

The Brazilian cohort of pulp and paper workers: the logistic of a cancer mortality study

Coorte brasileira de trabalhadores de indústria de celulose e papel: a logística de um estudo de mortalidade por câncer

Ana Claudia Gastal Fassa ¹
Luiz Augusto Facchini ¹
Marinel Mór Dall'Agnol ¹

¹ Departamento de Medicina Social, Faculdade de Medicina, Universidade Federal de Pelotas. C.P. 464, Pelotas, RS 96010-160, Brasil.

Abstract *The International Agency for Research on Cancer (IARC) proposed this international historical cohort study trying to solve the controversy about the increased risk of cancer in the workers of the Pulp and Paper Industry. One of the most important aspects presented by this study in Brazil was the strategies used to overcome the methodological challenges, such as: data access, data accuracy, data availability, multiple data sources, and the large follow-up period. Through multiple strategies it was possible to build a Brazilian cohort of 3,622 workers, to follow them with a 93 percent success rate and to identify in 99 percent of the cases the cause of death. This paper, has evaluated the data access, data accuracy and the effectiveness of the strategies used and the different sources of data.*

Key words *Epidemiology; Cohort Studies; Data Sources; Neoplasms*

Resumo *A Agência Internacional para Pesquisa sobre Câncer (IARC) propôs uma coorte histórica internacional para tentar resolver a controvérsia sobre o risco aumentado de câncer em trabalhadores das indústrias de celulose e papel. Um dos aspectos mais relevantes deste estudo no Brasil refere-se às estratégias utilizadas para superar desafios metodológicos, tais como: acesso aos dados, acurácia e disponibilidade dos dados, múltiplas fontes de dados e longo período de acompanhamento. Através de múltiplas estratégias, foi possível construir uma coorte brasileira de 3.622 trabalhadores, acompanhar 93% dos casos e identificar a causa de 99% dos óbitos. Este artigo avalia o acesso aos dados, sua acurácia e a efetividade das estratégias utilizadas para o acompanhamento da coorte.*

Palavras-chave *Epidemiologia; Estudos de Coorte; Fontes de Dados; Neoplasias*

Introduction

The cancer risk in the Pulp and Paper Industry has been controversial. Chlorine compounds are some of the most important chemical substances suspected to be carcinogenic. The studies already done have suggested an excess of certain cancers, especially lung cancer, malignant lymphomas, but for most part, they are inconclusive with regard to considerations of etiologic agents (Boffetta et al., 1991; Torén et al., 1996).

Thus, the International Agency for Research on Cancer (IARC) decided to perform the *Multicentric International Cohort Study of Workers in the Pulp and Paper Industry* which included 18 countries (Boffetta et al., 1991). The aim of this project was to investigate the mortality and cancer morbidity in personnel employed in plants producing pulp, paper and paper products, and in mills involved in the recycling, in relation to specific processes and chemical exposures in these industries (Boffetta et al., 1991). The historical cohort was planned to follow more than 100,000 workers (WHO, 1995).

The Brazilian Cohort is participating in the mortality study with 3,622 workers of one Pulp and Paper Industry. The studied industry is one of the 10 biggest in Brazil producing 300,000 tons per year of pulp and 2,000 tons per year of paper with a Kraft process. This is one of the first historical cohort studies carried out in Brazil and one of the few which had completed the follow-up with an acceptable lost.

This study is significant because of methodological challenges, such as data access, data accuracy, data availability, multiple data source, and the large follow-up period. The lack of models in the country and the lack of experience of the research team with this study design were also important challenges.

To people interested in learning more about historical cohort studies, or to implement them, it is instructive to examine in detail the methodology and operational issues of previous studies (Breslow & Day, 1987). Considering that, this paper describes the strategies used to complete the follow-up of the Brazilian Cohort, and discusses some critical issues to carry out a historical cohort study.

Methodology

The Brazilian cohort

The Brazilian cohort involves blue and white-collar workers, both sexes, with a minimum of one year continuous employment in the industrial area of a Pulp and Paper Industry, which is located in a city nearby the capital of the state of Rio Grande do Sul.

The exposition period of the cohort goes from June 1969 – when the industry starts – to December 1991. The workers dismissed before March 1972 were excluded, because a great part of them were working in the construction of the industry. The follow-up period stops in December 1994 (Figure 1).

Logistic

Cohort construction

To construct the cohort, the following identifying data were collected: worker's name, name of parents, date of birth and sex, and the occupational data, such as department, job titles and dates of starting and ending each occupational period.

