Radiation epidemiological studies in Russian National Medical and Dosimetric Registry with respect to Chernobyl accident: I. Cancer and non-cancer registration system among Chernobyl liquidators and the population living in the areas contaminated with radionuclides in Russia (history and the present status)

Ivanov V.K., Tsyb A.F., Maksioutov M.A., Pitkevich V.A., Birukov A.P., Matiash V.A., Korelo A.M., Chekin S.Yu., Nilova E.V.

Medical Radiological Research Center of RAMS, Obninsk

In May 1986 Ministry of Health of the former USSR made a decision about establishing the All-Union Distributed Registry of people exposed to radiation due to Chernobyl accident (AUDR). AUDR was operating until January 1, 1992, time of dissolution of the USSR. Individual medical and dosimetric information on 65,900 of exposed persons including 285,000 of clean-up workers came in Obninsk from all republics of the USSR. The basic principles of building of AUDR were further explored for building National Chernobyl registries in Belarus, Russia, the Ukraine and other independent states of the former USSR.

By beginning of 1992, national registries in Belarus, Russia and the Ukraine replaced the All-Union Distributed Registry. In Russia it was replaced by the Russian National Medical and Dosimetric Registry (RNMDR) and the Medical Radiological Research Centre, Obninsk, continued to be the central, responsible institution for the Russian registry.

At present RNMDR comprises data on 530,965 people from throughout the Russian Federation. All the RNMDR registries are divided into four primary registration groups (PRG): PRG 1 - liquidators (32,9%); PRG 2 - evacuated from 30-km zone (1,7%); PRG 3 - residents of contaminated territories (61,3%); PRG 4 - children born of people included in groups of higher risk (4,1%).

A large number of people included in the RNMDR reside in the Central and North Caucasian regions of Russia. In the Central region residents of the areas under surveillance in Bryansk, Kaluga, Tula and Orel oblasts constitute the majority, while in the North-Caucasian region liquidators prevail.

The majority of liquidators of 1986 and 1987 received doses exceeding 5 cGy (83.1% and 75.3% of the total number of liquidators who worked in the zone in 1986 and 1987, respectively), whereas in the following years their number did not exceed 15 % (for example, 1988 - 14.1%, 1989 - 4.5%, 1990-1993 - 4.4%). The average external dose for liquidators of 1986 is equal to 15.9 cGy, of 1987 - 9.0 cGy, of 1988 - 3.3 cGy, of 1989 - 3.2 cGy, of 1990 - 3.7 cGy.

The RNMDR is a multi-level information system. Primarily it is integrated in the existing health care system. In principle primary health care is provided at the levels of uchastoks (local), rayon's (district) and oblasts (province). Regional centers have been established for compiling and processing data from the oblast level. Thereby, four hierarchical levels exist federal, regional, oblast and rayon.

Regional centers have been set up for each of the 11 administrative, so-called economical regions of Russia. They have to be regarded as decentralized units of the central registry as their main task is to channel the information flow from the local to the federal level. According to the normal Russian administration these regions are further subdivided into republics, oblasts and krays. At present there are altogether 89 units at this level. These units are further subdivided into rayon's, altogether more than 2,000 in the Russian Federation. Four of the oblast units (Bryansk, Kaluga, Orel and Tula oblasts) have at the same time the logistical status of a regional center. These oblasts represent the most contaminated geographical areas in Russia.

The software for all levels of surveillance is used for computerization of primary information collection, uniform data base support and statistical and radio-epidemiological analysis. Thus the software employs advanced computer techniques to collect, store, process, transfer and analyze demographic, medical, dosimetric and epidemiological data and ensures the reliability of software products at all levels of surveillance, as well as their lack of complexity and ease of use for medical personnel.

The software falls into two functional parts - database support and analysis of information collected. Databases of large-scale and local registries store information in different ways, i.e. these bases are of different physical structure. In order to make the analytical part of the software flexible and multipurpose, i.e. independent of the specific features of particular bases, access to them should be uniform.

The system of statistical and radio-epidemiological data analysis (SDA) enables the basic epidemiological indexes to be calculated in the interactive mode. It also provides thorough statistical and radio-epidemiological analysis of data from the RNMDR primary documents. Standard and specialized software in use are integrated into a unified information system of storage and epidemiological, medical and statistical analysis of data.

1. History of organizing and structure of the Registry

Immediately after Chernobyl accident in May 1986, Ministry of Health of the former USSR convened an enlarged conference of experts in organizing dispensary check-ups of people exposed to radiation due to Chernobyl accident [1]. The conference issued a resolution on establishing the All-Union Distributed Registry of people exposed to radiation (AUDR). The registry was intended for information support or the dispensary check-ups in order to provide the necessary effective primary health care and clinical medical treatment as well as ensuring appropriate long-term radiological and epidemiological monitoring of the population exposed to radiation in order to evaluate the effects of radiation on health. The AUDR was a multi-level hierarchic information system. Medical and dosimetric information was collected at the rayon (district), oblast (province) and republic level. The information collected was passed on to the national level. The Medical Radiological Research Center of Russian Academy of Medical Sciences (Obninsk) was the central, responsible institution at this level. These issues were regulated by a governmental orders and orders of Ministry of Health of the USSR.

AUDR was operating until January 1, 1992, time of dissolution of the USSR. Individual medical and dosimetric information on 659 thousand of exposed persons including 285 thousand of clean-up workers came in Obninsk from all republics of the USSR. The basic principles of building of AUDR were further explored for building National Chernobyl registries in Belarus, Russia, the Ukraine and other independent states of the former USSR.

By beginning of 1992, national registries in Belarus, Russia and the Ukraine replaced the All-Union Distributed Registry. In Russia it was replaced by the Russian National Medical and Dosimetric Registry (RNMDR) and the Medical Radiological Research Center, Obninsk, continued to be the central, responsible institution for the Russian registry.

This is regulated by a governmental order (The Ministry of Health of the Russian Federation, Order N 173, 4 April 1991). It should be mentioned that these types of orders have to be regarded as the normal procedure because of the organization of the health care system in Russia as a state system. The major areas of activities defined in this order may be summarized as:

- Medical screenings of the population;
- Accompanying medical treatment and prophylactics;
- Assessment of morbidity structures, dynamics, trends and outcomes of the registered cohorts;
- Radio-epidemiological analyses to assess the dose-dependency of morbidity/mortality as well as to forecast long-term radiation effects in populations;
- Setting up of guidelines at the different levels of the RNMDR level in order to improve diagnosis, treatment and prognosis of diseases and thereby to improve the primary health care as a whole;
- Supporting and implementation of specific research programs of consequences of the Chernobyl accident.

The population groups (primary registration groups) to be registered have been defined as:

- Persons involved in cleanup works at the Chernobyl nuclear power station;
- Persons who left or were evacuated from the 30-km zone around the station;
- Residents of territories officially declared as areas of special surveillance;
- Natural children of registered persons.

At the end of 1993, the Government of Russia passed the act on the establishment of a uniform system of radio-epidemiological registration of people exposed to and affected by radiation due to the Chernobyl and other radiological accidents and incidents (The Government of the Russian Federation, Decree N 948, 22 September 1993), i.e. the registry population will now include any person having been incidentally exposed to radiation.

