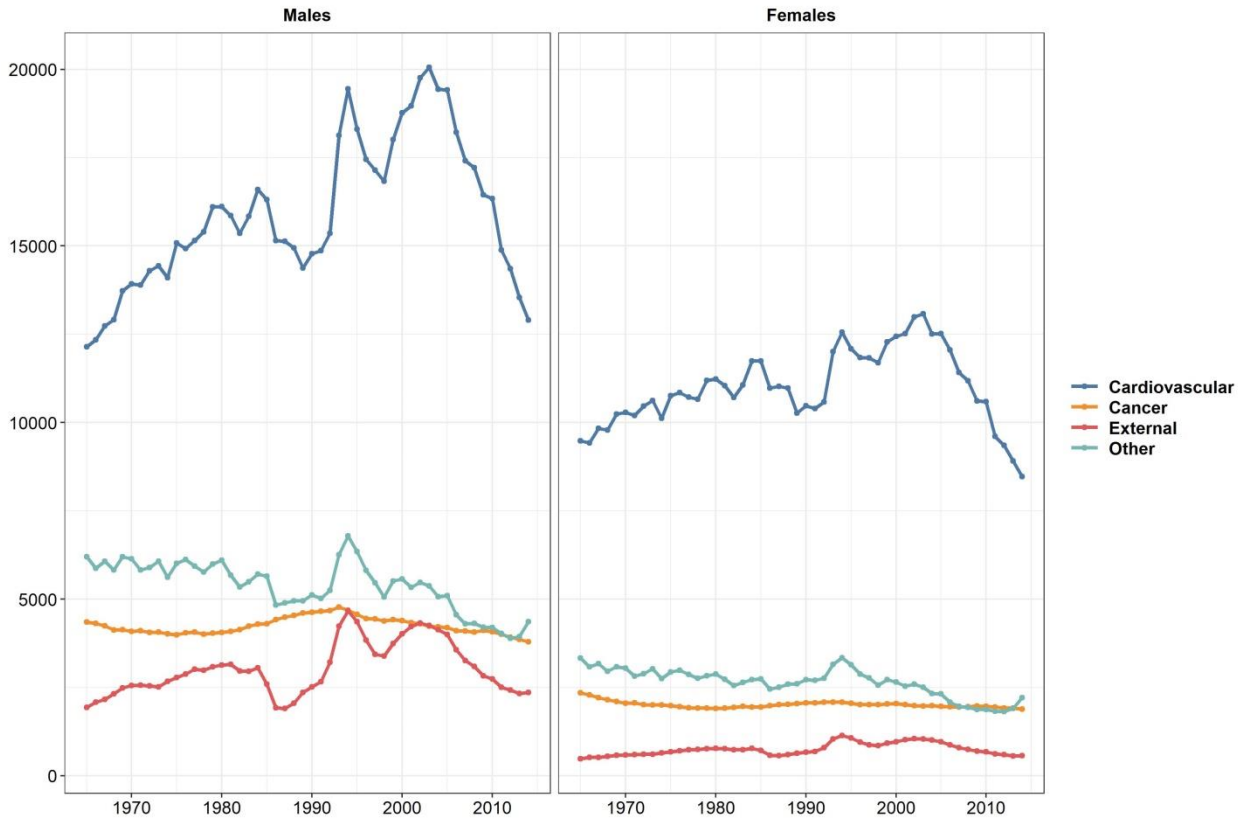

Figure 2 Standardized mortality per 1 000 000 for main groups of causes of death in Russia. European standard population 2013. **Source:** *Human Cause of Death Database*.