This information was extracted at the industry from three databases: the archives, the active workers files and the computerized records. The identification data was computerized to build the list of workers and to start the vital status identification. The list allowed us to check the workers who had been employed more than one time to join their data.

Four research assistants had worked in the cohort construction for two months. After this first data collection, it was realized that the parents' names were needed to link the identification data with other databases. One research assistant collected this extra information in six months.

Vital status identification

Electoral Court search

The central strategy to identify the vital status of the cohort was the Electoral Court search. This option was related to the legal obligation to vote for the Brazilian people over 18 years of age. Thus, all Brazilians over 18 years old who are formally employed are registered at the Electoral Court, and these records are computerized.

To establish this vital status, a list was printed with the identification data of the cohort

and sent to the Electoral Court. The workers' name, parents' names and date of birth were used by the Electoral Court to check the database of the state of Rio Grande do Sul. It identified the workers that were alive, some recent deaths and the ones who were not on the registers. A number of voting registrations were not used because it had changed in 1986, and for the most part the workers with voting registration numbers in the industry records were the old ones.

The Electoral Court completed the search in six months.

Identification, at the industry, of the Voting Registration Emission Place and the current number of the voting registrations

For the workers who were not in the Electoral Court records, they were identified at the industry, and the Voting Registration Emission Place. Also the voting registration that had the current numbers were sent to the Electoral Court, which did a new search. One industry worker was paid by the research to do this search overtime. It took one month.

Information from professionals of the industry's health center and from workers employed at the industry for a long time

Reseachers asked the professionals of the health center and the workers that had been at the industry for a long time to look at the list of workers that were not on the Electoral Court records and if they knew, to identify their vital status.

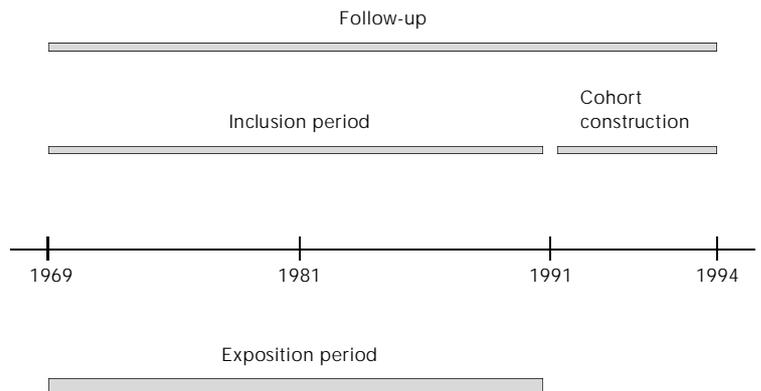
Household search in the city where the industry is located

This search was done for the 329 workers registered at the industry who were living in the city where the industry is located.

Five people were selected from the city where the industry is located to do the search. If the worker or his (her) family were living at the address, the vital status was identified. If there was another person living currently at the address, they were asked, or their neighbors were asked if they knew any information concerning the worker's whereabouts. People living in the city for a long time updated the names of the streets that had been changed. The household search took two months.

Figure 1

Design of the historical cohort. Brazilian pulp and paper study, 1998.



Phone search at the city where the industry is located and at the state capital

A phone search was done in the city where the industry is located and at the state capital. A similar search was done in the city where the industry is, in the capital of state, and in small cities around the area. It used the worker's name, the family name and the address (including neighbors) to search the phone numbers. At least one number was tried for each one of the 165 workers searched, but in some cases two or three phone numbers were tried. One research assistant and the coordination of the Brazilian cohort did the identification of the phones. The phone calls and the household search were done simultaneously.

Identification of the state to where the workers migrated: Electoral Court search in other states

Through the information gained during the household search and from workers employed at the industry for a long time, the place of residence was identified for 24 workers. After that, the Electoral Court was asked to search for them in the registers of the specified states.

Search in the Death Registration Office in the city where the industry is located, and in the oldest and largest Death Registration Office of the capital of the state

The Death Registration Office of the city where the industry is located and the oldest and largest Death Registration Office of the capital of the state were searched to find the 297 lost workers who were not found through other strategies.

In the Death Registration Office of the city where the industry is located, the data had two types of organization: books in order of date of death with an alphabetical resume at the end, and cards in alphabetical order, but not very well organized. The data was not computerized. One research assistant took 2 months to do the search.

In the oldest and largest Death Registration Office at the capital of the state, the data was computerized. The workers of the Death Registration Office had done this search in one week.

