EUROPEAN JOURNAL OF CANCER 46 (2010) 395–404

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

New opportunities for drug outcomes research in cancer patients: The linkage of the Eindhoven Cancer Registry and the PHARMO Record Linkage System

Myrthe P.P. van Herk-Sukel [a,*], *Lonneke V. van de Poll-Franse* [b,c], *Valery E.P.P. Lemmens* [b,d], *Gerard Vreugdenhil* [e], *Johannes F.M. Pruijt* [f], *Jan Willem W. Coebergh* [b,d], *Ron M.C. Herings* [a,g]

[a] *PHARMO Institute for Drug Outcomes Research, Utrecht, The Netherlands*
[b] *Eindhoven Cancer Registry, Comprehensive Cancer Center South, Eindhoven, The Netherlands*
[c] *Center of Research on Psychology in Somatic Diseases (CoRPS), Tilburg University, Tilburg, The Netherlands*
[d] *Department of Public Health, Erasmus University Medical Centre, Rotterdam, The Netherlands*
[e] *Department of Internal Medicine, Máxima Medical Centre, Veldhoven, The Netherlands*
[f] *Department of Internal Medicine, Jeroen Bosch Hospital, 's-Hertogenbosch, The Netherlands*
[g] *Department of Health Policy and Management, Erasmus University Medical Centre, Rotterdam, The Netherlands*

ARTICLE INFO

Article history:
Received 28 July 2009
Accepted 9 September 2009
Available online 5 October 2009

Keywords:
Cancer registry
Chemotherapy
Clinical laboratory measurements
Drug utilization
Hospitalisation database
Medical record linkage
Patient care
Population based
Pharmacoepidemiology
Validation

ABSTRACT

Background: Insight into co-morbidity and treatment effects is pivotal to improve quality of care for cancer patients.
Objectives: To determine whether linkage of the Eindhoven Cancer Registry (ECR) and the PHARMO Record Linkage System (RLS) was technically feasible and to assess which patient-centric data would result from this linkage.
Methods: The ECR records data on tumour stage and primary treatment of all newly diagnosed cancer patients in the southeastern Netherlands including co-morbidity at diagnosis, whereas the PHARMO RLS includes data from multiple linked observational databases such as data on drug utilisation (for both in- and out-patients, including chemotherapy), hospitalisations and clinical laboratory measurements. All patients who lived or had been living in the overlapping area served by the ECR and the PHARMO RLS during 1998–2006 were selected for linkage which was performed with probabilistic medical record linkage.
Results: The linkage resulted in an ECR-PHARMO cohort of 40,004 cancer patients with a total of 42,767 primary tumours. The cancer patients in the linked ECR-PHARMO cohort were representatives for the cancer patients included in the total ECR during 1998–2006. Cancer patients included in the cohorts had a mean history of 5 years and a mean follow-up ranging from 2 to more than 4 years (dependent on the survival rate of the specific cancer type).
Conclusions: Linkage of ECR and the PHARMO RLS creates the possibility to study patient-centric drug utilisation, health resources utilisation and their costs, in addition to the effectiveness and safety of pharmaceuticals in routine daily practice in cancer patients.

* *Corresponding author:* Address: PHARMO Institute for Drug Outcomes Research, P.O. Box 85222, 3508 AE Utrecht, The Netherlands. Tel.: +31 (0)30 2345 620; fax: +31 (0)30 2345 568.
E-mail address: Myrthe.van.Herk@pharmo.nl (M.P.P. van Herk-Sukel).
0959-8049/$ - see front matter
doi:10.1016/j.ejca.2009.09.010

1. Introduction

In Europe there are over two million incident cases of cancer every year and it is expected that this number will continue to rise.[1] With the ageing of the population and the increasing survival of cancer patients with co-morbid conditions, cancer is and will become an increasingly important factor in the global burden of disease in the decades to come.

A growing number of cancer patients are treated with novel treatment combinations and targeted therapies are being introduced at relatively early stages of disease progression. While effectiveness and safety of these new therapies are thoroughly studied in randomized clinical trials, little data are available on these parameters in a daily practice setting. Effectiveness of anti-cancer drugs may be different than expected as daily clinical practice differs from the experimental setting with respect to the heterogeneity of patients, their treatments and co-medication. Moreover, important safety issues may not be detectable in clinical trials since the frequency of many adverse events is low and adverse effects of cancer therapies may occur many years after drug administration.[2–4] Balancing effectiveness and adverse effects is therefore a major challenge in the treatment of cancer patients and monitoring post approval drug use is especially important for cancer drugs that receive accelerated FDA and EMEA approval.

Besides monitoring the effectiveness and safety of cancer treatments, pharmaceuticals used for a wide variety of other conditions may be associated with increasing or decreasing risk of cancer, such as non-steroidal anti-inflammatory drugs or lipid-lowering drugs.[5,6] Data to confirm, refute or elaborate on these hypotheses are sparse.

To shed light on the burden of cancer, disease management, safety aspects, co-morbidity patterns and outcomes assessments, follow-up of patients before and after cancer diagnosis in routine daily practice is pivotal. For this, large administrative databases or cancer registries are often used.[7] However, both type of data collections mostly lack detail either with respect to the incidence and staging of cancer or with respect to (pharmaco)treatments, morbidity and co-morbidity during follow-up.[8]

We therefore explored whether the linkage of a regional cancer disease register (Eindhoven Cancer Registry: ECR) and a patient-centric data network including multiple linked observational databases (PHARMO Record Linkage System: PHARMO RLS[9]) was feasible, valid and detailed enough to fulfil the abovementioned information needed. In this manuscript we describe the results of this linkage as well as the patient-centric data on drug treatment (out-patient and in-patient), hospitalisations and clinical laboratory measurements that become available during follow-up before and after cancer diagnosis for the cancer patients present in these linked databases.

