

Estimation of the coverage of Dutch municipalities by cancer registries and PALGA based on hospital discharge data

R. Alexandra Goldbohm, Piet A. van den Brandt, Elisabeth Dorant *

In a large-scale prospective cohort study on diet and cancer, which was initiated in 1986, follow-up for cancer is being conducted through record linkage with the regional cancer registries (CR) and with PALGA, a data base of pathology reports. During the first few years of the study, however, neither the CR nor PALGA operated on a nation-wide basis. Since the cohort was to be recruited from samples drawn from a large number of municipal population registries, selection of municipalities according to the degree to which their inhabitants were 'covered' by CR and PALGA would minimize loss to follow-up. Hospital discharge data, which include diagnosis, age, sex and residence of each discharged (or

deceased) patient, were used to estimate for each municipality the proportion of hospitalized patients admitted to those hospitals which were expected to participate in either the CR or PALGA at the date the cohort study was planned to start. A minimum coverage of 75% was used as criterion for selection of a municipality. Of the 204 municipalities selected, 188 attained more than 90% coverage; the anticipated mean coverage of the sampled cohort at the start of the study was 94.3%. The analysis was repeated several years after the start of the cohort study to assess retrospectively the true coverage of the cohort over time. Mean coverage increased from 98.5% at the start of the study to 100% in 1988.

Key words: cancer registries, medical registration, cohort study

In 1986, the Netherlands Cohort Study (NLCS) was initiated, which investigates the association between dietary habits and the risk for (stomach, colorectal, lung and breast) cancer among more than 120,000 men and women aged 55-69. After the baseline administration of a mailed questionnaire in September 1986, follow-up for cancer was to be accomplished by all nine regional, population-based cancer registries (IKN, IKO, IKMN, IKA, IKST, IKW, IKR, IKZ and IKL) and by PALGA (Pathologisch Anatomisch Landelijk Geautomatiseerd Archief), a Dutch data base of pathology reports.¹

Each cancer registry in the Netherlands is supposed to cover a region; all such regions together cover the whole country. The majority of cancer registries started to register in a limited number of hospitals in the period between 1982 and 1986 and gradually extended their registration activities to all hospitals in their region to attain complete coverage by 1989.² One registry (IKZ-SOOZ) dates from 1953. In January 1985, the PALGA data base included 28 of 70 pathology laboratories in the Netherlands, accounting for approximately 50% of all pathology reports.³ In June 1990, all laboratories had joined PALGA.

Considering the incomplete coverage of the Netherlands in 1986 as described above, it was essential for sufficient follow-up to recruit the cohort from geographic

areas in which either one of the cancer registries or PALGA was operational from the start of the study. Since it was decided for practical reasons to sample the cohort from the municipal population registries, the question arose how to determine the degree of coverage of each Dutch municipality by the CR and PALGA. The availability of a nation-wide data base of hospital discharge data enabled us to answer that question. This paper describes the methods to estimate the coverage of the municipalities, the subsequent selection of municipalities for cohort recruitment, and the actual coverage of the cohort sample as assessed retrospectively from the hospital discharge data.

METHODS

In 1985, during the planning stage of the cohort study, a list of hospitals was composed that were either participating in one of the registries already or very likely would do so in 1986. The information concerning participation of hospitals was obtained from each of the cancer registries. Since most of the pathology laboratories are connected to and working for one or more hospitals, a list of hospitals linked with PALGA was also drawn up.

