A FEW NOTES ON THE LEXIS DIAGRAM: THE 100th ANNIVERSARY OF THE DEATH OF WILHELM LEXIS[1)]

Klára Hulíková Tesárková, Olga Kurtinová

ABSTRACT

This paper provides a brief introduction to Wilhelm Lexis, his life, and the basic demographic tool that is named after him: the Lexis diagram. This topic is chosen to commemorate the centenary of the death of Wilhelm Lexis, who had an unquestionable influence on demographic research. The Lexis diagram, which is used to display demographic events in a dual time dimension, is an essential instrument for working correctly with demographic data. Therefore, it is not far from the truth to claim that demographers all around the world would be unable to imagine demography without the Lexis diagram, in spite of the fact that the diagram's ties to Lexis are not so direct.

Keywords: Wilhelm Lexis, lexis diagram, demographic model, demographic events, time

Demografie, 2014, **56: xxx–xxx**

1. INTRODUCTION

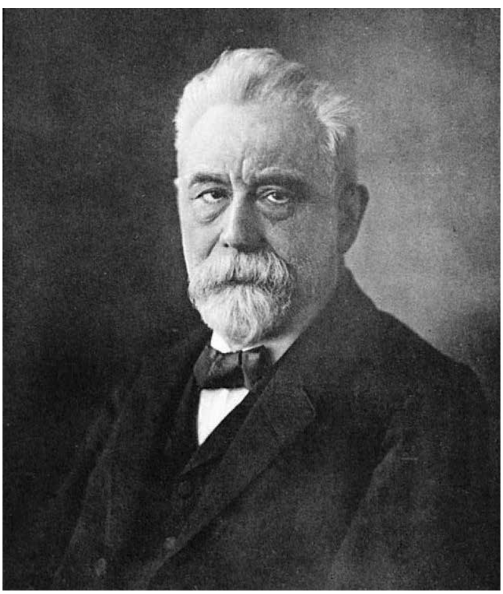
Wilhelm Lexis left an important legacy for demography in the form of what is today known as the Lexis diagram. The diagram allows us to deal, according to their structure or aggregation, with demographic events such as births, deaths, marriages, etc. in two basic time dimensions, age and calendar time (period), where a third dimension (cohort) can also be traced. It became popular above all in the second half of the 20th century due to the work of Pressat (*Caselli et al.*, 2005). Although the Lexis diagram is taken for granted in our research nowadays, the path to the diagram in the form in which it is used today was not so straightforward. It was quite the other way around and it may have escaped the notice of some demographers that the diagram we use widely today may not be entirely identical to the diagram proposed by Wilhelm Lexis. The roots of this graphic tool can be found in the work of other demographers and scientists in the second half of the 19th century, but, as we will show, the role of Lexis himself was very important. Therefore, the aim of this paper is to commemorate the 100th anniversary of Wilhelm Lexis's death and to introduce briefly the documented evolution of the Lexis diagram. The goal of this paper is also to demonstrate how the most common diagrams are used and how the probabilities and rates are calculated using the diagram; that means, how the Lexis diagram and its various forms can be used in practice and what the advantages and disadvantages of these different forms are. It is impossible within the scope of this brief paper to devote enough space to all the known variants of the diagram and detailed specifications of it. Therefore, the two versions of the diagram used in the Czech school of demography were selected for this article.

1) This article was written with the support of the Czech Science Foundation as part of work on project no. P404/12/0883 'Cohort life tables for the Czech Republic: data, biometric functions, and trends/Generační úmrtnostní tabulky České republiky: data, biometrické funkce a trendy'.

2. WILHELM LEXIS

Eschweiler in Germany is a town located not far from Aachen, which is known for several notable figures who were born or lived there. Wilhelm Lexis[2)] is one of them. He was born on 17 July 1837 into the family of physician Ernst Joseph Lexis and his wife Gertrud Stassen. Given the range of his interests he was known as a Renaissance man. He first studied law at the University of Bonn. Later on, he devoted himself to mathematics and natural sciences there. He graduated from the University of Bonn in 1859 and earned his doctoral degree in philosophy for his thesis on analytical mechanics (*Johnson – Kotz*, 1997; *Drechsler – Kattel*, 1997). He obtained a degree in mathematics also and for some time he taught mathematics at a gymnasium in Bonn and worked as a librarian and a journalist. Despite his background, Lexis also obtained experience in Bunsen chemical laboratory in Heildeberg (*Hertz*, 2001). Lexis married Pauline Emilie Lindenberg from Remscheid and had three children (*von Collani*, 2014).

Figure 1 Wilhelm Lexis



Source: Wikipedia: The Free encyclopaedia, 2014a.

In 1861 Wilhelm Lexis left for Paris, where he started to study the social sciences and political economy. This departure marked a turning point in his life and professional career, because it led to Lexis's first publication devoted to French export policies (*Die französischen Ausfuhrprämien im Zusammenhange mit der Tar-ifgeschichte und Handelsentwicklung Frankreichs seit der Restauration: volkswirthschaftliche Studien*, Bonn 1870). He pointed out that economic theory should base on quantitative data and expressed scepticism of 'pure economics' utilising only descriptive mathematical models (*Heiss*, 2014). The Franco-German War (1870–1871) forced him to return to Germany where later he became involved in the reform of German universities. In 1872 Lexis was appointed to the post of associate professor in political economy at the University of Strasbourg, which had just been established. At that time he published an introduction to the theory of statistical demography (*Einleitung in die Theorie der Bevölkerungsstatistik*, Strasbourg 1875), in which he included his well-known diagram. Lexis then left Strasbourg for Dorpat[3)] in the Russian Empire, where he was appointed as professor of geography, ethnology, and statistics. At the University of Dorpat, which belonged to the group of German-language universities till 1895, he spent only two years there and published a demographic essay on the sex ratio of births and probability theory (*Das Geslechtsverhältniss der Geborenen und die die Wahrscheinlichkeitsrechnung in: Jahrbuch für Nationalökonomie und Statistik*, 1876). In the following years, from 1876 to 1884, Lexis took up the position of chair of political economy at the University of Freiburg im Breisgau in Baden-Württemberg. Considering the list of his publications, this was a fruitful period in his career. Around this time he published works on statistics (e.g. *Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft*. Freiburg im Breisgau 1877; *Über die Theorie der Stabilität statistischer Reihen, in Jahrbch für Nationalökonomie und Statistik* 1879) and articles on economics (e.g. *Gewerkvereine und Unternehmerverbände in Frankreich: ein Beitrag zur Kenntniss der socialen Bewegung*, Leipzig 1879)

2) Full name: Wilhelm Hector Richard Albrecht Lexis.

3) Now Tartu in Estonia.

