UDC 519.8:614.39

Kopecká Lucie,
Ing., Ph.D. Student, Department of Mathematics and Quantitative Methods, Faculty of Economics and Administration, University of Pardubice, Czech Republic

BAYESIAN ESTIMATES OF PROBABILITY OF INCIDENCES AND MORTALITIES OF SELECTED SERIOUS DISEASES

Abstract. *Oncological diseases are really serious, because they are already occurring in childhood and their number is growing rapidly. This problem is very dangerous not only for the European population, but also for the population all over the world. The Bayesian theory of credibility, in particular the binomial/beta model, can be appropriate method to determine the probability of incidences and mortalities, especially in cases of insurance companies. The main aim of this article is to evaluate these probabilities using the binomial/beta model and compare the advantages of Bayesian estimates to the maximum likelihood estimations based on trachea, bronchus and lung cancers in the Czech Republic and Ukraine. Data are obtained from the databases of WHO and OECD.*

Key words: estimation methods, binomial/beta model, trachea, bronchus and lung cancers.

Копецка Луція,
інженер, аспірант, кафедра математики та кількісних методів, факультет економіки та управління, Університет Пардубіце,Чеська Республіка

БАЙЄСІВСЬКА ОЦІНКА ЙМОВІРНОСТІ ЗАХВОРЮВАНОСТІ ТА СМЕРТНОСТІ ВІД ОБРАНИХ НЕБЕЗПЕЧНИХ ХВОРОБ

Анотація. *Онкологічні захворювання є справді небезпечними, оскільки вони вже з'являються в дитинстві і їх кількість стрімко зростає. Ця проблема є не тільки небезпечною для населення Європи, але й для всього світу загалом. Байєсівська теорія ймовірності, зокрема біноміальна / бета-модель, може бути методом, придатним для визначення ймовірності випадків захворюваності та смертності, особливо стосовно діяльності страхових компаній. Основна мета цієї статті - оцінити ці ймовірності за допомогою біноміальної / бета-моделі та порівняти переваги байєсівських оцінок з оцінками максимальної ймовірності на основі даних про захворюваність та смертність на рак трахеї, бронхів та легень у Чеській Республіці та Україні. Показники було отримано з баз даних ВООЗ та ОЕСР.*

Ключові слова: методи оцінки, біноміальна/бета-модель, трахея, рак бронх та легень.

Statement of the problem. Serious oncological diseases belong to the diseases which affect population all over the world. Incidences of these diseases have tendencies to grow, specifically in many European countries, which can be cause by factors such as consumption of alcohol and tobacco, obesity, stress, age structure of population, availability of health care, socio-economics situation, etc.

It is important to find right method of estimations, mainly in case of institutions such as insurance companies because these companies should be able to estimate their risks correctly to be able to fulfil their obligations.

Analysis of the previous publications. As mentioned above, oncological diseases are caused by certain risk factors. For instance, these publications are focused on revealing factors which influence health of population [2], [10], [12].

Many publications describe possibilities of estimating of unknown parameters of probability distributions. For details see [1], [3], [4], [5], [7], [8], [9], [11] and [13].

Jindrová & Kopecká [3], Kopecká & Pacáková [4] and Pacáková & Kotlebová [11] focus on using Bayesian estimates, specifically binomial/beta model for insurance companies.

Aim of research. The main aim of this article is to estimate probability of incidences and mortalities by using binomial/beta model and to focus on comparison of advantages of Bayesian estimates in comparison with maximum likelihood estimations (MLEs) based on data

related to trachea, bronchus and lung cancers in the Czech Republic (CZE) and Ukraine (UKR).

Presentation of the main part of the research. Data which are used in this article come from two databases, namely database of World Health Organization (WHO) and Organisation for Economic Co-operation and Development (OECD). Specifically, data about incidences of trachea, bronchus and lung cancers for Czech Republic (CZE) and Ukraine (UKR) are obtained from database of WHO for period 1985-2014 and data related to mortalities of the same cancers for the same two countries are obtained for period 1985-2015. Next OECD database provides data about incidences of lung cancer in 2012 and mortalities of trachea, bronchus and lung cancers in 2015. These years are the latest available years for OECD countries. Both databases provide data related to incidences and mortalities of these cancers per 100 000 population (n = 100 000).

