V.P. KLIMENYK, O.M. OCHEREDKO

GEOSTATISTICAL MODELLING OF DEMOGRAPHICAL INPUT INTO CARDIOVASCULAR DISEASES RELATED DISABILITY RATES IN ZHYTOMYRSKA OBLAST, UKRAINE

Vinnytsya National Medical University, Ukraine

Purpose: disclosure of demographic environment impact on the spatial heterogeneity of disability risks due to cardiovascular diseases (CVD) in Zhytomyr region.

Materials and methods. Data were adapted to age-period-cohort (APC) compatible strip-split plot design, 11 birth cohorts from date of birth before 1940 and consequently by five year intervals («1941-45», «1946-50», … , «after 1985»), 13443498 adult-years totally. Incidence disability cases certified in 1999-2008 have been retrieved from records of medical expert committees. Methods: GLM mixed model with spatial covariance matrix processed by GLIMMIX procedure, SAS.

Results. Space heterogeneity in disability rates appeared to be significantly modified by irregularity in distribution of APC factors and density of population.

Conclusions. Significant impact of demographic factors on medical events has been established. However its studying is complicated with hierarchical data organization and presence of randomized effects that require multilevel mixed-approach. Another problem is collinearity APC-component. Geostatistical modelling can help in smoothing irregular local features and selection of regular local laws.

KEY WORDS: **geostatistical analysis, disability, cardiovascular diseases, medical and social environment.**

Health events proved to have certain geographical patterns in their incidence, prevalence, outcomes. This fact is helpful to capture determinants of counts and rates in their irregular space distribution. Much of avail is geostatistical modelling widely in use to assess and explain evident regional disparities. Space modelling is extremely relevant borrowing strength from information on neighboring space readings. The simplest approach is to integrate spatial correlations in covariance matrix (Littell, Ramon C., 2006) though advanced techniques exploit structural spatial priors (P.D. Congdon, 2010). Cardiovascular diseases are typecast of diseases with marked population distributions both across population groups and by localities (G.D. Smith, C. Hart, G. Watt et al., 1998). Strangely enough, CD related disability evaded attention of geostatistical studies. Being thin on the ground, list of papers includes Bayesian space modelling of stroke mortality and chronic ischemic heart disease in U.S. (Z. Hu et al., 2008, 2009), spatial analysis of CD mortality (A. van der Linde et al., 1995), regional trends in systolic blood pressure (G. Danaei et al., 2011), spatial relation of cardiovascular and stroke mortality rates to measures of neighbourhood deprivation (Murray M. Finkelstein, et al., 2005), area of living and risk of myocardial infarction (S.M. Kolegard et al, 2002), neighborhood socioeconomic environment and incidence of coronary heart disease (K. Sundquist, M. Winkleby, H. Ahlen et al., 2004), residential environments and cardiovascular risk (A.V. Diez Roux, 2003). Furthermore, we failed to find papers on geographical mapping of CD-related disability. So we proceeded with data covered all (26) counties of Zhytomyrska oblast, Ukraine. Demographic population data obtained from regional statistical agencies. Incidence disability cases certified in 1999-2008 were retrieved from records of medical expert committees.

Materials and Methods. Study design was APC construction compatible strip-split plot (Figure 1). Data also met requirements of spatial data organization. The principal unit is birth cohort. We studied 11 cohorts from date of birth before 1940 and consequently by five year intervals («1941-45», «1946-50», … , «after 1985»), 13443498 adult-years totally (Table 1). Each cohort captures unique combination of historical events (Fu, W. J., 2000). The other important component is time dimension that unfolds the succession of events. Age has intrinsic importance per se as well as indispensable covariate to solve ambiguity of time-age collinearity. Counties have been described by sociodemographic variables and geographical coordinates. Space density distribution of population was yet another important demographical entity under study.