2. Materials and methods

2.1. Data sources

The Eindhoven Cancer Registry (ECR) is a population-based registry (covering a demographic region with 2.4 million inhabitants) which is maintained by the Comprehensive Cancer Centre South and collates records on all newly diagnosed cancer patients in the southeastern part of the Netherlands.[10,11] The ECR is notified for new cases of cancer by 6 pathology departments, 10 general hospitals and two radiotherapy institutes. Trained registry personnel subsequently actively collect on site data on patient characteristics, diagnosis, tumour staging, co-morbidity at diagnosis and treatment received directly after diagnosis (e.g. chemotherapy (yes/no), radiation therapy and surgery).

The PHARMO RLS is a large, patient-centric data network including multiple linked observational databases designed for safety and outcomes research of drugs which collates patient records in 48 geographically defined areas in the Netherlands (covering a demographic region of three million inhabitants). The central patient database is linked to more than 10 databases using different medical record linkage algorithms.[12–15] Databases relevant for observational cancer research include virtual complete longitudinal data obtained from community pharmacies (out-patient), hospital discharge records (Dutch National Medical Registration: LMR), a mortality registration and a growing number of clinical laboratories, in-hospital pharmacies (in-patient) and general practitioners (these last three databases are available for a sub cohort of the patients included in the PHARMO RLS).

Both the ECR and the PHARMO RLS are recognised as high quality sources for epidemiological research that collect information in overlapping regions in the Netherlands for a period of at least 10 years.

2.2. Medical record linkage

The first step in combining the ECR and the PHARMO RLS was to identify in both databases all the patients who lived or had been living in overlapping zip code areas and were diagnosed with cancer in the period 1998–2006. This overlapping catchment area covered all inhabitants of 510 zip codes (approximately 1 million inhabitants).

Because these databases only contain de-identified information (i.e. all personal identifiers have been removed from the file), no unique personal identifiers were available and hence, the linkage of the ECR and the PHARMO RLS had to be based on probabilistic record linkage technology.[16–18] This technology is also used for the linkage of several databases in the PHARMO RLS and involves three major steps: (1) blocking, (2) matching and (3) linking. (1) In the blocking step[17], for each patient in the ECR, the data on patients from the PHARMO RLS with similar gender and date of birth were grouped into record pairs. An example of the approach would be as follows: If there were 2 patients with similar gender and date of birth in the ECR and there were 3 patients in the PHARMO RLS, then the blocking step would result in 6 (2×3) record pairs.

(2) In the matching step,[17] for each record pair the probability (match odds) that both records of the ECR and the PHARMO RLS belong to the same patient was calculated using Bayesian algorithms.[19] Multiple matching variables available in both databases were used: first initial, first letter last name, soundex code of last name,[20] first four characters of the patients' most recent zip code as well as singular variables.

The latter represent logical relation between two records, i.e. an extra weight was assigned if a cancer patient from the ECR had been hospitalised for cancer as notified in the hospital discharge records from the PHARMO RLS.

(3) Finally, linkage was performed based on the summarised match odds of all variables, the weight value, by applying a threshold weight value.[16,17] The record pair with the highest cumulative weight value above threshold was defined as positive link (same patient) and all other pairs were defined as negative links (different patients), as only one record could logically belong to the same patient.

2.3. Validation

The quality of the medical record linkage process was evaluated in a random sample in which additional person information was obtained from the original patient-centric data sources of the PHARMO RLS (in this case, community pharmacies participating in the PHARMO system) and the ECR, including the full name and address of the patients: the 'gold standard'. Record pairs where manually compared after being blinded with respect to outcome of the above described linkage process. This validation was approved by the privacy committee of the PHARMO Institute for Drug Outcomes Research and met the criteria of the health research code of conduct.[21]

The validation resulted in true and false record pairs and detailed information on the methodology and the validation of the used record linkage method can be found elsewhere.[12,13,16–19,22,23] A flowchart of the linkage of the ECR to the PHARMO RLS is presented in Fig. 1.

2.4. Representativeness

The linkage process resulted in a linked ECR-PHARMO cohort. Representativeness to the total ECR population was evaluated by comparing the distribution of gender, age at tumour diagnosis, year of diagnosis and the most common cancer sites. With respect to the latter, unlike other current cancer registries, basal cell carcinoma is also registered in the ECR.[24]

Characteristics per tumour cohort were shown for the total ECR and the linked ECR-PHARMO cohort for the four most common types of cancer (not accounting for basal cell carcinoma): breast cancer, colorectal cancer, lung cancer and prostate cancer.

2.5. Illustration of information for different cancer cohorts

An illustration of the information gained from linkage of the ECR and PHARMO RLS was presented for the patients included in the breast cancer, colorectal cancer, lung cancer and prostate cancer cohorts. For this, information on history (before cancer diagnosis) and follow-up (after cancer diagnosis) was determined for the patients included in the total linked ECR-PHARMO cohort and for the patients included in a sub cohort with information from the clinical laboratory database available (approximately 80% of the patients in the