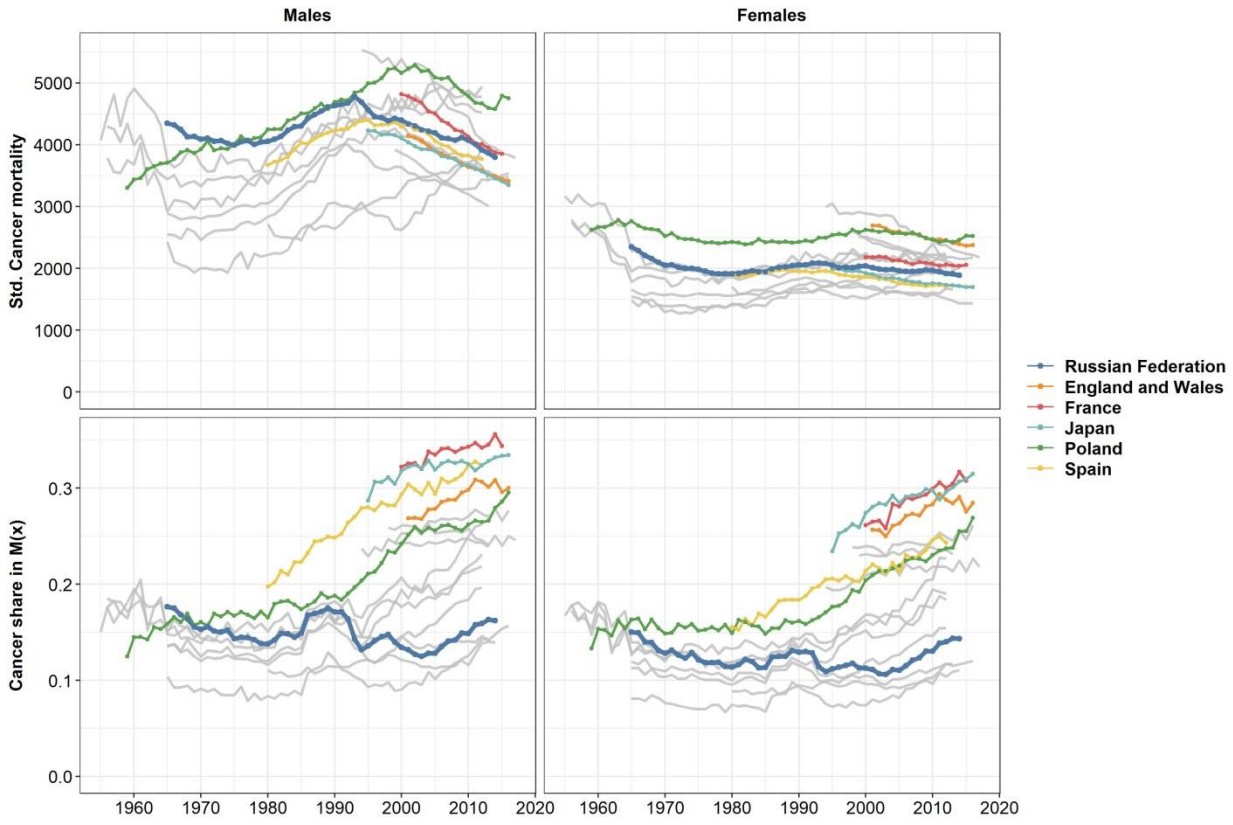

Figure 3 Standardized Russian cancer mortality per 1 000 000 compared to other populations and fraction of cancer in M(x). European standard population of 2013. **Source:** *Human Cause of Death Database*.