(*Drechsler – Kattel*, 1997; *Seneta*, 2014; *Heiss*, 2014). In 1884 Lexis was appointed for the chair of statistics at the University of Breslau[4] and remained there till 1887. Afterwards he moved to work at the University of Göttingen and from 1891 he became the chief editor of *Jahrbuch für Nationalökonomie und Statistik*. He died in Göttingen a few days after the beginning of the First World War, on 24 August 1914.

Wilhelm Lexis is famous for his contribution to theoretical statistics focusing on the application of the calculus of probabilities to statistical data. His pioneering work on dispersion resulted in a study of variance and his rejection of the assumption of statistical homogeneity in sampling supported by Adolphe Quetelet,[5] who is known as the author of the rules for modern population censuses, and his followers. Unlike them, Lexis stressed the fluctuations in different demographic time series and proposed a dispersion coefficient Q[6] as a ratio of the empirical variance of the series to the assumed theoretical variance. 'Normal' dispersion corresponds to Q equals 1 and is related only to chance. If the coefficient differs from 1, then fluctuations will be caused by the 'physical' component rather than chance. $Q > 1$ denotes, in his concept, *hypernormal dispersion* and $Q < 1$ denotes *hyponormal dispersion*. He also demonstrated that social data usually have hypernormal dispersion. The coefficient Q, which he discovered while studying qualitative changes in populations over time, was a forerunner of the statistics of K. Pearson and R. A. Fisher and the analysis of Chi-squared variance (*Seneta*, 2014; *Johnson – Kotz*, 1997).

In addition, Wilhelm Lexis was a scholar in the field of economics and finance and published several articles on the educational system (e.g. *Die neuen französischen Universitäten*, Munich, 1901; *Die Reform des höheren Schulwesens in Preussen*, Halle, 1902). In addition, he contributed several papers to the field of tuberculosis statistics (*Zur Statistik der Tuberkulose, Bericht über die Statistik der Tuberkulose*) and established the first German actuarial institute (*Königliches Seminar für Versicherungswissenschaften*) (*de Gans – van Poppel*, 2000).

3. THE LEXIS DIAGRAM AND ITS DEVELOPMENT

The Lexis diagram can, according to *Siegel* (2012, p. 945), be defined as '[a] *graph relating time and age and thus illustrating how, with the passage of time, the age references of a birth cohort change. If the diagram is shown in three dimensions, actual population values may be depicted.*' The importance of the Lexis diagram can be expressed by the fact that in demography it is often important to follow all three dimensions or to make the correct choice of methodological approach to the analysis (i.e. selecting the primary set to which import data are aggregated). In a certain point of view, the fact that all the dimensions (time/period, age and cohort/time of birth) are logically connected and tied simplifies a great deal, but on the other hand it could lead to some methodological problems (e.g. in the age-period-cohorts models). The interconnection of the three dimensions was expressed by *Keiding* (2011: 405): *'In demography and epidemiology, a central issue in studying fertility, morbidity, and mortality, is to keep track of the three different time variables age x of the individual, calendar time t at birth (cohort), and current calendar time $\tau = t + x$, often termed period.'*

Among demographers it is well known that graphical representation can be useful for answering some methodological question as well as for obtaining an initial orientation in data and also for presenting results. *'To track the simultaneous development of age, period, and cohort it is helpful to use graphs, nowadays usually the so-called Lexis diagram which is just a (time, age)-coordinate system in which individuals are represented by line segments of slope 1 starting at (time at birth, 0) and ending at (time at death, age at death).'* (*Keiding*, 2011: 406) However, the Lexis diagram, as mentioned by *Keiding* (2011), had its own development, and moreover, it cannot solely be associated with the name of Lexis. On the other hand, the diagram Leis proposed (see below) was pioneering in some aspects that are still important today, and it was also used in practice, above all before the 1960s, when Pressat introduced his modified version (see below).

4) Now Wroclaw in Poland.

5) His full name was Lambert Adolphe Jacques Quetelet.

6) His full name was Lambert Adolphe Jacques Quetelet.

The first works dealing with the graphical representations of demographic data often also included the first formal representations – usually related to the process of mortality. Those works could be traced to approximately the last quarter of the 19th century (*Keiding*, 2011). *Zeuner* (1869) was probably the first to deal with the issue of the graphical representation of population development in his work (*Keiding*, 2011). He focused on the possible ways of graphically expressing the life history of various cohorts, or simply speaking of groups of people born in a defined time period. Based on the graphs he developed, he defined the survival function as we know it today and many other relations. The basic equation behind it is the relation expressing the number of individuals born in the time interval from t_1 to t_2 who are still alive at age x (*Zeuner*, 1869: 12; *Keiding*, 2011: 406):

$$V(x) = \int_{t_1}^{t_2} f(t, x)\mathrm{d}t$$

In his work, *Zeuner* (1896) proposed a way in which to express the life history of a population graphically. He worked with a system of axes in a three-dimensional space, where the two horizontal axes stand for age and the time of birth and the vertical axis represents the number of individuals. The most important curves in the coordinate system illustrated the mortality process, which means the population size decreasing by mortality. The system of axes in this way is more useful for the study of birth cohorts. For the study of the population structure in a defined time, it is necessary to use the diagonal horizontal lines representing time (see Figure 2).

Moreover, in the diagram introduced above (the Zeuner sheet, Fig. 2) *Zeuner* (1869, Figures 4–6 in his original work) also showed the three primary sets of deaths defined by:

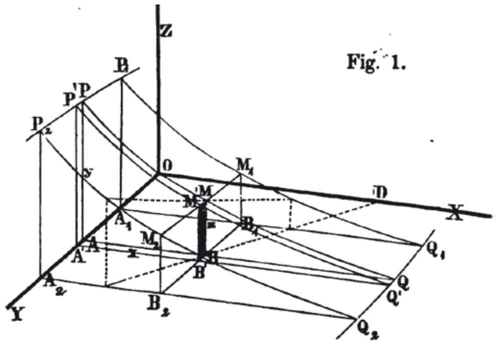
1) cohort and age (the 1st primary set of events),

2) cohort and period (the 2nd primary set of events) and

3) period and age (the 3rd primary set of events).

Those sets of deaths were found by projecting the corresponding change in 'mortality curves' to the bottom side of the graph. This means that the corresponding change in 'mortality curves' was bordered by cohort, age or period, depending on the selected primary set.

Figure 2 **Graphical representation of the survival curve according to *Zeuner* (1869) – the 'Zeuner Sheet'.**



Note: 'Vertical axis (Z) stands for the number of individuals born in time Y or still alive at age X. The horizontal lines denoted as P_1P_2 in the figure are called the 'birth curves' (*Keiding*, 2011: 407), which are the individuals who were born in the time period from Y = A_1 to A_2. The decreasing curves PQ represent the 'mortality curves', or survival functions in contemporary terminology. For each age X then it is easy to find the number of individuals who are still alive (curve BM in the Figure 2) (*Keiding*, 2011: 407). The dashed diagonal line BD represents the defined constant time.
Source: Zeuner, 1869: 9.