The Dutch Centre for Health Care Information (SIG/Zorginformatie) maintains the National Medical Registry (LMR), a data base of hospital discharge data.³ The data base contains (anonymous) data on each patient discharged from or deceased in the hospital. The following data were relevant to our study: hospital code, diagnosis (ICD-9 code), sex, age and municipality of residence, and, in case of malignancy, whether it concerned a first or a repeated admission for that specific diagnosis. For privacy reasons, we could not obtain the case-specific data. There-

* R.A. Goldbohm, TNO Toxicology and Nutrition Institute, Zeist, The Netherlands & Department of Epidemiology, University of Limburg, Maastricht, The Netherlands
P.A. van den Brandt, E. Dorant, Department of Epidemiology, University of Limburg, Maastricht, The Netherlands

fore, we provided the SIG with the two lists of registering hospitals and received in return the requested data aggregated by the municipality of residence. In agreement with the purpose and the study population of the cohort study, the data were restricted to the age group 55 to 69 and to diagnoses of digestive tract, lung and breast cancer (ICD-9 codes 151, 153, 154, 162, and 174-175 respectively). For each municipality, diagnosis and sex, SIG provided the following data: total number of discharges in a) CR hospitals, b) PALGA hospitals, c) CR and/or PALGA hospitals and d) all hospitals in the country. Similar data for first admission for the disease were also available.

From these data we calculated for each municipality the diagnosis- and sex-specific proportion of patients discharged from 'registering' hospitals, i.e. a, b, or c divided by d. These proportions can be considered as the coverage of a municipality by the respective registries. For example, 25 female patients with breast cancer inhabiting municipality X were admitted to any Dutch hospital. Five of these patients were admitted to hospitals that were not on the CR or PALGA lists. Thus, coverage of municipality X attained $(25-5)/25=80\%$. The proportion pertaining to coverage by either a CR or PALGA (c/d) was used as a selection criterion for cohort recruitment. Besides degree of coverage, the following criteria for eligibility of a municipality for cohort recruitment were applied:

- availability of a computerized population registry, administered by one of nine regional computer centres and
 - permission to draw a 40% sample from the men and women aged 55 to 69 in their population registry.
- A higher sampling fraction was thought to jeopardize the participation of municipalities. Of the 323 (from a total of 714) municipalities that met criterion 1, 23 (7%) refused participation, leaving 300 municipalities eligible for selection according to coverage.

Based on the number of cancer cases to be expected in the cohort study after five years of follow-up, the required sample size was set at 350,000. The aim of the analysis was to select municipalities in descending order of coverage until the required sample size was achieved, taking into account that the average coverage of the sample should not be lower than 90%. The calculations were performed for all data as well as for those pertaining to first admission only.

The selection of municipalities for the cohort study was based on the LMR data base of 1983, the most recent one available in the planning stage of the cohort study. After the start of the study, the analysis was repeated using the combined LMR data base of 1987/1988 and hospital lists updated retrospectively according to their actual registration status on 1 January 1987 and 1 January 1988. The situation at the start of the study (1 October 1986) differed from that in 1987 for three hospitals that were neither on the CR nor on the PALGA list. The contribution of two of these hospitals to eligible municipalities was negligible, whereas the third hospital contributed 20% of the hospital beds to a region, including a large city. Coverage of the municipalities in the latter region was corrected by multiplying their coverage degree in 1987 by 0.8. It was also estimated how migration to other municipalities would affect the coverage of the cohort. Based on an annual

Cut-off point (%)	Municipalities* (N)	Sample size**
No cut-off	300	491000
70	251	376000
80	244	339000
90	237	328000
100	219	256000

* Three hundred of the 714 municipalities were eligible for the analysis

** Assuming a sampling fraction of 40% (age group 55-69), except for one large municipality, which only permitted a 20% sample

Table 1 Number of municipalities and sample size according to cut-off point of anticipated coverage by the cancer registries and PALGA

migration rate of 1.55% in the age group 55-69⁴ and on the coverage degree calculated for the entire Dutch population (assuming a random migration pattern), the coverage of the cohort was adjusted.

An opportunity to check the validity of the analysis presented was provided by follow-up of the subcohort, a random sample of 5000 subjects from the entire cohort. Subcohort members reported biennially whether they had been diagnosed with cancer after the start of the study in 1986. The proportion of self-reported cases that were also retrieved from the cancer registries and/or PALGA can be considered as an independent measure of coverage of the cohort.