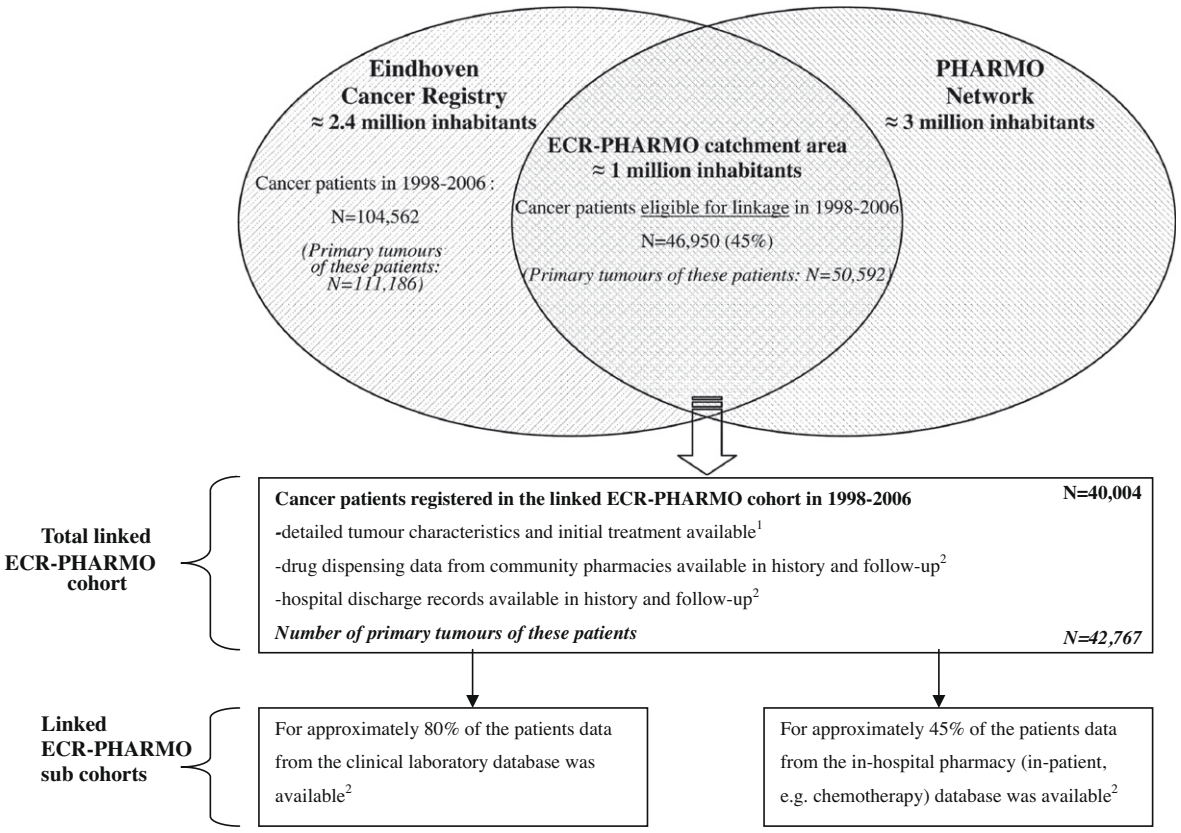

Fig. 1 – Flowchart of the linkage process and cohort formation. ECR: Eindhoven cancer registry; [1]Data obtained from ECR; [2]Data obtained from PHARMO RLS.

total linked cohort). History was defined as the time between the date of entering the PHARMO RLS to the date of cancer diagnosis as registered in the ECR. Follow-up was defined as the time between the date of cancer diagnosis until the end of data collection in the PHARMO Network (i.e. the patient moves out of the PHARMO Network catchment area), death, or end of the study period (31st December 2007), whichever occurred first.

Co-morbidities at diagnosis, medication use, hospitalisation and laboratory measurements during the 1 year before cancer diagnosis were determined for patients with at least 1 year history and with data from the clinical laboratory database available. Treatments received, medication use, hospitalisations and laboratory measurements during the 1 year after cancer diagnosis were determined for patients with at least 1 year follow-up and with data from the clinical laboratory database available. Co-morbidity at cancer diagnosis and initial therapy were extracted from the ECR and information on co-medication, hospitalisations and clinical laboratory measurement were extracted from the PHARMO RLS. For patients who received chemotherapy, data on the type of cytostatics (defined as ATC code L01) were extracted from both in-hospital pharmacies (in-patient, e.g. cytostatics administered intravenously) and community pharmacies (out-patient, e.g. cytostatics administered orally) for a sub cohort (approximately 45% of the patients) of the PHARMO RLS.

Data were subsequently analysed using SAS programs that are organised within SAS Enterprise Guide version 4.0 (SAS Institute Inc., Cary, NC, USA). Data management was conducted under UNIX using SAS version 9.1.

3. Results

Of the 104,562 cancer patients registered in the ECR in the period 1998–2006, 57,612 (55%) patients were not eligible for linkage of which 2339 (2%) were diagnosed in hospitals elsewhere in the Netherlands and 55,273 (53%) were not living in the ECR-PHARMO catchment area (Fig. 1). Of the remaining 46,950 patients who were eligible for linkage, 40,004 (85%) cancer patients were finally linked and included in the ECR-PHARMO cohort with a total of 42,767 primary tumours. Detailed tumour data and initial treatment data were extracted from the ECR. Drug dispensing data (from community pharmacies) and hospitalisation data were extracted from the PHARMO RLS for all patients included in this linked ECR-PHARMO cohort. Data from the clinical laboratory database were available for approximately 80% of the cancer patients in the linked ECR-PHARMO cohort and data on specific type

Table 1 – Representativeness of the linked ECR-PHARMO cohort compared to the total ECR in the period 1998–2006 for patients diagnosed with a primary tumour.