We followed two classical steps in the geostatistical modelling. We estimated the spatial variability and used the estimates to smooth observed spatial disability rates. Smoothing is especially propitious for small communities due to sporadic fluctuation of disability rates. Next, we used smoothed disability rates to examine plausible input to space heterogeneity of APC factors and consequently the density of population.
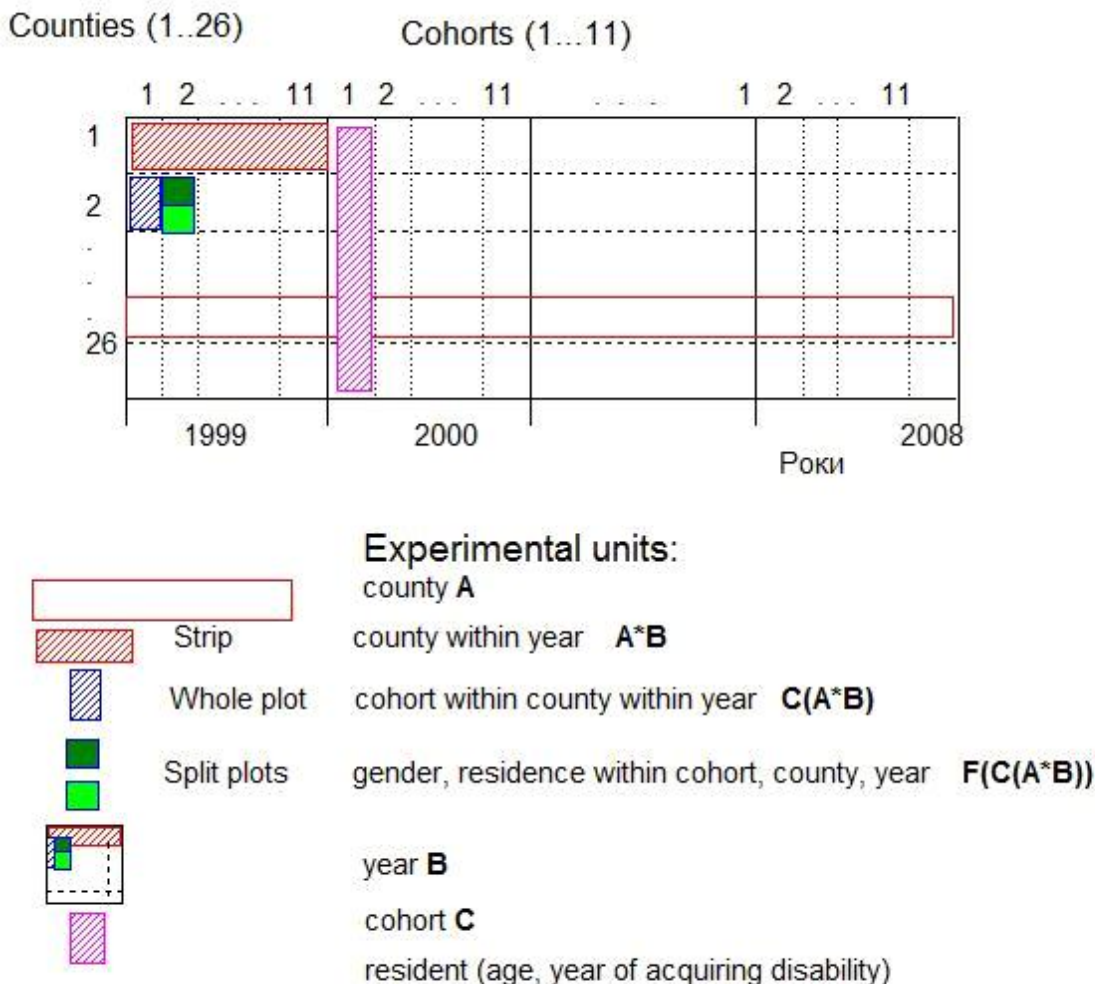
Figure 1. **Strip-split-plot design of studi**

The empirical semivariogram was computed in SAS by classical estimator from data residuals (r) using the formula:

$$\gamma(h) = \frac{1}{2m} \sum (r_i - r_j)^2$$

where m is the number of pairs of observations a distance, h, apart. $\gamma(h)$ is estimated for all distances at which pairs of observations exist or at a discrete set of lagged values within a tolerance to ensure that a sufficient number of observations contribute to each value of $\gamma(h)$. Two functions (exponential and Gaussian) regressed on values of sample semivariogram with parameters RANGE (R) =21 NUGGET=0.047 SILL=0.05 were fitted by NLIN procedure, SAS (Figure 2):

We used Gaussian function with these particular parameters in KRIGE2D procedure to produce a contour plot of the kriging estimates and the associated standard errors (r~2*range=100?=1˚40′ for Gaussian covariance structure):

proc krige2d data= variogramdata outest=est; pred var= r_ilink k r=100;

model   NUGGET=0.047   SCALE=0.052 RANGE=50.5 form=gauss;

coord xc= latitude yc= longitude;

grid x=2960 to 3100 by 5 y=1640 to 1800 by 5; run;

Grid values ranges from 49˚20′□ to 51˚40□′ for latitude (given in minutes) and from 27˚20′ to 30˚00′ for longitude. NUGGET, SCALE, RANGE defined by variogram (Figure 2).

