SCIENTIFIC ARTICLES

Estimation of morbidity dynamics among emergency workers based on the results of the annual health examinations

Mikhalsky A.I., Ivanov V.K.*, Maksyutov M.A.*, Morgenstern W.**

Institute of Control Science of RAS, Moscow;
* - Medical Radiological Research Center of RAMS, Obninsk;
** - Clinical Social Medicine, University Hospital, Heidelberg, Germany

The problem of estimation of morbidity among the emergency workers who, participated in the clean-up work on the Chernobyl nuclear power station after the 1986 accident, is considered taking into account the irregular participation of these workers in the annual health examinations. The link with the problem of the unobserved morbidity process estimation on the base of the diagnosis registration process is discussed. A methodology for morbidity estimation, stable in respect to the random fluctuations, is described.

The methodology has been used for of morbidity dynamics estimation among emergency workers during 1986-1993 period in 12 standard classes of diseases. Results show insignificant changes during this period for the class “Infectious and parasitic diseases”, a fourfold increase for class “Mental disorders” and an almost tenfold increase for “Diseases of the nervous system and sense organs”. The estimated increase is less than the “observed morbidity” increase because of the “accumulated morbidity” effect.

The work is devoted to the problem of morbidity dynamics estimation among participants in the clean up operations after the accident on the Chernobyl Nuclear power station. The morbidity estimates are based on the results of the annual health examinations of the cohort, registered in the Russian National Medical-Dosimetric Registry. The approach uses the mathematical methods for estimation of the stochastic processes in the presence of uncertainty: irregular participation in the annual health examinations of the registered people, observation not of the morbidity process, but the process of the disease registration, lack of the prior information about the morbidity dynamics model. These uncertainties force to use methods for estimation of unobserved processes and methods for optimal model selection on he base of the limited amount of empirical data.

Theoretical background

Theoretically morbidity is a rate of a person transition from the health state to the sick state. This rate depends on the age of the person x, time t, vector of different risk factors r. For the morbidity estimation it is more convenient to use not the transition rate but the probability to stay healthy during a time interval [t_1, t_2] under the condition, that in time t_1 the person of age x_0 was healthy. Using notation for the transition rate $\mu(x,t,r)$ one can write this probability as

$$S(x_0,r,t_1,t_2) = exp(-\int_{t_1}^{t_2} \mu(\tau - t_1 + x_0,\tau,r)d\tau).$$

For the small value of the exponent index this probability can be expressed as

$$p(x_0,r,t_1,t_2) = 1 - S(x_0,r,t_1,t_2) \approx$$
$$\approx \int_{t_1}^{t_2} \mu(\tau - t_1 + x_0,\tau,r)d\tau.$$

The expressions above give the relationship between the transition rate and the probability to become sick for a single person. In reality the morbidity estimation is made using the results of health examinations of a group of people, where every person has its own set of risk factors. In the result the estimated morbidity reflects the averaged morbidity for the given group of people, which depends on each individual rate of transition, and therefore it depends upon the group composition. Further, the morbidity is related not to the time moment, but to the time interval. For example, one can speak about the morbidity in a given year or in a five year interval. This means, that in the expression for the conditional probability to become sick during a time interval one should use the rate of transition, averaged over the distribution of the risk factors among healthy group members in the time interval [t_1, t_2]. The probability of the disease in the time interval takes the form

$$p(t_1,t_2) = (t_2 - t_1)\mu_{t_2} \; ,$$

where μ_{t_2} - the rate of transition, averaged over the distribution of the risk factors on the time interval [t_1, t_2]. The last expression one can consider as a definition for the morbidity on the time interval [t_1, t_2] in the given group.

The results of the annual health examinations of the emergency workers cohort are the diagnosed cases of the diseases. In this article for any person only the first diagnosed case is considered. This means that we are estimating the incidence rate. The probability to diagnose the disease in the given year can be estimated by the ratio between the number of diagnosed cases in this year investigated during the year. This estimate is not directly the estimate for the probability to become sick during the year. Different people, investigated in a particular year, could become sick during the year of and previous to the last health examination, when the person has been considered to be healthy, and the next examination. This means, that the probability of the disease diagnosis depends as on the morbidity dynamics of the previous years, and therefore upon the practice of participation in the annual health examinations. The real number of the new cases each year is unknown. So the problem of the morbidity dynamics estimation on the results of the annual health examinations leads to the problem of the unobserved process estimation - incidence rate, using observations of the other process - new cases of the disease registration during the annual health examinations.