	Total ECR (*N* = 111,186) *n* (%)	Linked ECR-PHARMO cohort (*N* = 42,767) *n* (%)	Difference (% total ECR minus% linked cohort) (%)
Gender			
Male	57,443 (52)	22,239 (52)	–0.3
Female	53,743 (48)	20,528 (48)	0.3
Age at tumour diagnosis			
⩽35	3659 (3)	1197 (3)	0.5
35–49	11,909 (11)	4277 (10)	0.7
50–59	20,322 (18)	7613 (18)	0.5
60–69	29,629 (27)	11,761 (28)	–0.9
70–79	30,895 (28)	12,231 (29)	–0.8
89–90	13,349 (12)	5175 (12)	–0.1
⩾90	1423 (1)	513 (1)	0.1
Year of diagnosis			
1998–2000	32,784 (30)	11,162 (26)	3.4
2001–2003	37,292 (34)	14,626 (34)	–0.7
2004–2006	41,110 (37)	16,978 (40)	–2.7
Cancer sites			
Skin, basal cell carcinoma	22,693 (20)	9403 (22)	–1.6
Breast	14,226 (13)	5545 (13)	–0.2
Colon and rectum	12,158 (11)	4804 (11)	–0.3
Lung, bronchus and trachea	12,105 (11)	4395 (10)	0.6
Prostate	9740 (9)	3835 (9)	–0.2
Haematolymphopoetic	6193 (6)	2285 (5)	0.2
Skin, other	4185 (4)	1594 (4)	0.0
Skin, melanoma	3469 (3)	1210 (3)	0.3
Primary site unknown	3129 (3)	1147 (3)	0.1
Urinary Bladder	2849 (3)	1030 (2)	0.2
Stomach	2545 (2)	935 (2)	0.1
Not further specified	17,894 (16)	6584 (15)	0.7

ECR: Eindhoven cancer registry.

Table 2 – Representativeness of breast, colorectal, lung and prostate cancer cohorts[a] extracted from the total ECR compared to the cohorts extracted from the linked ECR-PHARMO cohort in the period 1998–2006.

Characteristics	Breast		Colon and rectum		Lung, bronchus and trachea		Prostate	
	Total ECR (*N* = 14226) *n* (%)	Linked ECR-PHARMO (*N* = 5545) *n* (%)	Total ECR (*N* = 12105) *n* (%)	Linked ECR-PHARMO (*N* = 4804) *n* (%)	Total ECR (*N* = 12,158) *n* (%)	Linked ECR-PHARMO (*N* = 4,395) *n* (%)	Total ECR (*N* = 9740) *n* (%)	Linked ECR-PHARMO (*N* = 3835) *n* (%)
Gender								
Male	99 (0.7)	34 (0.6)	6419 (53)	2584 (54)	8720 (72)	3176 (72)	9740 (100)	3835 (100)
Female	14,127 (99)	5511 (99)	5686 (47)	2220 (46)	3438 (28)	1219 (28)	NA	NA
Age at tumour diagnosis								
⩽35	418 (3)	139 (3)	75 (0.6)	25 (0.5)	40 (0.3)	13 (0.3)	1 (<0.1)	0 (–)
35–49	3249 (23)	1178 (21)	650 (5)	252 (5)	816 (7)	252 (6)	75 (0.8)	26 (0.7)
50–59	3523 (25)	1375 (25)	1881 (16)	715 (15)	2253 (19)	781 (18)	1143 (12)	420 (11)
60–69	3105 (22)	1272 (23)	3409 (28)	1361 (28)	3969 (33)	1479 (34)	3576 (37)	1413 (37)
70–79	2507 (18)	1031 (19)	4036 (33)	1653 (34)	4056 (33)	1489 (34)	3778 (39)	1497 (39)
89–90	1265 (9)	488 (9)	1884 (16)	731 (15)	998 (8)	371 (8)	1104 (11)	454 (12)
⩾90	159 (1)	62 (1)	170 (1)	67 (1)	26 (0.2)	10 (0.2)	63 (0.6)	25 (0.7)
Tumour stage								
I	5710 (40)	2310 (42)	2279 (19)	931 (19)	2062 (17)	768 (18)	240 (3)	108 (3)
II	6137 (43)	2366 (43)	3776 (31)	1476 (31)	679 (6)	268 (6)	6185 (64)	2432 (63)
III	1426 (10)	556 (10)	2945 (24)	1244 (26)	3775 (31)	1389 (32)	1397 (14)	509 (13)
IV	621 (4)	225 (4)	2356 (20)	912 (19)	4039 (33)	1410 (32)	1589 (16)	663 (17)
Other/none/unknown	332 (2)	88 (2)	749 (6)	241 (5)	1603 (13)	560 (13)	329 (3)	123 (3)

ECR: Eindhoven cancer registry.
NA: not applicable.
[a] Four most common types of cancer not accounting for basal cell carcinoma.

of cytostatics administered to a patient in-hospital were available for approximately 45% of the cancer patients in the linked ECR-PHARMO cohort.

Validation of a random sample of the linked ECR-PHARMO cohort resulted in 2887 true positive links, 42 false positive links, 51 false negative links and 9009 true negative links, yielding a sensitivity of 98.3% (95% confidence interval (CI): 97.7–98.7%) and a specificity of 99.5% (95% CI: 99.4–99.7%).

The cancer patients included in the linked ECR-PHARMO cohort were representative of the cancer patients included in the total ECR in the period 1998–2006 as shown in Table 1. The difference in percentage between these two cohorts (% total ECR minus% linked ECR-PHARMO cohort) showed that patients who were linked tended to be somewhat older, were diagnosed in the more recent years (2004–2006) and were diagnosed with the more common types of tumours. Gender, age and tumour stage distribution for the patients included in the breast, colorectal, lung and prostate cancer cohorts were similar for the total ECR and the linked ECR-PHARMO cohorts (Table 2).