Precise estimations of space covariance matrix random effects NUGGET, SILL, RANGE we received from procedure MIXED. We recommend to use it with the scope of initial values around variogram estimates via PARMS option. Otherwise, procedure can converge to local maxima and provide grossly
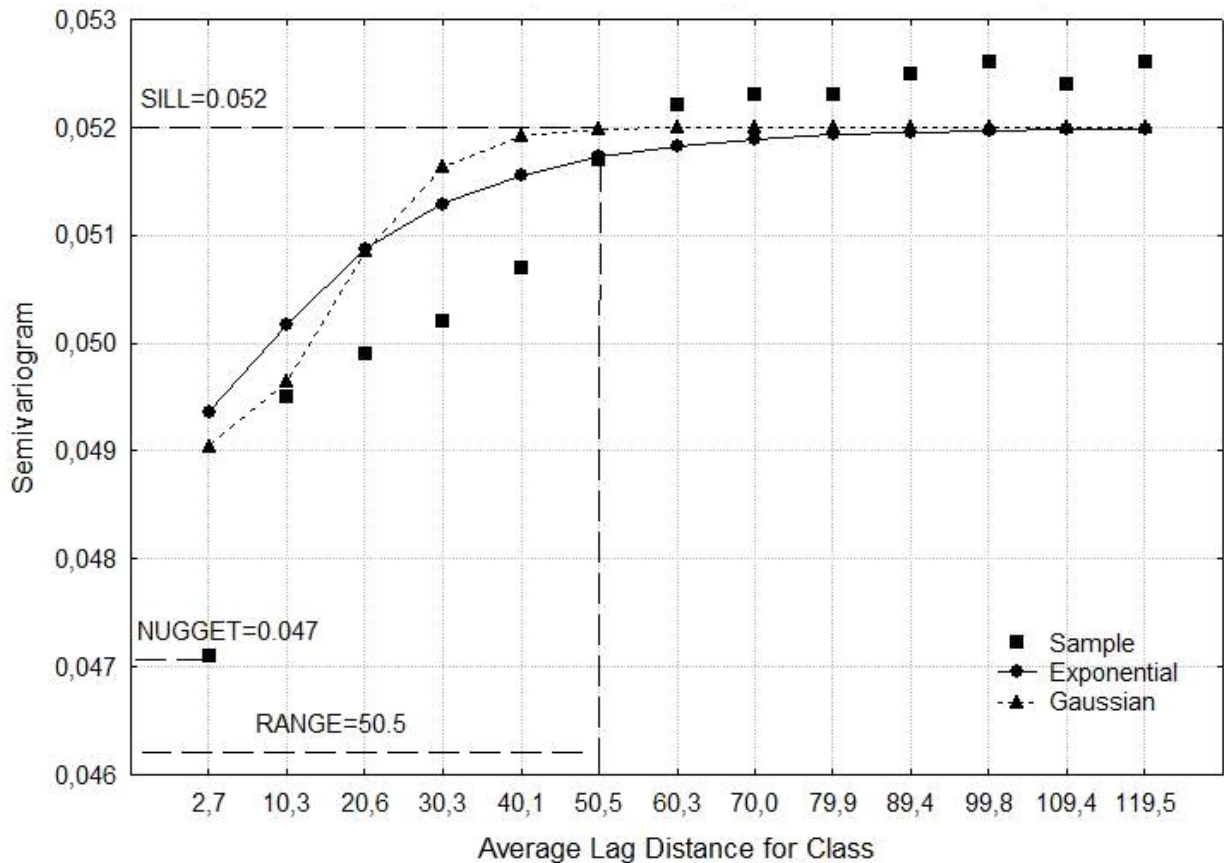
Figure 2. **Theoretical and Sample Semivariograms for CD Disability Data**

unreasonable estimates. The random effects estimates were: SILL=0,050, RANGE=20,7 NUGGET=0,047. These defined space covariance matrix.

Modifications to space heterogeneity in disability rates by irregularity in distribution of APC factors were assessed by comparison of likelihoods of nested models. Models differed by linear predictor (LP): shell model with only intercept in LP while LP of APC model was padded out with APC factors. The further LP modification was tailing it with density of population. All three models had the same spatial Gauss form covariance matrix. Binominal models with canonical logit link function were processed by procedure GLIMMIX:

proc glimmix data=spacedata;
class county cohort year Gender Residence;
model Disabilitynum/population=LP;
random _residual_ / subject = intercept
type = SP(GAU)(latitude longitude);
parms (0.050 20.7 0.047)
run;

Results. Heterogeneity in space distribution of CD-related disability risks in Zhytomyrska oblast is evident (Figure 3). Three picks are obvious. Ordered by height they are: northern (of latitude above 51° within longitude of the range 27°20′ – 29°), southern (of latitude up to 50°40′ within longitude of the range

27°40′ – 29°40′), and eastern (in ranges of latitude 50°20′ – 51°00′ and longitude from 29°30′).

The logic behind testing is straightforward. If changes in LP leaving the same covariance matrix induce significant modifications in space distribution of disability, then added factors indeed render input to space heterogeneity. We have used the difference between models double log-likelihoods as operational test statistics.

Shell model is composed just by space covariance structure. The « subject = intercept » option in procedure GLIMMIX treats all observations in the data set as potentially correlated. In fact, shell model depicts somewhat smoothed original space distribution of CD-related risks of disability. Negative double logarithm of shell model likelihood (-2 Log Likelihood) equals 5141512.

APC model. Including APC-related components to LP covariance matrix preserved helps to test the significance of induced changes to space distribution of risks. LP contains APC variables cohort, year, age, age squared. -2 Log Likelihood dropped to 3518939.