MLE belongs to the most often possibility how to estimate unknown parameters of probability distributions. This method of estimation of parameters is one of the best because of its advantageous properties. For instance, MLE has asymptotically normal distribution, is asymptotically unbiasedness, asymptotic efficient, consistent and invariant as describes [7].

On the other hand, MLE has some serious disadvantages. The most serious disadvantage is complicated calculation process. In case of some probability distribution it must be used sophisticated programs. Next disadvantage is the necessity of the large sample to be reached of the advantages of MLE mentioned above. Another disadvantage of MLE it is considered to use information only from one data source. However, the last mentioned disadvantage belongs to the more often disadvantages of classic methods of parameters estimates. For details see [3], [4], [7] and [11].

According to Pacáková [7] MLE of unknown parameter θ is vector Θ which maximizes the likelihood function $L(\Theta, x)$ where Θ is vector of unknown parameters and x is vector of random sample specified by density function $f(x, \Theta)$. In case of binomial distribution according to formula (1), MLE of parameter θ is:

$$\hat{\theta} = \frac{x}{n} \tag{1}$$

where x represents number of incidences or mortalities and n is state of population in this article.

Next possibility how to estimate unknown parameters of probability distribution is to use Bayesian theory of credibility. Bayesian theory of estimation is bases for Bayesian theory of credibility. The fundamental difference between classic estimates for instance MLE and Bayesian estimates is unknown, estimated parameter θ which is in case of classic estimates considered as a unknown constant but in case of Bayesian estimates the parameter θ is considered as random variable with own probability distribution.

Next Bayesian estimates don´t only include data from own resources as in case of method of classic estimates but they include data from foreign comparable risks (prior information) which are known before information from own resources. Sometimes, no information exists about prior distribution of estimated parameter so it must be considered that each value of estimated parameter is equally probable. This situation doesn´t reflect reality but it doesn´t happen very often. On the other hand poor prior information is gradually improved by information from own resources as describe [7].

Formula (2) shows posterior probability density function which combines prior information and information from random sample (data from own resources). For details see [5] and [7].

$$f(\theta/x) \propto f(x/\theta) \,.\, f(\theta) \tag{2}$$

Pacáková [7] describes important concept of statistical induction, namely conjugate prior distribution. When random sample comes from probability distribution R with unknown estimated parameter θ it is considered that probability distribution of type F is conjugate distribution for distribution R if prior distribution leads to posterior distribution in the same type but with different parameters. For instance beta distribution is conjugate prior distribution for binomial distribution with unknown estimated parameter θ. This model is suitable for using in insurance companies and it is called binomial/beta model. Problematics related to binomial/beta model is demonstrated for example in following publications: [3], [4], [5], [7], [8], [9], [11] and [13].

Within binomial/beta model formula (3) expresses prior probability density function called prior beta distribution with parameters α and β. Prior probability density function is distribution of estimated parameter θ which is parameter of binomial distribution mentioned above.

$$f(\theta) \propto \theta^{\alpha-1} \,.\, (1-\theta)^{\beta-1}, 0 < \theta < 1 \tag{3}$$

First prior information of estimated parameter θ has to be determined. The best way of determination of prior beta distribution parameters α and β is to employ basic characteristic of prior beta distribution, namely mean value μ (4) and dispersion σ^2(5).

$$\mu = \frac{\alpha}{\alpha+\beta} \tag{4}$$

$$\sigma^2 = \frac{\alpha \,.\, \beta}{(\alpha+\beta)^2 \,.\, (\alpha+\beta+1)} \tag{5}$$

Next formula (6) shows that random sample X has binomial distribution with estimated parameter θ but for our needs the constant $\binom{n}{x}$ is omitted.

$$\text{ч} f(x/\theta) \propto \theta^x \,.\, (1-\theta)^{n-x}, x = 0,1, \dots, n \tag{6}$$

As mentioned above posterior probability density function combines prior information and information from random sample. In this case posterior probability density function is called posterior beta distribution, see formula (7).