The RNMDR is a multi-level information system [2, 3]. Primarily it is integrated in the existing health care system. In principle primary health care is provided at the levels of uchastoks (local), rayon's (district) and oblasts (province). The lowest level of central policlinics is the rayon level. Thereby, it is the lowest one as well for the specific physical examinations conducted according to the given standards and guidelines of the registry. Data of the units at the rayon level are compiled at the responsible oblast hospital. Regional centers have been established for compiling and processing data from the oblast level. These are finally processed and documented at the central registry institution of the Medical Radiological Research Center, Obninsk. Thereby, four hierarchical levels exist:

- Federal,
- Regional,
- Oblast,
- Rayon.

Furthermore, there are five so-called departmental registries established at Russian ministries (Figure 1). This structure is basically followed by data reporting, compilation, processing and documentation procedures and is used for the normal, routine registry procedures as well (e.g. error checking procedures, education and training of medical staff, quality control procedures and similar). The rayon and oblast levels of monitoring are the primary links of the RNMDR multi-level information system as the annual medical check-ups (dispensary check-ups) are conducted there. They are organized and financed by the Ministry of Health of the Russian Federation.

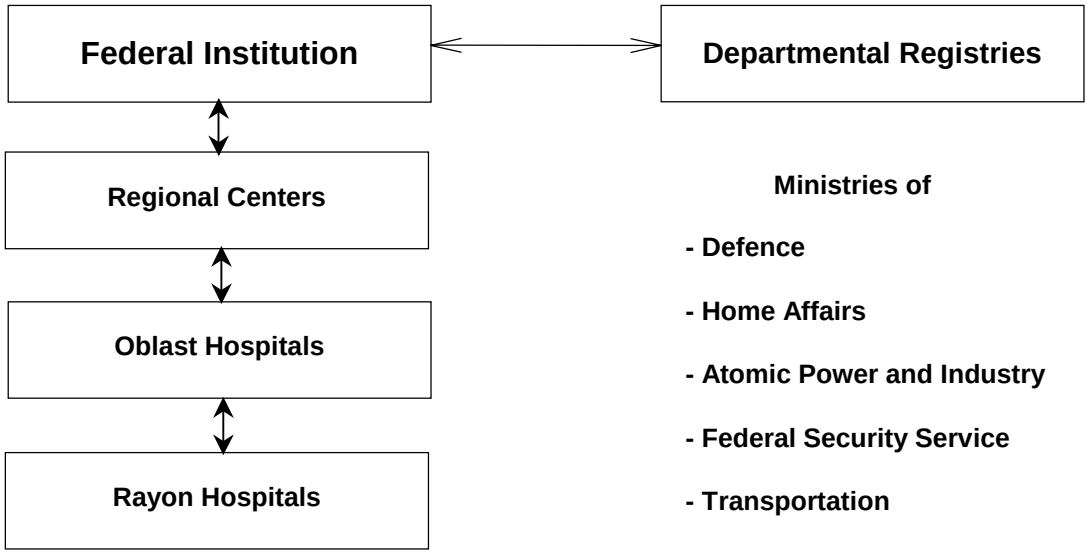

Fig. 1. The multi-level hierarchy of the Russian National Medical and Dosimetric Registry.

Regional centers have been set up for each of the 11 administrative, so-called economical regions of Russia. They have to be regarded as decentralized units of the central registry as their main task is to channel the information flow from the local to the federal level. According to the normal Russian administration these regions are further subdivided into republics, oblasts and krays. At present there are altogether 89 units at this level. These units are further subdivided into rayon's, altogether more than 2,000 in the Russian Federation. It has to be mentioned that due to the unstable situation in some regions of Russia these figures are subject to permanent changes. Four of the oblast units (Bryansk, Kaluga, Orel and Tula oblasts) have at the same time the logistical status of a regional center [4]. These oblasts represent the most contaminated geographical areas in Russia. With these exceptions the official governmental and administrative structure is completely mapped into the registry. The central registry institution supervises through its 11 plus 4 regional centers 3,000 examination and reporting units.

The five so-called departmental registries are established at the Russian Defense Ministry, the Ministry of Home Affairs, the Ministry of Atomic Power and Industry, Federal Security Service and Ministry of Transportation. These units have been established because civil servants involved in the rectification works at the nuclear power station after the accident have been mainly employed by these ministries and primary health care was provided through their medical units. The major administration of the registration system according to the governmental standards and guidelines given to the Russian National Medical and Dosimetric Registry are carried out under the responsibility of these ministries. The data of registered persons are directly passed to the federal, central institution, Obninsk, for as long as the individuals are employed by these ministries. The institution in Obninsk is only involved in the rather final stages of the routine reporting and controlling procedures and thereby these five administrative units are called registries as well (Figure 1).

From the organizational and logistic point of view the system of the data collection, processing, retrieval and documentation procedures may be divided into three subsystems (Figure 2). Major issues of the first subsystem are regulations and guidelines to implementation, maintaining and control of the managerial structures of the RNMDR as well as for the completion, documentation and final transfer of the primary examination and reporting forms at the rayon and oblast level.

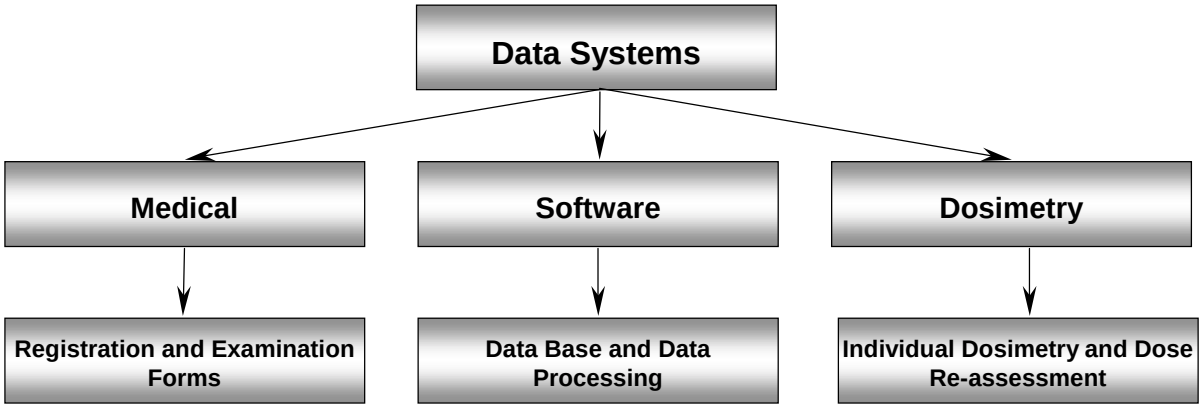

Fig. 2. The conception subsystems at the federal level of the RNMDR.

The second subsystem of the RNMDR covers the central database and its database services (e.g. updates, backups, queries etc.) as well as data analysis procedures.