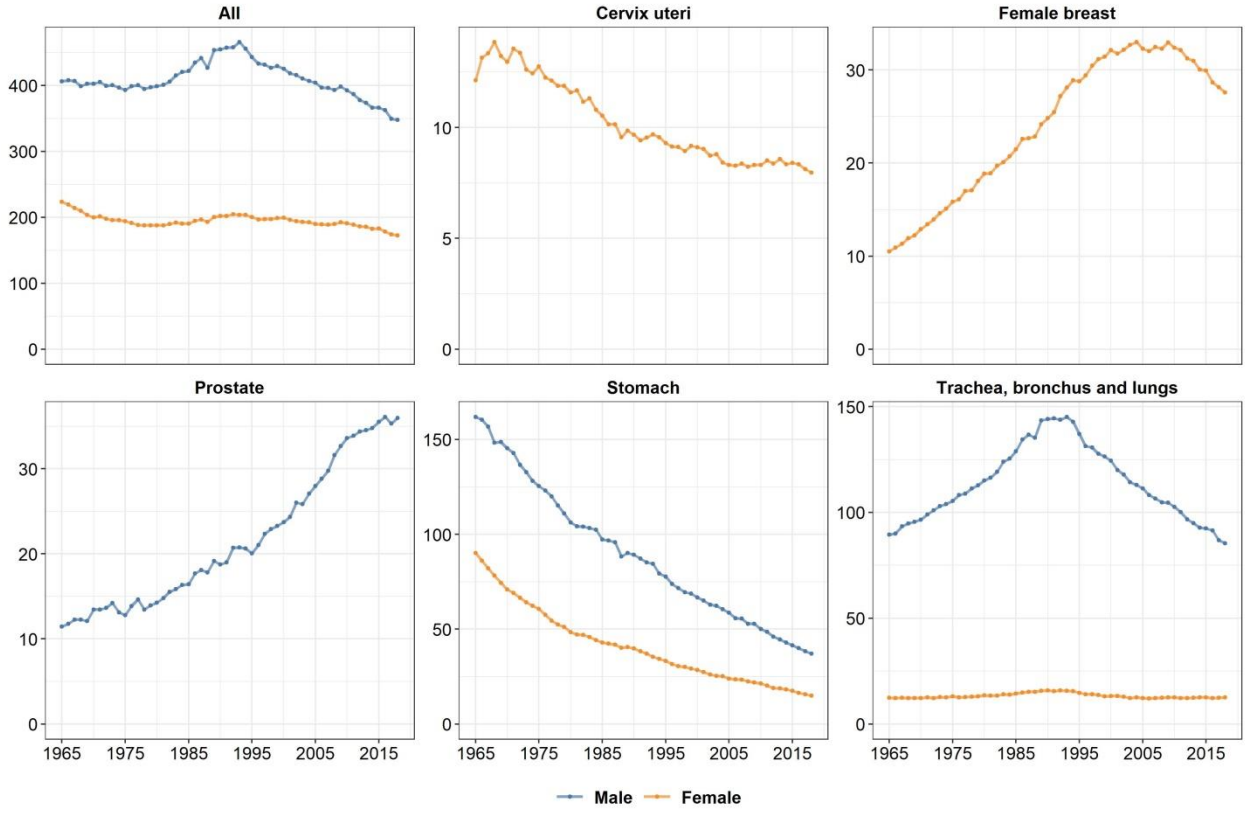

Figure 4 Age-standardized death rates by sex for all cancers combined and leading cancer sites per 100.000 person-years in Russia in 1965-2018. European standard population of 2013.
Source: Own calculations based on data from *RosBRiS*.

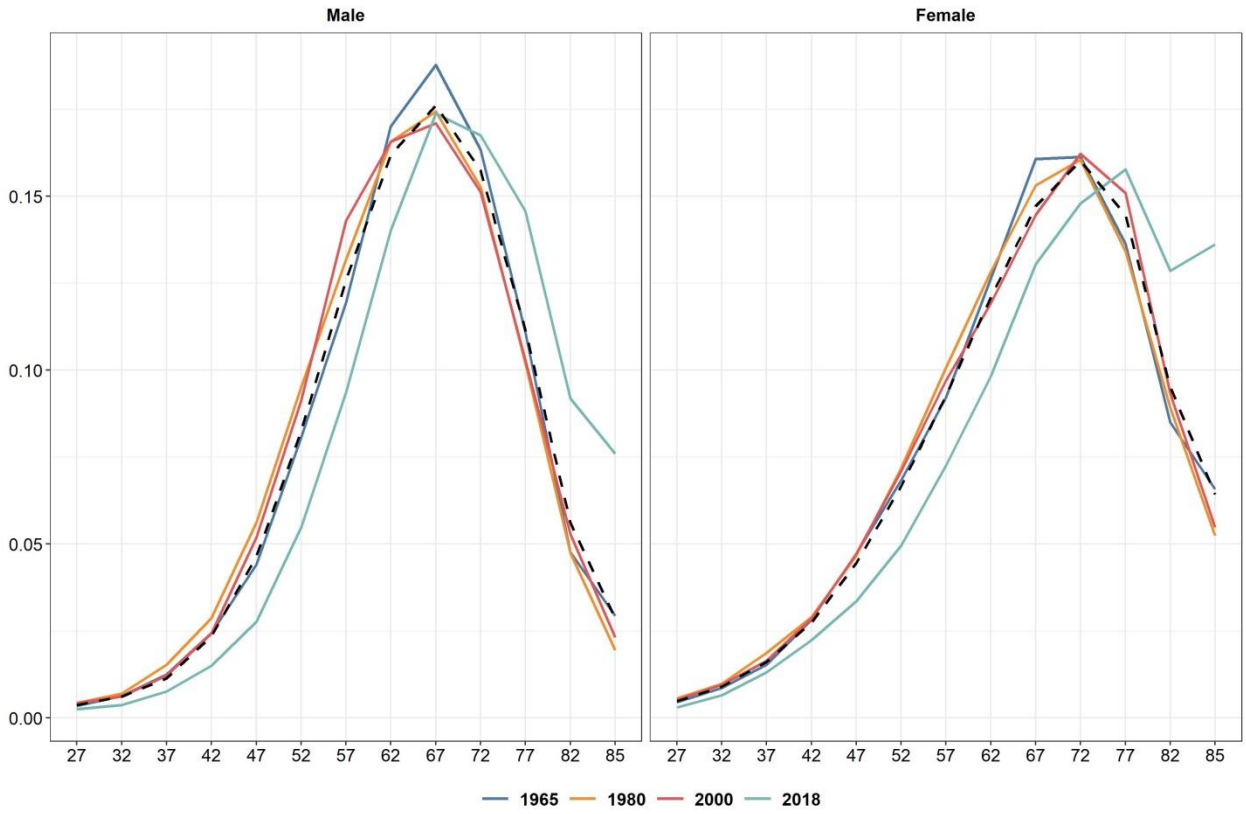

Figure 5 Cancer death distributions for chosen years for males and females in Russian Federation. Black dashed line represents the geometric mean across years. **Source:** Own calculations based on data from *RosBRiS*.