$$f(\theta/x) \propto \theta^{x}.(1-\theta)^{n-x}.\theta^{\alpha-1}.(1-\theta)^{\beta-1} = \theta^{\alpha+x-1}.(1-\theta)^{\beta+n-x-1} \quad (7)$$

According to formula (8) Bayesian estimate of estimated parameter θ is mean value of posterior beta distribution in case of minimizing squared loss function.

$$\theta_B = \frac{\alpha+x}{\alpha+\beta+n} \quad (8)$$

Formula (8) can be rewrite into credibility formula (9) according to [1], [4] and [9]. Credibility formula includes credibility factor $Z = \frac{n}{\alpha+\beta+n}$ which shows degree of reliability of random sample, $\frac{x}{n}$ is maximum likelihood estimation of parameter θ of binomial distribution which is used within this model as well and μ which is explained in formula (4).

$$\theta_B = Z.\frac{x}{n} + (1-Z).\mu \quad (9)$$

Now construction of maximum likelihood estimations and Bayesian estimates, specifically binomial/beta model is described based on data about incidences and mortalities of trachea, bronchus and lung cancers. Next comparison of these two estimation methods is indicated and graphically illustrated. Finally, development of these diseases is observed in two European countries, namely CZE and UKR in period 1985-2015 in case of incidences and in period 1985-2016 in case of mortalities.

As mentioned above, the first of all, it is necessary to determine prior information about estimated parameter θ of binomial distribution by employing basic characteristics of prior beta distribution. It means that the parameters of prior beta distribution (formula 3) are determined by solving system of equations formula (4) and (5). The same procedure of determination of prior information is used in publication [4].

Parameters of prior beta distribution are determined separately for incidences and mortalities by using data from OECD database. These database provides data about incidences of lung cancer in 2012. By solving system of equations mentioned above parameters $\alpha_{1985} = 13{,}38$ and $\beta_{1985} = 45\,967{,}74$. After that OECD database provides data about mortalities of trachea, bronchus and lung cancers in 2015. Parameters are obtained in the same way. Parameters $\alpha_{1985} = 14{,}44$ and $\beta_{1985} = 34\,874{,}91$.

After determination of prior beta distribution parameters Bayesian estimate of estimated parameter θ (probability of incidences of trachea bronchus and lung cancers) can be constructed for the first available year 1985. For the first estimated year Bayesian estimate is mean value of prior beta distribution formula (4), specifically $\theta_{B,1985} = 0{,}000291$. This prior information is used for both mentioned countries.

Next parameters of beta distribution must be recalculate for the following year 1986. It means that $\alpha_{1986} = \alpha_{1985} + x_{1985} = 13{,}38 + 55 = 68{,}38$; $\beta_{1986} = \beta_{1985} + n_{1985} - x_{1985} = 45\,967{,}74 + 100\,000 - 55 = 145\,912{,}7$ in case of CZE and $\alpha_{1986} = \alpha_{1985} + x_{1985} = 13{,}38 + 44 = 57{,}38\beta_{1986} = \beta_{1985} + n_{1985} - x_{1985} = 45\,967{,}74 + 100\,000 - 44 = 145\,923{,}7$ in case of UKR. In 1986 the Bayesian estimate equals to mean value of posterior beta distribution, see formulas (7) and (8).

In Fig. 1 we can see development of MLE and Bayesian estimates of probability of incidences of trachea, bronchus and lung cancers in CZE. In this Fig. 1 the advantages of Bayesian estimates are displayed in contrast with MLE.

.

