

UKRAINIAN NATIONAL CANCER REGISTRY: PROBLEM OF DATA QUALITY FOR CANCER EPIDEMIOLOGICAL STUDIES OF THE CONSEQUENCES OF CHERNOBYL ACCIDENT

Z.P. Fedorenko, L.O. Goulak, Y.L. Gorokh, O.V. Soumkina, A.Y. Ryzhov
Institute of Oncology of AMS of Ukraine, National Ukrainian Cancer Registry
03022, Ukraine, Kyiv, Lomonosova str., 33/43

НАЦИОНАЛЬНЫЙ КАНЦЕР-РЕГИСТР УКРАИНЫ: ПРОБЛЕМА КАЧЕСТВА ДАННЫХ ПРИ ИЗУЧЕНИИ ОНКОЭПИДЕМИОЛОГИЧЕСКИХ ПОСЛЕДСТВИЙ ЧЕРНОБЫЛЬСКОЙ АВАРИИ

З.П. Федоренко, Л.О. Гулак, Е.Л. Горох, Е.В. Сумкина, А.Ю. Рыжов
Институт онкологии АМН Украины, Национальный канцер-регистр
03022, Украина, Киев, ул. Ломоносова, 33/43

Abstract

This paper presents an experience of National Ukrainian Cancer Registry (UCR) which performed the evaluation of some factors that confound the incidence level, such as: deficient and duplicate cancer registration, discrepancies in medical and personal information about a patient from different sources. The disagreements found show the low level of quality and completeness of medical data available for research in Ukraine. Conducting the epidemiological studies of children incidence without preliminary scrupulous analysis of primary data, involving of all sources available and modern record linkage technology can cause wrong and groundless conclusions.

Keywords: cancer epidemiological study, cancer registry, data quality, record linkage.

INTRODUCTION

The data of state statistic reports, of past years in particular, as well as medical and registration documents from the archives of some medical institutions or special databases for specific categories of patients (the registries) can be used as source data for cancer epidemiological study in the field of the consequences of Chernobyl accident.

SUBJECT AND METHODS

All attempts to organize the permanent collection of data on incidence in some classes of disease of children population of the contaminated territories came into collision with the fact that children become adult, change their place of residence and get out of the field of activity and the interests of the researchers. Meanwhile the latent period of cancer diseases is rather prolonged, and those who were children at the moment of Chernobyl accident can reveal the disease in adult age. Monitoring of such diseases is only possible using the total permanently acting system of registration. The National Ukrainian Cancer Registry is such system.

The sample of new cancer cases in period of total coverage with the unified technology of UCR (since 1992 in Volynska oblast, since 1993 in Zhytomirska and Chernigivska, since 1997 in Kyivska and since 1998 in Rovenska oblast till 2002 in all these oblasts) in patients with year of birth 1972-1986 (aged 0-14 at the moment of accident) from rajons of higher radiation control of Rovenska, Volynska, Zhytomirska, Kyivska and Chernigivska oblasts has been selected from

ВВЕДЕНИЕ

Основой для проведения онкоэпидемиологических исследований в области последствий Чернобыльской катастрофы могут служить данные государственной статистики, особенно ретроспективные, а также медицинские и регистрационные документы из архивов лечебных учреждений, или специально созданные базы данных, содержащие информацию об отдельных категориях пациентов (регистры).

ОБЪЕКТ И МЕТОДЫ ИССЛЕДОВАНИЯ

Попытки организации долговременного сбора данных по заболеваемости отдельными классами заболеваний среди детского населения загрязненных территорий столкнулись с тем, что со временем дети взрослеют, меняют место проживания и выходят из сферы интереса исследователей и деятельности систем сбора данных. Между тем, латентный период для онкологических заболеваний продолжителен, и у лиц, которые были детьми на момент аварии на ЧАЭС, заболевание может проявиться во взрослом возрасте. Мониторинг таких заболеваний возможен только в рамках тотальной постоянно действующей системы регистрации, какой является Национальный канцер-регистр Украины.

Из общего массива данных Национального канцер-регистра была сформирована выборка случаев рака, зарегистрированных у лиц 1972-1986 гг.р. (возраста 0-14 лет на момент аварии) в районах радиационного контроля Ровенской, Волынской, Житомирской, Киевской и Черниговской областей за период, соответствующий тотальному охвату этих регионов единой технологией автоматизированной регистрации (с 1992 г. для Волынской области, с 1993 г. для Житомирской и

the Ukrainian Cancer Registry database. The sample included 1 104 cancer cases that were tested with complex procedures of automated logical checks.