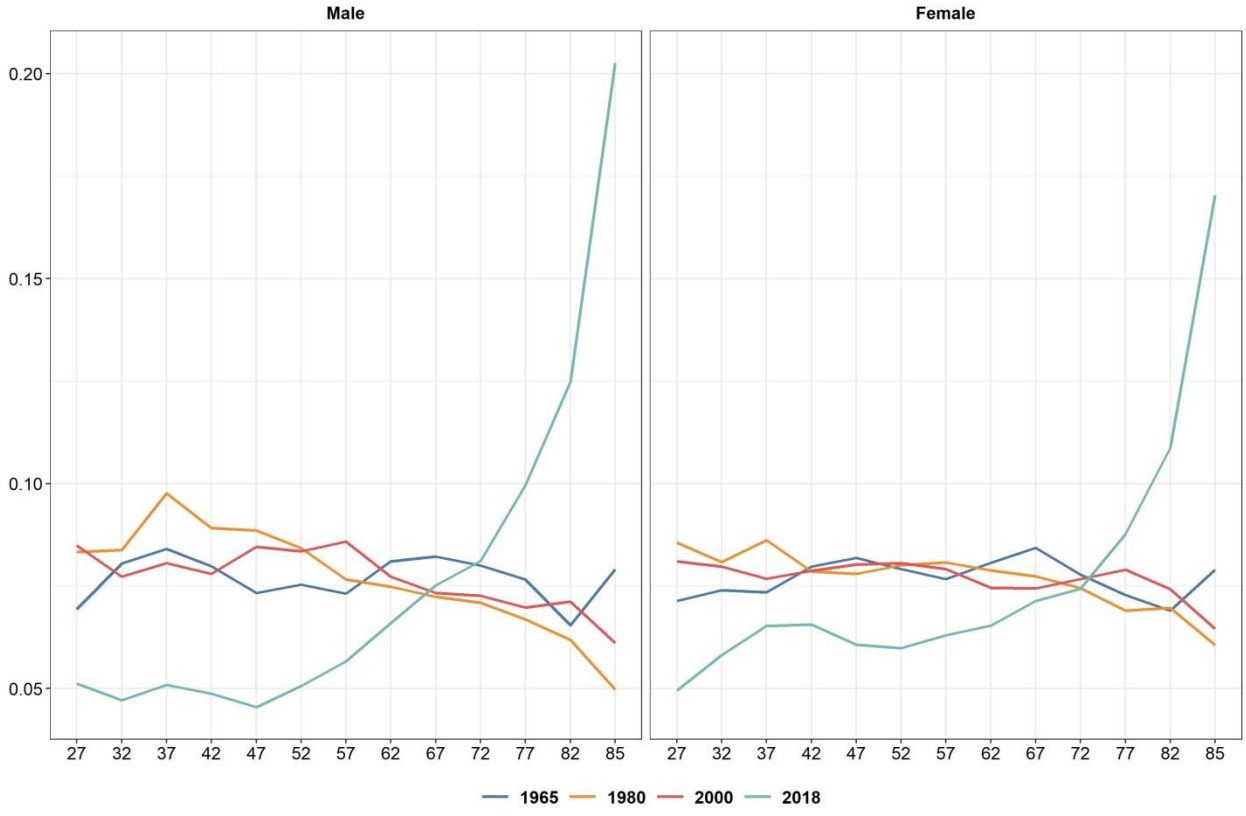

Figure 6 Log-transformed centered death distributions for chosen years. Males and females, Russian Federation. **Source:** Own calculations based on data from *RosBRiS*.