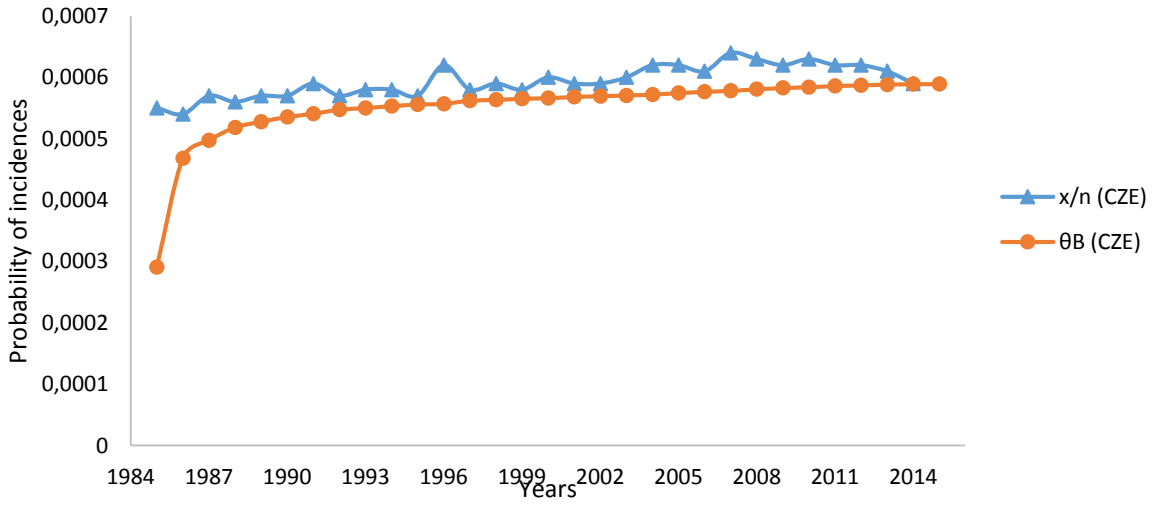

Fig. 1: Comparison of MLE and Bayesian estimates of probability of incidences of trachea, bronchus and lung cancers in CZE *(own calculation based on [6], [14])*

First advantage of Bayesian estimates is possibility of involving data from two databases. Bayesian estimate is determined from OECD database in 1985 in contrast with MLE which is only stated from database of WHO. For the first year Bayesian estimate is very low in contrast with MLE but this poor prior information is gradually improved by employing information from WHO database. This is the reason why the estimates are still closer. Next advantage of Bayesian estimates is smoother course of these estimates than in case of MLE and finally Bayesian estimates can be determined for the year following after the last known year (2015).

In Fig. 2 MLEs are displayed separately for CZE and UKR for period 1985-2014.

We can see that probability of incidences of trachea, bronchus and lung cancers is lower in case of UKR than in case of CZE over the whole period. On one hand trend of MLE is increasing within CZE in contrast with UKR where the trend of MLE is decreasing. It means that CZE is not only more affected by these diseases but the incidences are still higher. On the other hand Ukrainian population is in a better position in contrast with CZE.

In Fig. 3 Bayesian estimates are displayed again separately for CZE and UKR for period 1985-2015.

As mentioned above Bayesian estimate is determined based on prior information which is obtained from OECD countries (OECD database) for the first year (1985). We can see that this estimate of incidences of trachea, bronchus and lung cancers has closer to estimates of incidences in UKR than in CZE. It means that CZE and UKR belong to the countries which are more affected by these diseases in comparison with mean of OECD countries.