Number of additional fixed parameters against shell model is 21, see « df1» column of the table. The difference between values of double logarithms of model likelihoods follows in approximation chi-square distribution with degrees of freedom equals
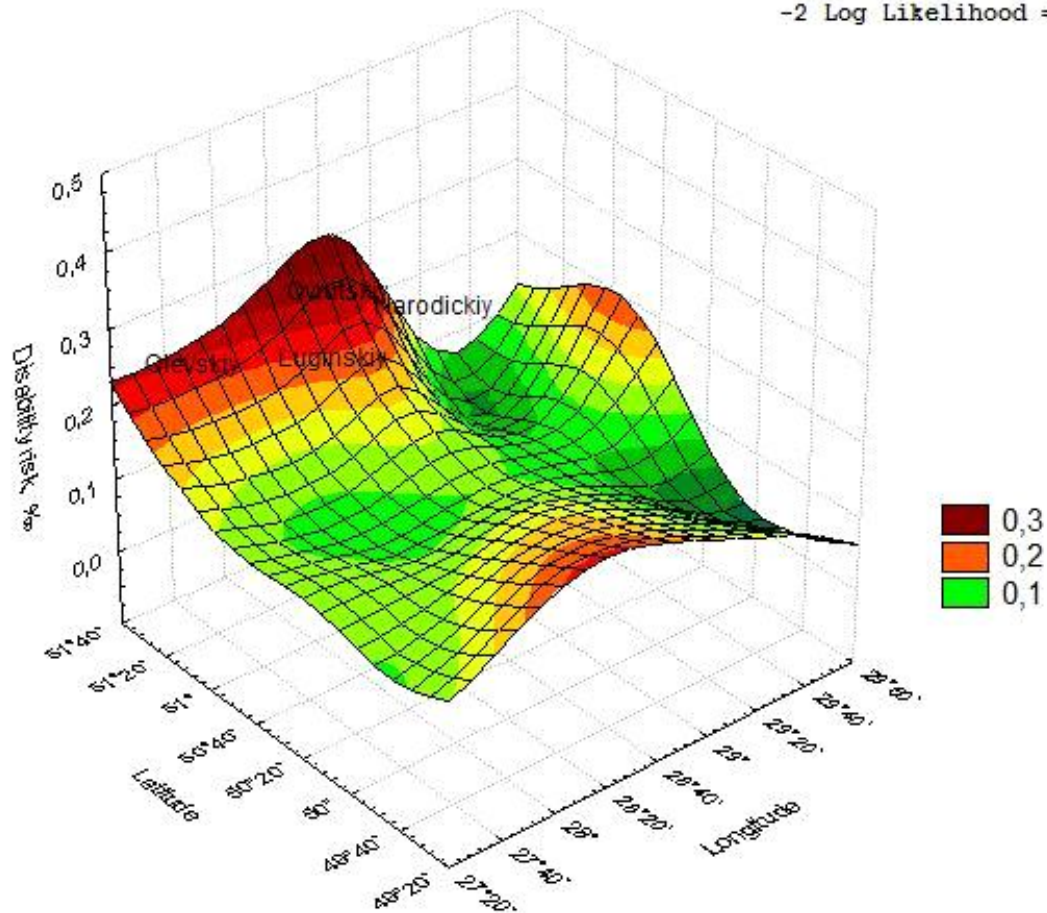
Figure 3. **Space distribution of CD-related disability risks (Zgytomyrska oblast, Ukraine)**

to number of additional parameters. Δ(2 Log Likelihood) = (2 Log Likelihood$_{shell}$ – 2 Log Likelihood$_{APC}$=5141512 – 3518939 =1622573). Chi-square (21) 0,999 centile = 54 that is considerably less then difference Δ(2 Log Likelihood). *Therefore, we stipulate significant p<0,0001 impact of APC factors on geographical distribution of CD-related risks of disability in Zhytomyrska oblast, Ukraine.*

Table. **Significance of impact of APC factors on geographical distribution of CD-related risks of disability in Zhytomyrska oblast, Ukraine**

Effect	df1	df2	F	Pr > F
Cohort	10	7117	63875	<.0001
Year	9	7117	25849	<.0001
Age	1	7117	82087	<.0001
Age*Age	1	7117	115548	<.0001

Moreover, every APC factor demonstrates significant p<0,0001 impact on geographical distribution of CD-related risks of disability. It's interesting that squared age renders the most substantial impact with largest F-value.

Detailed analysis of input of particular cohort witnesses that every consequent cohort in contrast with the last (youngest) demonstrates significant p<0,0001 impact on geographical distribution of CD-related risks of disability (t-test df=7117). Contrasts decrease monotonically from oldest cohort with birthdate before 1940 (β= -4.10) to β= -0.41 in next to youngest (Table 3). It corresponds to our other previous findings supporting the hypothesis of increase of disability rates in successive generations (V. Klimenyk, 2013).

Impact of years of observation on geographical distribution of CD-related risks of disability studied by their contrasts with year 2008. It appeared that each year reveals significant p<0,0001 impact on geographical distribution of CD-related risks of disability. On a whole contrasts decrease from β=0,99 in 2000 down to β=-0,06 in 2007. Again, data correspond with our previous findings (V. Klimenyk et al., 2012).