Cause of Death Identification

The search for cause of death was done in the Death Registration Office of the city where the industry is located and in all seven Death Registration Offices in the capital of the state. This search was done for the 95 workers identified as dead. The process took one month. Just two of the Death Registration Offices were computerized, the others were organized in two ways: a book ordered by date of death and cards in alphabetical order. Some Death Registration Offices were contacted by phone in other cities where some deaths were registered. The Office sent us a photocopy of the documents by mail or fax. For the cases that the search in the Death Registration Offices were unsuccessful, the family was contacted to get a photocopy of the Death Certificate. The workers of the Death Registration Offices did the search of cause of death and one research assistant got the Death Certificates with the families.

Results

Cohort construction

It has identified 3,622 workers that had been working in the industrial area for more than one year. It made a careful search to exclude repeated records, to join complementary records, and to preserve the homonymous (Figure 2).

Vital status identification

Electoral Court search

From the list of 3,622 workers of the Pulp and Paper Industry, the Electoral Court identified 81 percent alive, 0.3 percent dead and 19 percent who were not in their records. Once the Electoral Court eliminated from the records the deaths that occurred several years ago, it was

identified relatively more workers alive than dead. (Figure 2).

Identification, at the industry, of the Voting Registration Emission Place and the current number of voter registrations

At the industry, the Voting Registration Emission Place was identified for all 19 percent of the lost workers, but the current number of the voting registrations were available only for one percent of the lost workers. With this information, the Electoral Court had done a new search, identifying 85 percent alive, 0.3 percent dead and 15 percent were still not in their records (Figure 2).

Information from professionals of the industry's health center and from workers employed at the industry for a long time

This strategy allowed us to identify 11 lost workers, six workers were alive and five workers were dead. Thus, the proportion of lost workers was reduced to 14.6 percent (Figure 2).

Household search in the city where the industry is located

For the 329 workers registered as living in the city where the industry is located, the household search identified the vital status of 44.4 percent of them. This strategy decreases the study loss to 10.6 percent (Figure 2).

Phone search at the city where the industry is located and at the state capital

Through the phone list, it was searched 165 workers, identifying the vital status of 44.8 percent of them. The loss was reduced to 8.6 percent by the phone search (Figure 2).

Identification of the state for where the workers migrated: Electoral Court search in other states

The 24 lost workers identified as migrant were traced in the specified states by the Electoral Court. It has identified one worker as dead and 13 workers as alive. The proportion of lost workers was reduced to 8.2 percent through this strategy (Figure 2).

Search in the Death Registration Office in the city where the industry is located and in the oldest and largest Death Registration Office of the capital of the state

This strategy identifies only dead workers, recognizing 25 percent of the deaths in the cohort. After this procedure, the study loss decreased to 7.5 percent (Figure 2).

Cause of death identification

Despite the difficulties in the Death Registration Offices and the constraints in asking the family for the Registration of Death, the cause of death was identified in 99 per cent of the cases (Figure 2).

Discussion

After four years of developing multiple strategies and using a variety of sources we obtained the success to construct a cohort of 3,622 workers and to follow them until 1994 with an acceptable lost (Monson, 1990). Despite all logistical difficulties, it was possible to complete the follow-up of this historical cohort study. As Breslow & Day (1987) mentioned, the success with which the follow-up is achieved is probably the basic measure of the quality of cohort studies.

In Brazil, like other countries, the access to personal data is restricted by law (Hernberg, 1992). However, this access can be facilitated to researchers, if ethical requirements are observed. This study obtained an ethical approval from the IARC and the Ethical Committee of the Faculty of Medicine of the Universidade Federal de Pelotas (UFPel), who considered the relevance, the objectives of the study, and the confidentiality of the data. The industry and the Chief Judge of the Electoral Court, taking into account the ethical commitment of the study, allowed access for all the data. Furthermore, as soon as the vital status was identified, the worker's name and parents' names were excluded from the database. The greater problem regarding access to the data was the restriction to do quality control in the search realized by the Electoral Court and the Death Registration Offices.

The logistic was facilitated by the size and location of the involved cities. The studied workers mostly lived in the city where the industry is located or in the capital of state. In many cases, industries are located in industrial areas with several cities of different sizes

Figure 2

Strategies used to follow-up the cohort. Brazilian pulp and paper study, 1998.

