Use of Health Information Systems in the Russian Federation in the Assessment of Environmental Health Effects

Jouni J.K. Jaakkola,[1,2] Martin Cherniack,[1] John D. Spengler,[1] Halûk Özkaynak,[1] Bogdan Wojtyniak,[3] Andrey Egorov,[1] Pavel Rakitin,[3] Boris Katsnelson,[4] Sergei Kuzmin,[4] Larissa Privalova,[4] and Natalia V. Lebedeva[4]

[1]Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts, USA; [2]Department of Epidemiology, School of Hygiene and Public Health, Johns Hopkins University, Baltimore, Maryland, USA; [3]National Institute of Hygiene, Warsaw, Poland; [4]Environmental Epidemiology Component, Environmental Management Project, Center for Preparation and Implementation for Foreign Projects, Moscow, Russia

The Russian Federation has made an intensive effort to compile and use information on the environment and human health. In 1996–1997, we evaluated the information that was collected and analyzed on the local (raion), regional (oblast), and federal levels with reference to its usefulness in the assessment of environmental health effects. The Russian Federation maintains standardized nationwide institutions that routinely collect health data in polyclinics and hospitals and then report to the national offices. The allocations of the workforce and the broad range of surveyed health outcomes are extensive, but a lack of systematic control of information quality limits the ability to take full advantage of these efforts. On the other hand, the hierarchical system of data collection has advantages over more decentralized or commercial health systems. A major weakness in the current reporting is the aggregation and transformation of data. Although this may not disturb the generation of health statistics, it seriously limits the use of regional and federal level data in the assessment of health effects of environmental exposures. In spite of limitations, some revised approaches to the analysis of existing data may be both feasible and fruitful. Combining information from routine data and newly collected data is likely to be the most effective way to assess the relationship between environmental exposures and diseases. Although there is a strong and justifiable desire to rapidly translate information of environmental health effects into policy alternatives, at present, it seems more useful to emphasize data quality, completeness, and plans for the use of data. *Key words*: environmental health, morbidity, mortality. *Environ Health Perspect* 108:589–594 (2000). [Online 24 May 2000]
http://ehpnet1.niehs.nih.gov/docs/2000/108p589-594jaakkola/abstract.html

The Russian Federation covers the world's largest area, over 17,000,000 km^2, and has a population of 147 million. The end of the Soviet era revealed extensive pollution of air, water, and soil in vast areas of the country at levels deemed hazardous to human health. In 1995, the Government of the Russian Federation initiated an Environmental Management Project with a general aim to improve environmental conditions and associated human health (*1*). An Environmental Epidemiology Component of the Russian Environmental Management Project was initiated to perform policy-oriented environmental epidemiology and risk assessment, and to establish systems for information transfer to decision makers in environmental policy. The Harvard School of Public Health, representing a broad group of international investigators, was chosen as the principal international consultant to the Environmental Epidemiology Component.

A sense of urgency motivates the translation of collected data from health, demographic, and environmental databases and archives into public health policy. First, there is the widely recognized deterioration in adult survival in the Russian Federation. For example, while American male and female life expectancies have increased over the past 25 years by 5.1 years for men and 4.1 years for women, comparable indices in the Russian Federation have declined by 3.9 years for men and 2 years for women (*2*). This has been most striking for men; the life expectancies for American and Russian men in 1993–1994 was 72.2 and 57.7 years, respectively (*3*). Complete and valid information on population health and environmental conditions support assessment of relative roles of social, economic, and environmental conditions in this fast decline in public health. Second, the current fiscal constraints limit the optimal use of a centralized health and environment data collection system and complicate maintenance of data systems. Third, where data collection and preventive standards do exist, it is sometimes the case that systems are arcane that measurement may be inadequate, and that there is no effective strategy for the management of excursions beyond norms or for the management of adverse trends. Finally, the improvements in electronic data collection and analysis offer the opportunity to streamline processes and to encourage the use of local systems.