Covariate effect of age is positive (в=0,45) and significant (p<0.0001), that is the disability risk increases with aging. Meanwhile the covariate of squared age demonstrates even more significant negativity (в= -0,004). This corroborates open down parabolic feature of disability risk distribution by age

with apex in age group 51-55 with further levelling down found in previous research (V. Klimenyk, 2013).

After adjustment for APC-factors geographical distribution of CD-related risks of disability significantly altered, though shape of the surface still resembles previous (Figure 4). Narrower range of risk ordinates demonstrates more homogenous space distribution as was anticipated. The same conclusion goes from lower than in shell model value of -2 Log Likelihood, which is 3518939.
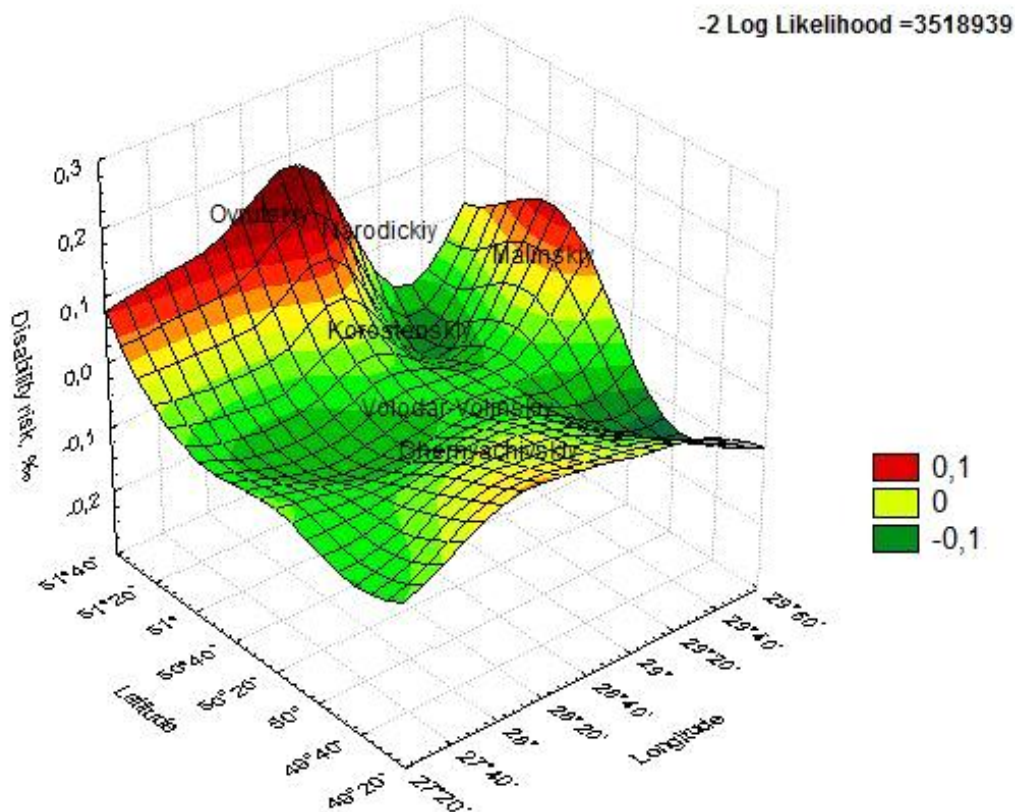

Figure 4. **Space distribution of CD-related disability risks after adjustment for APC factors (Zgytomyrska oblast, Ukraine)**

There is a chance that residual space heterogeneity may be due to irregular distribution of population. Population density is important integral indicator of development of territory and is the source of health events. So, we extracted the effect of population density from residual space distribution of disability.

As a result, -2 Log Likelihood further dropped to 3337453. Number of additional fixed parameters against APC model is 1Δ(2 Log Likelihood) = (2 Log Likelihood$_{APC}$ – 2 Log Likelihood$_{APC+HAC}$= 3518939 – 3337453=181486). Chi-square (1) 0,999 centile = 10,8 that is considerably less then difference Δ(2 Log Likelihood). *Therefore, we state significant p<0,0001 impact of population density on geographical distribution of CD-related risks of disability in Zhytomyrska oblast.* The conclusion is supported by regression coefficient (в=5.409E-6; t=15365; p<0,0001). Positive coefficient implies the more populated is the territory the higher disability risk is anticipated. The value of coefficient is small but if gradient in population is about 1000, the effect is tangible (exp(0,054)=1,005).

After adjustment for population density geographical distribution of CD-related risks of disability significantly altered again, though shape of the surface still resembles previous (Figure 5). Narrower range of risk ordinates demonstrates more homogenous space distribution. The same conclusion goes from lower than in APC model value of -2 Log Likelihood, which is 3337453.