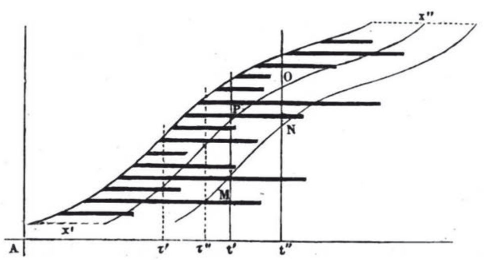
In his work, Zeuner (1869, cit. in *Keiding*, 2011) argues that in statistics or registers it is necessary also to collect information about the year of birth of a person, and not only the age and the year of death. In other words, using only the information of year of the event (death) and age, we cannot do any cohort analysis or define unambiguously an individual's cohort (the time or year of birth).

Based on the above, it is clear that Zeuner focused especially on graphical representations of the whole population and changes to the population in time or with age. He also developed formal expressions for these changes. On the other hand, there is a different way in which to study the changes in a population, namely, the study of individual life lines. Knapp (1869, 1874 cit. in *Keiding*, 2011: 408) was probably the first to use the graphical representations of individual life lines and also the first to study the individual lengths of life as the main goal.[7] He simply drew individual

7) Vandeschrick (2001) states that Knapp proposed his version of the diagram already in 1868.

lines in a graph using the time scale (see Figure 3). These lines represented the human's life from birth to death. He also took into account the number of births (*Vandeschrick*, 2001). Based on this graphical representation, Knapp (1874, cit. in *Keiding*, 2011) developed the calculus usable for the mortality analysis. However, given the way it is shown in the diagram (Figure 3), it is clear that using this graphical representation in practice was not very straightforward.

Figure 3 **The life lines proposed by Knapp (1874)**

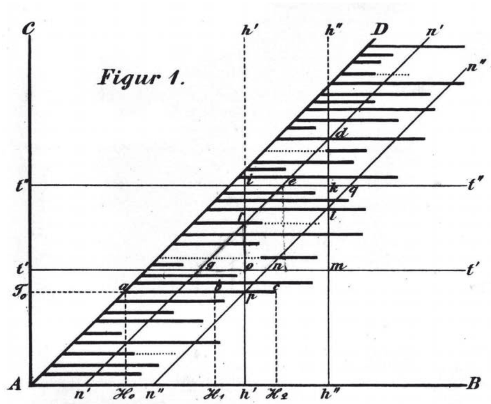


Note: *'This diagram has a line segment for each individual plotted against calendar time, ranging from birth to death of the individual. The right endpoints in the area MNOP represent the deaths during the time interval [t´;t´´] of individuals aged between x´ and x´´.'* (*Keiding*, 2011: 409)
Source: Keiding, 2011: 409.

In the diagrams above, in fact only two perspectives are used (time and age used by Knapp in 1874, or cohort and age in the diagram of Zeuner, 1869, which is combined with the population size). The other dimensions can be found in the diagrams, though they were not the main subject of study of their authors (like the time/period perspective in Zeuner, 1869, Figure 2).

Becker (1874) was probably the first to combine in some way all three perspectives – the time/period, cohort and age dimensions (*Keiding*, 2011). He followed the idea of Knapp (1874, cit. in *Keiding*, 2011), and the basic objects in the graph, as he proposed it, are the life lines. These are the horizontal lines, all of them start at the time of birth, and the age at birth is logically equal to zero. This means that age increases in a horizontal perspective going from left to right. In Figure 4 the constant values of age are represented by the diagonal curves, and the first of them (in Figure 4) is equal to age D = 0. Then the time of birth (cohort) can be read on the vertical axis and the calendar time on the horizontal axis. Because the time of birth has to be equal to the same calendar time, the diagonal curves have to be under a 45 degree slope (Figure 4).

Figure 4 **The period-cohort diagram proposed by Becker (1874)**



Note: 'Becker (1874) worked also with the possibility of emigration or immigration (the dotted lines in the diagram.
'The three primary sets of dead are indicated: (cohort, age) as the parallelogram eqng, (cohort, period) as the rectangle ikmo, (period, age) as the parallelogram dlpf.' (*Keiding*, 2011: 409)
Source: Becker, 1874: 77; Keiding, 2011: 409.

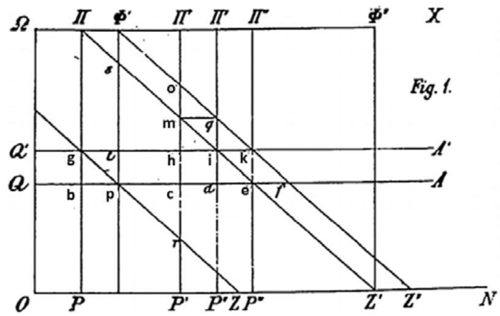
In Becker's diagram (1874) the ends of the life lines represent the event of death. Based on their occurrence in the diagram he was able to relatively easily define the three primary sets of events (deaths). In contrast to Zeuner (1869), Becker´s (1874) diagram is much easier to work with because it is defined as only a two-dimensional representation. The population size is not included in the diagram. This also means a greater possibility of this type of diagram being of practical use. This simple way of defining all three primary sets of events (deaths) could also be important today. However, when considering contemporary practical usage of this type of diagram, it must be mentioned that there is usually no need to draw the individual life lines, and moreover, particularly, for example, when studying the process of mortality, we need to express the sets of events for one time period or one cohort and to define the population size at risk of death. Drawing both these variables in Becker´s diagram would be quite spatially-demanding and it would probably be difficult for the user to gain an orientation in them (the deaths of persons in the youngest cohorts would be at the top of the diagram; age does not have its own axis, etc.).

Lexis (1875) studied (Lexis, 1880 cit. in *Vandeschrick*, 2001) the work of Zeuner independently (*Keiding*, 2011).[8] However, he did not work specifically with the life lines and their depiction in the diagram. The individual life histories or development of the population was not the main object in his study, which is obvious from his interests listed above. Instead of that he focused specifically on the graphical representation of the key information: the time of death (or other events), the age at the time of the event (the age at the time of death), and the time of birth of the studied individual (*Vandeschrick*, 2001). In his proposed diagram, the main feature is the possibility to read easily when a person (or group of people) was born (the time of birth/cohort – on the horizontal axis) and what is the age of the person/sub-population or age at the time of the event (the age at the time of death); he defined the 'birth points' and 'death points' and the lines separating them (*Keiding*, 2011). This feature also fully corresponds to the contemporary needs of demographic analysis and the data definitions we still need today. Instead of some time- or period- axis, he used diagonal lines with a slope of 135 degrees to represent fixed (constant) calendar time periods (e.g. moments of the census) (see Figure 5; *Vandeschrick*, 2001). However, if we wanted to draw a period a is in the diagram, it would lead from the bottom left to the upper right side of the diagram, diagonally, with a slope of 45 degrees (*Roubíček*, 1958).