A key element in the management of environmental problems is the availability of valid and relevant environmental and health information. There has been an intensive effort in the Russian Federation to compile and use information on the environment and human health. In 1996–1997, we assessed the information that was collected and analyzed on the local, regional, and federal levels. Environmental monitoring activities in the Russian Federation are dispersed among federal, regional, and local organizations, which loosely translate into the political and administrative divisions of the Russian Federation: regional oblasts and autonomous republics (of which there are 89), raions, and districts. Typically, pertinent information is collected by the Sanitary Epidemiological Service (Gossanepidnadzor; a nationwide public health entity), by Roshydromet, by the Ministry of Natural Resources, and by the State Committee for Environmental Protection. Gossanepidnadzor is responsible for the health protection of the general population. Its mandate includes environmental health, in particular water, food, and air quality. The State Committee for Environmental Protection is responsible for protecting the environment by monitoring and enforcement, particularly through the control of emissions from industry. There is considerable interaction between Gossanepidnadzor and the State Committee for Environmental Protection. Roshydromet is responsible for air monitoring in residential areas and targets short-term concentrations from industrial sources and monitoring of water quality. The pyramidal structure of primary data collection by agencies such as Gossanepidnadzor and Roshydromet produces a complex pattern: the federal level summaries and annual reports that are the bases of regional and local decision making are the digested and refined products of sequentially synthesized local information. They then become the substrate for more localized planning and policy.

In essence, our approach followed reciprocal bottom-up and secondarily top-down

Address correspondence to J.J.K. Jaakkola, The Nordic School of Public Health, PO Box 12133, S-402 42 Göteborg, Sweden. Telephone: 46-31-693980. Fax: 46-31-691777. E-mail: jouni.jaakkola@nhv.se

This study was supported by a World Bank loan to the Russian Federation.

Received 20 October 1999; accepted 21 January 2000.

strategies. We considered data systems, data quality, and data reporting on the local level in a single city (Cherepovets) and on the regional level in an oblast (Sverdlovsk). We also reviewed the variety and quality of federal databases. In this paper we focus on the evaluation of health information collection and use through illustrative examples from each of these administrative levels. An evaluation of environmental monitoring has been published elsewhere (*4*).

Information on diseases occurring in populations over time or space can be used to create hypotheses on health effects of environmental factors. An increase in disease occurrence in a defined population over time may inferentially implicate the influence of environmental factors and thus call for further evaluation. Also geographic differences in morbidity or mortality rates may suggest an environmental etiology. Geographic differences in patterns of exposure and in patterns of presumably environmentally influenced diseases are the most evident and useful features of large data systems, but they are also the sources of the most serious systematic errors, referred to as ecologic fallacy (*5,6*). There are always alternative explanations, such as differences in the characteristics of the compared populations, and differences in procedures for detecting and recording health events.

The assessment of the effects of environmental exposures on health requires both environmental and health information as well as additional information on other determinants of health. Routine monitoring of environmental and health information, as part of a broadly conceived surveillance program, can sometimes be used for the assessment of health effects from environmental factors. However, it is often necessary to assess a relationship between a given environmental factor and its health effect in specific and well-monitored settings where environmental factors, health effects, and other relevant factors can be measured validly and efficiently. Targeted and detailed observations in a limited number of selected localities can generate results that can be subsequently generalized to other settings or even for national application.

In this context, there is a broadly recognizable divergence between the environmental studies approach that evolved in the Soviet Union as compared, for example, to the experience in the United States. National and historical generalization provides a collection of potentially false premises; it is striking that the American approach to determining the health effects of air pollution, for example, has taken the direction of composite, meticulously constructed, community-based studies, whereas in the Soviet Union the objective was the construction of large geographic data systems with limited variables that deferred to local authority for the quality of data collection.

Methods

We began our study at the grass-roots level and followed the information flow toward the central offices in Moscow (Table 1). Principal generic health information comprises births, mortality, and morbidity, which is routinely collected by Russian health care personnel on standardized forms. Completion of these forms is required by law, and this is assimilated into the public health culture. We evaluated special sources of health information including individual disease registries and epidemiologic studies. The local-level observations were made in Cherepovets (population 320,000) in Vologda Oblast, and in Nizhni Tagil (population 450,000) and Yekaterinburg (population 1,400,000) in Sverdlovsk Oblast. The regional level observations came from Sverdlovsk Oblast. Our review of federal level information focused on institutions that receive primary information from the oblasts and on several important research institutes that made secondary use of this information.

Results

City of Cherepovets (local level). The city of Cherepovets is a major industrial center in European Russia. The modern city was developed in the 1950s around the Northern Steel production facility, the largest iron smelter in Russia and the main employer of approximately 60,000 people. In the 1960s two chemical factories were constructed to produce fertilizers. Currently there are 20 industrial facilities in Cherepovets. The city is located on the Sheksna River and the Rybinsk Upper Volga Reservoir.