Patients diagnosed with the four most common types of cancer included in the linked ECR-PHARMO cohort were followed before and after cancer diagnosis to assess patients' baseline covariate status before diagnosis and (adverse) events after diagnosis and treatment. History before cancer diagnosis was approximately 5 years and similar for all four cancer types, while follow-up after cancer diagnosis was dependent on patients' survival and ranged from a mean follow-up of 2 years for lung cancer patients to more than 4 years for breast cancer patients (Table 3).

Of the cancer patients included in the four major linked cancer cohorts who were registered 1 year or more in the PHARMO RLS before cancer diagnosis and with data from the clinical laboratory database available, co-morbidities at diagnosis and medication, hospitalisation and laboratory measurements in the 1 year before cancer diagnosis are presented in Table 4. For the most common co-morbidities: cardiovascular disease (including hypertension), diabetes and respiratory disease, the percentage of patients with these co-morbidities at diagnosis as registered in the ECR, and the percentage of patients using medications or who had been hospitalised for these co-morbidities as registered in the PHARMO RLS, were in the same range per type of co-morbidity. Moreover, to give an impression of the type of laboratory measurements available, the percentage of patients with laboratory measurements related to their co-morbidities in the year before cancer diagnosis are presented.

Of the cancer patients included in the four major linked cancer cohorts who were registered 1 year or more in the PHARMO RLS after cancer diagnosis and with data from the clinical laboratory database available, initial therapies, other received medications, hospitalisations and laboratory measurements are presented in Table 5. Type of cytostatics administered to a patient was available from the PHARMO RLS from in-hospital pharmacy data (in-patient) and the community pharmacy data (out-patient), for 41% (N = 465) of the breast cancer patients to 55% (N = 363) of the lung cancer patients who received chemotherapy. In Table 5 an overview of the most common types of cytostatics (categories of cytostatics) per cancer type are shown, not taking into account combinations of cytostatics used in the same treatment regimen. Next to the data on cancer treatment, the percentage of patients using medications or who had been hospitalised during the first year after cancer diagnosis for the three most com-

Table 3 – History before and follow-up after cancer diagnosis of patients included in the linked ECR-PHARMO cohorts and sub cohort in the period 1998–2006.

Characteristics	Breast (N = 5545)	Colon and rectum (N = 4804)	Lung, bronchus and trachea (N = 4395)	Prostate (N = 3835)
History (before cancer diagnosis)				
Mean ± SD (years)	4.9 ± 3.5	5.2 ± 3.6	5.1 ± 3.5	5.2 ± 3.5
Median (interquartile range, years)	4.6 (2.1–7.1)	4.9 (2.2–7.4)	4.9 (2.2–7.4)	4.9 (2.4–7.4)
Patients with ⩾1 year history[a] in total linked cohort	4687 (85%)	4153 (86%)	3776 (86%)	3324 (87%)
and with information from the clinical laboratory database available (sub cohort)[b]	3691 (67%)	3431 (71%)	3006 (68%)	2829 (74%)
Follow-up (after cancer diagnosis)				
Mean ± SD (years)	4.4 ± 2.6	3.4 ± 2.6	2.0 ± 2.2	3.9 ± 2.4
Median (interquartile range, years)	4.1 (2.3–6.3)	2.8 (1.3–5.2)	1.1 (0.5–2.6)	3.5 (1.9–5.6)
Patients with ⩾1 year follow-up[c] in total linked cohort	5164 (93%)	3910 (81%)	2403 (55%)	3541 (92%)
and with information from the clinical laboratory database available (sub cohort)[b]	4005 (72%)	3187 (66%)	1856 (42%)	2954 (77%)

[a] History was defined as time between the date of entering the PHARMO RLS to the date of cancer diagnosis.
[b] Information from the clinical laboratory database was available for approximately 80% of the patients in the total linked cohort.
[c] Follow-up was defined as the time between date of cancer diagnosis until end of data collection in the PHARMO RLS (i.e. the patient moves out of the PHARMO RLS catchment area), death, or end of the study period (31st December 2007), whichever occurred first.

Table 4 – Co-morbidities at diagnosis and medication, hospitalisation and laboratory measurements during the 1 year *before* diagnosis of patients with breast, colorectal, lung and prostate cancer, extracted from the linked ECR-PHARMO sub cohort in the period 1998–2006.[a]

Characteristics	Breast (*N* = 3691) *n* (%)	Colon and rectum (*N* = 3431) *n* (%)	Lung, bronchus and trachea (*N* = 3006) *n* (%)	Prostate (*N* = 2829) *n* (%)
Co-morbidity at diagnosis				
Data obtained from ECR				
Number of co-morbidities at diagnosis				
None	1781 (48)	1028 (30)	709 (24)	936 (33)
1	867 (24)	1012 (30)	949 (32)	870 (31)
⩾2	635 (17)	1090 (32)	1152 (38)	752 (27)
Unknown	408 (11)	301 (9)	196 (7)	271 (10)
Type of co-morbidities at diagnosis				
Cardiovascular disease	1030 (28)	1513 (44)	1333 (44)	1220 (43)
Diabetes	294 (8)	404 (12)	328 (11)	251 (9)
Respiratory disease	202 (5)	351 (10)	794 (26)	289 (10)
Medication and/or hospitalisations during 1 year of history				
Data obtained from PHARMO RLS				
Cardiovascular disease				
Hospitalisations	108 (3)	230 (7)	223 (7)	164 (6)
Medications	1586 (43)	1967 (57)	1663 (55)	1628 (58)
Hospitalisation and/or medication[b]	1606 (44)	1990 (58)	1691 (56)	1645 (58)
Diabetes				
Hospitalisations	5 (0.1)	8 (0.2)	6 (0.2)	6 (0.2)
Medications	273 (7)	365 (11)	273 (9)	245 (9)
Hospitalisation and/or medication[b]	274 (7)	367 (11)	276 (9)	248 (9)
Respiratory disease				
Hospitalisations	21 (0.6)	54 (2)	165 (6)	49 (2)
Medications	374 (10)	479 (14)	1026 (34)	424 (15)
Hospitalisation and/or medication[b]	382 (10)	500 (15)	1080 (36)	442 (16)
Clinical laboratory measurements during 1 year of history				
Data obtained from PHARMO RLS				
Cardiovascular related measurements, e.g.				
LDL cholesterol	688 (19)	940 (27)	899 (30)	938 (33)
HDL cholesterol	765 (21)	1024 (30)	972 (32)	1038 (37)
Diabetes related measurements, e.g.				
HbA1c, blood glucose	421 (11)	560 (16)	451 (15)	440 (16)