Cohort construction		
(N = 3,622)		
Vital status identification		
Electoral court search		
Alive = 2,939	Dead = 12	Lost = 671
Identification, at the industry, of the voting registration emission place and the current number of the voting registers		
Electoral court search		
Alive = 3,067	Dead = 13	Lost = 542
Information from professionals of the industry's health center and from workers employed at the industry for a long time		
Alive = 3,073	Dead = 18	Lost = 531
Household search in the city where the industry is located (n=329)		
Alive = 3,185	Dead = 52	Lost = 385
Phone search at the city where the industry is located and at the state capital (n=165)		
Alive = 3,241	Dead = 70	Lost = 311
Identification of the state where the workers migrated: electoral court search in other states (n=24)		
Alive = 3,254	Dead = 71	Lost = 297
Search in the death registration office in the city where the industry is located and in the oldest and largest death registration office of the capital in the state		
Alive = 3,254	Dead = 95	Lost = 273
Cause of death identification (n=95)		
Search in the death registration office of the city where the industry is located and in all (7) death registration offices of the capital in the state		
Identified = 94	Not identified = 1	
Cancer cases = 19	Other causes = 75	

around where the workers live. This can make it impossible or very expensive for the household search, and the search in the Death Registration Offices, which was essential to reduce the losses.

The most important problem to develop this historical cohort was the data quality. Generally, several problems can be found in the databases available in Brazil, such as under registration, missing-data and outdated data, once they are collected regularly for other purposes than research (Schilling, 1986). Hopefully, the state of Rio Grande do Sul has one of the best registers in the country. Otherwise, formal workers need to update their registers to get new employment. Both aspects contribute to the data quality of the cohort.

On the other hand, in Brazil, each person has several personal identification numbers,

which make difficult to link different sources of information (Hernberg, 1992). With the lack of a unique identification number, the link was done using the worker's name, parents' names, and date of birth.

The database of the industry used to construct the cohort had relatively accurate data on identification and exposure of workers, which is essential to the identification of the vital status, and the cause of death, but also to the data analysis (Schilling, 1986). One of the few problems found was missing data related to parents' names. However, the losses due this missing data were very low (Hernberg, 1992).

The Electoral Court was the central strategy to follow-up the cohort, once it identified the vital status for 81 percent of the cohort in the first search. Then, the other strategies needed to follow only 671 unidentified workers, instead of the whole cohort (3,622 workers) needing to be done. Despite this, there was a relatively greater identification of living people in relation to the dead people, because the Electoral Court did not maintain the records of deceased people a long time ago. Among the unidentified workers, there were deceased workers, migrants, or lost workers. This is a potential problem in any type of cohort study, particularly those involving a long follow-up period (Kleinbaum et al., 1982).

The Electoral Court updated their records based on the information sent by the Death Registration Offices, but they often did not have control of the ones that were responsible for this obligation and the others that were late.

Thus, there was a possibility of false living people. To manage this problem, it was decided to wait six months from the end of the follow-up to the beginning of the search. This was done to allow the Death Registration Offices time to send the data of death to the Electoral Court. There is no possibility of false dead people, once the death was confirmed by the Death Certificate. An up-to-dated number of the voting registers could be very useful in the vital status identification, making quicker and more precise the Electoral Court search.

The identification of the vital status of some workers by the professionals of the industry's health center and by workers employed at the industry for a long time was possible because they had a relationship with several ex-workers. In some cases, the information about ex-workers was obtained with their relatives working at the industry in the data collection period.

The household search and the phone search were the first tentative to deparature the greater concentration of dead people among the lost workers. One problem in this strategy was that some addresses were very old. In the last two decades, several streets changed their names, and several places did not exist anymore, like hotels and lodging houses that had been closed or changed their function, and households that changed to businesses. The most important problem in the phone search was that many workers did not have a telephone and that the women tended to list their phone numbers under the husband's name (Schonorr, 1993). Despite these obstacles, the mentioned strategies overcome the expectations. They were applied to sixty percent of the lost workers, and they made possible to find almost half of the contacted workers, reducing the losses by four percentage points.

The vital status of the identified migrant workers were also made by the Electoral Court of the state of Rio Grande do Sul, because they have access to the registers of other states. However, the Electoral Court is able to do the search only if the state of residence is known. The current number of the voting registrations could have enlarged this type of search.

The last strategy to establish the vital status was the search of lost workers in the Death Registration Offices. This search identifies just dead workers and compensates the greater identification of living people in relation to the dead people introduced by the first search at the Electoral Court.

The identification of the cause of death was done by the search in the Death Registration Offices and by the collaboration of some families. These strategies were very effective once it identified 99 percent of the causes of death. The unidentified case was the one which the cause of death were not found in the search at the Death Registration Offices, the place of registration of death was unknown and communication with the family was not possible.

The accuracy of the cause of death certificate is always a worry in mortality studies. But the authors evaluated that the broader the category, the better the reliability of the registered cause of death. In addition, they agree that the major categories of cause of death, such as cancer, are usually correctly registered. Otherwise, the cause of death reported on death certificates was coded by a trained professional in the rules specified by the International Classification of Diseases (ICD) and World Health Organization (WHO) based on the 10th revision (Breslow & Day, 1987; Checkoway et al.,

1989; Halperin et al., 1996; Hernberg, 1992; Monson, 1990).