Cherepovets initiated the computerization of maternal and children's health data in 1992. The health care delivery system is highly centralized. There are three maternity clinics, seven polyclinics administering care through 14 years of age, and one children's hospital. The decline in the birth rate by 50% over the last 20 years (currently 2,500 live births/year) has provoked major concerns over maternal and child health care. There are three major types of health information generated for the maternal and child health care systems: information on pregnancies and newborns from birth clinics; information on morbidity from polyclinics, generated by each physician visit; and information on causes of death maintained by the local administration unit of the Office of Population Statistics (ZAGS). Cherepovets is regarded as a model among Russian cities because the Center of Gossanepidnadzor has sponsored coding and direct computer entry through its own trained surveillance teams.

The maternal and child health surveillance system is highly organized and, with some modification, provides resources that cannot be mirrored in more decentralized health care systems. Triweekly, then biweekly, examinations are standardized through the first 29 weeks of pregnancy; weekly visits begin at week 30. Information on each woman admitted to a birth clinic and on her newborn is recorded systematically and entered into a database computer file in the Office of Gossanepidnadzor. This database includes relevant information such as maternal age, date of admission, address, mother's occupation and employer, number of previous pregnancies, complications during pregnancy and delivery in *International Classification of Diseases, Revision 9* (ICD-9) codes, malformations, birth weight, and other indices of size. The computerized database for information on newborn children provides a good population-based framework for studying effects of environmental factors during pregnancy. The most useful

Table 1. Examples of routinely collected health information in the Russian Federation.

Level	Data collected
Local level	
City of Cherepovets	Information on pregnancies and newborns from birth clinics
	Information on morbidity from polyclinics, generated by each physician visit
	Information on causes of death maintained by the Office of Population Statistics
Regional level	
Sverdlovsk Oblast	Aggregated (raion-level) data based on annual reports of polyclinic visits
	Aggregated (raion-level) data on hospital discharge
	Aggregated (raion-level) data on the health of pregnant women
	Aggregated (raion-level) data on the health of infants
Federal level	
The Ministry of Health (Russian Federation)	Morbidity data recorded at polyclinics and hospitals
Federal Center of Gossanepidnadzor	Primary information on infectious diseases from the oblasts (monthly and annual reports)
	Individually based registry of occupational diseases

health outcomes from the existing data are birth weight, gestational age, and mortality. Malformations are potentially useful, but at this time the malformations have not been clearly characterized. Currently, cases are reviewed by a specialist, but a formalized description of the major diagnostic groups and criteria is needed. Perinatal disease is a potentially useful outcome that is not considered in the current system.

The maternal and child health care data collection systems provide opportunities for sophisticated epidemiology and also for administrative dead ends. The level of detail exceeds the federal capacity to process and use such precise locally specific information. On the other hand, for studying health effects of environmental factors on birth outcome and perinatal health, additional information is needed on factors that influence health. These include factors in the home environment such as size and nature of the residence, members of the household, sources of indoor air pollution, occupational exposures, and factors such as parental smoking and alcohol consumption. Social variables and detailed behavioral information on maternal health during pregnancy are also relevant. However, these are within the capacities of the functioning Gosanepidnadzor data collection system. Accordingly, we recommend that for each pregnancy additional information should be collected routinely on maternal smoking, smoking of other family members in the home, number of people in the household, size of the residence, maternal and paternal education, and some indication of economic status.

Because the private sector is small and access to health care is generally unrestricted, the coverage of the entire pediatric population through polyclinic visit information is thought to be very high. Each visit is recorded in a standardized form using ICD-9 codes. However, the tendency of clinicians to proffer a limited number of diagnoses imposes distortions that are at least as remarkable as those found in an insurance-driven system. For example, a single respiratory diagnosis, "obstructive respiratory illness" (ICD-9 code 465.0), has become an evident default and something of a trade joke among professionals because it is used to describe > 50% of all pediatric visits in Cherepovets. Of 7,440 polyclinic visits in January 1996, 5,300 (71%) involved respiratory diseases, but > 95% were nonspecific obstructive respiratory disease described as ICD-9 code 465.0. Less than 1% of cases involved a diagnosis of asthma (ICD-9 code 493.0). Asthma is a highly restrictive diagnosis that requires specialty recognition, and it is essentially a hospital diagnosis. A 40-fold difference in the rates of pneumonia within the city indicates probable variation in diagnostic practice. Unexpected low rates of chronic otitis and allergic rhinitis, as well as asthma, indicated a possibility of underreporting of some diseases. Disease trends for conditions for which laboratory confirmation exists, such as hepatitis A and B and salmonellosis, are well documented, but there are few of these as compared to diagnoses of nonspecific gastroenteritis. Approximately 50% of reported acute gastrointestinal infections carry a specific etiologic diagnosis; this extensive effort to catalog pediatric gastrointestinal disease offers interesting opportunities for identifying waterborne disease patterns. There are plans to introduce modern cryptosporidium diagnostics, which will further increase the proportion of diagnosed cases.