Individual dosimetry is a major central part of the registry and therefore constitutes a separate subsystem in the organization and logistic of the registry. This subsystem covers the area of the reconstruction, re-assessment, calculating and quality control of individual absorbed external and internal doses for the purposes of the registry. The major issue in this subsystem is re-assessment and the validity and reliability of reconstructed absorbed doses, that is to say represents an object of research in dosimetry rather than an administrative or logistic entity.

2. The primary registration and examination forms

The frame for the basic structure of the primary registration and examination forms are defined and given by the 1991 governmental order (The Ministry of Health of the Russian Federation, Order N173, 4 April 1991). The issues addressed in this order may be summarized as follows:

- Passport and registration data of individuals exposed to radiation due to the Chernobyl accident (personal identification);
- Data on absorbed radiation doses (dosimetry);
- Information on pre-accident chronic diseases of registered persons (anamnesis);
- Records of examinations in the run of the personal, medical monitoring of registered persons (screening and follow-up);
- Reference data needed for evaluating, controlling and correcting the data of above.

These data have to be:

- Personalized in order to provide full, individual information needed for medical and administrative purposes;
- Computerized in order to allow direct, up-to-date, quick and visualized information retrieval (enquiries) of the personal information;
- Processed in order to make them accessible for research tasks.

This general frame is realized by the four primary registration and examination forms of the RNMDR (Figure 3):

- Registration form;
- Physical examination form;
- Dosimetric form;
- Correction form.

During the run of the registry from 1986 to the end of 1999 they were revised three times. Downward compatibility is fully ensured. In the following descriptions the addressing of these forms refers to the 3rd revision unless not explicitly stated otherwise.

The present forms are to a great extent a compromise between objective and subjective demands and logistical and organizational necessities. By designing forms for the objects and objectives of the registry one had and has to balance:

- Administrative demands and scientific interests;
- The scientifically desirable degree of specificity of single investigation objects and the reliability of data- depending on the educational level of the local examination staff;
- The coverage of heterogeneous registry population and thereby the complexity of forms versus the feasibility of their administration;
- The informative contents of medical examination versus their losses due to the needs in electronic data processing (e.g. coding schemes);
- Needs for standardization and quality control procedures versus human and financial resources for education, training and supervision of local staff members;
- Necessary individual engagement and motivation of registry staff versus the ballast of organizational overhead;
- And finally the theoretically feasible conception versus its public acceptance and financial support.

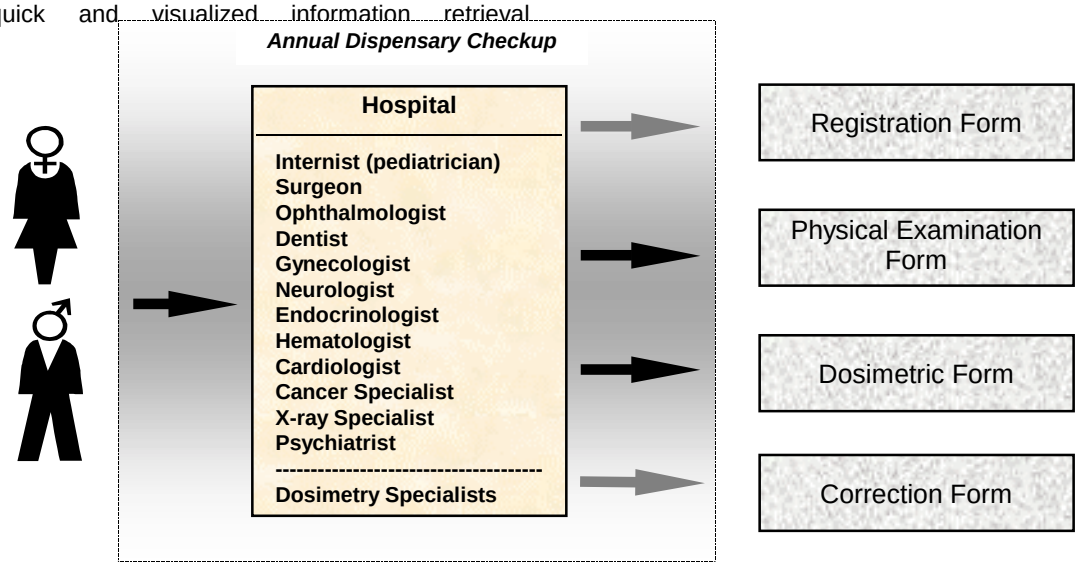

Fig. 3. The dispensary check-ups and the primary registration and examination forms.

These are common and well-known constraints in the optimization of registry designs. They may be more pronounced in the case of the RNDMR because it has to cover large geographical areas and the whole set of morbidity, disability and mortality measures and at the same time a territory, which covers a variety of different cultures and social traditions. Furthermore, it is well known that in central east and eastern European countries technical and financial resources are on a lower level as compared to western countries, especially in the ongoing period of political and economical transitions.

This background has to be given here, because it definitely has an impact on the overall conception of the project with regard to those parts, which are specific in terms of the application on the given data. In this book it has to be left on this rather verbal level of description because these conditions can be explicitly assessed only in the context of a specific application of mathematical models [5].

In the following the four primary registration and examination forms are only described and visualized by their major groupings of registry items.

The primary registration form (Figure 4) includes all items necessary for the personal identification according to the normal administrative regulations and procedures as well as those necessary for the RNMDR. Part of this information is used to identify one single examination forms of each person. From the point of view of the data processing this is the most sensitive area, because the whole structure of the data systems depends on the correctness of this unique record identification (updates of follow-up examinations, corrections, personalized information retrieval etc.). Furthermore this form contains the items of personal, medical history at the time of first registration and items necessary for linking the fourth primary population group of the registry (natural children of registered persons) to the other ones.

Registration Form

Identification
Place of Residence
Period Spent in the Exposure Zone
Dosimetry and Prophylactic Treatment
Chronic Diseases Diagnosed before 26.04.86
Pregnancy on 26.04.86
Parent of a new-born Child

Fig. 4. The primary registration form.

The medical investigation form (Figure 5) of the primary examination forms contains all items of the medical follow-up investigation. Its core is the occurrence of disease, disability and fatal events in the investigation period (time interval since the last documented follow-up investigation, resp. since first exposure to radiation for the first medical examination). Diseases and causes of death are coded according to International Classification of Diseases 9th revision (since 1999 in RNMDR are used 10th revision). Furthermore some other health status indicators are given as well as some items for describing the medical investigation itself.

Physical Examination Form

Identification
Health Status
Dispensary Examination

Dosimetry and Prophylactic Treatment
Diseases Diagnosed since Last Examination
Diseases Diagnosed during Dispensary Examination

Fig. 5. The primary examination forms: the physical investigation form.