RESULTS

Table 1 displays the number of municipalities and the achievable sample size according to different cut-off points of anticipated coverage. Complete coverage appeared to be attained for 219 municipalities, corresponding to a sample size of 256,000. A sample size of 350,000 could be achieved at a cut-off point between 70 and 80%. The eventually chosen cut-off point of 75% corresponded to a sample size of 342,000 and comprised 249 municipalities. As a result of consolidation of small municipalities into larger ones during the period between planning and sampling, the actual sample was drawn from 204 municipalities. The mean anticipated coverage of the sample amounted to 94.3 %. For the entire Dutch population a coverage of 78% was expected. When it was taken into account that part of the participants in the cohort study would move to municipalities that were not yet sufficiently covered, the anticipated coverage for the first year of follow-up decreased to 94.0%.

The coverage of the municipalities participating in the cohort study as determined from the actual registration situation on 1 January 1987 (three months after the start of the cohort study) is presented in figure 1 and 2. Figure 1 shows the coverage by the CR and PALGA combined, while figure 2 displays the coverage by the CR alone. The mean actual coverage of the cohort appeared to be 99.5%, much higher than the value of 94.3% anticipated before the start of the study. The actual coverage of three small municipalities, however, did not attain the initial cut-off point of 75%. Neither the CR nor PALGA attained sufficient mean coverage on its own (88.6% and 82.5%, respectively).

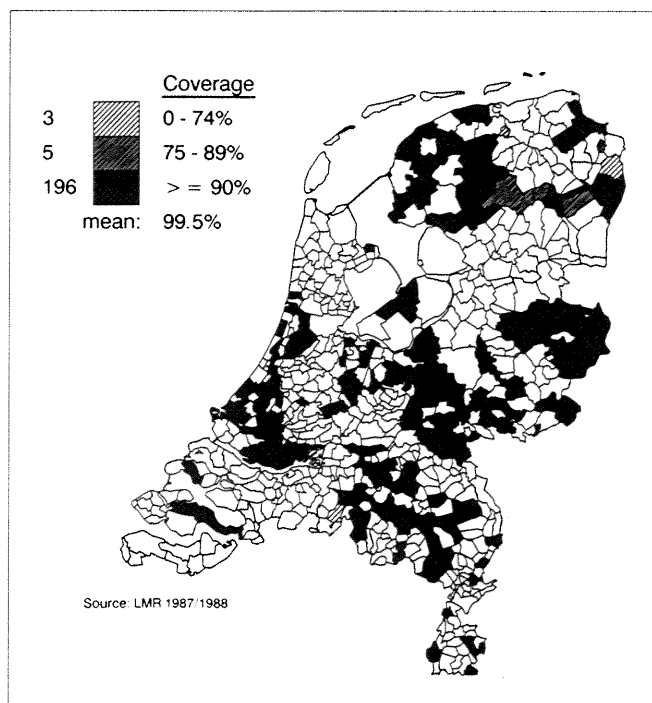


Figure 1 Actual coverage of the cohort municipalities by cancer registries and PALGA, 1 January 1987 (blank municipalities are not participating in the study)