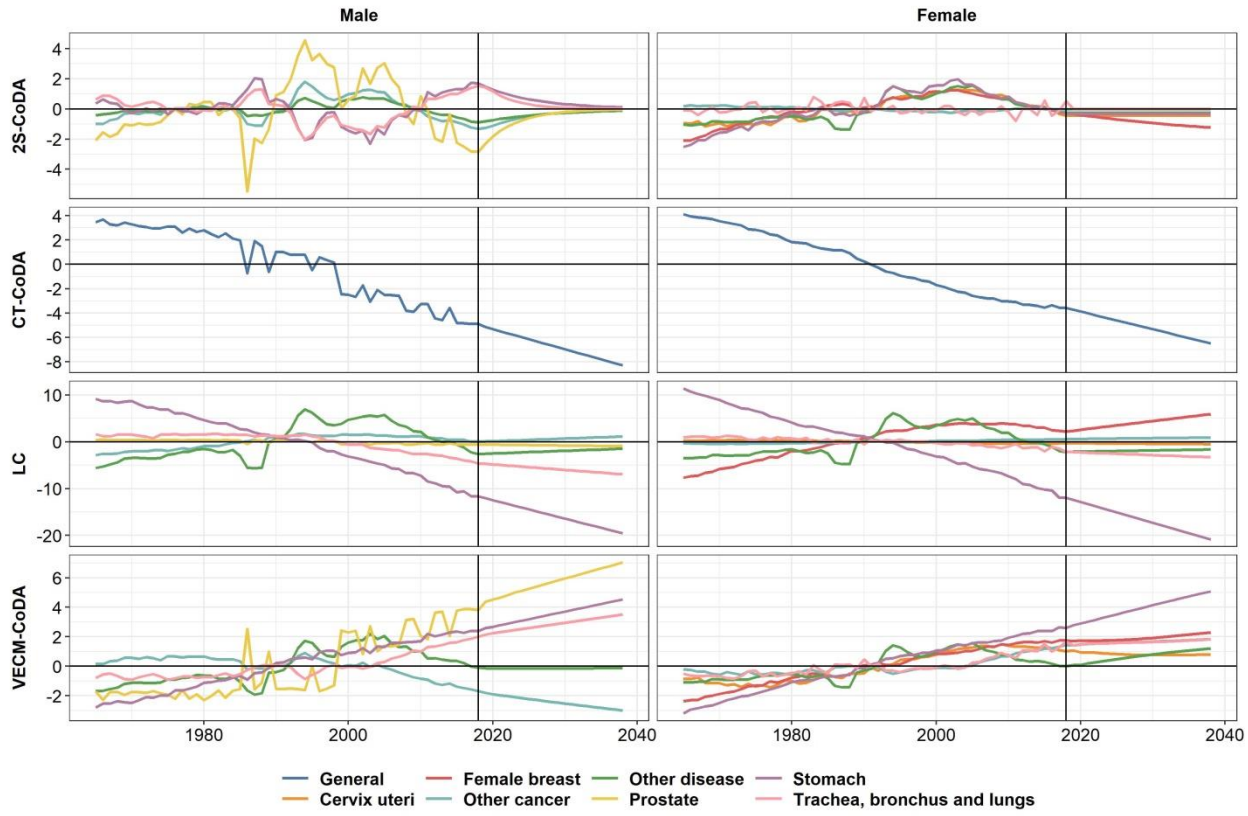

Figure 7 Observed and 20 year forecast of *k(t)* – the overall mortality development index for three compositional models and LC model. Males and females decomposed by causes of death (where possible), for Russian Federation. Forecast beginning is indicated with black vertical line. **Source:** Own calculations based on data from *RosBRiS*.