The computerized individual morbidity data has a good potential to be used for assessing health effects of environmental factors and as a basis for providing information for decision makers. To improve the usefulness of the information system, we recommend that *a*) diagnostic accuracy should be improved and more effective case definitions should be introduced for key health outcomes; *b*) additional information should be added on details of health behavior (smoking, alcohol), home environment (type and size of home, presence of pets, type of stove, etc.); and *c*) socioeconomic status should be included in the routine data collection to control confounding. Our study of Cherepovets indicates that routine quality control for accuracy of diagnostic practice, coding, and data entry should be established at the raion level.

The City Office of Gossanepidnadzor enters data from the death certificates in the Office of Population Statistics into a computerized database. The database records all deaths in children of 0–14 years of age (approximately 100–150 deaths/year). In general, children's mortality is a rather insensitive measure of the health effects of environmental factors such as air, water, or soil pollution because other factors such as congenital or early chronic diseases and accidents are responsible for the majority of these deaths. The relatively small number of deaths within Cherepovets limits the use of these data for testing hypotheses. Adult mortality would be a more useful measure for assessing the impact of environmental pollution, in particular, short-term effects of air pollution. Thus, computerization of individual death records is advisable. Interestingly, such a system of computerized death registration does exist in Sverdlovsk Oblast.

Sverdlovsk Oblast (regional level). Sverdlovsk Oblast, with a population of approximately 5 million, is located on the slopes of the Central Ural Mountains. It is a historical center of mining and metallurgy, a status that was further reinforced by industrial relocation of major machine producing facilities during World War II. Accordingly, the population is heavily concentrated in industrial cities, with large proportions of the population employed in a small number of industrial complexes. Limited agriculture further contributes to the overall urbanized nature of the oblast, with limited population areas outside city limits. The capital of Sverdlovsk Oblast, Yekaterinburg, is a historically important city that has strong traditions as a center of higher education. It was also a major scientific and technology production center in Soviet times.

The Sverdlovsk Oblast Gossanepidnadzor is located in Yekaterinburg. The Department of Social Hygiene maintains a social and hygienic monitoring system consisting of databases and routing computer programs. There are > 1,500 items of information collected by Gossanepidnadzor for the entire oblast. The databases are organized into five blocks: *a*) environmental monitoring, *b*) air pollution, *c*) water quality, *d*) soil pollution, and *e*) general morbidity data. General morbidity, which includes raion-level data based on annual reports of polyclinic visits, hospital discharges, health of pregnant women, health of infants, and information on sanitary-hygienic conditions, is linked to this database.

All polyclinics and hospitals in the 56 raions of Sverdlovsk Oblast record the diagnosis at each visit or admission on a standardized form, which includes name, address, employment (yes/no), place of work, age group (0–14 years, 15–17 years, ≥ 18 years), diagnosis, and first episode (yes/no). The number of visits and morbidity rates (per 1,000) are reported annually to the Raion Central Polyclinic. The Raion office sends aggregate data to the oblast Department of Health in annual reports. The raions, included in Yekaterinburg and Nizni Tagil, the two largest cities, send their reports to the City Department of Health, which then forwards the transmission to the Department of Health in the oblast. The oblast level information is then forwarded to the Ministry of Health, Russian Federation, and to the Oblast Center of Gossanepidnadzor. There is no oblast-based system for refining or checking morbidity data entry, for which reporting is required on the federal level. Personnel from the Sverdlovsk Oblast Center of Gossanepidnadzor believe that data accuracy falls in the 70% range, based on verification studies carried out in the 1980s. There seems to be a large discrepancy between urban and rural data quality.