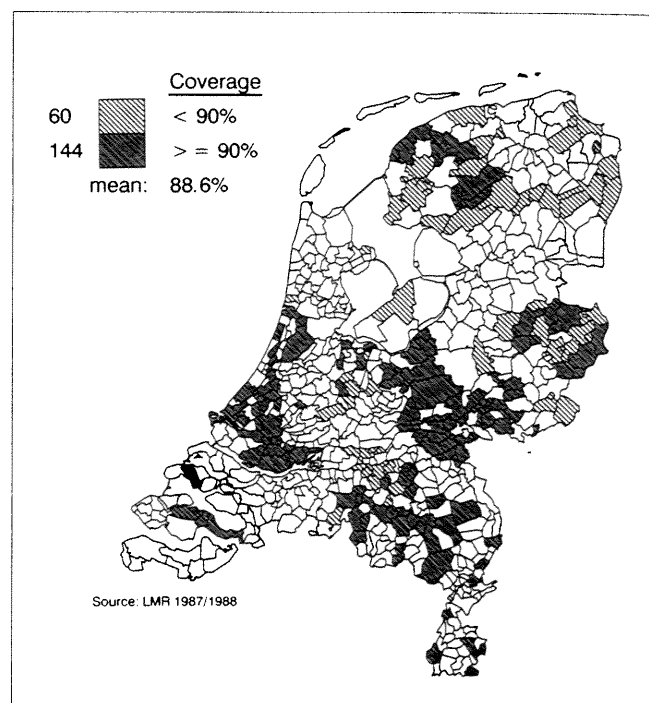


Figure 2 Actual coverage of the cohort municipalities by the cancer registries, 1 January 1987 (blank municipalities are not participating in the study)

Table 2 gives the site- and sex-specific coverage on 1 January 1987. No large differences were detected between any of the cancer sites nor between men and women. Furthermore, mean coverage did not differ between all admission data and those pertaining to first admissions (table 2). Table 3 shows, however, minor differences within municipalities between coverage degrees calculated from all admissions and those calculated from first admissions only. Figure 3 summarizes the anticipated coverage of the cohort (94.3%) and the development of the actual coverage for three reference dates: the start of the study (98.5%), 1 January 1987 (99.5%) and 1 January 1988, when 100% coverage was reached.

Of the subjects in the subcohort who reported to have a cancer (145 by the end of 1989), 115 had also been matched independently in the record linkage with the CR and PALGA. Almost all of the missing cases (29) had reported skin cancer. This can be explained by the fact that basal cell carcinoma of the skin is not recorded routinely

Table 2 Actual coverage (%) by cancer registries and PALGA on 1 January 1987 of the cohort sampling population specified for cancer site and sex, calculated for all admissions and first admissions only

Site	ICD code	Men		Women	
		All admissions	First admissions	All admissions	First admissions
Stomach	151	99.7	99.5	99.3	99.3
Colon	153	99.5	99.5	99.7	99.7
Rectum	154	99.2	99.2	100.0	100.0
Lung	162	99.1	98.8	99.3	99.2
Breast	174+175			99.4	99.5
All sites (both sexes)		99.5	99.4		

by the CR. One self-reported case with another cancer type, however, was not matched to a record in the PALGA data base because of disagreement as to residence. After exclusion of the missed subjects reporting skin cancer, the proportion of cancer cases retrieved thus amounts to $115/116 = 99\%$ with a lower 95% confidence bound of 96%.

DISCUSSION

The value of the presented analysis for the evaluation of the follow-up for cancer in the cohort study depends on the reliability of the data and the underlying assumptions.

From 1986 onwards, all general and university hospitals in the Netherlands supply the required registration data to the LMR data base. As for the 1983 data base, a few hospitals were lacking. Their share in the hospital discharges was 2.5%.⁵ An important hospital that neither contributed to the 1983 nor to the 1987/1988 data base was the Dr Daniel den Hoed Cancer Centre, the oncology hospital in Rotterdam. The missing hospitals, however, supplied the most crucial data, i.e. the number of admitted patients and their residence. The distribution of diagnoses in the missing general hospitals was estimated from that in other general hospitals. The diagnoses in the Dr Daniel den Hoed Cancer Centre were assumed to have the same

Table 3 Classification of cohort municipalities (n=204) according to coverage on 1 January 1987: all admissions versus first admissions only

First admissions	All admissions		
	0-74%	75-89%	≥ 90%
0-74%	3	0	0
75-89%	0	3	4
≥ 90%	0	2	192

