Research on the influence factors of technogenic environmental pollution on population health based on statistical panel data analysis

S N Trostyanskiy¹, S A Kurolap², I O Baklanov¹, O V Klepikov¹ and O Ja Kravets³

¹Russian Air Force Military Educational and Scientific Center “Air Force Academy named after Professor N.E Zhukovsky and Y.A. Gagarin”, 54a, Old Bolsheviks str., Voronezh, 394064, Russia
²Voronezh State University, 1, Universitetskaya sq., Voronezh, 394018, Russia
³Voronezh State Technical University, 14, Moscow ave., Voronezh, 394026, Russia

E-mail: trost58@rambler.ru

Abstract. A statistical analysis of the incidence rate of various types of diseases in a population for different age groups was carried out: children (up to 14 years old inclusive), adolescents (15-17 years old), adults (18 years old and older), depending on environmental factors such as concentration of atmospheric impurities, as well as drinking water quality indicators. A dynamic autoregressive model for panel data was used to determine the quantitative dependence of the incidence rate in a population on environmental pollution factors. Panel data on population morbidity indicators, air pollution and the drinking water quality were built on the basis of official statistics for the districts of the Voronezh region and the city of Voronezh over a multi-year period. Considering that the incidence levels in the districts of the Voronezh region and in the city of Voronezh linearly depend on determining factors, linear dynamic models were constructed to describe the level of the disease classes recorded for the year, depending on the level of the corresponding disease classes for the previous year, as well as those registered for the analyzed year of concentrations of chemical impurities in the atmosphere or, accordingly, the values of drinking water quality indicators. As a result of the analysis, statistically significant quantitative dependences on the recorded factors of technogenic environmental pollution were obtained for certain classes of morbidity.

1. Statistical data and methods for their analysis
One of the significant factors that pose risks to public health is environmental pollution with chemical impurities of technogenic origin. To analyze the statistics on the incidence rate of various types of diseases in a population, depending on the concentrations of chemical impurities in the atmosphere and on the values of drinking water quality indicators, panel data were used for districts of the Voronezh region and for the city of Voronezh over a long-term period.

Panel data related to industrial pollution of the atmosphere were analyzed for the period from 2004 to 2011. Data on atmospheric impurity concentrations were compiled on the basis of official statistics of atmospheric air pollution levels, which covered 32 districts of the Voronezh region (without statistics on the city of Voronezh, as not typical on atmospheric pollution of an object among the considered set
of administrative-territorial units) and were registered at 33 route posts of the Voronezh Regional Center for Hygiene and Epidemiology. The main substances (by the number of studies) monitored were: nitrogen dioxide, lead, formaldehyde, sulfur dioxide, suspended solids, phenol, carbon monoxide.

Panel data related to technogenic pollution of drinking water were analyzed for the period from 2004 to 2012. Data on drinking water quality are compiled on the basis of official statistics on monitoring the status of centralized water supply sources in the region recorded by the accredited testing laboratory center Voronezh Regional Center for Hygiene and Epidemiology, in the territories of 32 districts of the region and in the city of Voronezh. Monitoring of the drinking water quality was carried out at 64 monitoring points of control from sources of centralized water supply in a quarterly mode and 181 monitoring points of the distribution network every month. In monitoring points of centralized drinking water supply, the quality of drinking water was determined by priority indicators: the content of iron, manganese, nitrates (by $\text{NO}_3^-$), nitrites (by $\text{NO}_2^-$), total hardness, the content of chlorides (by $\text{Cl}^-$), sulfates (by $\text{SO}_4^{2-}$), total mineralization (dry residue).

Data on the incidence rate in a population are based on official statistics from the Department of Health of the Voronezh Region for 32 districts of the Voronezh region and the city of Voronezh for the period from 2004 to 2012, separately for children (up to and including 14 years old), adolescents (from 15 to 17 years), adults (18 years and older). The statistical data covered the following thirteen classes of diseases for these populations: general morbidity, blood diseases, neoplasms, diseases of the nervous system, diseases of the musculoskeletal system, diseases of the circulatory system, diseases of the genitourinary system, diseases of the endocrine system, skin diseases, diseases of the digestive system, diseases of the reproductive system, respiratory diseases, infectious diseases.