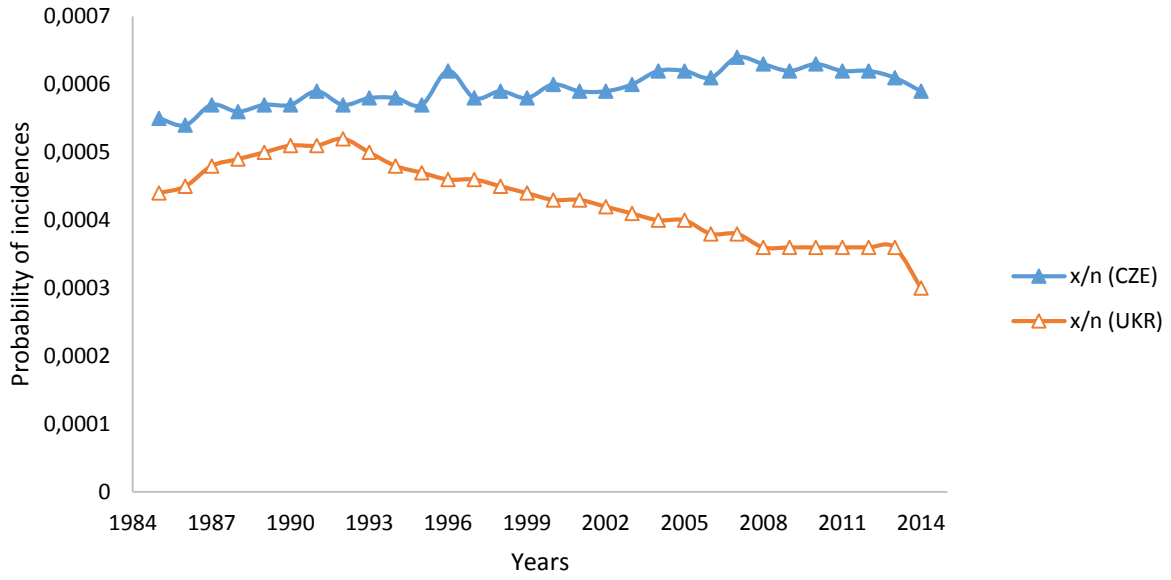

Fig. 2: Comparison of MLEs of probability of incidences of trachea, bronchus and lung cancers for CZE and UKR *(own calculation based on [6], [14])*

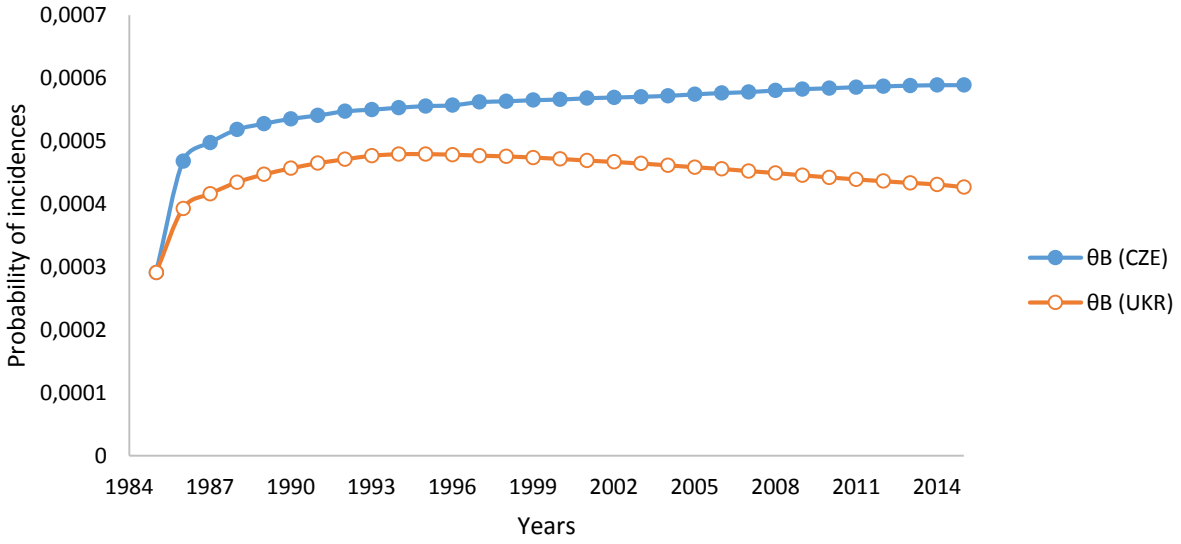

Fig. 3: Comparison of Bayesian estimates of probability of incidences of trachea, bronchus and lung cancers for CZE and UKR *(own calculation based on [6], [14])*

As in case of incidences after determination of prior beta distribution parameters Bayesian estimate of estimated parameter θ (probability of mortalities of trachea bronchus and lung cancers) can be constructed for the first available year 1985. For the first estimated year Bayesian estimate is mean value of prior beta distribution formula (4), specifically $\theta_{B,1985} = 0{,}000414$. This prior information is used for both mentioned countries. Calculation process of Bayesian estimates is the same as in case of incidences mentioned above.

In Fig. 4 we can see MLEs which are displayed separately for CZE and UKR for period 1985-2015.

We can see that probability of mortalities of trachea, bronchus and lung cancers is lower in case of UKR than in case of CZE again over the whole period. However, the trends of mortalities are decreasing in case of both countries. It means that cure of these diseases is more and more effective. Ukrainian population is in a better position again, which can be connected with lower incidences of these diseases.