ECR: Eindhoven cancer registry; SD: standard deviation; LDL: low density lipoprotein; HDL: high density lipoprotein; HbA1c: glycated haemoglobin.

[a] Patients with at least 1 year of history, defined as time between the date of entering the PHARMO RLS to the date of cancer diagnosis.

[b] Numbers do not add up as hospitalised patients mostly also have received medications.

mon co-morbidities are shown. Moreover, an example is shown of the type of laboratory measurements patients had related to patients' cancer diagnosis and co-morbidities during the 1 year follow-up.

4. Discussion

In this study, the ECR and the PHARMO RLS were linked creating a novel ECR-PHARMO population-based and patient-centric cohort that provides an excellent opportunity to gain real-life insights into drug utilisation (for both in- and out-patients), health resources utilisation and their costs. Moreover, the effectiveness and safety of pharmaceuticals in cancer patients outside the clinical trial setting can be assessed using hospitalisation data and clinical laboratory measurements. This linkage also enables research to study cancer as an adverse event of a wide range of pharmaceuticals. By widely communicating study results and providing feedback to clinicians on therapeutic choices made and subsequent outcomes, research performed with the ECR-PHARMO cohort can lead to a better understanding of the burden of cancer and offer potential for improvements in disease management in the field.

Of the 46,950 patients who were eligible for linkage, 85% could be accurately linked with an overall specificity of 99.5% and a sensitivity of 98.3%. The quality of the linkage is comparable to linkage based on a unique patient identifier. The 15% eligible ECR cancer patients who were not linked can be explained by the overestimation of the number of eligible cancer patients as it is extremely difficult to define the exact overlapping catchment area of the ECR and the PHARMO RLS:

Table 5 – Received treatments, medications, hospitalisations and laboratory measurements during the 1 year *after* diagnosis of patients diagnosed with breast, colorectal, lung and prostate cancer extracted from the linked ECR-PHARMO sub cohort in the period 1998–2006.[a]

Characteristics	Breast (*N* = 4005) *n* (%)	Colon and rectum (*N* = 3187) *n* (%)	Lung, bronchus and trachea (*N* = 1856) *n* (%)	Prostate (*N* = 2954) *n* (%)
Initial therapies				
Data obtained from ECR				
Received initial therapies				
Surgery	3828 (96)	2973 (93)	569 (31)	462 (16)
Radiation therapy	2767 (69)	822 (26)	757 (41)	1242 (42)
Hormonal therapy	1635 (41)	0 (–)	0 (–)	1383 (47)
Chemotherapy	1127 (28)	902 (28)	656 (35)	15 (0.5)
Of those who received chemotherapy, type of cytostatics received during 1 year of follow-up				
Data obtained from PHARMO RLS for a sub cohort [b,c]				
	(*N* = 465)	(*N* = 411)	(*N* = 363)	(*N* = 7)[d]
Pyrimidine analogues	304 (65)	387 (94)	151 (42)	–
Platinum compounds	0 (–)	174 (42)	271 (75)	–
Cyclophosphamide	425 (91)	4 (1)	70 (19)	–
Anthracyclines	359 (77)	0 (–)	67 (18)	–
Etoposide	0 (–)	0 (–)	136 (37)	–
Methotrexate	78 (17)	0 (–)	0 (–)	–
Monoclonal antibodies	31 (7)	33 (8)	0 (–)	–
Taxanes	44 (9)	0 (–)	36 (10)	–
Irinotecan	0 (–)	58 (14)	0 (–)	–
Vinca alkaloids	0 (–)	0 (–)	18 (5)	–
Medication and/or hospitalisations during 1 year of follow-up				
Data obtained from PHARMO RLS				
Cardiovascular disease				
Hospitalisations	132 (3)	185 (6)	136 (7)	198 (7)
Medications	1821 (45)	1896 (59)	1122 (60)	1824 (62)
Hospitalisation and/or medication[e]	1842 (46)	1922 (60)	1136 (61)	1843 (62)
Diabetes				
Hospitalisations	6 (0.1)	4 (0.1)	3 (0.2)	1 (<0.1)
Medications	317 (8)	340 (11)	175 (9)	285 (10)
Hospitalisation and/or medication[e]	319 (8)	340 (11)	175 (9)	285 (10)
Respiratory disease				
Hospitalisations	42 (1)	52 (2)	197 (11)	58 (2)
Medications	405 (10)	409 (13)	779 (42)	460 (16)
Hospitalisation and/or medication[e]	434 (11)	430 (13)	857 (46)	478 (16)
Clinical laboratory measurements during 1 year of follow-up				
Data obtained from PHARMO RLS				
Tumour markers, e.g.				
PSA	NA	392 (12)	212 (11)	2528 (86)
CEA	476 (12)	1862 (58)	92 (5)	82 (3)
CA 15.3	412 (10)	15 (0.5)	10 (0.5)	NA
Blood level/count, e.g.				
Hemoglobin	3502 (87)	2986 (94)	1631 (88)	2161 (73)
Leukocytes	2807 (70)	2804 (88)	1578 (85)	1665 (56)
Thrombocytes	2307 (58)	2449 (77)	1451 (78)	1053 (36)
Liver enzymes, e.g.				
ALT	2512 (63)	2538 (80)	1450 (78)	1172 (40)
AST	2102 (52)	2334 (73)	1368 (74)	909 (31)
ALP	1545 (39)	1798 (56)	1066 (57)	924 (31)
GGT	1811 (45)	2014 (63)	1074 (58)	734 (25)
LDH	684 (17)	717 (22)	409 (22)	181 (6)