Demographical input to health events is considerable. Yet its study poses some difficulties in part due to hierarchical data organization and random variables that calls for involved multilevel mixed modelling. One may encounter problem with collinearity of APC components (W.J. Fu, 2000). Inconsistency of registries also requires adjustment. Geostatistical modelling may be helpful in adjustment of local irregularities by smoothing as well as in explanation gross discrepancies. Behind seemingly stationary time distribution of CD-related disability proved to mask significant trends by cohorts, years, age. For instance, we unveiled the dramatic decrease in CD-related disability risks over 1999-2008 that cannot be explained by successes
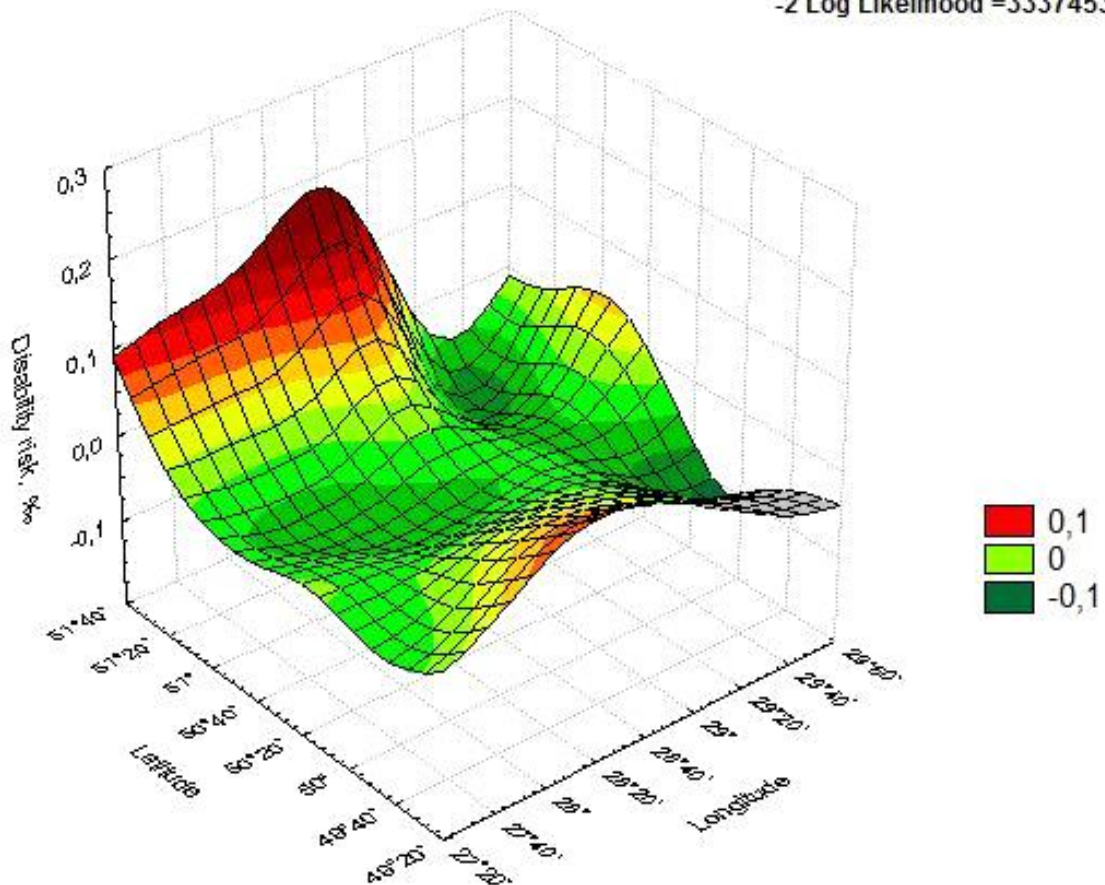
Figure 5. **Space distribution of CD-related disability risks after adjustment for APC factors and population density (Zgytomyrska oblast, Ukraine)**

in preventive or clinical medicine. We have traced formidable increase in CD-related disability risks from 0,45‰ in cohort birthed before 1941 up to 1,87‰ in cohort birthed after 1985 року. Probably this is due to deterioration of health in coming generations. We also revealed that the maximal values of risk pertains to age group 51-55 with regularity for prevalent groups of CD. The answer to elicited discrepancies might be look for in social aspect of disability certification. On a whole data support the increase in disability rates matter of fact while simultaneous decrease of opportunity to be certified. It still in wait to unclose whether it is a state policy (regulation of pensions and related benefits to disabled), or rather pecuniary motivation of officials proceeded disability expertise, yet possible tidings in labor market or even society values. Population density is another important demographical indicator being oversight as often as not. We observed that the more populated is the territory the higher disability risk is anticipated. All in all we advocate the advances and opportunities of geostatistical modeling of health events.

Conclusions

1. By geostatistical hierarchical nonlinear mixed model we ascertained the peculiarities of disability

distributions across cohorts, age groups and time periods (APC).

2. Results bear witness to enhanced local heterogeneity in disability rates due to irregularity in distribution of APC factors. Analysis suggests significant $p < 0,0001$ impact of APC factors on geographical distribution of CD-related risks of disability in Zhytomyrska oblast.