Relationship between the probability of the disease onset and the probability of the disease detection

Consider the case, when a person has been examined at time t_{k1} and the disease has not been detected. After this the person skipped the health examinations at times t_j, ***k1<j<k2***, and had the next health examination in time t_{k2}. The probability to find the disease in time t_{k2} in this case is equal to the probability to become sick in the time interval [t_{k1}, t_{k2}] under the condition, that the person has been healthy in time t_{k1}. This probability equals to

$$p(t_{k1},t_{k2}) = \sum_{j=k1+1}^{k2} (t_j - t_{j-1})\mu_j .$$

The probability to find the disease in time t_{k2} among N_{k2} people, which have been healthy in the last health examination is

$$p(t_{k2}) = \sum_{i=1}^{N} \frac{1}{N_{k2}} p(t_{k1},t_{k2}) =$$
$$= \frac{1}{N_{k2}} \sum_{i=1}^{N} \sum_{j=k1i+1}^{k2} (t_j - t_{j-1})\mu_j ,$$

where t_{k1i} is the time of the last health examination of the i-th person, when the disease has not been found.

The probability to find the disease in time t_i one can rewrite in form

$$\sum_{j=1}^{k2-1} (t_j - t_{j-1}) n_{ij}\mu_j = N_i p(t_i), \qquad (1)$$

where n_{ij} is the number of people among N_i people, examined in time t_i, which has had the last health examination before time t_j, $t_j \le t_i$, during which the disease has not been found. For convenience the starting point of the follow-up is denoted as t_0. It is important to say, that in the frame of the model it is supposed, that at the very beginning of the follow-up period t_0 all members of the emergency workers cohort has been healthy. To fit this supposition to the reality one is to construct the morbidity estimates only for those people, which has no diseases detected at the moment of the registration in the registry. If a person had the first health examination several years after time t_0 one can consider this as a case of skipping all previous health examinations. So, the approach described can be used as in the case of the follow-up drops out, so in the case of the follow-up drops in.

Equation (1) for different times t_i composes the system of linear equations. The element of the system matrix, corresponding to the i-th line and j-th column ($j < i$) is the number of people, examined in the time t_i and which could became sick in the time interval [t_{j-1}, t_i]. The right part of the system is the vector of the mathematical expectations for the number of diagnosis, found in different years. For the morbidity estimation one has to use the numbers of diagnosis instead of unknown mathematical expectations. The solution of the linear system is the vector of yearly morbidities, which eliminates the influence of the irregular health examinations participation on the morbidity estimation.

It is interesting to compare the solution of the system (1) with the " observed morbidity" - the ratio between the number of the first diagnosed cases of the disease in the year of examination and the number of examined in this year people. The " observed morbidity" is the solution of the system (1) in a case of diagonal matrix, which in tern is the case of regular 100% participation in yearly health examinations. It is easy to see, that in the case of irregular participation in yearly health examinations the " observed morbidity" is higher, than the system (1) solution. This is a result of the " morbidity accumulation" effect resulting from the undiagnosed sick cases because of skipped health examinations.

The other morbidity estimate - the ratio between the number of the first diagnosed cases of the disease in the all period of the follow-up and the person-years under the risk can be derived from the system (1) as well. To do this one can suppose, that the morbidity estimates for different years of health examinations should be equal. The solution of the system then is the ratio of the sum of the right part vector elements to the sum of the system matrix elements. This estimate reflects the averaged in the whole period of the follow-up morbidity in contrast to the estimate, described in the article, which obtains the dynamics of the morbidity.

The morbidity estimate stabilization

The morbidity estimate is the solution of the system of linear equations (1) with the number of new diagnosed cases in different years as a vector in the right part of the system. The number of diagnosed cases is effected by the stochastic fluctuations, which causes irregular changes in the system solution.