The existing information system seems most useful for administrative purposes, such as allocation of health resources. The broad age categories (0–14 years, 15–17 years, and

≥ 18 years) limit the evaluation of the health status of populations. The causal inference based on the available information has serious limitations typical for ecologic studies where the unit of observation is a group of individuals rather than the individual (5). The exposure within the units of observation (raions) is not expected to be homogenous, and differences in exposure between units of observation are possibly smaller than the individual differences within each unit of observation. There is very little information on potential confounding factors.

The Sverdlovsk Department of Health has computerized mortality information from 1992 and from cancer registries in six cities. The mortality data could be used to assess temporal and, to some extent, spatial variation in relation to environmental factors. The usefulness of the death registry would be improved through the inclusion of all contributory causes of death and the place of residence. Other useful additions include socioeconomic classification and coding of occupations to reflect possible occupational exposures.

Cancer registration, although nationally mandated in Russia, is a primary example of what can occur when local resources are limited. When a primary health care physician suspects that a patient may have a malignant disease or when there are positive findings from cancer surveillance, the patient is referred to an oncologist. The oncologist, together with other specialists, makes the confirmatory diagnosis and, according to regulations, completes a special form (090-y) within 3 days of the diagnosis. The treatment and follow-up of cancer patients in Russia is concentrated in oncological hospitals (dispansers) located in cities and towns. Each hospital is responsible for a population from a geographically defined area. The hospitals make annual reports to the oblast dispansers that report to the Ministry of Health and to the Statistical Department of the oblast. Two different forms are used: a statistical report on individuals with tumors (Form 35) and a statistical report on cancer cases (Form 7). Both forms summarize cancer cases in groups based on ICD-9 diagnostic codes. Form 35 displays cancer cases separately for new and for all cases, for cases verified morphologically, and by stage of cancer. There is a separate section for deceased cases and types of treatment. Form 7 provides age (18 5-year periods) and sex distributions. The Ministry of Health prepares a general annual report and 89 annual regional reports to the State Committee of Statistics (Goscomstat) using Forms 7 and 35. The effort by Oblast Center of Gossanepidnadzor personnel to obtain supplemental data for risk factor identification is compromised by serious problems of ascertainment. Because federally required data provide too little detail for inference generation and because local supplemental data collection is undermined by resource limitations, cancer registration is not sufficiently useful, despite a significant data acquisition effort and the existence of mandatory reporting.

Existing mortality data can be used for time–series analyses of the impact of short-term variation of environmental conditions on mortality and for analyses of spatial differences in mortality in relation to environmental factors. In the latter type of studies, further information on potential confounders is necessary for valid assessment of the role of the environment.

Polyclinics and hospitals are required to immediately report all cases of infectious diseases. The reports are made by telephone to the office of the Raion Center of Gossanepidnadzor. Information is recorded in the office either directly to computer file or on paper. All diseases are reported and recorded on an individual basis, except acute respiratory infections and influenza, which are reported and recorded in aggregated form on a daily basis for the following age categories: 0–2, 3–6, 7–14, and ≥ 15 years. These individual databases are maintained in the raion offices, and aggregated data are reported on a monthly basis to the Oblast Center of Gossanepidnadzor and to the Ministry of Health.

The nearly real-time information system is very sophisticated and, in principle, this type of immediate provision of key data can be highly useful for conducting environmental health studies. The validity of the infectious disease diagnoses is unknown.

The existing data are suitable for studying the relationship between infectious disease distributions in space and time and their potential influence from environmental factors. Studies of disease related to microbiologic water contamination are one appropriate use of this type of database.

Federal level. The Ministry of Health (Russian Federation) is responsible for collecting and assessing morbidity data. Health information is recorded at polyclinics and hospitals on special forms using ICD-9 codes. The information is reported annually through local (raion and city hospital centers), and regional (oblast) departments of health to the General Computing Center in the Ministry of Health. In addition, infectious disease reporting takes place through Gossanepidnadzor's local network, and both monthly and annual recording is performed.

There are three principal federal offices that maintain health and health-related information: the General Computing Center, the Federal Center of Gossanepidnadzor, and GoscomStat. The General Computing Center receives all morbidity information from the oblasts and is the principal repository of national health data. GoscomStat receives statistical reports on births and mortality. Primary data from the three federal offices is distributed to specific research institutes.

The data management of aggregated federal level health information is very well organized in the MedStat information system, which is prepared on a yearly basis by GoscomStat. There is systematic, partly automated quality control of information with a feedback system linked to producers of information at the regional level.