It is theoretically possible to draw individual life lines in the Lexis (1875) diagram very easily. They would be vertical lines, all of them starting on the horizontal axis (age = 0), each would end at different age (could be read on the vertical axis), and the time of death would be bordered by the two following time lines. All the primary sets of events defined earlier could be used in this diagram and drawn again very easily (this will be described below in this text).

It could be said that this type of diagram could also be fully useful for contemporary needs of analysis – the lowest ages would be drawn in the bottom part of the graph (this means that the less certain higher ages do not need to be drawn in the diagram), the time lines precisely define the population size at the beginning as well as in the middle of the year (or more generally, in the studied time period; the time lines could also be drawn according to the defined studied time period, not necessarily for each calendar year), and all the primary sets of events could be drawn very easily. On the other hand, from Figure 5 it is already clear that this type of diagram would be especially suitable for cohort analysis – the cohorts are straight vertical lines, which can be easily followed across their history. However, in mortality (and the same holds also for most of the other demographic processes as well) we have to deal mostly with the period perspective. For example, if the studied time period should be one year of time, using the *Lexis* (1875) diagram (shown in Figure 5) we would need to define the calendar year with two time lines, which run from the bottom right to the upper left – the beginning of these time lines (the lowest ages) would lie at the extreme right

Figure 5 **The Lexis diagram (Lexis, 1875)**



Note: The primary set defined by cohort and age can be illustrated by a rectangle within the diagram (for example, rectangle bhcg). The parallelogram (for example, ekom) stands for the primary set defined by cohort and period and the parallelogram (for example, peig) represents the primary set defined by age and period (*Keiding*, 2011: 410).
Source: Lexis, 1875: 139; Keiding, 2011: 410, author´s modification of the letters in the figure.

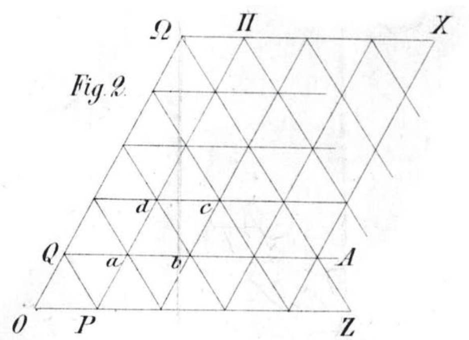
8) The construction and specific features of the Lexis diagram can be found in the work of Vandeschrick (2001).

of the graph; this method of graphical visualisation would be again relatively spatially-demanding. On the other hand, when illustrating the data structure only for the lowest ages, then the diagram is fully useful and practical (see Part 4 of this article).

However, *Lexis* (1875) noticed that not all the time dimensions (age, period and cohort) are represented symmetrically in the first proposed form of his diagram (described above, see Figure 5). If we consider the range of birth times (horizontal axis) as equal to 1 year, and the age interval (vertical axis) also as equal to 1 year, this means that both these lines would be of a length equal to 1. In this situation, the line that represents the time dimension (the distance between the two time lines in Figure 5) would be equal to $\frac{\sqrt{2}}{2}$. For this reason he proposed a slightly modified diagram (see Figure 6), an 'equilateral diagram' (*Keiding*, 2011: 410), where all the axes have the same scale. In this version of the diagram he 'only' changed the slope of the vertical axis (to 60 degrees). In the diagram itself then, not the axes, but the age-, time- (or period-) and cohort- (or time of birth-) lines are important. Those lines separate the different completed ages, time periods and cohorts (in today's terminology). In this modified Lexis diagram all the differences between two exact ages, periods or cohorts (which are all equal to the same value) are represented by equally long lines.

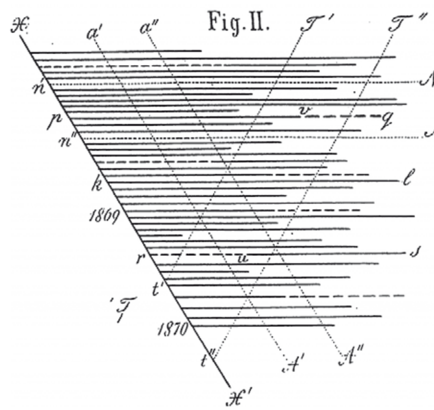
In this equilateral version of the Lexis diagram (Figure 6) it is also possible to draw life lines. This could be important for identifying the state of the population at the beginning of the year and in the middle of the year. These life lines would be parallel to cohort lines (a slope equal to 60 degrees). One year later, Lewin (1876, cit. in *Keiding*, 2011) continued to work with the Lexis diagram and added life lines to the diagram. At the same time, he also rotated the whole diagram so that the life lines run horizontally (see Figure 7). Moreover, he precisely defined the three primary sets of events as well as the two elementary sets defined by only one generation, age and calendar year. However, his work was criticised by his counterparts because of its unnecessarily excessive formal complexity (e.g. Zeuner, 1886 cit. in *Keiding*, 2011).

Figure 6 **Lexis 'equilateral' diagram (Lexis, 1875)**



Note: 'the horizontal lines represent age-lines (QA), the vertical lines running up and to the right represent the different times of birth, the cohort-lines (e.g. 0Q), and the vertical lines running up and to the left represent time-lines (PQ). Each age-line stands for an exact age, time-line and cohort-line corresponding to one exact moment in time.
Source: Lexis, 1875: 139; Keiding, 2011: 411.

Figure 7 **The diagram proposed by Lewin (1876)**



Note: 'The life lines are horizontal and parallel with the cohort-lines, the age-lines lead downwards and the time-line runs upwards and to the right. This is only the rotated version of the equivocal *Lexis* (1875) diagram (Figure 6).
Source: Keiding, 2011: 411.