To diminish these changes in the solution of the system - to stabilize the estimate, one can use the regularization technique [1] combined with the procedure for the regularization parameter value selection on the base of empirical data of limited size. To do this one has to apply a restriction on the vector-solution μ of the system (1) in form

$$\Psi(\mu) = \|B\mu\|^2 = \mu^T B^T B\mu \le \gamma . \qquad (2)$$

Matrix B is selected in such a way, that the stabilization functional $\Psi(\mu)$ gets low values for not mush changing functions and get large values for high oscillating functions. If one takes the value of increase in morbidity estimate in subsequent years as a measure for the estimate fluctuations, the matrix B can be constructed as two diagonal matrix with $m-1$ lines and m columns

$$B = \begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & . & . & . & \\ & & & -1 & 1 & \\ & & & & -1 & 1 \end{bmatrix},$$

where m - the total number of the health examinations. It is clear, that if the morbidity estimates do not change for different years, the value of the stabilization functional is zero. In the case of high fluctuations in the morbidity estimate this functional gets large values. The restriction value γ controls the degree of fluctuations in the system (1) solution. It is easy to rewrite the problem of the system (1) solution under the restriction (2) in the form of the unconditional minimization problem

$$\|Y - C\mu\|^2 + \alpha\|B\mu\|^2 \xrightarrow[\alpha]{} \min, \qquad (3)$$

where Y - vector in the right part of the system (1), elements of matrix C are calculated in correspondence with the formula $C_{ij} = (t_j - t_{j-1})n_{ij}$, value of parameter α - the regularization parameter, has to be in consistency with the value of random disturbance in the right part of the (1) and the smoothness property of the solution.

In practice the techniques are used for selection of the regularization parameter α value, which provide the stable solutions in the case of the infinite small disturbances in the data [2]. In the case of finite disturbances in the empirical data one can select the value of the regularization parameter using the principle of optimal model selection [3]. To implement this principle consider the solution of the problem (3) under the fixed value of the regularization parameter α as a result of an operator A_α action on the random vector of the system (1) right part. The operator A_α is defined by the problem (3) and can be written as the matrix operator

$$A_\alpha = (C^T C + \alpha B^T B)^{-1} C^T .$$

For $\alpha = 0$ this operator corresponds to the least squares method.

The result of an operator A_α action on the random vector of the system (1) right part is the morbidity estimate and the operator A_α can be considered as a " morbidity model" . The set of operators A_α for different values of the parameter α forms the class of F. In this terms the problem of the parameter α selection can be considered as a problem of a model selection, which is more adequate to the empirical data, in the set $\boldsymbol{F}$.

To formalize the model selection procedure in the set $\boldsymbol{F}$ one is to define the model performance functional as the mean value of the loss function

$$J_\alpha = M\|Y - CA_\alpha Y\|^2. \quad (4)$$

The averaging is made on the random vector $\boldsymbol{Y}$ of the system (1) right part distribution - distribution of he number of the disease diagnoses. The functional J_α has a meaning of the mathematical expectation of the error, which produces the model for a fixed value of the parameter α in the respect to the empirical data. The distribution of the vector $\boldsymbol{Y}$ is unknown. In this case one has to use not just the functional (4), but its estimate, derived in [3] using the functional of empirical losses. The last functional is the performance of the model, calculated on the same data, which have been used for the morbidity estimation. The specific feature of the estimate, derived in [3], is that it is a uniform estimate. In the other words this estimate is valid with the given probability for all models from the set $\boldsymbol{F}$ at the same time. The last means, that this estimate can be used to select the model, which supplies in the model set $\boldsymbol{F}$ the “guaranteed minimum” value for the functional J_α. The condition for selection of such a model is formulated in [3] as minimization, in respect to parameter α, the criterion

$$K_\alpha = \frac{J_{e\alpha}}{1 - 2sp(CA_\alpha)/m},$$

where $J_{e\alpha} = \|Y - CA_\alpha Y\|^2$ - the square residual value for the problem (3) solution under the fixed value of parameter α. Minimization of the criterion K_α is made for the values of the regularization parameter, which satisfy the restriction $0<2sp(CA_\alpha)<m$.