The information system can produce fast and efficient time trends and spatial distributions of disease occurrence, indices of health care systems, and other relevant information. This information is useful for administrative purposes when allocating resources for health care. The quality control between oblast and federal levels is well organized. However, there is no systematic quality control of information originating from the health care providers from raion and city levels to the oblast level. Further, the data are aggregated and transformed two or three times between the grassroots level and the federal level, resulting in the loss of important information. Consequently, the oblast level averages are not useful for assessing health effects of environmental factors.

The Federal Center of Gossanepidnadzor has three activities that deal with collection and analysis of health information: the Federal Center of Gossanepidnadzor receives the primary information on infectious diseases from the oblasts; carries out special studies; and the staff is responsible for the development of sociohygienic monitoring in the Russian Federation, which also covers environmental and health information of interest. The Federal Center of Gossanepidnadzor also maintains an individually based registry of occupational diseases.

The Federal Center of Gossanepidnadzor receives monthly reports that describe the frequency of 65 infectious diseases from all of the 89 oblasts and autonomous republics. Annual reports of 78 infectious diseases are also received. Furthermore, the Federal Center of Gossanepidnadzor receives annual reports describing the sanitary conditions in the territories.

All data were computerized in 1992 and are pooled into a general information system. The software used in the interface allows both numerical and graphic presentation and analysis of the data. All of the routine tables for the annual reports are produced by existing programs.

These data represent counts, means, or proportions prepared at the level of the

oblasts. The oblast level information is aggregated from the raions of each oblast. However, the Federal Center of Gossanepidnadzor has access to all the infectious disease data collected by the Gossanepidnadzor network. The raion level data were in computerized form in approximately 60–70% of the oblasts.

There is excellent management of information from the oblasts, and the information systems are well suited for the monitoring of infections. Local health records are used for studying infectious disease epidemics. These records are usually maintained on paper, and a team of investigators cooperate with the regional and local Gossanepidnadzor staff. The computerized information on monthly and annual counts and rates is not useful for assessing effects of environmental factors, although the occurrence of infectious diseases per se can be influenced by air, water, and soil pollution.

Federal level health information based on oblast level data is not useful for assessing the health effects of environmental factors because of similar but more pronounced limitations of the raion level. At the same time, the Russian Federation maintains a strong network and routines for collecting information on health phenomena, which forms a good nationwide framework.

Based on our assessment, we recommend that the the federal level agencies should develop health information for the purposes of environmental health as follows:

- Federal agencies should develop access to community (raion) level and individual health data. The federal agencies have a legitimate access to any local or regional level information. Although it may not be feasible or desirable at this stage to try to concentrate all of the individual information in the federal agencies, it is useful to develop the capacity to access this information; the current standardized statistical forms provide an excellent basis for such a database. An additional approach would be to develop standardized software to be used in handling routine individual health data.
- Federal agencies should focus on the development of selected information to maintain high quality. This focus would improve production of valid and relevant health information with optimal costs. The World Health Organization and the European Union are both developing suitable indices for monitoring population health and environmental conditions. It is advisable to follow the international experience in this field, although each country should develop the most suitable indices based on their special needs.
- Federal agencies should support and coordinate development of individual-based disease registries, such as existing registries for cancer, malformation, and occupational disease. Information on health and exposure of individuals is needed to assess the relationship between environmental exposures and the risk of disease. Often studies carried out in selected, optimal settings (cities or small areas) are the most efficient way to learn about exposure–response relationships; this knowledge can then be used both nationally and internationally. This is especially true with relationships between common diseases and well-defined exposures. However, in the study of environmental and other effects on rare diseases such as cancers and malformations, the possibility of bringing together information from large populations crucially improves the possibilities of studying these effects. Thus, the federal support and coordination of the currently emerging cancer and malformation registries would essentially increase the possibilities of studying the role of environmental factors in the causation of disease. Studies of occupational disease would add to our knowledge of possible effects of high-level environmental exposures. Coordination of the registries of common diseases would also be likely to improve the quality of information, although pooling of the information would not be necessary.
- Federal agencies should develop national health surveys that collect information on relevant health outcome and environmental factors in the home, work, and other environments. Besides routine health information flows and special epidemiologic studies, the use of national health surveys could provide important information on the relationship between environmental conditions and health. An appropriate sampling frame would also allow the assessment of public health impact from some of the most important factors.
- Federal agencies should establish and train task forces to maintain a center of excellence in environmental health and to provide consultation and disseminate information in the regions. There is a common belief that development of sophisticated information systems and automated analytical approaches are central to improving the use of empirical data in decision making. However, we believe that the primary issue is to improve the knowledge of Russian specialists in solving the diverse problems related to environmental health and rational environmental management. At the first stage, there is an urgent need for knowledgeable federal task forces consisting of experts with multidisciplinary backgrounds who can follow current international development, address complicated ad hoc environmental problems, and develop long-term strategies for improving the quality of environmental and health information and its appropriate use in the decision-making processes. These task forces could form centers of excellence capable of advising regional groups and disseminating information to the regions. Second, to ensure future development, the universities should be involved in teaching and research in the fields of environmental health and epidemiology.

