ORIGINAL ARTICLE

European Journal of Cancer Care WILEY

A population-based linked cohort of cancer and primary care data: A new source to study the management of cancer in primary care

Josephina G. Kuiper[1,2] | Myrthe P. P. van Herk-Sukel[3] |
Valery E. P. P. Lemmens[2,4] | Ernst J. Kuipers[5] | Ron M. C. Herings[1,6]

[1]Research, PHARMO Institute for Drug Outcomes Research, Utrecht, The Netherlands

[2]Department of Public Health, Erasmus University Medical Center, Rotterdam, The Netherlands

[3]Department of Internal Medicine and Dermatology, University Medical Center Utrecht, Utrecht, The Netherlands

[4]Department of Research and Development, Netherlands Comprehensive Cancer Organisation, Utrecht, The Netherlands

[5]Department of Gastroenterology and Hepatology, Erasmus University Medical Center, Rotterdam, The Netherlands

[6]Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands

Correspondence
Josephina G. Kuiper, PHARMO Institute for Drug Outcomes Research, van Deventerlaan 30-40, 3528 AE Utrecht, The Netherlands.
Email: josine.kuiper@pharmo.nl

Abstract

Objective: Insight into the management of cancer in the primary care setting is pivotal to improve early recognition and survival of cancer patients. Therefore, the Netherlands Cancer Registry (NCR) was linked to the General Practitioner (GP) Database of the PHARMO Database Network to make this research possible.

Methods: The NCR collects tumour data on all newly diagnosed cancer patients, whereas the GP Database comprises data from electronic patient records registered by GPs. Databases were linked using a probabilistic record linkage technology.

Results: Through record linkage of the NCR and the GP Database, we have established a large population-based cohort (NCR-PHARMO GP cohort) of 135,868 cancer patients. Data are available on demographics, tumour characteristics, primary health care use before and after cancer diagnosis including medication use, medical conditions, laboratory tests, and referrals. Data can be used for a number of different studies, for example, to study the diagnostic pathway in the primary care setting in order to identify possibilities for early recognition.

Conclusion: The NCR-PHARMO GP cohort provides rich data on the primary care management of cancer facilitating large-scale observational cancer research in the primary care setting. The patient-level linkage allows for long-term follow-up of cancer patients, with ongoing annual updates.

KEYWORDS
cancer management, general practitioner, NCR-PHARMO GP cohort, primary care

1 | INTRODUCTION

Globally, cancer is one of the leading causes of death and its incidence is still increasing (Global Burden of Disease Cancer Collaboration, 2017). In the Netherlands and other countries, the general practitioner (GP) is the first point of contact for those who experience health problems. GPs play a key role in the patient journey and are the gatekeepers to specialised cancer care (Knottnerus & ten Velden, 2007). Often, time to diagnosis and time in primary care is too long; in the Netherlands this is true for 10 to 25% of cancer patients which could lead to a potential unfavourable treatment outcome {Helsper, 2017 #26}. Research with the aim to improve early

recognition and survival is therefore important, but up to now a large population-based database facilitating this research is hardly available. Linkages that have been established with cancer registries in Europe are mainly with drug registries (Furu, 2009; Henson et al., 2018; Pottegard et al., 2017; Teppo et al., 1994; van Herk-Sukel et al., 2010; Wettermark et al., 2007). This allows research on, for example drug, exposure-cancer associations but will not make it possible to study the cancer management in the primary care setting. In the Netherlands the GP Database of the PHARMO Database Network contains information relevant to the patients' primary care pathway before and after diagnosis and referrals to specialists. The Netherlands Cancer Registry (NCR) contains information on the actual diagnosis of cancer and specific tumour and treatment characteristics. We linked these to databases and created the NCR-PHARMO GP cohort with the aim to provide a valid source for future studies on the management of cancer in the primary care setting with annual updates. In this paper, we describe both databases, the linking process, the resulting cohort, data access, and provide an example of the utilisation of the data.

2 | METHODS

2.1 | Setting

The NCR-PHARMO GP cohort is a linkage of the NCR and the GP Database of the PHARMO Database Network. With data collection starting in 2006, the linkage of these population-based data sources facilitates large-scale observational cancer research in the primary care setting. It contains detailed information on the cancer itself (tumour characteristics), the diagnostic pathway in the primary care setting, and long-term follow-up and outcomes after cancer diagnosis, with ongoing annual updates of the routinely collected data.