Results of the morbidity dynamics estimation among emergency workers

The approach, described in the previous part of the article, has been used to analyze the morbidity dynamics for different classes of diseases in the period 1986-1993 among participants in the clean up operations after the accident on the Chernobyl Nuclear power station - emergency workers, living in Russia. The analysis has been conducted in the frame of the collaborated research methodological project “Effects of Radiation on Health - Risk and Projection Models”, run by the Russian Academy of Sciences, the Russian Academy of Medical Sciences and the Heidelberg Academy for the Humanities and Sciences (Germany) [4]. The data represented a random sample of records about 11043 emergency workers, registered in the Russian National Medical-Dosimetric Registry [5]. This information includes the registration data, results of the annual health examinations from 1986 till 1993, disability information. In the analysis only results of the annual health examinations and information about chronic diseases at the time of registration have been used. Dosimetric information, time and duration of work in 30-km zone and disability information have been skipped from the analysis.

The morbidity dynamics estimation has been produced for 12 classes of diseases, listed in the Table 1.

In the Table 2 for the 12 classes of diseases are given the numbers of persons, examined in every year, which have not had the disease before, the number of the first diagnosed cases of the disease in the year, the “observed morbidity” per 100,000 population. It is to be mentioned, that the table is constructed using the information only about 11043 emergency workers, randomly sampled from the Russian National Medical-Dosimetric Registry data base. This is the reason, why the numbers in the table does not reflect the state of the whole Registry. They are provided for the morbidity estimation better understanding.

The results of the morbidity dynamics estimation for the mentioned above classes of diseases per 100,000 population are presented in Table 3.

Figures 1 - 12 show the “observed morbidity” (dashed curve) and the morbidity estimate, for each of the 12 classes of diseases obtained using the described approach, (solid curve).

Table 1

Classes of diseases

N	Class of diseases	ICD-9 codes
1	Infectious and parasitic diseases	001 - 139
2	Neoplasms	140 - 239
3	Malignant neoplasms	140 - 208
4	Endocrine, nutritional and metabolic diseases	240 - 279
5	Diseases of blood and blood-forming organs	280 - 289
6	Mental disorders	290 - 319
7	Diseases of the nervous system and sense organs	320 - 389
8	Diseases of the circulatory system	390 - 459
9	Diseases of the respiratory system	460 - 519
10	Diseases of the digestive system	520 - 579
11	Diseases of the genitourinary system	580 - 629
12	Diseases of the skin and subcutaneous tissue	680 - 709

Table 2

Number of examined people without the disease before, number of the first diagnosed cases, " observed morbidity" per 100,000 population for each class