3. Each APC factor rendered significant modification to geographical distribution of disability rates.

4. Birth cohort effects decreased monotonically from oldest cohort with birthdate before 1940 ($\beta=-4.10$) to $\beta=-0.41$ in next to youngest. It suggested hypothesis of increase of disability rates in successive generations

5. Each year revealed significant $p < 0,0001$ impact on geographical distribution of CD-related risks of disability. On a whole contrasts decreased from в=0,99 in 2000 down to в=-0,06 in 2007.

6. Covariate effect of age was positive (в=0,45) and significant ($p < 0.0001$), that is the disability risk increases with aging. Meanwhile the covariate of squared age demonstrated even more significant negativity (в=-0,004).

7. Findings also state significant p<0,0001 impact of population density on geographical distribution of CD-related risks of disability in Zhytomyrska oblast. The more populated was the territory the higher disability risk was anticipated.

Perspectives outline. The method is extendable by taking individual unobservables or frailties in consideration, but such enhancement most efficiently relies on MCMC samplers. The method is sensitive to assess the population treatment impact by incorporating DD approach in linear predictor. The discreet space modelling is yet another option that is less sensitive to distancing but more relevant to bluff shifts on the boundaries of communities. Moreover, space structured priors are of utmost importance in this set up, that again can be implemented by MCMC samplers only. Mixed distribution composition like semiparametric «Stick Breaking» approach is promising as well as jumping MCMC algorithm, both suggest broader opportunities in vague knowledge of underlying space heterogeneity drivers.

References

1. Littell Ramon C., George A. Milliken, Walter W. Stroup, Russell D. Wolfinger and Oliver Schabenberger. 2006. SAS® for Mixed Models, Second Edition. Cary, NC: SAS Institute Inc.
2. Congdon P. D. Applied Bayesian Hierarchical Methods. Chapman and Hall/CRC – 2010. – 604 p.
3. Individual social class, area based deprivation, cardiovascular disease risk factors, and mortality: the Renfrew and Paisley Study / Smith G. D., Hart C., Watt G. [et al.] // J. Epidemiol. Community Health. – 1998. – Vol. 52. – P. 399405.
4. Hu Z. Linking stroke mortality with air pollution, income, and greenness in northwest Florida: an ecological geographical study / Z. Hu, J. Liebens, K. R. Rao // International Journal of Health Geographics. – 2008. – Vol.7-20.
5. Hu Z. Particulate air pollution and chronic ischemic heart disease in the eastern United States: a county level ecological study using satellite aerosol data / Z. Hu, K. R. Rao // Environmental Health. – 2009. – Vol. 8. – P. 26.
6. van der Linde Spatial-Temporal Analysis of Mortality / van der Linde, K.-H. Witzko, K.-H. Jockel // Using Splines Biometrics. – 1995. – Vol. 51, No. 4 (Dec.). – P. 1352–1360.
7. National, regional, and global trends in systolic blood pressure since 1980: systematic analysis of health examination surveys and epidemiological studies with 786 country-years and 54 million participants / Danaei G., Finucane M. M., Lin J. K. [et al.] // Lancet. – 2011. – Vol. 6736. – P. 62036–62033.
8. Murray M. Finkelstein Environmental inequality and circulatory disease mortality gradients / Murray M. Finkelstein, Michael Jerrett, Malcolm R. Sears // J. Epidemiol. Community Health. – 2005. – Vol. 59. – P. 481–487 doi:10.1136/jech.2004.026203
9. Socioeconomic context in area of living and risk of myocardial infarction: results from Stockholm heart epidemiology program (SHEEP) / Kolegard S. M., Diderichsen F., Reuterwall C. [et al.] // J. Epidemiol. Community Health. – 2002. – Vol. 56. – P. 29–35.
10. Neighborhood socioeconomic environment and incidence of coronary heart disease: a follow-up study of 25,319 women and men in Sweden / Sundquist K., Winkleby M., Ahlen H. [et al.] // Am. J. Epidemiol. – 2004. – Vol. 159. – P. 655–62.
11. Diez Roux A. V. Residential environments and cardiovascular risk / A. V. Diez Roux // J. Urban Health. – 2003. – Vol. 80. – P. 569–89.
12. Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates / Piel F. B., Patil A. P., Howes R. E. [et al.] // Lancet. – 2013. – Vol. 381. – P. 142–51.
13. Fu W. J. Ridge Estimator in Singular Design with Application to Age-Period-Cohort Analysis of Disease Rates / W. J. Fu // Communications in Statistics—Theory and Method. – 2000. – Vol. 29. – P. 263–278.
14. Klimenyk V. The study of cardiovascular disability risks by the birth cohorts / V. Klimenyk // East European J. of Pub. Health. – 2013. – Vol. 1 (21). – P. 155–56.
15. Analysis of CD related disability risks by APC construction on the basis of strip-split-plot design / Klimenyk V., Galachenko O., Ocheredko O., Andrievsky I. // East European J. of Pub. Health – 2012. – Vol. 2–3 (18–19). – P. 155–63.
16. Klimenyk V. The investigation of historical trends of stroke disability rates based on APC-decomposition / V. Klimenyk // East. European J. of Pub. Health. – 2013. – Vol. 1 (21). – P. 63–71.