2.2 | Data sources

The PHARMO Database Network is a Dutch population-based network of electronic health care databases which contains data linked on a patient level from general practitioners, out-patient and in-patient pharmacies, hospitals, and clinical laboratories. Data are linked by the "Stichting Informatievoorziening voor Zorg en Onderzoek" (STIZON). STIZON is an ISO/IEC 27001 and NEN 7510 certified foundation, compliant to data protection regulations. Part of the PHARMO Database Network is a GP database. Records include information on diagnoses and symptoms, laboratory test results, referrals to specialists, type of drug prescriptions, their date; strength; dosage regimen; quantity; and route of administration. Data are available from 2006 onwards and covers a catchment area representing 3.2 million residents (∼20% of the Dutch population). Diagnoses and symptoms are coded according to the International Classification of Primary Care (ICPC) and mapped to ICD codes or entered as free text

("International Classification of Primary Care (ICPC)," https://www.nhg.org/themas/artikelen/icpc). Prescriptions are coded according to the WHO Anatomical Therapeutic Chemical (ATC) Classification System.("WHO Anatomical Therapeutic Chemical Classification System," [https://www.whocc.no/atc_ddd_index]).

The NCR is a unique population-based registry maintained by the Netherlands Comprehensive Cancer Organisation (IKNL) and contains individual patient data from all Dutch hospitals and radiotherapy institutes. The NCR contains data on cancer diagnosis, tumour stage, WHO ICD-O3 classification of tumour site (topography) and morphology (histology), and treatments received.

2.3 | Data linkage

The GP Database and the NCR do not contain unique patient identifiers. Linkage performed by STIZON was therefore based on probabilistic record linkage technology. Identified were patients with a similar NCR and GP database zip code and cancer diagnosis between 2006 and 2014. Linkage was based on first name (first initial), last name (first letter), zip code (first four characters), and singular variables. The latter represent a logical relation between two records, that is, an extra weight was assigned if a patient also had a GP recorded cancer diagnosis. The linkage process resulted in a cohort (NCR-PHARMO GP cohort) that is depleted of patient identifiable information. Details on the methodology of the used record linkage method can be found in the Supporting Information.

Linkage of both databases will be repeated annually to increase individual follow-up periods and capture new cancer patients.

2.4 | Data access

Access to the NCR-PHARMO GP cohort is granted by the independent Compliance Committees of the PHARMO Institute and the IKNL and based on Dutch legislation. Users only have access to anonymized data via a remote server in accordance with the General Data Protection Regulation (GDPR). Request for access can be made by academic affiliates via the PHARMO or IKNL websites (https://www.pharmo.com, www.iknl.nl).

3 | RESULTS

3.1 | The NCR-PHARMO GP cohort

Of the 957,401 patients diagnosed with cancer identified in the total NCR, 135,868 were linked to the GP Database of the PHARMO Database Network and included in the NCR-PHARMO GP cohort. The NCR-PHARMO GP cohort covers a catchment area of approximately 3.2 million inhabitants (∼20% of the Dutch population). This cohort currently includes 135,868 patients diagnosed with cancer in the period 2006–2014 and enables follow-up in the primary care setting

TABLE 1 Key information available of patients included in the NCR-PHARMO GP cohort

	Variable description	Explanation
Netherlands Cancer Registry	*Cancer details*	
	Unique patient identifier	Unique patient identifier, which corresponds to the patient identifier in the GP Database
	Date of diagnosis	Month-year of cancer diagnosis[a]
	Tumour staging	Tumour staging according to the TNM-classification developed and maintained by the Union for International Cancer Control (UICC, [https://www.uicc.org])
	Tumour site	Topography
	Morphology	Histology
GP Database	*Patient details*	
	Unique patient identifier	Unique patient identifier, which corresponds to the patient identifier in the NCR
	Year of birth	Patient birth year
	Diagnoses details	
	GP recorded diagnosis/symptom	Diagnoses and symptoms are coded according to the International Classification of Primary Care (ICPC, [https://www.nhg.org]), which can be mapped to ICD codes, but can also be entered as free text.
	Date of diagnosis/symptom	Month-year of GP recorded diagnosis/symptom[a]
	Type of consult	The type of GP consultation (GP home visit, office visit or telephone contact)
	Referral	Information on referrals by a GP to a specialist, including type of specialist
	Prescription details	
	Date of prescription	Month-year of prescribed medication[a]
	ATC code	WHO-defined Anatomical Therapeutical Chemical code
	Label description	Product name
	Amount	Number of tablets/units per package
	Unit	Unit of the prescribed drug
	Dose	Number of tablets/units prescribed
	Strength	Numerical strength per tablet/unit
	Examination details	
	Date of testing	Month-year of testing[a]
	Description test	Type of test performed
	Test result	Result of the test
	Unit	Unit of the test result