The dosimetry forms (Figure 6) give the absorbed external and internal radiation dose in three equivalents. In most of the cases the annual dose equivalent of external, whole-body gamma-radiation exposure is given. Furthermore, this measure is reconstructed in most of the cases (in about 15% this measure is based on readings from personal dosimeters). This reconstruction may be subject to necessary re-assessments. These are documented in the correction forms (see below). However, the information used for reconstructing the absorbed doses is not documented. This has to be regarded - not only with respect to this project - as a major lack of information, which is difficult to control and will need rather exhaustive efforts in methodological consideration. It cannot be avoided because the reconstruction of absorbed doses needs information, which in general cannot be operationalized on a level necessary for electronic data processing. This seems to be a general logistical limitation in registries, which are not exclusively designed for clinical studies. At least this holds true for the RNMDR, with its administrative functions, which are indispensable to become realized, that is to say, other objectives have to be adjusted for this prerequisite.

Dosimetric Form

Identification
Dosimetry Data
- Dose of Equivalent of Gland Exposure To Iodine Radioisotopes
- Estimated Annual Dose Equivalent of Whole-body Intern Exposure to Cesium Radioisotopes
- Annual Dose Equivalent of External Whole-body Gamma-Exposure

Fig. 6. The primary examination form: the dosimetric form.

In the correction form (Figure 7) each correction is documented, which has been necessary during the monitoring of registered persons. One has to consider that with the given objectives of the registry the data has to be up-to-date at any given time because it is used for administrative and medical purposes by personalized queues. From the point of data processing this implies special attention and efforts in the maintenance procedure for the database and a considerable amount of administrative overhead is indispensable. Thereby, this form has a key function in data processing as well. Secondly, this part of a database, which is subject to permanent changes, needs special attention and efforts in research as well. At least explorative data analyses are technically difficult to conduct because of continuous changes in the base sets of empirical data, especially in iterative procedures for determining data structure consistency. This has been the major determining factor in the decision to use a frozen data set in the first phase of the project.

Correction form

Identification
Correction Lines

Correction Line

Line N	Codes of correction form				New entry
	Form no.	Item no.	Line no.	Corrected entry	

Fig. 7. The primary examination form: the correction form.

The primary RNMDR forms are usually completed in the course of the general dispensary check-ups of

the residents of Russia at the rayon hospitals and out-patient clinics. Registered persons are personally invited to the yearly follow-up investigations by letters. In many local areas nurses are making home visits to those persons not having followed this first invitation. However, up to now this procedure is not part of the mandatory instructions. In the physical examinations medical specialists are involved according to the medically defined needs with regard to the diagnosis, therapy and prognosis of diseases in the individual care of patients (Figure 3).

Data are collected, accumulated, stored, updated and documented at the rayon level according to a specific administrative order of the Russian Ministry of Health. Detailed instructions and guidelines by manuals of operations have been set up and are routinely controlled by the central registry institution in Obninsk and its regional centers (with the exception of the five departmental registries) and are part of the above-mentioned governmental order as well. As mentioned earlier, issues related to the implementation, maintaining and control of the given standards of the registry follow as far as possible the managerial structure of the existing system for providing primary health care at local levels.

At the oblast level the data reported by the district units are collected, coded, updated, stored and documented following technical procedures similar to those at the district level. Differences are due to differences in local conditions at the district levels. Details of these technical issues are not given here. The completed forms have to pass automated (software designed) error checking procedures as well as the approval of advisory commissions. These commissions are composed of experts in the fields of medicine, dosimetry and computer sciences and this control is conducted at the oblast level. Documents, which fail this control, are returned to the district level for correction.

At the regional level these data are compiled and processed. Regional centers are responsible for the primary registration that is to say for the administration of register numbers and issues related to personal identification. In case of any doubts concerning these data provided by the lower logistical levels of the registry, the regional center verify and correct primary registration data in direct contact with governmental institutions (mainly the Public Health Departments of oblasts) and public organizations (mainly oblast branches of "Soyuz Chernobyl", Russia).

At the Federal level data are finally compiled and processed for the purpose of the central database of the registry. This includes data submitted by departmental registries. The final, comprehensive checking for data consistency (including epidemiological and demographic features) is conducted here as well. The documents, which fail, are sent back to the regional level and to the departmental registries for correction.

Data are reported unprocessed (plain raw data files of individual records) and transferred between the four levels mainly on magnetic media (floppy-disks). The data format is given by the primary forms, which at the same time could be used as coding schemes and thereby as standardised reporting formats. Software necessary for the different tasks of electronic data processing has been designed at the central registry institution in Obninsk from where technical support is provided as well, including the hardware equipment at lower logistical levels. Thereby, uniform, standardised and transparent data processing procedures in the registry system are in principle ensured (the conception is outlined in chapter below). However, it has to be mentioned that the organizational management of the registry system is faced with major logistical problems in the maintaining and control of given standards and quality control procedures. This is due to the system and the objectives itself, which implies a logistical conception fitting to the considerable differences in local conditions. The constantly changing external conditions given in the period of economical and social transition in Russia add to these problems, which in their accumulation have to be taken into account as major constraints in assessing the overall objectives of the registry in the quality aimed at.

3. RNMDR software subsystem

The software for all levels of surveillance is used for computerization of primary information collection, uniform data base support and statistical and radio-epidemiological analysis [6].

Listed below are the principal requirements for the development of the RNMDR software:

- Data base formation, information control and analysis;
- Uniform computer technology of data input, editing, control and correction at all levels of surveillance;
- Use of standard reference and regulation information (territorial codes of municipalities, ICD-9 and ICD-10 codes of diseases, etc.);
- Unified statistical packages for radio-epidemiological analysis;
- Information exchange between different surveillance levels;
- Exchange of information obtained within other national programs studying Chernobyl accident consequences;
- Information exchange on international level and comparison of results;
- Reliability, susceptibility to updating and modification;
- Efficient use of computers and minimum keyboard input.

The following specific features of software application were taken into account in its design:

- Limited period of primary data collection;
- Medical personnel and operators without special computer training are employed at primary information registration, completion and input stages;
- Information entered is subject to correction at all stages of processing.

Thus the software employs advanced computer techniques to collect, store, process, transfer and analyze demographic, medical, dosimetric and epidemiological data and ensures the reliability of software products at all levels of surveillance, as well

as their lack of complexity and ease of use for medical personnel [7].

3.1. RNMDR software concept

In order to provide unified organization of data storage at district, oblast, regional and Federal level, as well as to combine an interface with standard statistical analysis software a common dbf file format was used.

The software falls into two functional parts - database support and analysis of information collected. Databases of large-scale and local registries store information in different ways, i.e. these bases are of different physical structure. In order to make the analytical part of the software flexible and multipurpose, i.e. independent of the specific features of particular bases, access to them should be uniform.

Many database management systems (DBMS) make databases accessible via inquiry language SQL. Software designed using C, FORTRAN, SQL, CLIPPER, VISUAL BASIC, etc. can function in computers having different operating systems (DOS, WINDOWS, OS/2, UNIX) and with different DBMS.

Software for district, oblast, regional and Federal levels contains applications designed for data base support.

At oblast and region levels data are analyzed via standard statistical software EPICURE, Systat and specialized programs using SQL interface or dBase III files directly. These programs are used to calculate basic epidemiological indexes. Computers at Federal level employ statistical software application packages SAS, EPICURE, Systat. SAS application uses SQL enquiries to gain access to data stored both via IBM Database Manager and in the aBase III files. Text editors functioning in IBM DOS as well as more powerful Aldus Page Maker and Word for Windows functioning in the Windows operating system are used to prepare texts and plots for reports.