ECR: Eindhoven cancer registry; SD: standard deviation; NA: not applicable; PSA: prostate specific antigen; CEA: carcinoembryonic antigen; CA: carcinogen antigen; ALT: alanine aminotransferase; AST: aspartate aminotransferase; ALP: alkaline phosphatase; GGT: gamma glutamyl transpeptidase; LDH: lactate dehydrogenase.

Table 5 – (*Continued*)

[a] Patients with at least 1 year of follow-up, defined as the time between date of cancer diagnosis until end of data collection in the PHARMO RLS (i.e. the patient moves out of the PHARMO RLS catchment area), death, or end of the study period (31st December 2007), whichever occurred first.

[b] For a sub cohort of approximately 45% of the patients, both in-hospital pharmacy (e.g. cytostatics administered intravenously) and community pharmacy (e.g. cytostatics administered orally) data available.

[c] Numbers do not add up as patients may have received multiple cytostatics, pyrimidine analogues included: fluorouracil, gemcitabine and capecitabine; platinum compounds included: oxaliplatin (colorectal cancer) and cisplatin (lung cancer); anthracyclines included: doxorubicin and epirubicin; monoclonal antibodies included: trastuzumab (breast cancer) bevacizumab (colorectal cancer) and cetuximab (colorectal cancer); taxanes included: docetaxel and paclitaxel.

[d] Or further specified, due to low numbers.

[e] Numbers do not add up as hospitalised patients mostly also have received medications.

eligible ECR patients might not turn up in the PHARMO RLS when (1) they migrate out of the catchment area, (2) are institutionalised or (3) fill their prescriptions in pharmacies outside the catchment area of the PHARMO RLS while they are living inside the defined catchment area. Next to eligibility, patients might be lost in the blocking step of the linkage as the blocking variables date of birth and gender are subject to recording errors, which will directly lead to a non-link. This is not the case with the variables used in the probabilistic matching step, however, in this step it is inevitable that there is some degree of uncertainty by which true links might have been missed.

Currently, cancer patients included in the cohorts had a mean history of 5 years and a mean follow-up ranging from 2 to more than 4 years (dependent on the survival rate of the specific cancer type) to study utilisation-related outcomes, such as medication, clinical laboratory measurements and (re)hospitalisation for complications following cancer diagnosis and treatment. This follow-up period will increase for the cancer survivors as both the ECR and the PHARMO RLS are prospective data collection systems. The longitudinal and patient-centric nature of these data is a major benefit, because they allow for evaluation of health care utilisation both before and after cancer diagnosis, as well as enable evaluation of long-term outcomes often missed in clinical trials.

The ECR and the PHARMO RLS currently collect data of more than 2 and 3 million inhabitants, respectively. However, only 45% of the catchment areas (approximately 1 million inhabitants) overlap. The size of the linked ECR-PHARMO cohort can be doubled in the future by either expanding the PHARMO RLS in the ECR region or by linkage of the PHARMO RLS to other cancer registries in the Netherlands.

At the time of this study, recruitment of all clinical laboratories in the ECR-PHARMO region was not yet completed, resulting in clinical laboratory findings for around 80% of all patients. Moreover, mid-2009, the recruitment process of in-hospital pharmacy data (including intravenous chemotherapy data) was still going on. The availability of cytostatics data was dependent on conditions such as when and in which hospital the patient was treated and whether there was access to data from the in-hospital pharmacy in that specific time period. Recruitment of in-hospital pharmacies supplying the hospitals in the ECR-PHARMO region to obtain information on specific type of cytostatics administered intravenously was for about 45% complete in the study period 1998–2006. In this manuscript, a description of the data becoming available after linkage of two existing databases, and the most common categories of cytostatics per cancer type were shown. For a specific study using these data, one could present the various cancer treatments by tumour stage or stratify treatment regimens in first, second and third line therapy as was performed in a previous study in breast cancer patients.[25]

The availability to select a cancer-free control group using data from the PHARMO RLS enables comparative studies between cancer patients and their cancer-free controls that would not be possible if there were only data on cancer patients. For example, the controls can be used to compare the rate of adverse events in cancer patients following cancer treatment with the rate of similar adverse events found in the general population as identified from the controls. These comparisons are of major importance as they allow quantification of the occurrence of adverse effects and accompanying risk factors.

Finally, additional information on quality of life measurements, information on health beliefs or health behaviours, psychological measurements, detailed patient characteristics, tumour biomarkers, disease progression and diagnostics not standard collected via the ECR or the PHARMO RLS can be collected additionally via medical files or questionnaires by going back to the original data sources or the patient.

In conclusion, the results of our study show that the linkage of data from the ECR and the PHARMO RLS is feasible and accurate; yielding a new database that facilitates broader research possibilities. The size of the cohort is still relatively limited and expansion is needed to facilitate the study of more rare cancers. However, it offers a unique opportunity for detailed pharmacoepidemiologic cancer-related treatment and outcomes research, eventually allowing improvement in future patient care.