[a]Due to privacy regulations only month-year of dates are made available for research purposes.

currently until December 2017. Table 1 contains a list of the key data in the NCR-PHARMO GP cohort.

In the Netherlands, the incidence of cancer is approximately 550 per 100,000 inhabitants per year. The coverage of the GP Database is approximately 3.2 million inhabitants, so the expected number of cancer patients was approximately 158,400 in a period of 9 years. The incidence may fluctuate over the years and not all patients could be matched, so the total number of 135,868 cancer

TABLE 2 Representativeness of the linked NCR-PHARMO GP cohort compared to the total NCR population for patients diagnosed with a primary tumour in the period 2006–2014

Characteristics	Total NCR population N = 957,401 n (%)	Linked NCR-PHARMO GP cohort N = 135,868 n (%)	Difference (% total NCR population minus % linked NCR-PHARMO GP cohort) %
Gender			
Male	493,254 (51.5)	69,497 (51.2)	−0.3
Female	464,147 (48.5)	66,371 (48.8)	+0.3
Age at tumour diagnosis			
<35	24,315 (2.5)	3242 (2.4)	−0.1
35–49	87,193 (9.1)	13,422 (9.9)	+0.8
50–59	151,003 (15.8)	22,367 (16.5)	+0.7
60–69	265,415 (27.7)	39,527 (29.1)	+1.4
70–79	263,289 (27.5)	37,531 (27.6)	+0.1
80–89	147,337 (15.4)	17,945 (13.2)	−2.2
≥90	18,849 (2.0)	1834 (1.3)	−0.7
Mean (±SD)	66.3 (±14.1)	65.6 (±13.7)	
Year of diagnosis			
2006–2008	290,349 (30.3)	36,681 (27.0)	−3.3
2009–2011	326,298 (34.1)	46,729 (34.4)	+0.3
2012–2014	340,754 (35.6)	52,458 (38.6)	+3.0
Common cancer sites			
Breast	133,912 (14.0)	16,933 (12.5)	−1.5
Colon & rectum	118,014 (12.3)	14,205 (10.5)	−1.8
Lung, bronchus	110,420 (11.5)	12,253 (9.0)	−1.5
Prostate	94,614 (9.9)	11,397 (8.4)	−1.5
Haematolymphopetic	74,160 (7.7)	8289 (6.1)	−1.6
Skin, other	72,346 (7.6)	8448 (6.2)	−1.4
Skin, melanoma	45,080 (4.7)	5464 (4.0)	−0.7
Urinary bladder	26,010 (2.7)	2925 (2.2)	−0.5
Kidney	19,846 (2.1)	2463 (1.8)	−0.3
Pancreas	19,334 (2.0)	1976 (1.5)	−0.5
Oesophagus	17,503 (1.8)	1884 (1.4)	−0.4
Unknown primary site	17,103 (1.8)	1758 (1.3)	−0.5
Stomach	16,367 (1.7)	1872 (1.4)	−0.3
Uterus	16,306 (1.7)	1824 (1.3)	−0.4

Abbreviations: GP, General practitioner; NCR, Netherlands Cancer Registry; SD, standard deviation.

patients in the linked NCR-PHARMO GP cohort is in line with expectations. Patients in the linked NCR-PHARMO GP cohort were representative for the cancer patients included in the total NCR (Table 2). The difference in percentage between these two populations showed that patients who were linked tended to be somewhat younger (65.6 years vs. 66.3 years) and were more often diagnosed in the period 2012–2014 (38.6% vs. 35.6%). The most common cancer site was breast cancer, followed by colorectal cancer, lung cancer, and prostate cancer.