To determine the quantitative dependence of the morbidity rate on the concentration of pollutants in the atmosphere and on the values of drinking water quality indicators, a dynamic autoregressive model for panel data was used [1]. Estimates of parameters in dynamic models were carried out on the basis of a generalized method of moments within the framework of the Arellano-Bond estimator [1, 2]. In this case, the DPD 98 program was used [2].

Considering that the incidence rates in the districts of the Voronezh region and in the city of Voronezh linearly depend on determining factors, linear dynamic models were constructed to describe the level of the disease classes recorded for the year, depending on the level of the corresponding disease classes for the previous year, as well as those registered for the analyzed year of concentrations of chemicals polluting the atmosphere or the values of drinking water quality indicators.

2. Results on the influence of atmospheric impurities on population health

Proceed from the consideration of models linking the incidence rate with concentrations of impurities in the air. Equations of the form (1) for the various analyzed classes of diseases differed by the “$j$” indices for the variables and coefficients, where, respectively: $j = 1$ - the general incidence; $j = 2$ - blood disease; $j = 3$ - neoplasms; $j = 4$ - diseases of the nervous system; $j = 5$ - diseases of the musculoskeletal system; $j = 6$ - diseases of the circulatory system; $j = 7$ - diseases of the genitourinary system; $j = 8$ - diseases of the endocrine system; $j = 9$ - skin diseases; $j = 10$ - diseases of the digestive system; $j = 11$ - diseases of the reproductive system; $j = 12$ - respiratory diseases; $j = 13$ - infectious diseases.

$$Z_{ijt} = a_{ij}Z_{ij(t-1)} + a_{i1}V_{1_{it}} + a_{i2}V_{2_{it}} + a_{i3}V_{3_{it}} + a_{i4}V_{4_{it}} + a_{i5}V_{5_{it}} + a_{i6}V_{6_{it}} + a_{i7}V_{7_{it}} + C_{1_{ij}}$$ (1)

In these equations, the indices $i$ and $t$ denote the area and year, respectively. Index $k$ corresponds to three age groups of the population: $k = 1$ - children (up to 14 years old inclusive), $k = 2$ - adolescents (15-17 years old), $k = 3$ - adult population (18 years and older). Dependent variables in equations (1): $Z_{ij}$ - the number of diseases of the $j$-th class for 1 year per 1000 people. The independent variables in equations (1) correspond to the concentrations of the following atmospheric impurities in [mg/m$^3$]: $V_{1_{it}}$ - sulfur dioxide; $V_{2_{it}}$ - suspended solids (dust); $V_{3_{it}}$ - nitrogen dioxide; $V_{4_{it}}$ - carbon monoxide; $V_{5_{it}}$ - phenol; $V_{6_{it}}$ - formaldehyde; $V_{7_{it}}$ - lead. The constants $C_{1_{ij}}$ in equations (1) correspond to unaccounted...
factors in the corresponding equations. Based on the results of a regression analysis of equations (1), it was found that a statistically significant dependence of the incidence rate of the population on the concentration of atmospheric impurities, with a significance level of error \( p < 1\% \) taking into account the permissible quality of the model, which is characterized by the value of the Wald test \( (\chi^2 > 40) \), manifests itself only for morbidity levels in children in the following classes: general morbidity, diseases of the genitourinary system, blood diseases (table 1).

Table 1. The results of a panel data statistical analysis on the dependence of the rate incidence in the population (up to 14 years inclusive) in the districts of the Voronezh region on the concentration of atmospheric impurities.