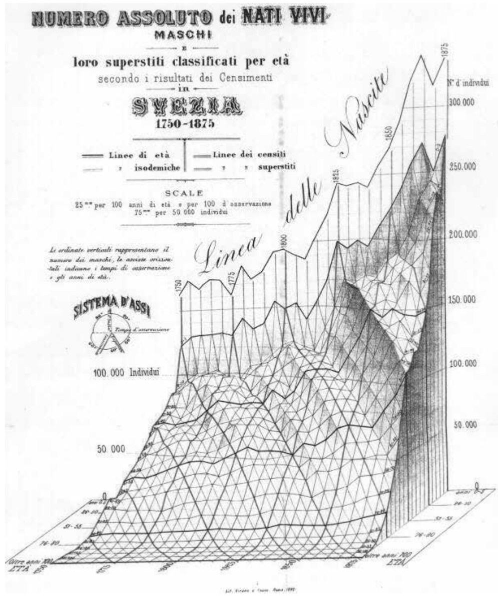
Based on the information above, it could be said that *Lexis* (1875) was probably the first author of a graphical expression of demographic processes and events who did not specifically consider life lines (although it is possible to draw them in his versions of the diagram), but instead used all three dimensions, which are still considered in the diagram: the age-, cohort- and period- (time-) dimension.[9)] He moreover devoted consideration to the different lengths of the lines representing these dimensions in rectangular version of the diagram. His equivocal version of the diagram could even today be regarded as very precisely designed, reflecting its inner logic. It is very easy to draw all the primary sets of events as well as the elementary sets of events and also to define the beginning of the year (the initial population at risk) and the middle of the year (the exposed population). However, the equivocal version of the diagram would even today be more complicated for technical processing, so the rectangular version is more useful in this respect. This was probably the reason why the rectangular version was applied in practice more often.

Work on the graphical representation of demographic events and demographic development continued in the late 19th century and in the 20th century. There is a quite famous colour version of one form of the Lexis diagram representing the demographic development of Sweden (Perozzo, 1880a, cit. in *Keiding*, 2011; see Figure 8).

At the beginning of the 20th century, graphical representation of studied demographic events or processes was not very popular among demographers. If they considered using graphical presentation of data at all, they more often focused on three-dimensional illustrations (as shown, for example, in Figure 8). For example, Verrijn Stuart (1928, cit. in *de Gans – van Poppel*, 2000) did not expect the Lexis diagram to spread and to be used widely in demography at all. It could be said that the Lexis diagram is still known and used because of the work of Roland Pressat, although even before Pressat's work, published in the 1960s, the Lexis diagram was used occasionally. One example of this can be found also in Czech demographic literature.

Roubíček (1958) dealt with the Lexis diagram and stressed that using this graphical tool could simplify many problems involved in demographic analysis or in the construction of life tables. More specifically, in his work he mentioned the 'demographic grid' (*Roubíček*, 1958: 171), and the Lexis diagram as its most illustrative type. He introduced the original Lexis diagram, as shown in Figure 5. Roubíček also drew the life lines into the diagram as vertical lines, as described above. In this case, the life lines started on the horizontal axis (axis of cohorts) and crossed the age-lines proceeding from the left axis (the axis of age) to the right. The time lines (also called isochrones) are lines with a slope of 135 degrees. As already derived above, this type of the diagram is especially useful for cohort analysis, because cohorts are marked in this diagram by particular vertical lines. It is then easy to read the age of a studied group of people or set of events, but defining the time of the event is not so straightforward.

Figure 8 Graphical representation of the demographic development of Sweden, 1750–1875



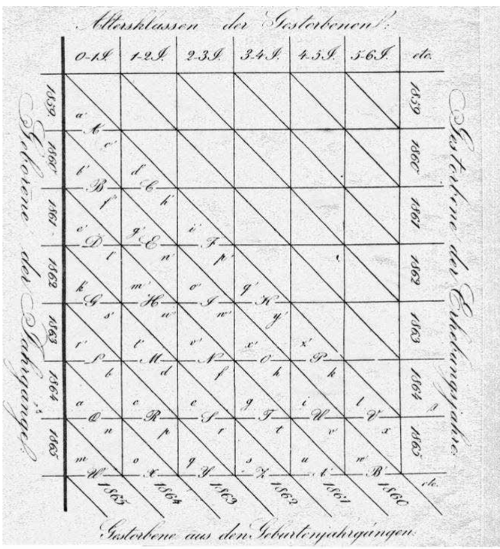
Note: *'This representation of the development of the Swedish population 1750–1875 was redrawn by Perozzo from a similar (black) stereogram by Berg (1860). The representation corresponds to Zeuner's stereograms with a (cohort, age) diagram in the horizontal plane. The bold lines leading up represent censuses (cross-sections at fixed calendar years) and the diagonal lines represent the survival of cohorts.'* (*Keiding*, 2011: 414)
Source: Wikipedia: The Free Encyclopaedia, 2014b.

9) Arguments against this statement are mentioned, for example, by Vandeschrick (2001), but with a particular uncertainty owing to the absence of certain historical documents that were not available or known of at that time.

Roubíček (1958) also showed the possibility of drawing the time axis in the diagram (as a line with a slope of 45 degrees). However, he also pointed out that this time axis is not usually used in the diagram because the time can be calculated using the cohort axis or using the known values of the age at the time of the event and the cohort.

Pressat's version of the Lexis diagram, simplified to two rectangular axes (age and time) and cohort lines with a slope of 45 degrees, was presented in Czech demographic literature by Pressat in translation of *Pavlík* (1968). However, *Vandeschrick* (2001) mentions that the version of the diagram corresponding to Pressat's work, and the one contemporarily used most often in the world, can already be traced in the work of Brasche (1870, cit. in *Vandeschrick*, 2001; see Figure 9), and that this author and his work was nearly forgotten in demography. This may be due to the fact that Brasche was not very explicit in his work and did not explain all the aspects of the diagram. Taking this into account, Brasche in 1870 (*ibid.*) was the first to propose the version of the diagram most often used today and Pressat in the 1860s 'only' reinvented it.

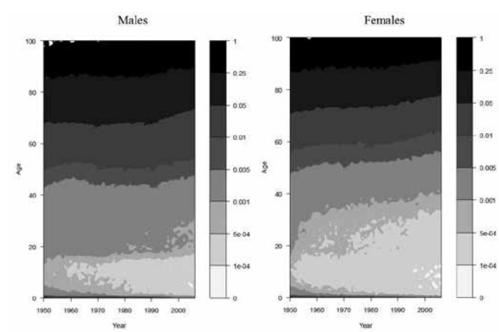
Figure 9 **The diagram proposed by Brasche (1870)**



Note: In fact this version of the diagram is nearly the same as the one we are used to today or that was described by Pressat (e.g. in 1968) – the only difference is its rotation (90 degrees to the right – the vertical axis in the figure is today traditionally depictured as a horizontal axis).
Source: Vandeschrick, 2001: 111.