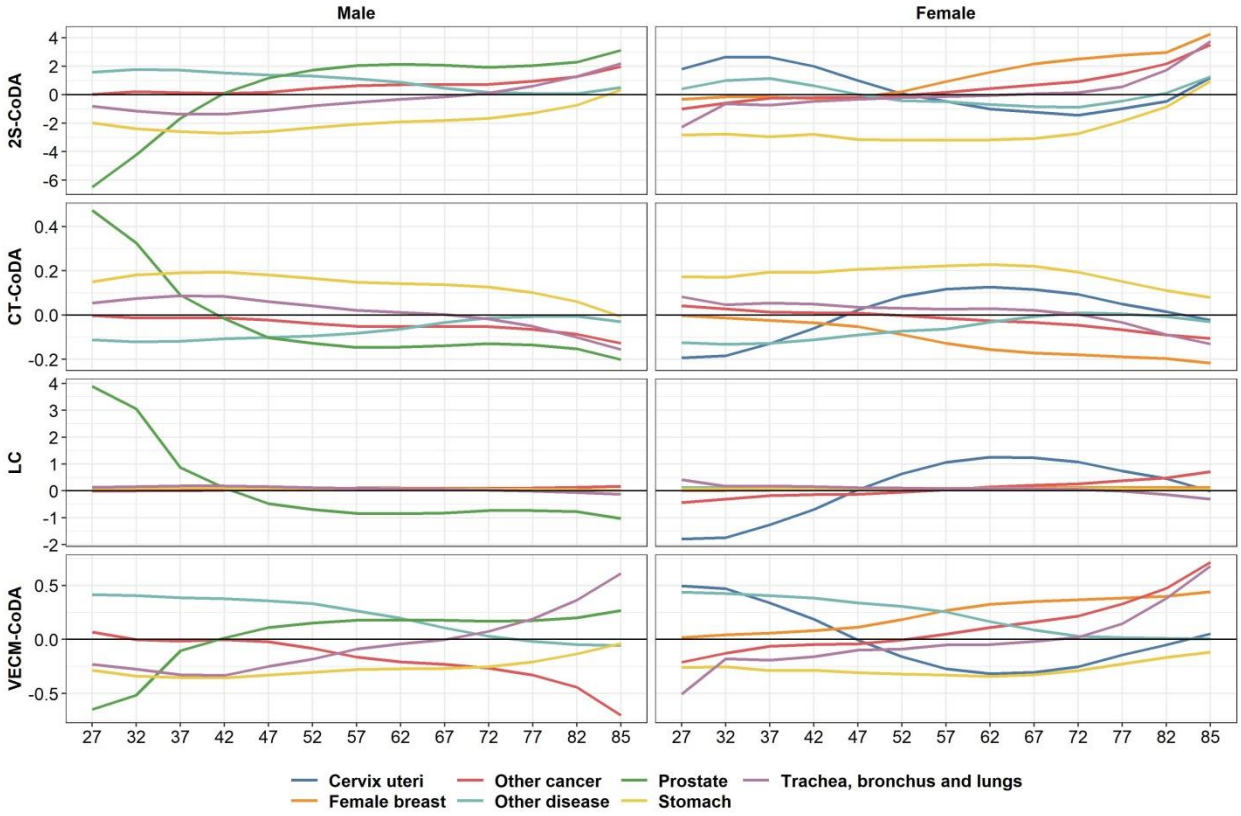

Figure 8 Observed and 20 year forecast of *b(x)* - index for the redistribution of deaths between ages and causes for three compositional models and LC model. Males and females decomposed by causes of death, for Russian Federation. **Source:** Own calculations based on data from *RosBRiS*.