Standard and specialized software in use are integrated into a unified information system of storage and epidemiological, medical and statistical analysis of data.

3.2. Software for database support

Requirements to specialized software for initial data input, control and correction as well as for database (DB) support follow the general outline of database support (Figure 8).

Irrespective of the level of surveillance primary information from completed primary documents either on paper or on magnetic media (floppy-disks, streamer cassettes) constitutes the input into DB support system. Information from paper documents is entered into the computer via input and editing subsystem (IES), after that it is brought to the control and correction subsystem (CCS). Documents from the lower level get into CCS via information reception subsystem (IRS). The next stage of information processing depends on the results of the automated control:

1) Positive result - information is loaded into database;

2) Negative result - erroneous and duplicate documents containing protocols (on paper or magnetic media) are subjected to erroneous information evaluation that is performed by specialists, e.g. district expert commissions. Upon correction information is again submitted to IES.

Information loaded into DB is checked for consistency. The staff working with DB performs this check-up. In case some dubious information is detected it is subjected to special examination.

Information is transferred to a higher level (from the district level to that of oblast, from the oblast level to the regional one, etc.) via the information unloading subsystem (IUS).

DB support system incorporates the subsystems of database inquiry (DBES) and of information organization (IOS) for further processing within the system of statistical and radio-epidemiological analysis of data.

Listed below are the essential features and requirements for DB support systems.

- Input and editing subsystem (IES) ensures correct input of initial information and controls:
- Presence of obligatory fields;
- Format;
- Allowable limits of parameter values;
- Consistency of document fields.

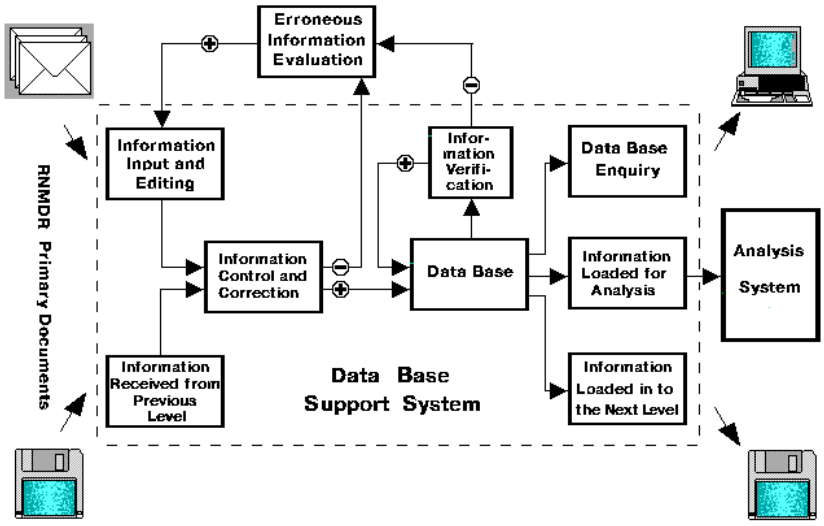

Fig. 8. Database support outline.

The subsystem of information reception from the lower level (IRS) and the subsystem unloading information to the higher level (IUS) ensure the correctness of information exchange between levels on floppy disks or streamer cassettes if necessary.

Information control and correction subsystem (CCS) ensures the quality of information accumulated in the DBs. Control and correction procedures in the CCS constitute a single stage. Corrections are made in the interactive mode using a special document editor.

The CCS checks for:

- Consistency of all fragments and fields of the document;
- Recurrent registration of one person;
- Recurrent registration number;
- Occurrence of duplicates;
- Sufficient amount of data.

Control and correction procedures having been executed information are loaded into the DB. Simultaneously, protocols of information loaded into the DB are generated for the lower level suppliers of documents.

Database inquiry subsystem is used to obtain individual and generalized information on any fragment or line of the documents via standard and special inquiries to DBs. This subsystem makes it possible to form inquiries of "query by example" type in the interactive mode. Results of the inquiries take the form of lists or reference information.

The subsystem that organizes information for further processing (IOS) is used to prepare special data arrays submitted to the system of statistical and radio-epidemiological data analysis (SDA). IOS is a set of interfaces in the form of a menu system where the user first chooses the subsystem of analysis and then specific document fragments. After that information is selected from the DB to be analyzed in one of SDA subsystems.

DB support system ensures high efficiency of large data array processing, it is flexible, adaptable and convenient to use.

3.3. Software for statistical and radio-epidemiological data analysis

The system of statistical and radio-epidemiological data analysis (SDA) enables the basic epidemiological indexes to be calculated in the interactive mode [8]. It also provides thorough statistical and radio-epidemiological analysis of data from the RNMDR primary documents.

SDA consists of the following basic modules (Figure 9):

1) Subsystem of epidemiological express-analysis;
2) Subsystem of mortality, morbidity and prevalence dynamics analysis;
3) Subsystem of general statistical analysis;
4) Subsystem of competing risk factor analysis;
5) Subsystem of stratified data analysis;
6) Subsystem of long-term radiation effect analysis.

Each SDA subsystem has its own interface used to accept data from the DB support system. In contrast to other subsystems, data express-analysis subsystem interface addresses files of integrated information. In subsystems employing standard statistical software applications SAS and EPICURE commonly used in medical statistics and radio-epidemiology, interfaces are realized in procedure languages of these applications.

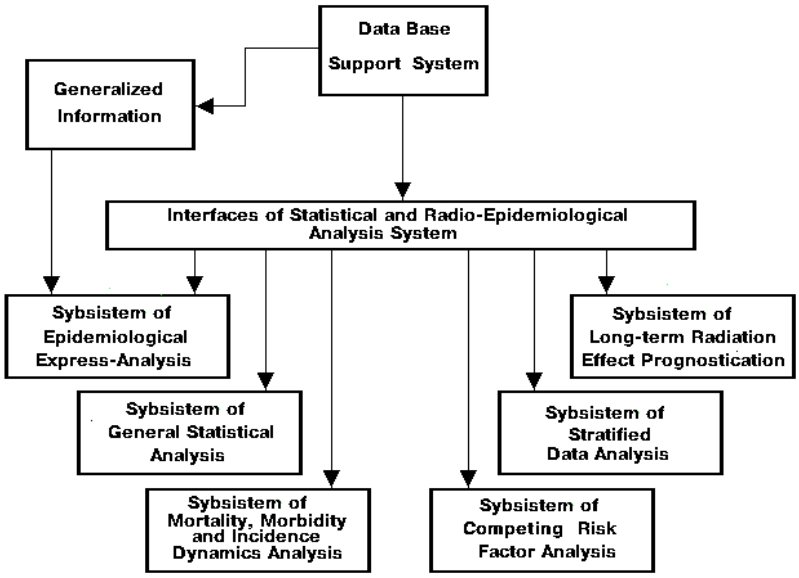

Fig. 9. SDA general outline.