Historical cohorts on mortality data may still be useful to investigate undiscovered occupational hazards that are fatal. The most important study limitations are related to the quality of the information and the completeness of the follow-up (Breslow & Day, 1987; Hernberg, 1992; Schilling, 1986). In this study, the selected strategies made it possible to identify the vital status of 93 percent of the cohort, and almost 100 percent of the cause of death of the deceased workers, thus achieving the required completion of the follow-up.

Acknowledgements

This study was supported by the Financiadora de Estudos e Projetos (FINEP – Convênio 6694032100), the International Agency for Research on Cancer (IARC), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the Fundação de Amparo à Pesquisa do Rio Grande do Sul (FAPERGS).

We are grateful for the collaboration of the Riocell S.A. (pulp and paper industry), the professionals of the industry's health center (Marcelo Silveira and Mara Chagas), the Tribunal Regional Eleitoral do Rio

Grande do Sul (the Electoral Court), the computer coordinator of the Electoral Court Jorge Lheureux de Freitas and the research assistants (Marcelo Zardo, Cristiane Rios Petrarca, Monia Di Lara Dias, Igor Gastal Grill, Patricia Manzoli, Eneu Borges Teixeira Junior, Rodrigo Sanderson, Luiz Fernando Teixeira, Angélica Chagas, Andréia Severo Leite, Daiane Padula Paz, Isabel Cristina Padula Paz, Regina Maria Pohasky, Carlos Roberto Pacheco Moreira) which made possible the conclusion of this study.

References

- BOFFETTA, P.; KOGEVINAS, M.; SARACCI, R. & VAINIO, H., 1991. *IARC International Cohort Study on Cancer Risk Among Workers in the Pulp and Paper Industry. Protocol of the Study*. Lyon: International Agency for Research on Cancer (mimeo).
- BRESLOW, N. E. & DAY N. E., 1987. The role of cohort studies in cancer epidemiology. In: *Statistical Methods in Cancer Research: The Design and Analysis of Cohort Studies* (N. E. Breslow & N. E. Day, N. E., eds.), pp. 2-46. Lyon: International Agency for Research on Cancer.
- CHECOKWAY, H.; PEARCE, N. E. & CRAWFORD-BROWN, D. J., 1989. Cohort studies. In: *Research Methods in Occupational Epidemiology* (H. Checokway; N. E. Pearce & D. J. Crawford-Brown, eds.), pp. 103-169. New York: Oxford University Press.
- HALPERIN, W.; BEAUMONT, J.; BROWN, D.; CLAPP, D.; HARING, M.; McCAMMON, C.; MEINHARDT, T. J. & OKUN, A., 1986. The cohort study. In: *Epidemiology of Occupational Health* (M. Karvonen & M. I. Mikheev, eds.), pp. 149-180. Copenhagen: World Health Organization.
- HERNBERG, S., 1992. Sources of information. In: *Introduction to Occupational Epidemiology* (S. Hernberg, ed.), pp. 39-59. Chelsea: Michigan Lewis Publishers.
- KLEINBAUM, D. G.; KUPPER, L. L. & MORGENSTERN, H., 1982. Typology of observational study designs. In: *Epidemiologic Research: Principles and Quantitative Methods* (D. G. Kleinbaum & H. Morgenstern, eds.), pp. 62-95. New York: Van Nostrand Reinold Company Inc.
- MONSON, R. R., 1990. The collection of epidemiologic data: retrospective cohort studies. In: *Occupational Epidemiology* (R. R. Monson, ed.), pp. 51-56. Boston: CRC Press, Inc.
- SCHNORR, T., 1993. Video display terminals and adverse pregnancy outcomes. In: *Case Studies in Occupational Epidemiology* (K. Steenland, ed.), pp. 7-20. Oxford: Oxford University Press.
- SCHILLING, R. S. F., 1986. Sources of data. In: *Epidemiology of Occupational Health* (M. Karvonen & M. I. Mikheev, eds.), pp. 81-96. Copenhagen: World Health Organization.
- TOREN, K.; PERSSON, B. & WINGREN, G., 1996. Health effects of working in pulp and paper mills: malignant diseases. *American Journal of Industrial Medicine*, 29:123-130.
- WHO (World Health Organization), 1995. *Biennial Report 1994/1995*. Lyon: International Agency for Research on Cancer, World Health Organization.