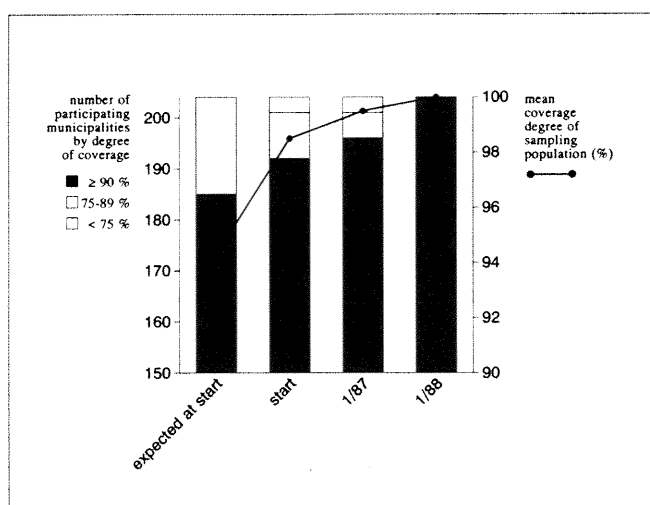


Figure 3 Anticipated and actual coverage of the cohort sampling population in 204 selected municipalities by the cancer registries and PALGA

distribution as those in the Antoni van Leeuwenhoek Hospital, the oncology hospital in Amsterdam. Thus, the overall error due to missing hospitals in the LMR data base could not have been very large.

Comparison of the LMR data with data on the number of discharged patients from another source has shown that the LMR data base is also virtually complete within each hospital (SIG, personal communication). However, even if incompleteness would be substantial, it would not invalidate the present analysis, which is based on proportional, not absolute, coverage. Some information is available on the accuracy of the recorded data: the residence appeared to be satisfactorily recorded, but the diagnosis was not quite correct in 7% of the cases.⁶ Although it should have been recorded whether a patient was admitted for that specific (malignant) disease for the first time, this item is presumed to be inaccurate, in particular for university hospitals (SIG, personal communication). The analysis based on all admissions and that based on first admissions only produced similar results, however.

A substantial change over time in referral patterns, for example as a result of closing and merging of hospitals, may threaten the representativeness of the data base for the population in the cohort study. This argument is valid for the 1983 data base, which was used to estimate coverage expected in 1986, but not for the 1987/1988 data base, which referred to exactly the same period as the follow-up of the cohort and which must actually include the cohort members diagnosed with cancer during these years of follow-up.

The results of the presented analysis are useful only if all eligible cases who were admitted to a hospital affiliated with a cancer registry or PALGA were actually included in these registries. A study conducted in the IKMN cancer registry, in which the 1986 cancer registry data were compared to the LMR data, has shown that 11% of the cases were initially missed by the cancer registry.⁷ In 52% of these missing cases, however, the diagnosis was not histologically confirmed. The percentage of cases missed was much lower for cancer of the breast (2.5) and digestive tract (6.8) than for lung cancer (12.9). These percentages reflect the proportion of cases not confirmed by histological examination. From 1986 onwards, LMR was also introduced in most cancer registries as a check for com-

pleteness and as additional source of cancer cases. A more recent study on completeness, performed by the IKL cancer registry (1988-1990), revealed that on average 3.8% of the cases that should have been registered were not.⁸ Again, the percentages were higher for lung cancer (4.4%) than for breast and digestive tract cancer (1.8% and 1.4% respectively).

We conclude from the data presented here that in the cohort study loss to follow-up is likely to be very small. This conclusion is confirmed by the high proportion (99%) retrieved of subcohort members who reported cancer. A second conclusion concerns the use of LMR data for this type of problem. These data provided a quick, efficient and apparently reliable way to solve an important problem in the planning stage of the cohort study, i.e. how to minimize loss to follow-up for cancer. Although it will not be necessary to repeat this type of analysis for other (prospective) epidemiological studies on cancer, since the CR and PALGA have attained national coverage, it may be used to check coverage by other local disease registries.