The subsystem of epidemiological express-analysis of data (SEEAD) is designed for interactive processing of generalized information and is intended for users without specialized training in computer science.

We shall describe this subsystem in more detail because it is a very convenient interactive tool for processing large amounts of data, which cannot be reviewed by standard means within a reasonable period of time.

In the interactive mode the above subsystem:

- Calculates mortality and morbidity rates;
- Calculates prevalence of diseases;
- Evaluates mortality and morbidity structure;
- Calculates other demographic rates;
- Compares mortality and morbidity in contaminated and control areas;
- Compares mortality and morbidity rates for registration groups' different exposure doses;
- Checks the consistency of statistical material with known data.

In order to cope with the above tasks in the interactive mode we suggested an approach heretofore referred to as personal data aggregation (PDA). This approach proceeds from storing and processing integrated information on people arranged into subgroups. Integrated information forms a clean-cut hierarchical structure and assumes the form of a multi-dimensional matrix. Each of L upper levels (with the total number of hierarchical structure levels being L+1) is determined by its information field, the number of elements of the level corresponding to the number of grades in these fields. L+1 level of the structure consists of elements corresponding to the number of grades of one or more than one information fields. Elements at each level may consist of a set of parameters characterizing the subgroup of this level of the hierarchical structure.

Setting all the elements of one level by a certain sequence of parameters each of which forms a set together with the corresponding parameters of other elements at this level, we can perform various operations on them. The parameters of a combination of two or more elements can be defined by means of such operations on sets of parameters of different elements at the same level. The combination parameters characterize the subgroup of people contained in one of the two (or several) grades.

The following parameters characterizing a subgroup of people can be chosen as those having the value of information field corresponding to a certain grade: the total number of people included in the subgroup, some minimum (maximum) value among the people of the subgroup in question, other characteristics which can be subjected to addition, averaging, root-mean- square sum, etc.

Therefore, with the help of certain calculation techniques realizing the PDA method we can get a set of integrated data in the form of multi-dimensional matrix out of a personal data array.

The set of procedures realizing in software, the PDA method is called the unit of integrated data (ID) generation. This software set allows the personal data included in the RNMDR primary documents (registration form, code card and dosimetric data record) to be transformed into a specific structure of integrated characteristics for the subgroups concerned.

Fig. 10. The example of work of the inquiry interface of the system SEEAD.

Files of personal data from the RNMDR database containing all the information necessary for obtaining integrated data on one or more primary documents are entered into the ID generation unit.

Each entry of these files corresponds to a set of personal data on one person included in the RNMDR.

Input information processing can be compared to sorting of letters by postal codes with simultaneous counting of their number (one of characteristics) in each accumulating pile. Proceeding from the PDA method the following fields were selected as the ones determining the levels of the hierarchical structure: 1) territorial code (municipality classifier); 2) primary registration group; 3) sex; 4) age.

The above four information fields determine the structure of integrated data under processing. The subgroup data integrated over different information fields and their grades are formed at the bottom (5-th) level. Integrated information ready to be processed is stored on hard disk of the PC in special format files.

ID generation unit makes it possible to collect the full set of integrated data arrays within 1-2 days and to receive up-to-date information on the registry for a long time (until a new series of documents is loaded into the RNMDR database). Files of integrated data for all years occupy about 30 Mbyte of hard disk space (the total space occupied by the personal base is 4 Gbyte).

After the ID generation unit finishes operation SEEAD can be used directly. The user makes an inquiry in the interactive mode by means of a "friendly interface" simultaneously defining: 1) the form of the output; 2) information set under study (regions, PRG, sex, age, date(s) of dispensary examination(s) or dose assessment).

The next stage of SEEAD operation consists in loading the file selected into the computer memory. All the hierarchical structures that define the multi-dimensional arrays were designed taking into account the necessity of loading the corresponding files into the computer memory. Thus, loaded integrated information is processed at a very high rate.

Upon obtaining all the information of interest on the information set chosen the user can repeat the above procedure, etc. It should be noted that the entire cycle of operation from the inquiry to the result takes not more than 1 minute on personal computer.

The outcomes of the analysis in a system SEEAD can be represented in table forms, maps and graphics. In a Figure 10 the example of work of the inquiry interface of the system SEEAD is shown. The Figure 11 shows an example of submission of outcomes after the analysis of incidence rate for different classes of diseases in liquidators as the graphics or the map of selected region. These two pictures illustrate work of a system SEEAD.

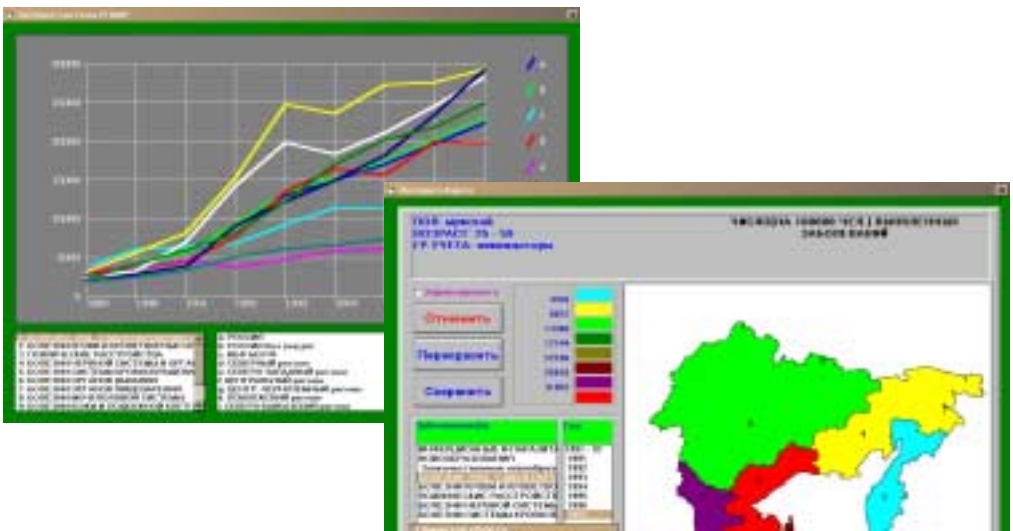

Fig. 11. The example of submission of analysis outcomes of incidence rate for different classes of diseases in liquidators as the graphics or the map of selected region.

4. Current status of the RNMDR

Figure 12 presents the information on the dynamics of the RNMDR registree number growth in 1986-1998. As the figure shows during the entire period of its existence the data bank of the Federal level of the RNMDR kept accumulating medical dosimetric information. At present it comprises data on 530965 people from throughout the Russian Federation. As described above, all the RNMDR registrees are divided into four primary registration groups (PRG):

PRG 1 - liquidators (32,9%);

PRG 2 - evacuated from 30-km zone (1,7%);

PRG 3 - residents of contaminated territories (61,3%);

PRG 4 - children born of people included in groups of higher risk (4,1%).