For estimation of migration and degree of duplicate records the sample of cases of lymphoma or leukemia (ICD-10 codes C81-C96) in patients with year of birth 1970-2001 has been selected from the UCR database. Size of this sample was 13 126 records. The same criterion was used for selecting the sample of 2 209 records of thyroid cancer cases. These samples were tested with the automatic duplicate checking procedure using the record linkage technology.

RESULTS AND DISCUSSION

The study helps to reveal the most important classes of problems concerning the generally established practice of collection and processing of personified information that often remain aside of attention of the researchers.

Delay in registration

In cancer registration the principle of centralized registration is established: information from all the places where the cancer diagnosis has been established or a cancer patient received the treatment is directed to the oncological dispensary at place of residence of the patient in form of regulated informational document. Not always it is done in the prescribed time. Many children receive treatment not in the oncological dispensaries at place of their residence but in specialized clinics. In 1996 only 86% of new cancer cases in patients of 1970+ year of birth living in the territories with higher dosimetric control was registered in the same year. This percentage has a tendency to growing, and in 2000 95% of the new cancer cases were registered in the same year. However we have no certainty that we observe the real improvement of the terms of registration but not the falsification of date of diagnosis and substitution it with the date of registration of the case.

Doubling in registration

The main source of the personified information for any research are the medical records – these are various registration medical forms, as well as lists and samples produced from the information of these documents. The quality of manual

Черниговской, с 1997 г. для Киевской, с 1998 г. для Ровенской области и по 2002 год). Выборку составили 1104 случая злокачественных новообразований различных локализаций, которые были подвергнуты комплексным процедурам автоматизированного логического контроля.

Для оценки миграции и уровня дублирования информации была сформирована выборка случаев злокачественных заболеваний лимфатической и кроветворной системы (код МКБ-10 C81-C96) у лиц 1970-2001 года рождения, находящихся в Национальном канцер-регистре Украины. Мощность этой выборки составила 13126 записей. Такая же выборка из данных канцер-регистра объемом в 2209 записей была сформирована для злокачественных заболеваний щитовидной железы. Для этих выборок был проведен автоматизированный контроль наличия дублирующихся записей с применением технологий и средств линкиджа (связывания пар) данных.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Исследование позволило выявить наиболее значительные классы проблем, возникающих при сложившейся практике сбора и обработки персонифицированной информации и зачастую остающихся вне сферы внимания исследователей.

Задержка в регистрации информации

При регистрации злокачественных новообразований (ЗН) применяется принцип централизованного учета: информация из всех мест, где выявлен или получает лечение больной со злокачественным заболеванием, направляется в онкологический диспансер по месту жительства больного в виде регламентированных информационных документов. Не всегда это делается достаточно оперативно. Многие дети получают лечение не в онкологических диспансерах по месту жительства, а в специализированных центрах. По состоянию на 1996 год только 86% новых случаев ЗН у лиц 1970+ г.р. на территориях радиационного контроля зарегистрировано в том же году. Этот процент имеет тенденцию к повышению, в 2000 году было зарегистрировано уже 95% всех заболевших в этом году, однако нет уверенности в том, что мы наблюдаем реальное улучшение оперативности регистрации, а не фальсификацию дат установления диагноза и подмену их датами регистрации.

Дублирующий учет

Основным источником персонифицированных данных для любых исследований являются записи медицинских документов – различных регистрационных форм, а также всевозможные списки и выборки, формируемые на основании

writing of these forms resulted in numerous mistakes in identification of patients, and, therefore, in duplicate records arising at time of receiving the information on the same patient from different sources by the researcher. In addition to spontaneous faults the non-observance of the registration technology takes place (registration with not indicating the real place of residence, but the place of diagnosis, receiving the treatment or burial, registration of multiple tumours in one patient as several patients, etc.). In case of manual processing of documents the revealing and eliminating of duplicate records is very difficult and unreliable, if the group under study is large it is scarcely feasible.