| Factors                          | Model 1: general incidence | Model 2: blood diseases | Model 3: diseases of the genitourinary system |
|---------------------------------|-----------------------------|-------------------------|---------------------------------------------|
|                                 | equation (1) for \( Z_{kij} \) at \( k=1, j=1 \) [sick/10^5 people-year] | equation (1) for \( Z_{kij} \) at \( k=1, j=2 \) [sick/10^5 people-year] | equation (1) for \( Z_{kij} \) at \( k=1, j=7 \) [sick/10^5 people-year] |
| \( Z_{kij(t-1)} \) - the number | 0.6777 **                 | 0.484 ***               | 0.646 ***                                  |
| [sick/10^5 people-year]         | \(( p=0.000)\)             | \(( p=0.000)\)          | \(( p=0.000)\)                             |
| \( V_{1j} \) - sulfur dioxide  | 1687.01 ***               | 2.422                   | 71.727 ***                                 |
| [mg/m^3]                        | \(( p=0.008)\)             | \(( p=0.912)\)          | \(( p=0.000)\)                             |
| \( V_{2j} \) - dust [mg/m^3]    | -414.76 ***               | -47.860 ***             | 8.103                                      |
|                                 | \(( p=0.188)\)             | \(( p=0.000)\)          | \(( p=0.375)\)                             |
| \( V_{3j} \) - nitrogen dioxide| -203.95                   | 3.004                   | 12.128                                     |
| [mg/m^3]                        | \(( p=0.464)\)             | \(( p=0.757)\)          | \(( p=0.136)\)                             |
| \( V_{4j} \) - carbon monoxide | -50.024                   | -2.925                  | 0.792                                      |
| [mg/m^3]                        | \(( p=0.401)\)             | \(( p=0.173)\)          | \(( p=0.639)\)                             |
| \( V_{5j} \) - phenol           | -102330.9 *               | -476.40                 | -20.088                                    |
| [mg/m^3]                        | \(( p=0.065)\)             | \(( p=0.809)\)          | \(( p=0.990)\)                             |
| \( V_{6j} \) - формальдегид    | -15672.27                 | -532.66                 | -51.527                                    |
| [mg/m^3]                        | \(( p=0.226)\)             | \(( p=0.249)\)          | \(( p=0.894)\)                             |
| \( V_{7j} \) - lead            | -311374.2                 | -17243.8 *              | 5233.54                                    |
| [mg/m^3]                        | \(( p=0.266)\)             | \(( p=0.076)\)          | \(( p=0.525)\)                             |
| \( C_{1j} \) - constant        | 1196.83 ***               | 42.191 ***              | -1.418                                     |
|                                 | \(( p=0.001)\)             | \(( p=0.00)\)           | \(( p=0.877)\)                             |
| Wald test, \( \chi^2 \)        | 59.83                     | 67.80                   | 45.62                                      |

In table 1 and below, asterisks indicate the statistical significance of the regression coefficients: * - significance at the error level of 10%, ** - significance at the error level of 5%, *** - significance at the error level of 1%. The acceptable quality of the models reflects the values of the Wald test: \( \chi^2_i = 59.83 \) for model 1; \( \chi^2_2 = 67.80 \) for model 2; \( \chi^2_3 = 67.80 \) for model 3.

Table 1 indicates that an increase in the incidence in children is significantly affected by an increase in the concentration of sulfur dioxide in the air. Given that the value of the concentration of sulfur dioxide \( SO_2 \) in the air in the 32 districts under consideration from 2004 to 2011 was on average \( V_{1_{max}} = 0.1282 \) [mg/m^3], (which corresponds to 25.6% of the maximum allowable maximum single concentration of MAC, and 256% of the daily average MAC [3]), the increase in the total incidence per
3. Results of the influence of the drinking water quality on population health

Proceed to consider models linking the level of various classes of morbidity with the values of drinking water quality indicators. Corresponding to these models, equations of the form (2) differ by the indices “j” and “k” for the variables and coefficients, while the indices “j” denote the same classes of diseases that were considered for equations of the form (1), and the indices “k” denote the same age groups of the population that were considered for equations of the form (1).

\[
Z_{ij} = d_{i1}Z_{ij(1-1)} + d_{i2}W_{1j} + d_{i3}W_{2j} + d_{i4}W_{3j} + d_{i5}W_{4j} + d_{i6}W_{5j} + d_{i7}W_{6j} + d_{i8}W_{7j} + d_{i9}W_{8j} + C2j.
\]  

Dependent variables in equations (2): \(Z_{ij}\) – the number of diseases of the \(j\)-th class for 1 year per 1000 people. Independent variables in equations (2) correspond to the values of priority indicators of the quality of drinking water in [mg dm\(^{-3}\)] or in [mg-equivalent/dm\(^3\)], characterized by the content of: \(W_{1j}\) - iron; \(W_{2j}\) - manganese; \(W_{3j}\) - nitrates; \(W_{4j}\) - nitrites; \(W_{5j}\) - total hardness of water; \(W_{6j}\) - chlorides; \(W_{7j}\) - sulfates; \(W_{8j}\) - general mineralization. Constants \(C2j\) in equation (2) correspond to unaccounted factors in the corresponding equations. Obtained as a result of a regression analysis of equations (2), significant statistical dependences of the levels of the analyzed classes of diseases for various age groups of the population on the concentration of chemical impurities and the quality of drinking water indicators are presented in table 2.