Discussion

The strengths and weaknesses of Russian health information are discussed in the context of assessing environmental health effects. The availability of existing environmental information or new environmental information naturally influences this assessment.

In general, the Russian Federation maintains an extensive standardized nationwide organization that routinely collects health data in polyclinics and hospitals. The information is collected on standardized forms and reported through two or three stages to the national offices. The system forms a strong basis for acquiring descriptive health data, which can be used for administrative purposes such as allocation of resources and assessment of disease trends, with certain limitations due to dramatic changes in society during the 1990s. Although extensive in its allocations of manpower and breadth of surveyed health outcomes, health assessment in the Russian Federation is prone to the twin threats of overly ambitious expectations and erratic control of information quality. On the other hand, the hierarchical system of data collection has advantages over more decentralized or commercial health systems. Some of these advantages are underused. In particular, the current standardized data collection system with potential for completeness is a potentially excellent basis for surveillance techniques.

Problems of data quality, such as the deterioration in completeness of death reporting or the underestimation of the population at risk, have been considered as partial explanations for the dramatic rise in mortality rates in Russia and the decrease in life expectancy (*3,7,8*). In the 1970s and 1980s there was a deliberate underreporting of infant mortality because definition of infant mortality in the Soviet-era excluded all infants who died within 7 days of birth and were substantially preterm (< 28 weeks gestation, < 1,000 g birth weight, or < 35 cm birth length) (*9*). The Russian definition was changed in 1993, but the traditional reporting practice remains common. In a recent study, Leon et al. (*7*) analyzed the age-specific and cause-specific patterns of mortality in Russia between 1984 and 1994. On the basis of stable mortality rates from neoplasm in contrast to other causes over the same period, the authors inferred that the changes in life expectancy could not be an artifact related to underestimation of

the population. In diagnosing morbidity or defining the causes of death, the changes over time or the differences in location may seriously influence assessment of time trends or comparison between regions and countries. However, these problems do not necessary impair the use of the information in assessing the effects of environmental exposures. Understanding these limitations is the key in designing the optimal analytical approach.

Registries for cancer and other diseases evolve out of national reporting guidelines, where emphasis for completeness can be introduced starting from strategically important oblasts. Basic cancer registration appears to be feasible within a 5-year interval, provided that registry goals are well defined and that completeness takes precedence over detail. National or oblast level health surveys directed toward particular measures of health status and specific disease patterns are another underused strategy.

A major weakness in the current method of reporting is the aggregation and transformation of data. This may not disturb the production of health statistics, but it seriously limits the use of oblast and federal level data in assessing health effects of environmental exposures. Aggregation of data eliminates the use of an individual as a unit of observation and limits the use of information because it may introduce potential bias. The exposure parameters are not likely to represent a homogenous exposure over the unit of observation (raion or oblast). The measures of health outcomes are likely to be influenced by factors that affect the detection and diagnosis of cases as well as other determinants of disease that may not have been included in the study (i.e., confounders).

The individual visits to polyclinics and hospitals are recorded systematically on paper, and ICD-9 codes are generally used in disease registration. In some areas, part of the individual information is computerized. This creates a basis for estimating disease occurrence over time and place. However, there is no general practice of quality control for data collection, entry, and handling, and there is little information on the variations of the diagnostic practices over space and time. Diagnostic practices for some diseases may vary considerably by region; thus spatial comparison of disease distributions may not be meaningful. Also, limited information is available on potential confounders on either the individual or group level.

In spite of these limitations, the following approaches to the analysis of existing data appear to be both feasible and fruitful.