According to the estimates of Ukrainian Cancer Registry staff during the inputting of paper registration forms from regional oncology dispensaries into the database from 5% to 15% of paper cards were found as duplicates and eliminated. However, automated database in itself does not guarantee the absence of duplicate records. The additional work of revealing and joining the duplicates with using special means is needed. Such means were used in UCR software both for regular checks and at the time of data input.

If the source of data is manually filled-in registration medical documents or a list of data selected by hand from these documents then check for absolute coincidence of passport data is not sufficient. The special data linkage means are necessary for effective solving this problem.

With the help of enhanced automated data linkage procedure the checks for duplicate records in the sample of 13 126 patients of Ukraine with lymphoid and hemopoietic malignancies with year of birth 1970-2001 have been done. 441 double records were revealed and analysed. It estimated 3.8% level of doubling in registration for this group. 220 duplicates were found within the bounds of the same regional registry. 221 duplicates were the result of inter-oblasts migration. The similar result was obtained for the thyroid cancer sample.

Such low levels of doubling was achieved owing to the long-time efforts of regional cancer registries and because of permanent using the means of data check in registration in Ukrainian Cancer Registry. Nevertheless, many doubled records were revealed only after using the updated algorithms specially developed in Ukrainian Cancer Registry and designed for the conditions of Ukraine, taking into account the peculiarities of circulation and the quality of medical information. In the most suffered from the Chernobyl accident regions and in Kyiv city

этих документов. Качество ручного заполнения таких форм приводит к множественным ошибкам при идентификации больных, и, следовательно, к дублирующимся записям, создающимся при поступлении к исследователю информации об одном и том же пациенте из различных источников. Помимо спонтанных ошибок существуют нарушения технологии регистрации (учет больных не по месту прописки, а по месту выявления диагноза, получения лечения или захоронения; регистрация множественных раков как отдельных больных и др.). В условиях ручной обработки информации выявление и устранение дублирующихся записей очень трудоемко и ненадежно, а при большой исследуемой группе – практически нереально.

По оценкам Национального канцер-регистра Украины при перенесении бумажных картотек региональных онкологических диспансеров в компьютер выявлялось и устранялось в качестве дубликатов 5–15% общего объема картотеки. Однако автоматизированная база данных не является гарантией отсутствия дубликатов, необходима дополнительная работа по их выявлению и устранению с применением специализированных средств. Такие средства использовались в программном обеспечении канцер-регистра для регулярного контроля и при вводе новых записей.

Если в качестве источника данных применяются вручную заполненные документы или же вручную сформированные из них списки-выборки, то контроль полного совпадения паспортных данных недостаточен. Необходимы специализированные средства линкиджа, способные эффективно решать эти проблемы.

С помощью усовершенствованных автоматизированных средств линкиджа данных был проведен контроль наличия дублирующихся записей в выборке из 13126 больных ЗН лимфатической и кроветворной системы 1970-2001 г.р. в Украине. Была выявлена и проанализирована 441 пара дублирующихся записей, что позволило оценить уровень дублирующего учета для рассматриваемой группы в 3,8%. Из них в пределах одного регионального регистра были выявлены 220 пар, и 221 пара явились результатами межобластной миграции. Аналогичные результаты были получены для выборки ЗН щитовидной железы.

Такой низкий уровень дублирования достигнут в результате многолетних усилий региональных канцер-регистров и наличия штатных средств контроля при регистрации данных. Тем не менее, значительное число пар дублирующихся записей были выявлены только после применения усовершенствованных алгоритмов, специально разработанных в канцер-регистре для применения в условиях Украины с учетом специфики и

the level of doubling in cancer registration is sufficiently higher: in Kyivska oblast it is above 13%. It can be partly explained with migration of the population

Migration factor

The Chernobyl accident caused considerable dislocations of Ukrainian population. Revealing of such facts is possible in a centralized way, finding this information in the regional registry at the place of original registration of the patient. Children and young people are the most mobile cancer patients and for them such a work is necessary.

Whereas factor of migration from the contaminated areas is still taken into consideration with some researchers, but during the last years the inverse tendency for children with lymphoma or leukemia has become apparent: the “migration of incidence” to the contaminated regions, especially to Kievska oblast.

Quality of registration of medical information

According to the data of Ukrainian Cancer Registry, near 15% of cancer diagnoses in children in Ukraine have no morphological (pathomorphological) verification.