The difference between the original *Lexis* diagram (proposed by Lexis, 1875, used, e.g., by *Roubíček*, 1958, and shown in Figure 5) and the modification described by Pressat is illustratively shown by *Dupâquier* (1999: 33). Pressat used the rectangular form of the Lexis diagram, and the values of age remained on the vertical axis. On the horizontal axis, Lexis originally proposed showing the values of cohorts (moments of birth), which corresponded more to the cohort approach to demographic analysis. Pressat used the horizontal axis for the time dimension, then the isochrones started on this horizontal axis and rose vertically. In the original Lexis diagram the life lines, like the cohort lines, went straight up. In Pressat's version, the life lines as well as the cohort lines have a slope of 45 degrees (*Dupâquier*, 1999). It could be said, that this version of the Lexis diagram is currently the one best known in the world (see Figure 11). This version of the diagram fully corresponds to the period approach to demographic analysis (often used because of data availability, etc.) and is also more practical for use with modern graphical and other types of software. Using the vertical line for the time dimension also enables an illustrative description of the development of some processes in time. This version of the Lexis diagram is also involved in the work of *Preston et al.* (2001).

Figure 10 **Mortality surfaces, age-specific mortality rates, Czech Republic, by sex, 1950–2006**



Note: 'Prepared in R software.
Source: Hulíková Tesárková, 2012.

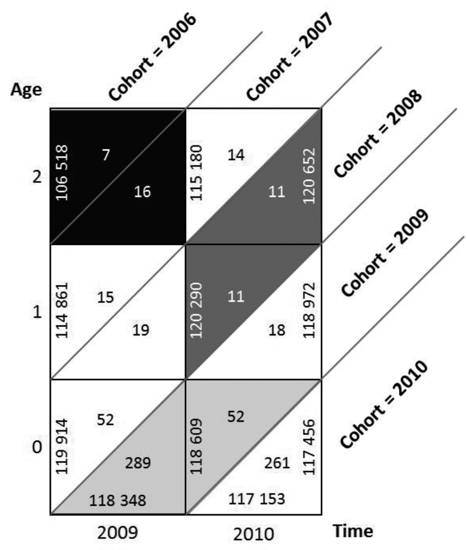
Nowadays, research in this field of study does not focus specifically on any changes or modifications of the Lexis diagram itself, but rather on the possible ways of constructing the diagram electronically using any information technologies or software. This is also the result of the greater availability of data in demographically developed countries, because it raised the problem of how to present the long development of any process briefly and illustratively. In such cases, the Lexis diagram was 'transformed' into a Lexis surface (surface or contour graphs in general, see Figure 10). The Lexis surface is often associated with the work of Arthur and *Vaupel* (1984). Another advantage of the contour map is its variability, it can be used to illustrate probably all demographic and other similar events and processes when they are organised by time and age (*Vaupel et al.*, 1998). Moreover, in this data structure, when using the contour diagram, the cohort effects can be showed. Some independent software has been published for the purpose of constructing the Lexis diagram (e.g. in *Vaupel et al.,* 1998). However, it would seem more effective for this topic to use more general statistical software – for example, with specific packages designed for constructing the Lexis diagram.[10] The possibilities of a graphical analysis of demographic data have also been described by *Hulíková Tesárková* (2011; 2012), and these focused on traditional mortality surface graphs and on three-dimensional graphs, where the third dimension stands for the population size or the intensity of any process.

4. COMPARISON OF THE TWO MOST COMMONLY USED LEXIS DIAGRAMS IN THE CZECH SCHOOL OF DEMOGRAPHY

As mentioned above, in the second half of the 20th century there were two types of Lexis diagram mentioned in Czech demographic literature and also used in practice. The first of them can be found in the work of *Roubíček* (1958), and this is in fact the version of the diagram originally proposed by Lexis (1875). The second type of diagram used is the version proposed by *Pressat* (e.g. 1968). In this part of the text a brief comparison of the usage of both of these approaches will be presented. This will help to better illustrate the practical application of each of them and also to describe their advantages and disadvantages.

By way of illustration, data for the Czech Republic and two calendar years were selected, 2009 and 2010,[11] for a period analysis of mortality at age 0, 1 and 2. Data were structured into all three primary sets and the calculation of age-specific mortality rates and probabilities will be described.

Figure 11 The Lexis diagram currently used most often in the world, data for the model calculation are imputed



Note: The three primary sets of data are marked: the 1st set (age-cohort) is light grey, the 2nd set (cohort-period) is darker grey, and the 3rd set (period-age) is black.
Source of data: Czech Statistical Office, 2013; 2014.

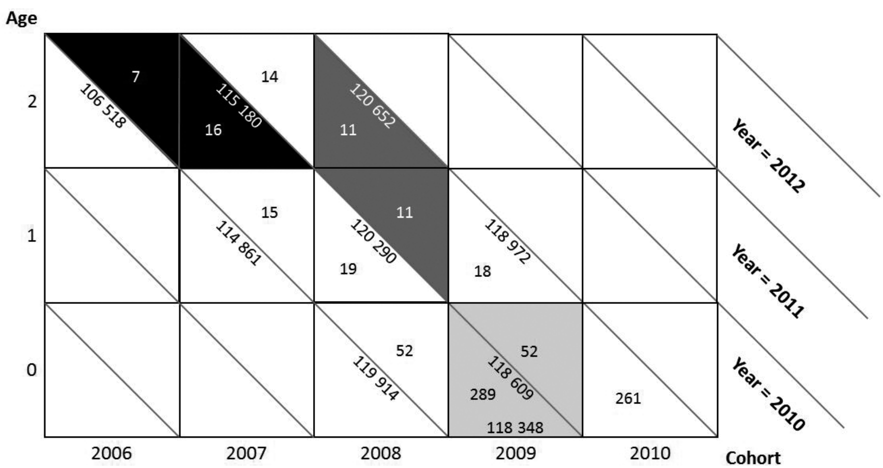
10) For example, LexisSurface package for software R (https://sites.google.com/site/timriffepersonal/r-code/lexissurface).

11) These two years were selected on purpose, because they are not influenced by the presence of the population census or any data corrections, like the year 2011 (where the initial number of inhabitants does not correspond to the number at the end of the year 2010). Selection of any years with corrections would complicate the figures below.

Figure 11 shows the most commonly used version of the Lexis diagram nowadays. As noted above, this version of the diagram is especially useful for period data analysis. In the figure, three primary sets of data are also marked – the first set (defined by one cohort and one age) is light grey, the second set (defined by one cohort and one calendar year) is darker grey, and the third set (defined by one age and one calendar year) is black. The same primary sets are also marked in Figure 12, where the *Lexis* (1875) diagram is shown. In both diagrams (Figure 11 and 12) the same data were used and also the primary sets of events are defined and marked in the same way

This demonstrates that Pressat's version (Figure 11) fully corresponds to period analysis and that the Lexis version (Figure 12) is more useful for cohort analysis. In this version of the Lexis diagram more space was needed to illustrate the data from only two calendar years. Also, in the original Lexis diagram (Figure 12) it is harder to define the initial population size (the population at the beginning of the year) and the exposed population (the population in the middle of the year).