In Fig. 5 Bayesian estimates are displayed again separately for CZE and UKR for period 1985-2016.

We can see that CZE has still closer to the estimate of mortalities of trachea, bronchus and lung cancers in comparison with mean of OECD countries and UKR is in better situation than OECD countries.

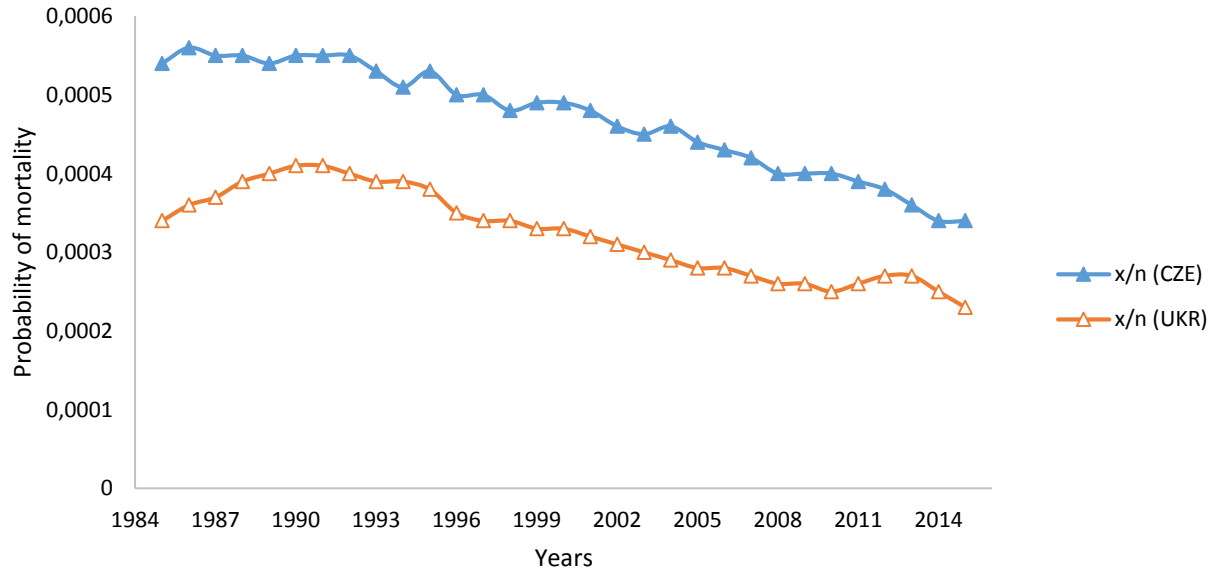

Fig. 4: Comparison of MLEs of probability of mortalities of trachea, bronchus and lung cancers for CZE and UKR *(own calculation based on [6], [14])*