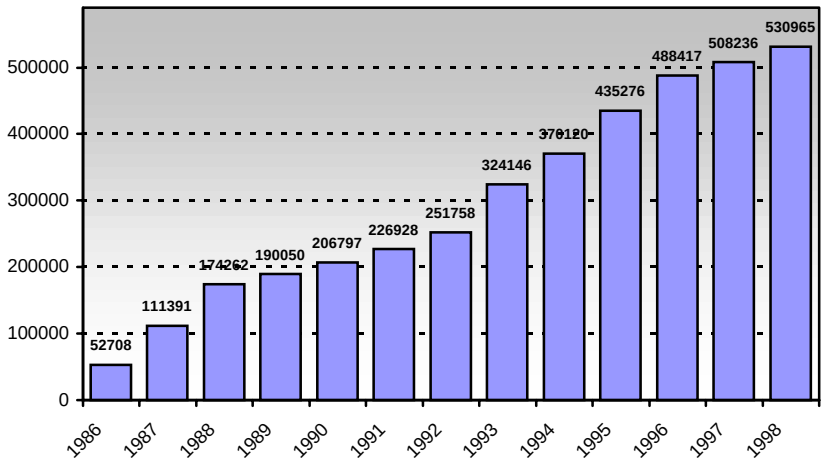

Fig. 12. Dynamics of RNMDR registree number growth in 1986-1998.

Figure 13 shows the dynamics of the structure changes in the RNMDR primary registration groups in 1986-1998. In the period of 1987-1992 emergency workers (PRG 1) constituted more than a half of those registered in the RNMDR. But it should be noted that in 1993-1998 the number of residents of contaminated areas significantly grew due to the improvement of operation of Bryansk, Kaluga, Orel and Tula regional centers.

Figure 14 illustrates the regional distributions of people registered in the RNMDR. Table 1 shows the number of RNMDR registrees in regions of Russia and departmental registries. A large number of people included in the RNMDR reside in the Central and North Caucasian regions of Russia. In the Central region residents of the areas under surveillance in Bryansk, Kaluga, Tula and Orel oblasts constitute the majority, while in the North-Caucasian region liquidators prevail.

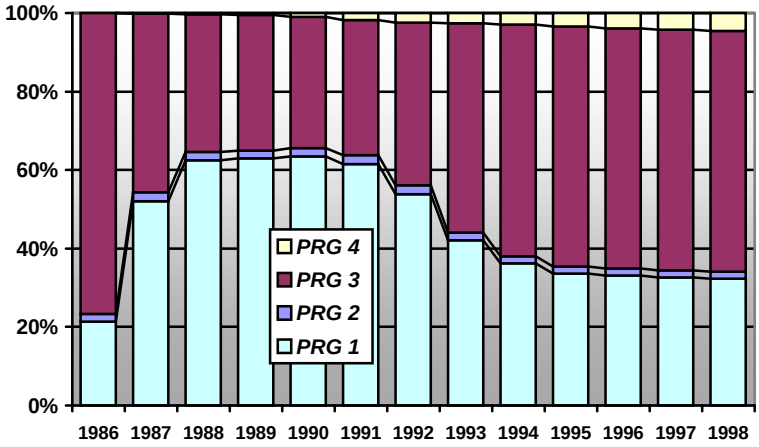

Fig. 13. Dynamics of changes in the RNMDR registree structure according to primary registration groups in 1986-1998.

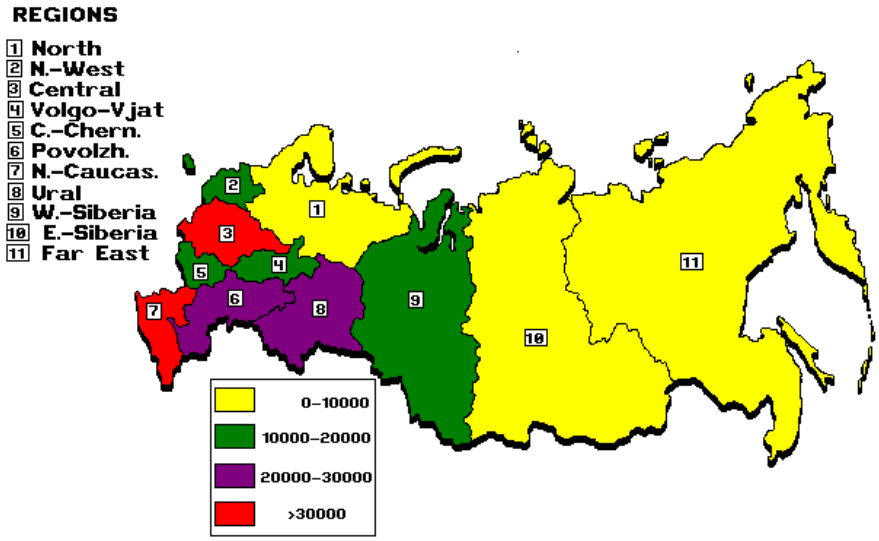

Fig. 14. The regional distribution of people registered in the RNMDR.

Table 1

The number of RNMDR registrees in regions of Russia and departmental registries (01/01/1999)

Region	Total	Liquidators	Evacuated from 30-km zone	Residents	Children of liquidators
Russia (All RNMDR)	530,965	174,916	9,212	325,377	21,460
Regions of Russia:					
North	7,668	5,771	415	669	813
North-West	13,934	10,004	1,289	1,625	1,016
Central (without 4 most contaminated oblasts)	27,547	19,103	1,193	3,789	3,462

Volgo-Vyatsky	11,647	9,130	263	501	1,753
Central-Chernozem	11,512	8,798	517	456	1,741
Povolzhsky	22,982	19,257	690	919	2,116
North Caucasus	33,432	26,732	1,559	2,397	2,744
Urals	25,748	22,056	568	411	2,713
West Siberia	12,476	10,013	568	555	1,340
East Siberia	3,416	2,307	390	333	386
Far East	2,439	1,059	239	962	179
Most contaminated oblasts:					
Bryansk oblast	199,530	2,864	132	196,331	203
Kaluga oblast	53,219	1,894	63	50,874	388
Orel oblast	17,916	1,315	52	16,307	242
Tula oblast	47,524	2,468	133	44,513	410
Departmental registries:					
Ministry of Internal Affairs	8,311	4,237	38	4,020	16
Ministry of Defense	5,138	5,138	0	0	0
Federal Security Service	1,464	996	13	455	0
Ministry of Transportation	1,703	1,575	51	36	41
Ministry of Atomic Energy	23,359	20,199	1,039	224	1,897

If the distribution of external exposure doses by liquidators depending on the date arrival to the zone is considered. Figure 15 shows that you can see that the majority of liquidators of 1986 and 1987 received doses exceeding 5 cGy (83.1% and 75.3% of the total number of EWs who worked in the zone in 1986 and 1987, respectively), whereas in the following years their number did not exceed 15 % (for example, 1988 - 14.1%, 1989 - 4.5 %, 1990-1993 - 4.4%). The average external dose for EWs of 1986 is equal to 15.9 cGy, of 1987 - 9.0 cGy, of 1988 - 3.3 cGy, of 1989 - 3.2 cGy, of 1990 - 3.7 cGy [9].