After the duplicate checks the analysis of doubled records has been conducted. The important discrepancies in medical information about the same patient from different sources have been found: in 45% the year of diagnosis differs, in 24% even the type or variant of lymphoid or hemopoietic malignancy did not coincide (the third figure of ICD-10 code).

CONCLUSIONS

After the Chernobyl accident a considerable number of studies concerning the cancer incidence of children population in the contaminated areas was initiated. Frequently different editions announce contradictory and discrepant information. Non-critical manner of using data from different sources can lead to obtaining such results of study that reflect not real changes in level of cancer incidence, mortality, etc., but rather the bias errors or differences in registration or changes in principles and ways of gathering and processing the information.

As a main ground for a group of problems a low

качества циркулирующей здесь медицинской информации. В наиболее пострадавших вследствие аварии на ЧАЭС регионах и городе Киеве уровень дублирования значительно выше - в частности, в Киевской области – более 13%. Это может частично объясняться миграцией населения.

Фактор миграции

Чернобыльская катастрофа вызвала значительные перемещения населения Украины. Выявить такие факты можно централизованно, найдя информацию об интересующем нас лице в региональном регистре по первоначальному месту учета. Дети и лица молодого возраста являются наиболее мобильной из всех категорий онкологических больных, и для них такая работа просто необходима.

Если фактор миграции с загрязненных территорий еще принимается во внимание некоторыми исследователями, то в последние годы для пораженного ЗН лимфатической и кроветворной системы детского населения проявилась обратная тенденция: миграция заболеваемости в чернобыльские регионы, в частности, в Киевскую область.

Качество регистрации медицинской информации

По данным канцер-регистра Украины около 15% онкологических диагнозов, выявленных у детского населения Украины, не имеют морфологической верификации.

По результатам исследования дубликатов был проведен анализ записей, признанных относящимися к одному и тому же лицу. Выявлены существенные расхождения данных о больном из различных источников: в 45% случаев отличался год установления диагноза, в 24% случаев не совпадал даже вид или вариант заболевания лимфатической и кроветворной ткани (третий знак кода МКБ-10).

ЗАКЛЮЧЕНИЕ

После аварии на ЧАЭС было инициировано значительное число исследований, посвященных проблеме онкологической заболеваемости детского населения на загрязненных территориях. Зачастую различные публикации содержат не согласующиеся и противоречивые сведения. Некритичное использование данных различных источников может привести к получению таких результатов исследования, которые отражают не реальные изменения уровня заболеваемости, смертности и т.д., а систематические ошибки или отличия в учете информации, изменение принципов и путей ее сбора и обработки.

Основной причиной целого класса проблем можно

culture of work with the personified information both in researchers and, above all, in persons who perform the initial registration and processing of the information – medical registrars, physicians and medical statisticians - can be considered.

Frequently the study is conducted using not the complete data (the data of the first year of registration or a separate source of data). It is impermissible in epidemiological study of children incidence, because the most frequent violations of data registration principles happen exactly for this group of patients.

The level of doubling of registration has a substantial influence on the results of the study in present practice of conducting the epidemiological studies. Effective estimation of this influence can be done only with the use of specialised means of data linkage.

считать низкую культуру работы с персонифицированной информацией, как у исследователей, так и в первую очередь у лиц, проводящих первичную регистрацию и обработку информации – медрегистраторов, врачей, статистиков.

Зачастую исследования проводятся на не полных данных (данные первого года учета или отдельный источник данных). Это недопустимо при эпидемиологических исследованиях детской заболеваемости, поскольку именно для этой группы заболевших наиболее часты нарушения принципов регистрации информации.

Уровень дублирующего учета в существующей практике проведения эпидемиологических исследований имеет значительное влияние на результаты исследования. Эффективно оценить это влияние можно только с применением специализированных средств линкиджа данных.

REFERENCES

Parkin D.M., Chen V.W., Ferlay J., Galcrean J., Storm H.H., Whelan S.L. Comparability and Quality Control in Cancer Registration. IARC Technical Report No.19. International Agency for Research on Cancer (WHO) and International Association of Cancer Registries, Lyon, 1994: 169 p.

Jensen O.M. et al. Cancer Registration: Principles and Methods. IARC Scientific Publications No.95. International Agency for Research on Cancer, Lyon, 1991: 288 p