- Analysis of the effects of short-term exposure to environmental conditions on mortality and morbidity. Daily monitoring of air pollution is performed in most industrial cities in the Russian Federation. The routine monitoring usually comprises three or four daily measurements of 20 min. Follow-up of daily number of episodes or events in a defined dynamic population eliminates the confounding because of personal factors and a number of environmental factors. In modern time–series analysis, the latency period of the effects, seasonal trends of disease occurrence, and effects of other pollutants and meteorologic factors can be taken into account. The accuracy and precision of the disease measures vary, and some measures are sufficiently valid. The possible health outcomes for time–series analysis include general and disease-specific mortality, occurrence of respiratory and gastrointestinal infections, episodes of asthma, and acute cardiovascular problems. This approach has already been applied to study the relationship of general and cause-specific mortality to daily levels of air pollutants in Yekaterinburg and Nizhni Tagil (*4*).
- Analysis of the effects of long-term exposure to environmental conditions. Using the place of residence as the basis of exposure assessment could be feasible when the population is expected to be stable and when there is information on past exposures that can be allocated to residential areas. In such a situation, associations between disease occurrence and cumulative or time-specific levels of exposure can be estimated in a retrospective cohort study using either fixed or dynamic cohorts. However, routine data on potential confounders are insufficient. Age and sex standardization can be carried out in most locales, and some disease registrations include crude information on occupation.

The combination of information from routine data and new data collection is likely to be the most effective way to assess the relationship between environmental exposures and diseases. The limited resources allocated for health care may, in the future, require a reduction in the breadth of routine health data collection. The emphasis would be on careful planning of data collection and performance of well-designed and accurately focused epidemiologic cohort or case–control studies. In general, small homogenous countries, such as the Nordic countries, have been able to make effective use of standardized disease registry data in assessing potential effects of environmental factors on birth outcomes and cancer (*10–12*). It is unlikely that these approaches will transfer easily to Russian conditions. Although there is a strong and justifiable desire to rapidly translate information on environmental health effects into policy decisions, it seems more important at this time to emphasize data quality, completeness, and plans for the use of data.

References and Notes

1. Konygin E. The largest IBRD-financed project is being implemented in Russia. In: Environment Management Newsbulletin No 1. Center for Preparation and Implementation of International Projects on Technical Assistance. Moscow:Russian Ecological Federal Information Agency Publishers, 1996;6–14.
2. National Center for Health Statistics. Vital and Health Statistics: Russian Federation and United States, Selected Years 1980–1993. PHS 95-1485. Hyattsville, MD:National Center for Health Statistics, 1995.
3. Cherniak MG, Chorba TL, McNabb SJN. Demographic features of premature mortality in the countries of Central and Eastern Europe and the new independent states. Curr Issues Public Health 2:78–89 (1996).
4. Özkaynak H, Spengler JD, Jaakkola JJK, Ford T, Xue J, Egorov A, Schwartz J, Kuzmin S, Rakitin P, Privalova L, et al. Evaluation of existing environmental information systems in Russia applicable for human health effects assessment. In: Air Pollution in the Ural Mountains: Environmental, Health, and Policy Aspects (Linkov I, Wilson R, eds). Boston, MA:Kluwer Academic Publishers, 1998;195–217.
5. Rothman KJ. Modern Epidemiology. Boston:Little, Brown and Company, 1986.
6. Morgenstern H. Ecologic studies in epidemiology. Annu Rev Public Health 16:61–81 (1995).
7. Leon DA, Chenet L, Shkolnikov VM, Zakharov S, Shapiro J, Rakhmanova G, Vassin S, McKee M. Huge variation in Russian mortality rates 1984–94: artefact, alcohol, or what? Lancet 350:383–388 (1997).
8. Notzon FC, Komarov YM, Ermakov S, Sempos CT, Marks JS, Sempos EV. Causes of declining life expectancy in Russia. JAMA 279:793–800 (1998).
9. Anderson B, Silver B. Infant mortality in the Soviet Union: regional differences and measurement issues. Popul Dev Rev 12:705–738 (1986).
10. Rylander L, Stromberg U, Hagman L. Decreased birthweight among infants born to women with a high dietary intake of fish contaminated with persistent organochlorine compounds. Scand J Work Environ Health 21:368–375 (1995).
11. Koivusalo M, Pukkala E, Vartiainen T, Jaakkola JJK, Hakulinen T. Drinking water chlorination and cancer - a historical cohort study in Finland. Cancer Causes Control 8:192–200 (1997).
12. Magnus P, Jaakkola JJK, Skrondal A, Alexander J, Becher G, Krogh T, Dybing E. Water chlorination and birth defects. Epidemiology 10:513–517 (1999).