In conclusion, we note that the experience gained in creating, implementing and operating the Russian National Medical and Dosimetric Registry since 1986 with a variety of complex organizational, scientific and practical problems solved can be of use in organizing and operating large-scale information systems intended for study of adverse effects on the health status of large populations.

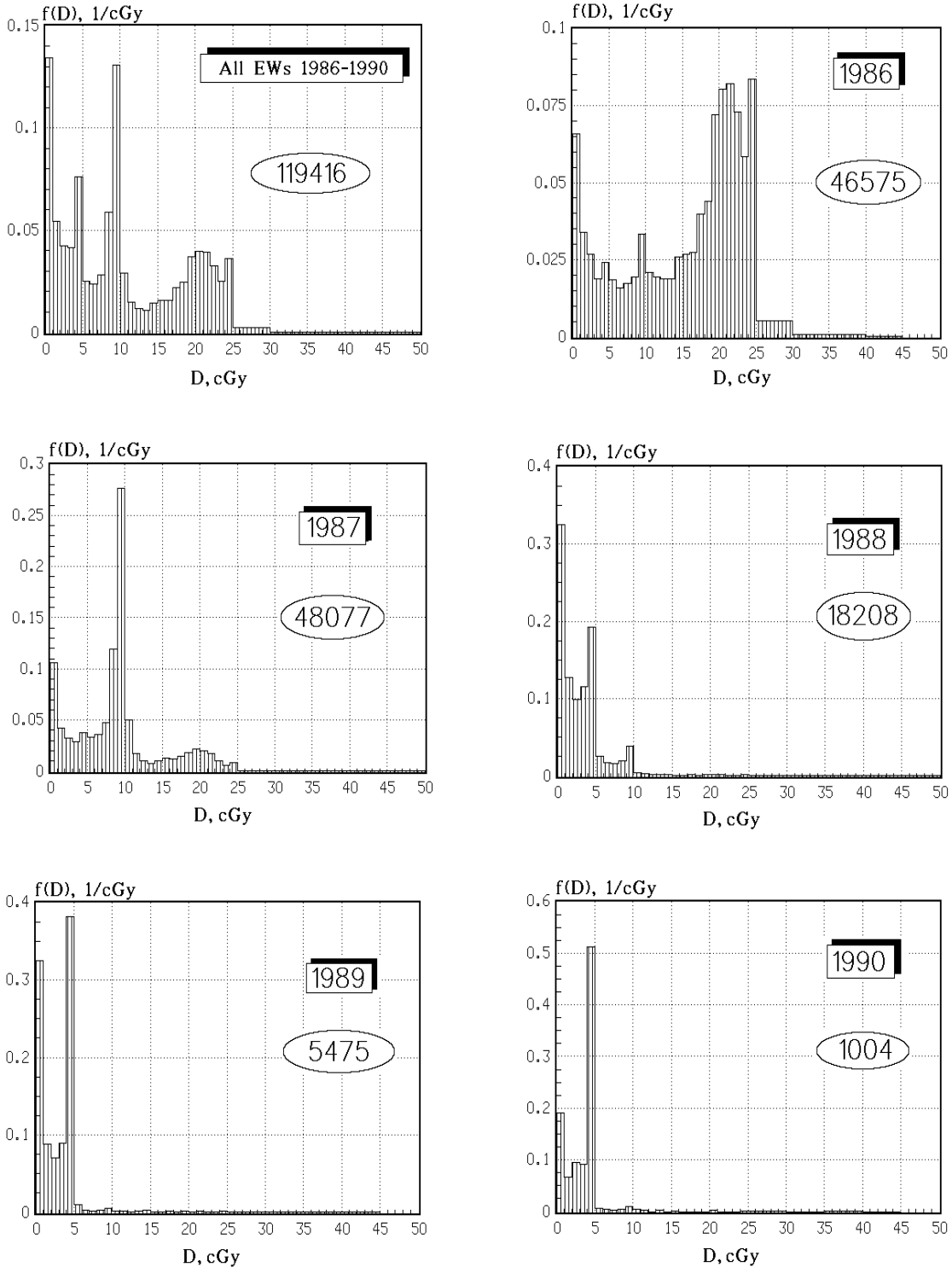

Fig. 15. Distributions of external exposure doses ***D*** for EWs (liquidators) registered in the RNMDR.
Distributions for different dates of arrival to the 30-km zone (1986, 1987, 1988, 1989, 1990) are demonstrated separately.
Inside the oval the number of EWs employed in the 30-km zone in the given year is indicated on each picture.

References

1. **Tsyb A.F., Ivanov V.K., Pozhidaev V.V. et al.** Software and All-Union distributed register data management. Report of the JAPAN-USSR seminar on radiation effect research. - Tokyo, 1990. - P. 43-47.
2. **Tsyb A.F., Ivanov V.K., Airapetov S.A. et al.** Software and information system of Russian National Medical and Dosimetric Registry/Radiation and Risk. Bulletin of the National Radiation and Epidemiological Registry. - 1992. - Issue 1. - P. 132-146 (in Russian).
3. **Ivanov V.K., Maksioutov M.A., Tsyb A.F. et al.** Information system and modeling: Data and organizational aspects of the Chernobyl State Registry. In: Proc. First Conf. of Int. Simulation Societies. - Zurich, 1994. - P. 579-583.
4. **Ivanov V.K., Tsyb A.F., Maksioutov M.A. et al.** Assessment of morbidity parameters and their dose relationships for the population living in areas of Kaluga region contaminated with radionuclides/Radiation and Risk. Bulletin of the National Radiation and Epidemiological Registry. - 1994. - Issue 4. - P. 18-27 (in Russian).
5. **Morgenstern W., Ivanov V.K., Mikhalski A.I., Tsyb A.F., Shettler G. et al.** Mathematical modeling with Chernobyl Registry data. Registry and concepts. - Berlin-Heidelberg: Springer, 1995.
6. **Ivanov V.K., Tsyb A.F., Rastopchin Eu.M., Maksioutov M.A. et al.** Planning of long-term radiation and epidemiological research on basis of the Russian National Medical and Dosimetric Registry. In: Nagasaki symposium on Chernobyl update and future. - Amsterdam: Elsevier, 1994. - P. 203-216.
7. **Ivanov V.K., Starinsky V.V., Maksioutov M.A. et al.** Information and mathematical software for conducting the cancer registration on radio contaminated territories of Russia/Radiation and Risk. Bulletin of the National Radiation and Epidemiological Registry. - 1995. - Issue 5. - P. 26-77 (in Russian).
8. **Ivanov V.K., Tsyb A.F., Maksioutov M.A., Pitkevich V.A. et al.** Radiation epidemiological analysis of the data of the National Chernobyl Registry of Russia: prognostication and facts nine years after the accident/Radiat. Prot. Dosim. - 1996. - V. 64. - P. 121-128.
9. **Pitkevitch V.A., Ivanov V.K., Tsyb A.F., Maksioutov M.A. et al.** Exposure levels for persons involved in recovery operations after the Chernobyl accident. Statistical analysis based on the data of the Russian National Medical and Dosimetric Registry/Radiat. Environ. Biophys. - 1997. - V. 36, N. 3. - P. 149-160.