Table 1 indicates that there is a significant relationship between an increase in the level of diseases of the endocrine system in children and an increase in the concentration of nitrates in drinking water. Given that the value of the concentration of nitrates in drinking water for the period from 2004 to 2012 in the districts of the Voronezh region and in the city of Voronezh averaged \(W_{3j}\) =14 [mg/dm\(^3\)] (which corresponds to 31.1% of the MAC), associated with such a concentration of nitrates, the increase in endocrine system diseases per year per 1000 children averaged: \(AZ_{18j} = 10.24\) [sick/10\(^3\) people-year]. These conclusions are consistent with the results obtained in [6].

Table 2. The results of a statistical panel data analysis on the dependence of the incidence rate in various age groups in the areas of the Voronezh region and in the city of Voronezh on the values of priority drinking water quality indicators.
From the results of table 2, a statistically significant relationship was found between a decrease in the level of blood diseases in children and an increase in the concentration of iron in drinking water. Considering that the value of iron concentration in drinking water for the period from 2004 to 2012 in the districts of the Voronezh region and in the city of Voronezh averaged $W_{1_{av}} = 0.189$ [mg/dm$^3$] (63% of the MAC), associated with this concentration of iron, the decrease in blood diseases per year per 1000
The study was carried out with the financial support of the Russian Foundation for Basic Research, project No. 20-05-00779yu.

4. Conclusion
A study of the influence of environmental pollution factors on public health, conducted on the basis of statistical analysis of panel data. For the age group of the population (up to 14 years old inclusive) significant quantitative dependencies were obtained: 1) for the level of the general incidence of the concentration of sulfur dioxide in the atmosphere; 2) for the level of blood diseases: on the concentration of suspended particles in the atmosphere; on the concentration of iron in drinking water; 3) for the level of diseases of the endocrine system on the concentration of nitrates in drinking water; 4) for the level of diseases of the genitourinary system on the concentration of sulfur dioxide in the atmosphere. For the age group of the population (from 18 years and older) significant quantitative dependencies were obtained: 1) for the level of blood diseases on the concentration of nitrates in drinking water; 2) for the level of diseases of the digestive system on the concentration of iron in drinking water.

Acknowledgments
The study was carried out with the financial support of the Russian Foundation for Basic Research, project No. 20-05-00779yu.

References
[1] Arellano M and Bond S 1991 Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations Review of Economic Studies 58 277–97
[2] Arellano M and Bond S 1998 Dynamic Panel Data Estimation Using DPD98 for Gauss: a Guide for Users (Institute for Fiscal Studies, London) p 46
[3] Federal Law of the Russian Federation “On the Sanitary and Epidemiological Well-Being of the Population” No. 52-FZ of March 30 1999
[4] Bochkaryov V V 2012 Theoretical foundations on environmetal protection technological processes (Tomsk: TPU) p 320
[5] Shakirov F M, Koziy S S and Koziy T B 2009 Assessment of dustiness of the air and the effects of dust on the human body (Samara: Samara State Aerospace University) p 36
[6] Gatseva P D and Argirova M D 2005 Iodine status of children living in areas with high nitrate levels in water Arch. Environ. Occup. Health. 6 317-19
[7] Beinner M A, Lamounier J A and Tomaz C 2005 Effect of iron-fortified drinking water of daycare facilities on the hemoglobin status of young children J. Am. Coll. Nutr. 24(2) 107 - 14
[8] Klein S V, Vekovshchina S A and Shoev A S 2016 Priority risk factors for drinking water and associated economic damage Hygiene and Sanitation 95(1) 10-4
[9] Aamodt G, Bulholm G, Jahnsen J et.al. 2008 The association between water supply and inflammatory bowel disease based on a 1990 - 1993 cohort study in southeastern Norway *Am. J. Epidemiol.* **168**(9) 1065-72