	1986	1987	1988	1989	1990	1991	1992	1993
				Class 1				
# of people examined w/o disease before	549	2942	6319	6441	7043	6984	7424	7264
# of the first diagnosed cases	0	1	19	24	32	36	44	49
" Observed morbidity" per 100,000/class	0	34	301	373	454	515	593	675
				Class 2				
# of people examined w/o disease before	552	2950	6336	6456	7069	7012	7438	7259
# of the first diagnosed cases	0	1	13	24	33	50	60	72
" Observed morbidity" per 100,000/class	0	34	205	372	467	713	807	992
				Class 3				
# of people examined w/o disease before	554	2955	6343	6472	7100	7061	7512	7367
# of the first diagnosed cases	0	1	3	5	6	11	14	24
" Observed morbidity" per 100,000/class	0	34	47	77	84	156	186	326
				Class 4				
# of people examined w/o disease before	547	2941	6314	6402	6952	6807	7104	6682
# of the first diagnosed cases	0	8	51	90	153	230	383	462
" Observed morbidity" per 100,000/class	0	272	808	1410	2200	3380	5390	6910
				Class 5				
# of people examined w/o disease before	549	2946	6339	6465	7090	7037	7472	7316
# of the first diagnosed cases	0	0	9	9	18	27	27	24
" Observed morbidity" per 100,000/class	0	0	142	139	254	384	361	328
				Class 6				
# of people examined w/o disease before	546	2912	6223	6260	6717	6470	6718	6181
# of the first diagnosed cases	3	39	118	199	323	342	474	503
" Observed morbidity" per 100,000/class	549	1340	1900	3180	4810	5290	7060	8140
				Class 7				
# of people examined w/o disease before	517	2855	6096	6107	6540	6294	6468	5849
# of the first diagnosed cases	0	7	147	199	277	390	723	1008
" Observed morbidity" per 100,000/class	0	245	2410	3260	4240	6200	11200	1720
				Class 8				
# of people examined w/o disease before	527	2885	6181	6252	6726	6569	6838	6449
# of the first diagnosed cases	0	12	78	163	198	260	381	452
" Observed morbidity" per 100,000/class	0	416	1260	2610	2940	3960	5570	7010
				Class 9				
# of people examined w/o disease before	523	2838	6123	5997	6179	5750	5750	5272
# of the first diagnosed cases	0	23	293	511	550	645	683	657
" Observed morbidity" per 100,000/class	0	810	4790	8520	8900	11200	11900	12500
				Class 10				
# of people examined w/o disease before	520	2825	6070	6153	6593	6377	6553	6167
# of the first diagnosed cases	0	6	64	186	249	328	496	627
" Observed morbidity" per 100,000/class	0	212	1050	3020	3780	5140	7570	10200
				Class 11				
# of people examined w/o disease before	546	2941	6311	6430	7035	6966	7357	7141
# of the first diagnosed cases	0	2	19	26	51	75	116	167
" Observed morbidity" per 100,000/class	0	68	301	404	725	1080	1580	2340
				Class 12				
# of people examined w/o disease before	546	2940	6312	6419	7009	6911	7323	7144
# of the first diagnosed cases	0	1	30	46	69	77	92	80
" Observed morbidity" per 100,000/class	0	34	475	717	984	1110	1260	1120

Table 3

The morbidity dynamics estimation for the Classes of diseases per 100,000 population

	1986	1987	1988	1989	1990	1991	1992	1993
Class 1	36	96	197	276	325	360	388	414
Class 2	20	76	180	297	393	499	564	621
Class 3	13	24	40	62	85	119	150	184
Class 4	96	335	764	1340	2020	2850	3740	4300
Class 5	15	44	96	140	191	229	226	218
Class 6	621	9487	1580	2550	3380	3930	4540	4930
Class 7	232	790	1810	2880	4100	5850	8110	9890
Class 8	183	537	1150	1910	2450	3090	3770	4250
Class 9	645	1770	3730	5630	6390	6950	7010	7110
Class 10	82	487	1270	2350	3210	4200	5290	6100
Class 11	34	112	253	424	646	903	1180	1410
Class 12	46	160	365	556	686	747	756	726

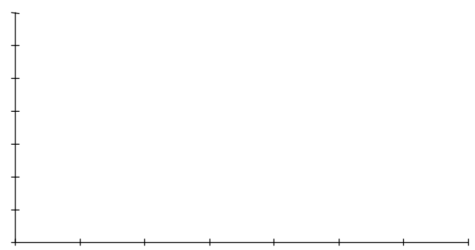

Figure 1. ICD-9: 001-139.

Figure 2. ICD-9: 140-239.

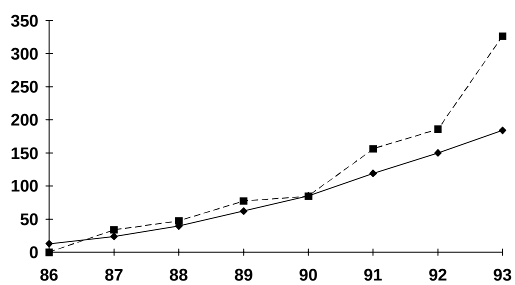

Figure 3. ICD-9: 140-208.

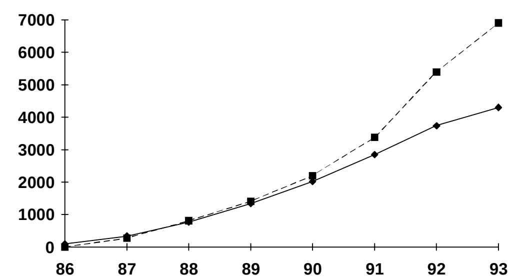

Figure 4. ICD-9: 240-279.