Conflict of interest statement

No conflict of interest declared for the authors Dr. L.V. van de Poll-Franse, Dr. V.E.P.P. Lemmens, Dr. G. Vreugdenhil, Dr. J.F.M. Pruijt and Prof. Dr. J.W.W. Coebergh. Mrs. M.P.P. van Herk-Sukel and Dr. R.M.C. Herings are employees of the PHARMO Institute for Drug Outcomes Research. This research institute performs financially supported studies for several pharmaceutical companies. This study however, was performed as

part of the PhD programme of Mrs. M.P.P. van Herk-Sukel, initiated by Dr. L.V. van de Poll-Franse and Prof. Dr. J.W.W. Coebergh of the Comprehensive Cancer Centre South.

Acknowledgements

The authors would like to thank the participating pharmacies for the opportunity to validate ECR-PHARMO linkage. The authors thank Mark AGM Tinga, PhD [employee of the PHARMO Institute for Drug Outcomes Research], for performing the medical record linkage. This work was partially supported by the Netherlands organization for health research and development [ZonMw 80-82500-98-8227].

REFERENCES

1. Ferlay J, Autier P, Boniol M, et al. Estimates of the cancer incidence and mortality in Europe in 2006. *Ann Oncol* 2007;**18**:581–92.
2. Harpe SE. Using secondary data sources for pharmacoepidemiology and outcomes research. *Pharmacotherapy* 2009;**29**:138–53.
3. Ladewski LA, Belknap SM, Nebeker JR, et al. Dissemination of information on potentially fatal adverse drug reactions for cancer drugs from 2000 to 2002: first results from the research on adverse drug events and reports project. *J Clin Oncol* 2003;**21**:3859–66.
4. Viale PH, Moore S. Postmarketing surveillance for oncology drugs. *Clin J Oncol Nurs* 2008;**12**:877–86.
5. Alsheikh-Ali AA, Trikalinos TA, Kent DM, Karas RH. Statins, low-density lipoprotein cholesterol, and risk of cancer. *J Am Coll Cardiol* 2008;**52**:1141–7.
6. Zhao YS, Zhu S, Li XW, et al. Association between NSAIDs use and breast cancer risk: a systematic review and meta-analysis. *Breast Cancer Res Treat* 2009;**117**:141–50.
7. Ayanian JZ, Jacobsen PB. Enhancing research on cancer survivors. *J Clin Oncol* 2006;**24**:5149–53.
8. Schrag D. Enhancing cancer registry data to promote rational health system design. *J Natl Cancer Inst* 2008;**100**:378–9.
9. Herings RMC. The PHARMO system. *CNS Spectr* 2007;**12**:14–7.
10. Houterman S, Janssen-Heijnen ML, van de Poll-Franse LV, Brenner H, Coebergh JW. Higher long-term cancer survival rates in southeastern Netherlands using up-to-date period analysis. *Ann Oncol* 2006;**17**:709–12.
11. Janssen-Heijnen ML, Houterman S, Lemmens VE, et al. Prognosis for long-term survivors of cancer. *Ann Oncol* 2007;**18**:1408–13.
12. Herings RMC. *PHARMO: a record linkage system for postmarketing surveillance of prescription drugs in The Netherlands*. Thesis [Doctorate]. Utrecht: Utrecht University; 1993.
13. Herings RMC, Bakker A, Stricker BH, Nap G. Pharmaco-morbidity linkage: a feasibility study comparing morbidity in two pharmacy based exposure cohorts. *J Epidemiol Community Health* 1992;**46**:136–40.
14. Herings RMC, Stricker BH, Bakker A. Pharmaco-morbidity linking: a pilot study of its technical possibilities in The Netherlands. *Ned Tijdschr Geneeskd* 1990;**134**:1903–7.
15. Stricker BH, Herings RM, Bakker A, Valkenburg HA, Sturmans F. Pharmaco-morbidity linking: a potential instrument for post-marketing surveillance. *Ned Tijdschr Geneeskd* 1990;**134**:1886–9.
16. Blakely T, Salmond C. Probabilistic record linkage and a method to calculate the positive predictive value. *Int J Epidemiol* 2002;**31**:1246–52.
17. Meray N, Reitsma JB, Ravelli AC, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. *J Clin Epidemiol* 2007;**60**:883–91.
18. Newcombe HB. *Handbook of record linkage: methods for health and statistical studies, administration, and business*. Oxford: Oxford University Press; 1988.
19. Howe GR. Use of computerized record linkage in cohort studies. *Epidemiol Rev* 1998;**20**:112–21.
20. Fenna D. Phonetic reduction of names. *Comput Programs Biomed* 1984;**19**:31–6.
21. The Good Conduct code [Code Goed Gedrag. Gedragscode Gezondheidsonderzoek]. Rotterdam: FMWV [Federation of Medical Research Associations]. <http://www.federa.org/>; 2005.
22. Dal Maso L, Braga C, Franceschi S. Methodology used for "software for automated linkage in Italy" (SALI). *J Biomed Inform* 2001;**34**:387–95.
23. Roos LL, Walld R, Wajda A, Bond R, Hartford K. Record linkage strategies, outpatient procedures, and administrative data. *Med Care* 1996;**34**:570–82.
24. de Vries E, Louwman M, Bastiaens M, de Gruijl F, Coebergh JW. Rapid and continuous increases in incidence rates of basal cell carcinoma in the southeast Netherlands since 1973. *J Invest Dermatol* 2004;**123**:634–8.
25. Sukel MP, Breekveldt-Postma NS, Erkens JA, et al. Incidence of cardiovascular events in breast cancer patients receiving chemotherapy in clinical practice. *Pharmacoepidemiol Drug Saf* 2008;**17**:125–34.