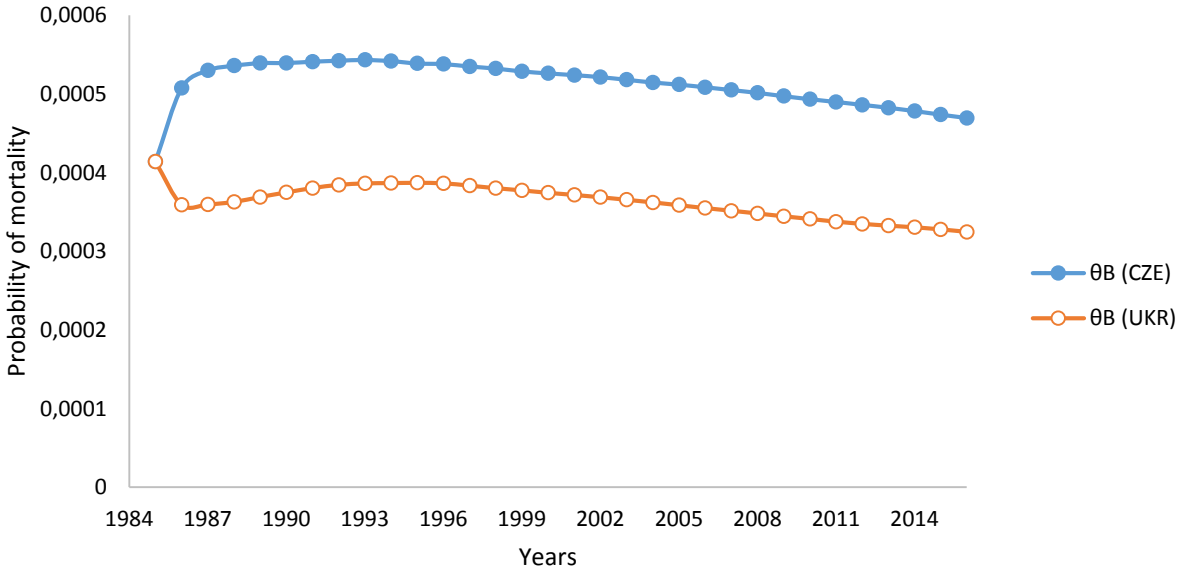

Fig. 5: Comparison of Bayesian estimates of probability of mortalities of trachea, bronchus and lung cancers for CZE and UKR *(own calculation based on [6], [14])*

Conclusion and perspectives of future investigations. The main aim of this article was to estimate probability of incidences and mortalities of trachea, bronchus and lung cancers by using binomial/beta model and to focus on comparison of advantages of Bayesian estimates in comparison with MLE in the CZE and UKR. Bayesian estimates appear to be suitable because of their advantages, mainly in case of insurance companies because these companies don´t have to only use data from own resources but they can focus on data from foreign comparable risks, which is one of reasons for employing Bayesian estimates.

Despite the fact that CZE is member country of OECD and EU, UKR has better results of probability of incidences and mortalities of trachea, bronchus and lung cancers which belong to the most serious cancers in Europe. Both countries show decreasing trends in case of mortalities. However in case of UKR this decrease is slower than in CZE. On the other hand CZE shows increasing trend in case of incidences of these diseases in contrast with UKR which has decreasing trend of incidences. In case of incidences both countries move above the mean of OECD countries and in case of mortalities CZE is moving above this mean but it is still closer to mean value.

The most of causes causing incidences or mortalities of oncological diseases are detected. Nevertheless the main causes making disparities in development of oncological disease among individual countries are not clear and this is the reason for the next research in this field.

LITERATURE

1. Gogola, J. (2013). Spôsob permanentnej úpravy výšky poistného v neživotnom poistení. *E+M Ekonomie a Management, 4*/2013, str. 134-142.

2. Jindrová, P., Kopecká, L. (2017a). Assessment of risk factors of serious diseases in OECD countries. *Ve:* Proceedings 11. Międzynarodowa Konferencja Naukowa im. Profesora Alexandra Zeliasa na temat „Modelowanie i prognozowanie zjawisk spoleczno-gospodarczych. Krakow: Uniwersytet Ekonomiczny w Krakowie, str. 123-132.

3. Jindrová, P., Kopecká, L. (2017b). Kvantifikace rizik pro úrazové pojištění. *Scientific Papers of the University of Pardubice – Series D. Univerzita Pardubice: Fakulta ekonomicko-správní. 39*(XXIV), str. 75-86.

4. Kopecká, L., Pacáková, V. (2017). Bayesian Estimation of Probability of Incidences of the Most Serious Oncological Diseases in the Czech Republic. *Ve Finanční řízení podniků a finančních institucí: Sborník 11. Mezinárodní conference Finanční řízení podniků a finančních institucí. Ostrava. VŠB – Technická Univerzita Ostrava. str*. 407-414.

5. Kotlebová, E. (2009). *Bayesovská štatistická indukcia v ekonomických aplikáciách*. Bratislava: EKONÓM.

6. OECD. (2017). *OECD statistics*. Dostupné z http://stats.oecd.org/.

7. Pacáková, V. (2004). Aplikovaná poistná štatistika. 3. vyd. Bratislava: IURA EDITION.

8. Pacáková, V. (2012). Bayesian Estimations in Insurance Theory and Practice. V: *Advances in Mathematical and Computational Methods*. *Proceedings of the 14th WSEAS International Conference on Mathematical and Computational Methods in Science and Engineering (MACMESE'12), Sliema, Malta,* str. 127-131.