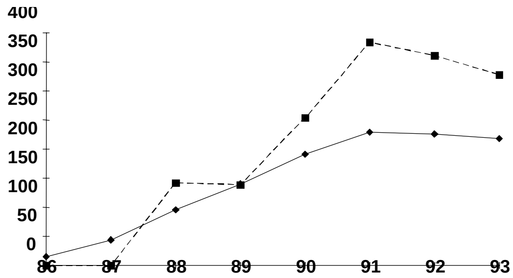

Figure 5. ICD-9: 280-289.

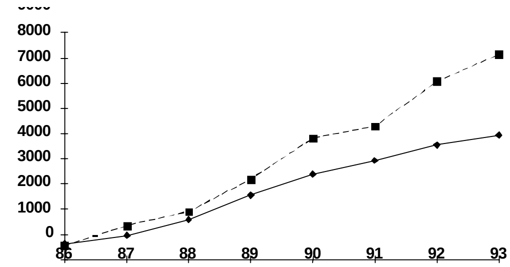

Figure 6. ICD-9: 290-319.

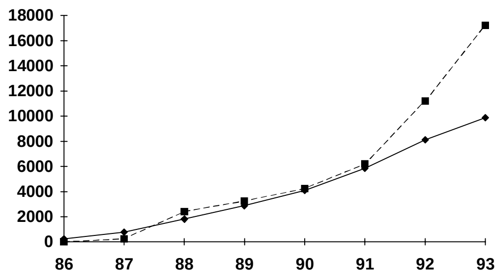

Figure 7. ICD-9: 320-389.

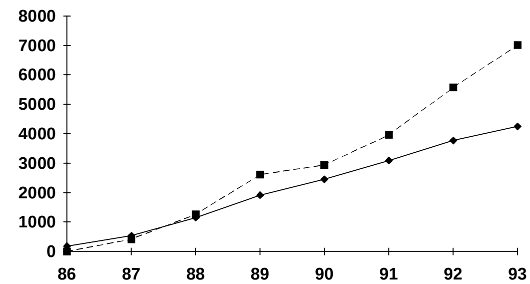

Figure 8. ICD-9: 390-459.

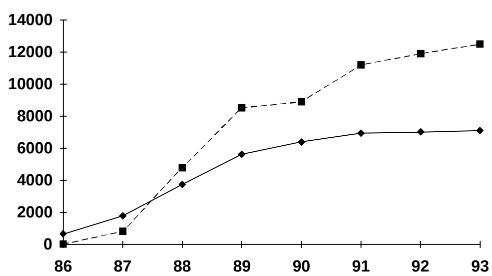

Figure 9. ICD-9: 460-519.

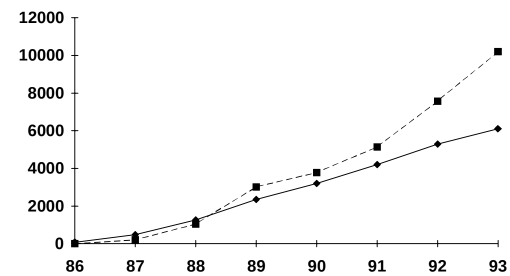

Figure 10. ICD-9: 520-579.

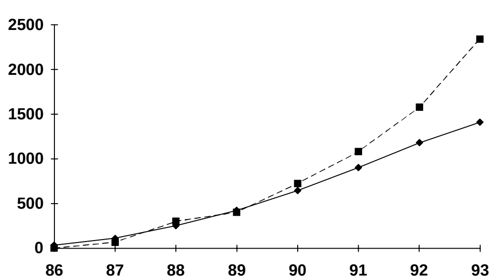

Figure 11. ICD-9: 580-629.

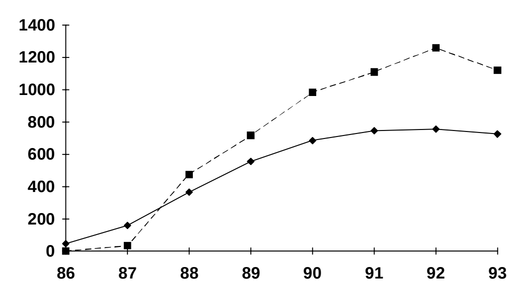

Figure 12. ICD-9: 680-709.