ГЕОСТАТИСТИЧНЕ МОДЕЛЮВАННЯ ВПЛИВУ ДЕМОГРАФІЧНОГО СЕРЕДОВИЩА НА ІНВАЛІДИЗАЦІЮ ВНАСЛІДОК СЕРЦЕВО-СУДИННИХ ЗАХВОРЮВАНЬ У ЖИТОМИРСЬКІЙ ОБЛАСТІ, УКРАЇНА

В.П. Клименюк, О.М. Очередько
Вінницький національний медичний університет ім. М.І. Пирогова, Україна

Мета: розкриття впливу демографічного середовища на просторову гетерогенність ризиків інвалідизації внаслідок серцево-судинних захворювань (ССЗ) у Житомирській області.

Матеріали і методи. Дані організовані в адаптований до конструкції вік-період-когорта (АРС) стріп-спліт-плот-дизайн, основу якого склали 11 когорт за народженням від років народження до 1940 і далі за п'ятирічними інтервалами («1941-45», «1946-50», ... , «після 1985»), всього 13443498 людино-років.

Враховувались усі випадки первинної інвалідності, сертифіковані у 1999–2008 роках медико-соціальними експертними комісіями. Методи: GLM мікст-модель з просторовою коваріаційною матрицею, параметри обчислені за процедурою GLIMMIX статистичної системи SAS.

Результати. Просторова гетерогенність рівнів інвалідизації суттєво обумовлювалась нерегулярністю просторового розподілу АРС-факторами та щільністю проживання населення.

Висновки. Встановлено значний вплив демографічних факторів на медичні події. Проте його вивчення затруднене ієрархічною організацією даних і присутністю рандомізованих ефектів, які вимагають багаторівневого мікст-підходу. Проблемою також є колінеарність АРС-компонент. Геостатистичне моделювання може допомогти у згладженні нерегулярних локальних особливостей та у виділенні регулярних місцевих закономірностей.

КЛЮЧОВІ СЛОВА: **геостатистичний аналіз, інвалідизація, серцево-судинні захворювання, медико-соціальне середовище**.

ГЕОСТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВЛИЯНИЯ ДЕМОГРАФИЧЕСКОЙ СРЕДЫ НА ИНВАЛИДИЗАЦИЮ ВСЛЕДСТВИЕ СЕРДЕЧНО-СОСУДИСТЫХ ЗАБОЛЕВАНИЙ В ЖИТОМИРСКОЙ ОБЛАСТИ, УКРАИНА

В.П. Клименюк, А.Н. Очередько
Винницкий национальный медицинский университет им. М.И. Пирогова, Украина

Цель: раскрытие влияния демографической среды на пространственную гетерогенность рисков инвалидизации вследствие сердечно-сосудистых заболеваний (ССЗ) в Житомирской области.

Материалы и методы. Данные организованны в адаптированный к конструкции возраст-период-когорта (АРС) стрип-сплит-плот-дизайн, основу которого составили 11 когорт по рождению от года рождения до 1940 и далее по пятилетним интервалам («1941-45», «1946-50 »...,«после 1985»), всего 13443498 человеко-лет. Учитывались все случаи первичной инвалидности, сертифицированные в 1999–2008 гг. медико-социальными экспертными комиссиями. Методы: GLM микст-модель с пространственной ковариационной матрицей, параметры вычислены по процедуре GLIMMIX статистической системы SAS.

Результаты. Пространственная гетерогенность уровней инвалидизации существенно обуславливалась нерегулярностью пространственного распределения АРС-факторами и плотностью проживания населения.

Выводы. Установлено значительное влияние демографических факторов на медицинские события. Однако его изучение осложнено иерархической организацией данных и присутствием рандомизированных эффектов, требующих многоуровневого микст-подхода. Проблемой также является колинеарность АРС-компонент. Геостатистическое моделирование может помочь в сглаживании нерегулярных локальных особенностей и в виделении регулярных местных закономерностей.

КЛЮЧЕВЫЕ СЛОВА: **геостатистический анализ, инвалидизация, сердечно-сосудистые заболевания, медико-социальная среда.**

Відомості про авторів:

Клименюк Володимир Петрович – к.мед.н., директор Вищого комунального навчального закладу «Бердичівський медичний коледж»; асистент кафедри соціальної гігієни та організації охорони здоров'я Вінницького національного медичного університету ім. М.І. Пирогова; тел. (04143) 2-02-17.

Очередько Олександр Миколайович - д.мед.н., доц., завідувач кафедри соціальної медицини та організації охорони здоров'я Вінницького національного медичного університету; тел. службовий: 0432-56-28-89.