ONDERSTEUNING

De auteurs danken SIG/Zorginformatie, alle Integrale Kankercentra en PALGA voor het verstrekken van de benodigde gegevens en J. Holl voor het idee LMR-gegevens te gebruiken.

SAMENVATTING

Schatting van de dekking van Nederlandse gemeenten door de kankerregistratie en PALGA met behulp van ziekenhuisontslaggegevens

Voor het in 1986 gestarte landelijke cohortonderzoek naar voeding en kanker, wordt de follow-up van kanker uitgevoerd door de regionale kankerregistraties (KR) en door het Pathologisch Anatomisch Landelijk Geautomatiseerd Archief (PALGA). Aan het begin van de follow-up periode echter hadden noch de KR noch PALGA een landelijke dekking. Aangezien het cohort gewonnen zou worden via steekproeven uit de bevolkingsregisters van een groot aantal Nederlandse gemeenten, zou selectie van die gemeenten op grond van de mate van dekking door de KR en PALGA de 'loss to follow-up' tot een minimum kunnen beperken. Gegevens van de Landelijke Medische Registratie (LMR) - een bestand van uit het ziekenhuis ontslagen patiënten dat onder meer diagnose, leeftijd, geslacht en woongemeente van de patiënt bevat - werden gebruikt om voor iedere gemeente te berekenen welk deel van de patiënten was opgenomen in ziekenhuizen die naar verwachting zouden deelnemen aan de KR of PALGA op het moment van de start van het cohortonderzoek. Een minimale dekking van 75% werd gebruikt als selectiecriteria voor een gemeente. Van de 204 aldus geselecteerde gemeenten bereikten 188 een dekking van 90% of hoger; de verwachte gemiddelde dekkingsgraad van de steekproef werd geschat op 94,3%. De analyse werd enkele jaren na het begin van het cohortonderzoek herhaald om te bepalen hoe hoog de dekking van het cohort in werkelijkheid was geweest. De werkelijke gemiddelde dekking bleek te zijn gestegen van 98,5% ten tijde van de start van het onderzoek tot 100% in 1988.

Trefwoorden: kankerregistratie, medische registratie, cohortonderzoek

REFERENCES

- 1 Brandt PA van den, Goldbohm RA, Veer P van 't, Volovics A, Hermus RJJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in the Netherlands. *J Clin Epidemiol* 1990;43:285-95.
- 2 Landelijk Overlegorgaan Kankercentra. Progress report Dutch Cancer Registry 1988. Utrecht, 1991.
- 3 SIG/Informatiecentrum voor de Gezondheidszorg. Onderzoek en medische registraties: een bruikbaarheidsbeoordeling in het kader van sociaal-economische status en gezondheidsverschillen. Utrecht, 1991.
- 4 Centraal Bureau voor de Statistiek. Binnenlandse migratie, Jaarwerk 1982. Voorburg, 1983.
- 5 SIG/Informatiecentrum voor de Gezondheidszorg. LMR Jaarboek Ziekenhuizen. Utrecht, 1983.
- 6 Hoogendoorn D. Stichting Informatiecentrum voor de Gezondheidszorg (SIG). In: Maas PJ van der, Hofman A, Dekker E, eds. *Epidemiologie en gezondheidsbeleid*. Alphen aan den Rijn: Samsom Stafleu, 1989: 31-41.
- 7 Berkel J. Volledigheid van de kankerregistratie [Completeness of the cancer registry]. *Ned Tijdschr Geneesk* 1989;133:2027-30.
- 8 Schouten LJ, Höppener P, Brandt PA van den, Knottnerus JA, Jager JJ. Completeness of cancer registration in Limburg, the Netherlands. *Int J Epidemiol* 1993;22:369-76.

CORRESPONDENCE

R.A. Bausch-Goldbohm, TNO Toxicology and Nutrition Institute, P.O. Box 360, 3700 AJ Zeist, The Netherlands

Accepted for application 13 December 1993