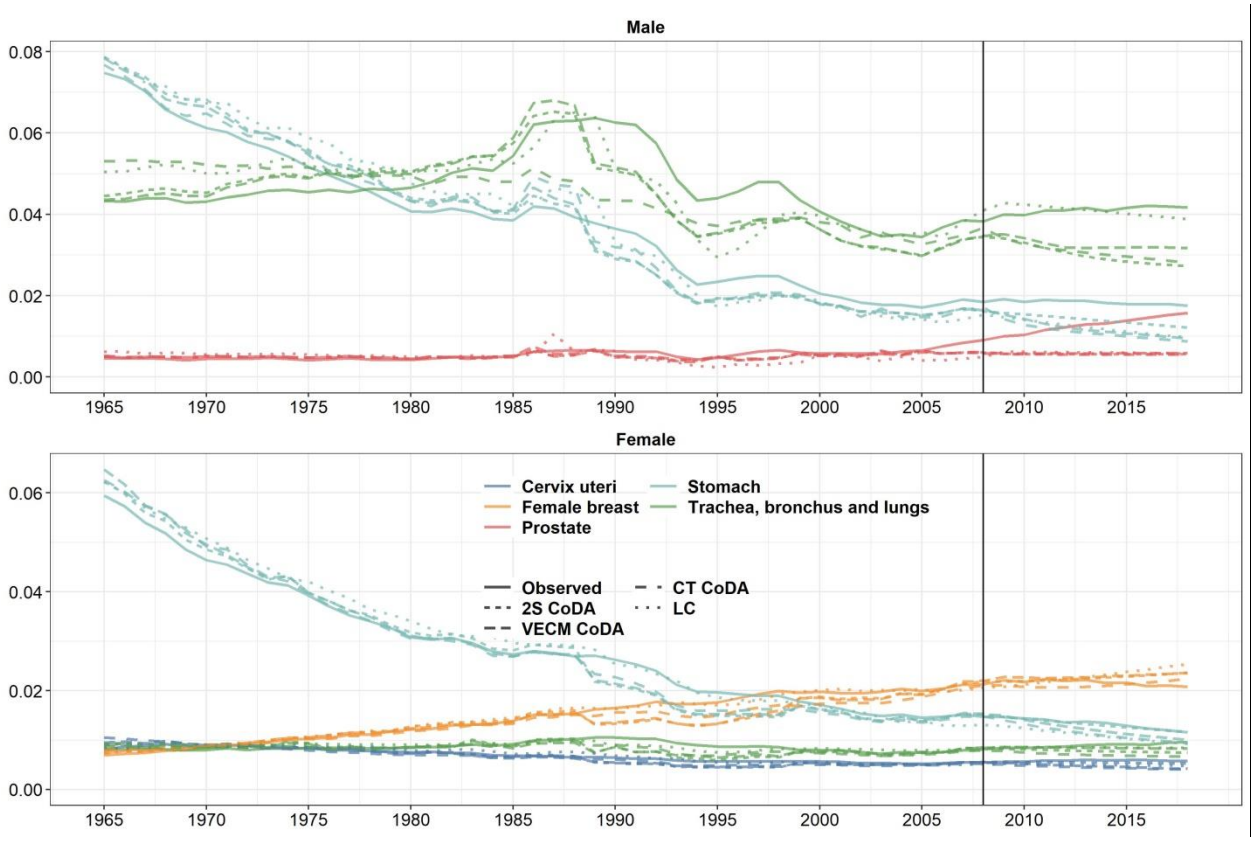

Figure 9 Ten years in sample forecast of life-table cancer death for chosen groups of cancers for males and females in Russian Federation. Three compositional models and LC model. Black vertical line indicates the beginning of forecast period. **Source:** Own calculations based on data from *RosBRiS*.

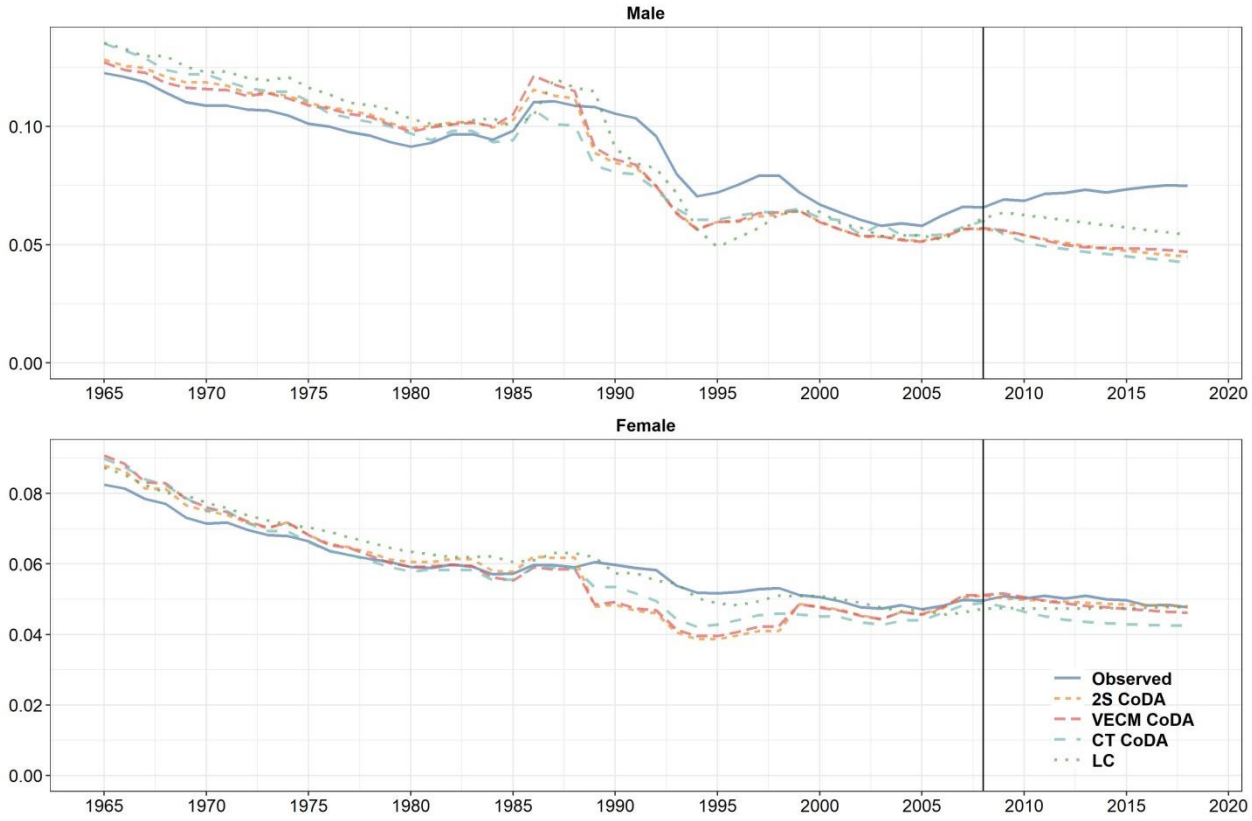

Figure 10 Ten years in sample forecast of sum of cancer death for chosen groups of cancers displayed in figure 9, for males and females in Russian Federation. Three compositional models and LC model. Black vertical line indicates the beginning of forecast period. **Source:** Own calculations based on data from *RosBRiS*.