The exceeding of the “ observed morbidity” over the morbidity estimate for all classes of diseases demonstrate, that the approach described in the article adjusts the phenomena of the “ morbidity accumulation” between the health examinations, which is the result of the skipping of some examinations by some people. The excess is as higher as higher is the real morbidity.

Figure 13 shows morbidity estimates for three classes of diseases: “ Infectious and parasitic diseases” (the lower curve), “ Mental disorders” (the middle curve) and “ Diseases of the nervous system and sense organs” (the upper curve).

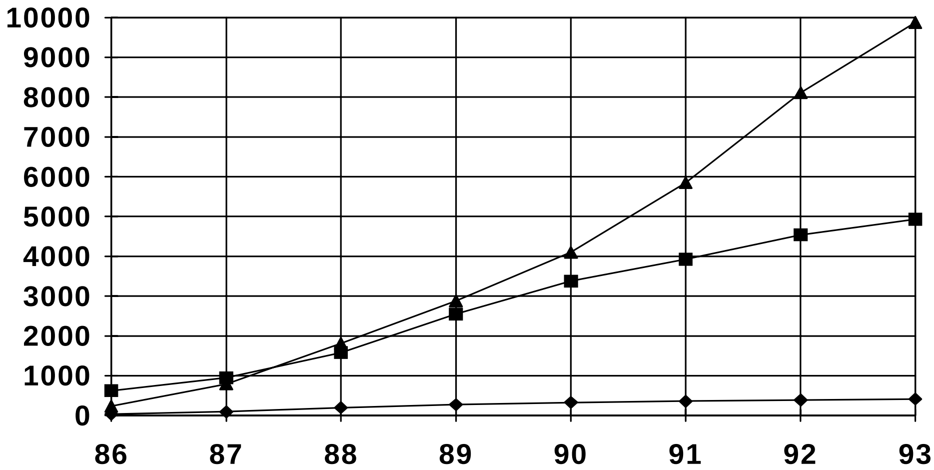

Figure 13. Morbidity estimates among emergency workers for three classes of diseases.

From the chart one can see, that the morbidity in the class “ Infectious and parasitic diseases” changes in time insignificant, the morbidity in the class “ Mental disorders” during 1986-1993 has grown up approximately 4 times, morbidity in the class “ Diseases of the nervous system and sense organs” for the same period has grown up 10 time and tends to grow in the future.

Conclusion

The consideration allows one to make a conclusion about the effectiveness of our approach, described in the article, for estimation of the morbidity dynamics on the results of the annual health examinations. It is demonstrated, that the “ observed morbidity” exceeds the proposed estimates. The combination of the method for regularization of the estimates and the method for regularization of parameter selection on the limited amount of empirical information gives the morbidity estimates, which are stable in respect to the stochastic fluctuation in empirical data.

Morbidity estimates for 12 classes of diseases, calculated on the sample of data from the Russian National Medical-Dosimetric Registry in 1986-1993, show the different dynamics of morbidities in this period. This can either be a result of the different risk factors that are present, or of improvements made in the diagnostsis of the diseases.

References

1. **Tikhonov A.N., Arsenin V.Ya.** Methods of solutions of ill-posed problems. Moscow: Nauka, 1979 (in Russian).
2. **Morozov V.A**. On selection of regularization parameter for solution of functional equations by the regularization method//DAN SSSR. 1967, N6., P. 175 (in Russian).
3. **Mikhalsky A.I.** Choosing an algorithm of estimation based on samples of limited size. Automatization and Remote Control. 1987, Vol. 48, N 7., P. 91-102 (in Russian).
4. Mathematical Modelling with Chernobyl Registry Data/Eds. Morgenstern W. et al., Springer-Verlag, 1995.
5. **Tsyb A.F., Ivanov V.K., Airapetov S.A., Gagin Eu.A., Maksyutov M.A., Rozhkov O.V., Stadnik O.E., Saakyan A.K. and Chekin S.Yu.** Hardware and software architecture of the All-Russia Medical and Dosimetric State Registry//Radiation and Risk. - 1992. - Issue 1. - P. 132-146 (in Russian).