9. Pacáková, V. (2013). Credibility models for permanently updated estimates in insurance. *International Journal of Mathematical Models and Methods in Applied Sciences, 7*(3), str. 333-340.

10. Pacáková, V., Jindrová, P., Zapletal, D. (2016). Comparisons of Public Health Systems of Selected Countries. *V: Evropské finanční systémy 2016. Sborník 13. Mezinárodní vědecké konference. Brno: Masarykova Univerzita,* str. 534-541.

11. Pacáková, V., Kotlebová, E. (2014). Bayesian Estimation of Event Probability in Accident Insurance. *V: Evropské finanční systémy 2014: Sborník 11. Mezinárodní vědecké conference. Brno: Masarykova Univerzita, str.* 462-468.

12. Pacáková, V. & Papoušková, M. (2016). Multidimensional Comparisons of Health Systems Functioning in OECD Countries. *International Journal of Mathematical Models and Methods in Applied Sciences. 10*(12), str. 388-394.

13. Šoltés, E. (2009). *Modely credibility na výpočet pojistného*. 1. vyd. Bratislava: EKONÓM.

14. WHO. (2018). *European Health Information Gateway*. Dostupné z: https://gateway.euro.who.int/en/

REFERENCES

1. Gogola, J. (2013). Spôsob permanentnej úpravy výšky poistného v neživotnom poistení. *E+M Economics and Management, 4*/2013, pp. 134-142.

2. Jindrová, P., Kopecká, L. (2017a). Assessment of risk factors of serious diseases in OECD countries. In: *Proceedings of the 11th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*. Cracow: Cracow University. pp. 123-132.

3. Jindrová, P., Kopecká, L. (2017b). Kvantifikace rizik pro úrazové pojištění. *Scientific Papers of the University of Pardubice – Series D. University of Pardubice: Faculty of Economics and Administration*. 39(XXIV), pp. 75-86.

4. Kopecká, L., Pacáková, V. (2017). Bayesian Estimation of Probability of Incidences of the Most Serious Oncological Diseases in the Czech Republic. *In Financial Management of Firms and Financial Institutions: Proceedings of the 11th International Conference Financial Management Firms and Financial Institutions. Ostrava: VŠB-Technical University of Ostrava*. pp. 407-414.

5. Kotlebová, E. (2009). *Bayesovská štatistická indukcia v ekonomických aplikáciách*. Bratislava: EKONÓM.

6. OECD. (2017). *OECD statistics*. Available from http://stats.oecd.org/

7. Pacáková, V. (2004). Aplikovaná poistná štatistika. 3rd edition. Bratislava: IURA EDITION.

8. Pacáková, V. (2012). Bayesian Estimations in Insurance Theory and Practice. In: *Advances in Mathematical and Computational Methods*. Proceedings of the 14th WSEAS International Conference on Mathematical and Computational Methods in Science and Engineering (MACMESE'12), Sliema, Malta, pp. 127-131.

9. Pacáková, V. (2013). Credibility models for permanently updated estimates in insurance. *International Journal of Mathematical Models and Methods in Applied Sciences*, *7*(3), pp. 333-340.

10. Pacáková, V., Jindrová, P., Zapletal, D. (2016). Comparisons of Public Health Systems of Selected Countries. In: *European Financial Systems 2016*. Proceedings of the 13th International Scientific Conference. *Brno: Masaryk University,* 534-541.

11. Pacáková, V., Kotlebová, E. (2014). Bayesian Estimation of Event Probability in Accident Insurance. *In European Financial Systems 2014: Proceedings of the 11th International Scientific Conference. Brno: Masarykova Univerzita*. pp. 462-468.

12. Pacáková, V. & Papoušková, M. (2016). Multidimensional Comparisons of Health Systems Functioning in OECD Countries. *International Journal of Mathematical Models and Methods in Applied Sciences*. *10*(12), pp. 388-394.

13. Šoltés, E. (2009). *Modely credibility na výpočet pojistného*. 1st edition. Bratislava: EKONÓM.

14. WHO. (2018). *European Health Information Gateway*. Available from https://gateway.euro.who.int/en/