Model	**Males**	**Females**
2S-CoDA	2358	102
VECM-CoDA	2289	182
CT-CoDA	2597	575
LC	1448	257

Table 1 Point estimate of total root mean squared errors of life-table death for ten year in sample forecast for three compositional models and LC model. For males and females and chosen groups of cancer localizations in Russian Federation. **Source:** Own calculations based on data from *RosBRiS*.

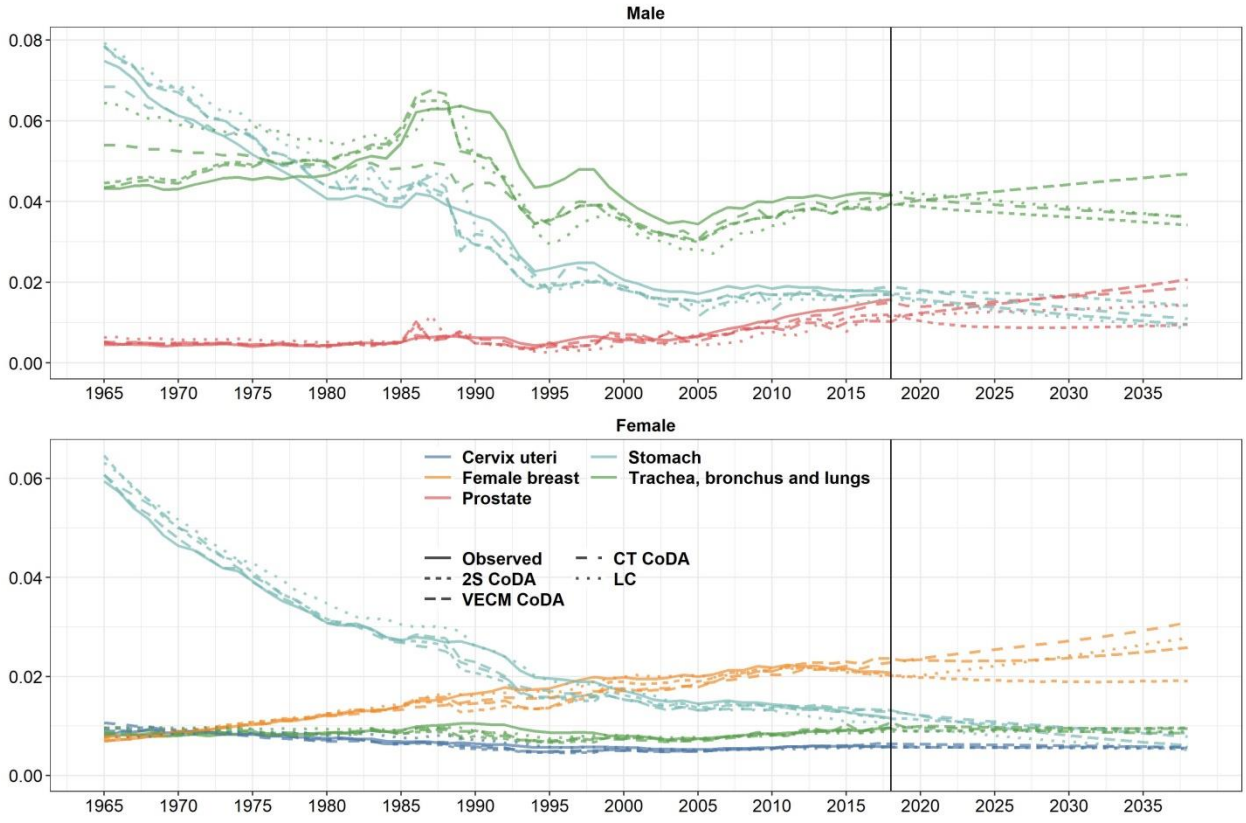

Figure 11 Twenty year out of sample forecast of life-table death densities for chosen groups of cancers for males and females in Russian Federation. Black vertical line indicates the beginning of forecast period. **Source:** Own calculations based on data from *RosBRiS*.