Figure 12 **The Lexis diagram proposed by Lexis (1875) and introduced in the Czech school of demography by Roubíček (1958), data for the model calculation are imputed**



Note: The three primary sets of data are marked: the 1st set (age-cohort) is light grey, the 2nd set (cohort-period) is darker grey, and the 3rd set (period-age) is black.
Source of data: Czech Statistical Office, 2013; 2014.

To better illustrate usage of the diagrams, the ways of calculating the most important relations are shown below. For the first, second, and third primary sets of data the mortality rates and probabilities are calculated. Of course the results have to be the same; the purpose is only to help to provide a better orientation in the diagrams.

The first sets of data, marked in light grey in Figures 11 and 12, are defined by one year of birth (cohort) and one completed age. Because this set of data is more cohort-oriented, in the original Lexis diagram (Figure 12) it corresponds to a rectangle. In both diagrams shown this set corresponds to completed age 0. The probability of death in this set of data must be the total number of deaths under exact age 1 divided by the initial population size, which is the total number of live births. That means:

$$q_0^{\,I} = \frac{289 + 52}{118348}$$

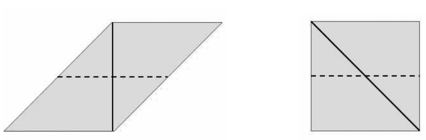
A slightly more complicated task could be the definition of the population size in the middle of the studied period (commonly representing the exposed population) in this set of data. There are two possible ways (in Figure 13 the solid line stands for the middle of the two calendar years covered by the first set of events, on the other side the dashed line better represents the supposed exposed population, on the assumption that the events are equally distributed within the age interval). Using the numbers from Figure 11 and 12, we can then define the mortality rate as:

$$m_0^{I} = \frac{289 + 52}{118609} = 0{,}00287499$$

or alternatively

$$m_0^{I*} = \frac{289 + 52}{[118348 + (118609 - 52)]/2} = 0{,}00287879$$

Figure 13 The first primary set of data in the Lexis diagram according to Pressat (left) and according to Lexis (right), illustrating the two possible exposed populations (the solid line and the dashed line)



For the second set of data the calculation is quite easy: in the case of probabilities we have to divide only the number of events by the initial population, i.e. the population size at the beginning of the year. Using the numbers from Figure 11 and 12:

$$q^{II}_{2010} = \frac{11 + 11}{120290}$$

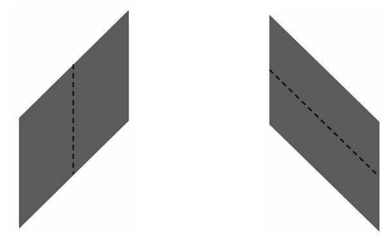
Age-specific mortality rates for the second set of data are also quite easily defined as:

$$m^{II}_{2010} = \frac{11 + 11}{(120290 + 120652)/2}$$

The exposed population (the population in the middle of the year) for the second set of data is defined clearly (see Figure 14).

In the third set of data the calculation is slightly more complicated in the case of the probability of death. The reason can be found in the different corresponding populations (cohorts) for each of the elementary sets of data (a triangle defined by one age, the cohort and the year of the event). Because the probabilities can be calculated for each triangle,

Figure 14 The second primary set of data in the Lexis diagram according to Pressat (left) and according to Lexis (right), illustrating the exposed population (the dashed line)



and they are independent probabilities, their multiples have to be subtracted:

$$q^{III}_{2} = \frac{7}{106518} + \frac{16}{115180 + 16} - \left(\frac{7}{106518} * \frac{16}{115180 + 16}\right)$$

Calculating the age-specific mortality rate is again quite easy for the third set of events (Figure 15):

$$m^{III}_{2} = \frac{7 + 16}{(106518+ 115180)/2}$$

Figure 15 The third primary set of data in the Lexis diagram according to Pressat (left) and according to Lexis (right), illustrating the exposed population (the dashed line)

5. CONCLUSION

The year 2014 marks the 100th anniversary of the death of Wilhelm Lexis, the German statistician, economist, journalist, and teacher who significantly influenced demographic research with the diagram that bears his name, which helps us to display demographic events reflecting a given time, age and cohort. Wilhelm Lexis's life was devoted to science and research, which is obvious not only from the bibliography but also from the list of all the places where he worked: the University of Bonn, the University of Strasbourg, the University of Tartu, the University of Freiburg im Breisgau, the University of Wroclaw, and the University of Göttingen. However, nowadays Lexis is known above all thanks to the 'Lexis diagram'. In everyday demographic practice we do not usually use the original form of the diagram proposed by Lexis, but this does not undermine the fact that Wilhelm Lexis has a firm place in the history of demography.

From the article, it is clear that work to devise some graphical tool that could be used to illustrate population development or demographic processes had already started several years before Lexis's work (1875). However, Wilhelm Lexis was among the first to combine all three dimensions (age, cohort and period) in one diagram. Also, the diagram proposed by Lexis (1875) did not change significantly during the development. In the 1960s Pressat 'only' changed the axes in the diagram (*Dupâquier*, 1999), or maybe publicised the historical version of the diagram originally proposed by Brasche in 1870 (*Vandeschrick*, 2001). As was shown above, both versions of the diagram (proposed by Lexis and by Pressat) could be used easily in practice and logically the results (rates or probabilities) have to be the same. The illustration of the practical application of the diagrams proved that both versions have their advantages as well as their disadvantages. The Lexis version of the diagram is more useful in the case of cohort analysis, where the cohorts are defined easily in the diagram. On the other hand, the diagram proposed by Pressat fully corresponds to the needs of the period approach to data analysis.

This paper presented not only some information about the life of Lexis, one of the best known demographers in the world and in the history of demography, but also about the life and development of the 'Lexis diagram'. To sum up it could be said that although it is now 100 years since the time Lexis passed and approximately 140 years since the Lexis diagram was proposed (regardless of who actually invented it), Lexis as well as the Lexis diagram still have their place in the field of demography and demographic analysis.

References

- Arthur, W. B. – Vaupel, J. W. 1984. Some General Relationships in Population Dynamics, *Population Index*, 50 (2), pp. 214–226.
- Dupâquier, J. 1999. La démographie. In: Bardet, J.-P. – Dupâquier, J. 1999. *Histoire des populations de l'Europe, tome 3: Les temps incertains, 1914–1998.* Fayard. 1999. ISBN 978-2213598819.
- Becker, K. 1874. *Zur Berechnung von Sterbetafeln an die Bevölkerungsstatistik zu stellende Anforderungen.* Verlag des Königlichen statistischen Bureaus, Berlin, 1874 (September 4, 2014).
  Available at: <http://gdz.sub.uni-goettingen.de/dms/load/pdf/?PPN=PPN527528277>.
- Caselli, G. – Vallin, J. – Wunsch, G. 2005. *Demography – Analysis and Synthesis: A Treatise in Population.* Academic Press, 2005. Pp. 2 976. ISBN 012765660X, 978012765660.
- Czech Statistical Office, 2013. *Czech Demographic Handbook 2012.* Code of the publication: e-4032-13. (September 8, 2014).
  Available at: <http://www.czso.cz/csu/2013edicniplan.nsf/engpubl/4032-13-eng_n_2013>.
- Czech Statistical Office, 2014. *Demografické ročenky (pramenná díla)* 2009, 2010. (September 8, 2014).
  Available at: <http://www.czso.cz/csu/redakce.nsf/i/casova_rada_demografie>.
- de Gans Henk, A. – van Poppel, F. 2000. Contributions from the margins. Dutch statisticians, actuaries and medical doctors and the methods of demography in the time of Wilhelm Lexis. Paper prepared for presentation at the workshop on *'Lexis in Context: German and Eastern& Northern European Contributions to Demography 1860–1910'* at the Max Planck Institute for Demographic Research, Rostock, August 28 and 29, 2000.
- Drechsler, W. – Kattel, R. 1997. Karel Bücher in Dorpat. *A Journal of the Humanities and Social Sciences.* Trames. 1 (4), pp. 322–368.
- Heiss, K. P. 1968. Wilhelm Lexis, in International Encyclopedia of the Social Sciences. *Encyclopedia.com.* (September 10, 2014).
  Available at: <http://www.encyclopedia.com/doc/1G2-3045000708.html>.
- Hertz, S. 2001. Wilhelm Lexis. Pp. 204–207. In *Statisticians of the Centuries.* New York: Springer.

- Hulíková Tesárková, K. 2011. Možnosti zobrazení vývoje úmrtnosti v České republice a na Slovensku ve statistickém software SAS a R. Česko a Slovensko ve společném státě a samostatně, podobnosti a odlišnosti – *XLI. konference České demografické společnosti.* Příspěvek v posterové sekci na konferenci, Přírodovědecká fakulta Univerzity Palackého v Olomouci, Olomouc, 26.–27. května 2010.
- Hulíková Tesárková, K. 2012. *Selected methods of mortality analysis focused on adults and the oldest age-groups.* Dizertační práce (Ph.D.). Univerzita Karlova v Praze, katedra demografie a geodemografie, Praha, 2012, 259 s.
- Johnson, N. L. – Kotz, S. 1997. Wilhelm Lexis. Pp. 305–306. In *Leading Personalities in Statistical Sciences from the seventeen century to the present.* John Wiley & Sons.
- Keiding, N. 2011. Age–period–cohort analysis in the 1870s: Diagrams, stereograms, and the basic differential equation. *The Canadian Journal of Statistics / La revue canadienne de statistique, Statistical Society of Canada / Société statistique du Canada,* 39 (3), pp. 405–420.
- Lexis, Wilhelm. 1875. Einleitung in die Theorie der Bevölkerungsstatistik. Trübner, Strassburg, 1875. (September 4, 2014). Available at: <http://gdz.sub.uni-goettingen.de/dms/load/pdf/?PPN=PPN527522651>.
- Pressat, R. 1968. *Základy demografické analýzy.* Státni pedagogické nakladatelství Praha, 1968. Translated to the Czech language by Pavlík, Zdeněk.
- Preston, S. H. – Heuveline, P. – Guillot, M. l. 2001. *Demography: Measuring and Modeling Population Processes.* Oxford: Blackwell Publishers Inc. ISBN 1-55786-214-1.
- Roubíček, V. 1958. *Demografická statistika*. Pedagogické nakladatelství, n.p. Praha, 1958. Učební texty vysokých škol, Vysoká škola ekonomická v Praze, Fakulta všeobecně ekonomická.
- Seneta, E. W. 2014. Wilhelm Lexis (version 5). *StatProb: The Encyclopedia Sponsored by Statistics and Probability Societies.* (September 10, 2014). Available at: <http://statprob.com/encyclopedia/WilhelmLEXIS.html>.
- Siegel, J. S. 2012. The *Demography and Epidemiology of Human Health and Aging.* Springer. ISBN 978-94-007-1314-7.
- Vandeschrick, Ch. 2001. The Lexis Diagram, a misnomer. *Demographic Research.* Max-Planck-Gesellschaft, 4 (3), pp. 97–124.
- Vaupel, J. W. – Zhenglian, W. – Andreev, K. F. – Yashin, A. I. 1998. *Population Data at a Glance: Shaded Contour Maps of Demographic Surfaces over Age and Time.* Odense University Press, ISBN 87-7838-338-2. Available at: <http://www.demogr.mpg.de/Papers/Books/Monograph4/start.htm>.
- von Collani C. 2014. Wilhelm Lexis. In *Encyklopedia Stochastikon GmbH*. (September 10, 2014). Available at: <http://132.187.98.10:8080/stochastikon/frameset.jsp?app=stochastikon&lang=en&start=default.jsp>.
- Wikipedia: The Free Encyclopaedia. 2014a. *Portrait of Wilhelm Lexis.* (10 September 2014). Available at: <http://sv.wikipedia.org/wiki/Wilhelm_Lexis#mediaviewer/File:Voit_172_Wilhelm_Lexis.jpg>.
- Wikipedia: The Free Encyclopaedia. 2014b. Stereogram (three-dimensional population pyramid model on actual data (Swedish census, 1750-1875). (12 September 2014). Available at: <http://it.wikipedia.org/wiki/Luigi_Perozzo#mediaviewer/File:Stereogram_(three-dimensional_population_pyramid)_modeled_on_actual_data_(Swedish_census,_1750-1875).jpg>.
- Zeuner, G. 1869. *Abhandlungen aus der Mathematischen Statistik*. Leipzig: Felix, 1869. (September 9, 2014). Available at: <http://www.forgottenbooks.com/download_pdf/Abhandlungen_aus_der_Mathematischen_Statistik_1100111194.pdf>.

KLÁRA HULÍKOVÁ TESÁRKOVÁ

completed her doctoral study in demography in 2012 at the Faculty of Science of Charles University in Prague, where she has worked since 2008 in the Department of Demography and Geodemography. Her research focuses of demographic methodology (methods of mortality analysis, application of mathematical and statistical methods in demography) and applied demography (focused on education and life insurance).

OLGA KURTINOVÁ

is a graduate of the doctoral study programme in Demography at the Faculty of Science of the Charles University in Prague. Since 2011 she is a research assistant at the Department of Demography and Geodemography of given faculty. Her research is focused on applied demography with respect to economic interdependence of demographic processes.