3.2 | Illustration of available data

As an example, the primary health care use before a diagnosis of colorectal cancer patients have been assessed. A total of 14,205 patients with colorectal cancer were identified of whom 68.0% were diagnosed with colon cancer, 29.5% with rectum cancer, and 2.5% with rectosigmoid cancer (Table 3). These patients were representative for the colorectal cancer patients included in the total NCR population. The difference in percentage between these two

TABLE 3 Representativeness of colorectal patients in the total NCR population compared to colorectal cancer patients in the NCR-PHARMO GP cohort

Characteristics	Colorectal cancer patients in the total NCR population (N = 118,014) n (%)	Colorectal cancer patients in the NCR-PHARMO GP cohort (N = 14,205) n (%)	Difference (% total NCR population minus % linked NCR-PHARMO GP cohort) %
Gender			
Male	64,541 (54.7)	7937 (55.9)	+1.2
Female	53,473 (45.3)	6268 (44.1)	−1.2
Age at tumour diagnosis			
<35	510 (0.4)	57 (0.4)	0.0
35–49	5412 (4.6)	692 (4.9)	+0.3
50–59	15,053 (12.8)	1893 (13.3)	+0.5
60–69	32,963 (27.9)	4228 (29.8)	+1.9
70–79	38,440 (32.6)	4737 (33.3)	+0.7
80–89	23,238 (19.7)	2408 (17.0)	+2.7
≥90	2398 (2.0)	190 (1.3)	−0.7
Mean (±SD)	69.9 (11.5)	69.1 (11.2)	
Tumour stage			
I	21,012 (17.8)	2654 (18.7)	+0.9
II	31,445 (26.6)	3828 (26.9)	+0.3
III	34,339 (29.1)	4145 (29.2)	+0.1
IV	25,839 (21.9)	3061 (21.5)	−0.4
Unknown	5379 (4.6)	517 (3.6)	−1.0
Tumour site			
Colon	80,654 (68.3)	9663 (68.0)	−0.3
Rectum	33,301 (28.2)	4192 (29.5)	+1.3
Rectosigmoid	4059 (3.4)	350 (2.5)	−0.9

Abbreviations: NCR, Netherlands Cancer Registry; SD, standard deviation.

populations showed that patients who were linked tended to be slightly more often male (55.9% vs. 54.7%) and somewhat younger (69.1 years vs. 69.9 years).

Within a year before diagnosis, patients diagnosed with colorectal cancer at age 85 years or older consulted their GP more frequently than those diagnosed at a younger age (Figure 1a). Especially among the younger population, the proportion of GP consultations especially started to increase from 4 months before colorectal diagnosis, which was also seen when stratifying by GP home or office visits and GP phone consultations (Figure 1b–d).

It was also seen that the use of beta-blockers, oral antidiabetics, proton pump inhibitors, and statins increased with increasing age, while for nonsteroidal anti-inflammatory drugs, the use decreased with increasing age (Figure 2).

Future research will compare these results with a cancer-free control population which is possible to select from the PHARMO Database Network. This enables comparative studies between cancer patients and their cancer-free controls. Thereby, also other patient characteristics, such as less severe comorbidities and symptoms or laboratory measurements, are available in the NCR-PHARMO GP cohort which can be taken into account when

studying the relationship between the use of specific drugs and cancer risk.

4 | DISCUSSION

Through record linkage of the NCR and the GP Database of the PHARMO Database Network, we have established a large population-based and patient-centric cohort including data on demographics, tumour characteristics, primary health care use before and after cancer diagnosis including medication use, medical conditions, laboratory test results, and referrals to specialised care. This cohort provides excellent opportunities to gain insights in the management of cancer in the primary care setting.

To our knowledge, there are limited large population-based databases available in Europe with such detailed information on the tumour itself and the diagnostic pathway and management of cancer survivors in the primary care setting as the NCR-PHARMO GP cohort. One similar resource is available in England in which a subset of the UK Clinical Practice Research Datalink (CPRD) is linked to cancer registry data (Herrett et al., 2015). Similar to our GP Database, the CPRD
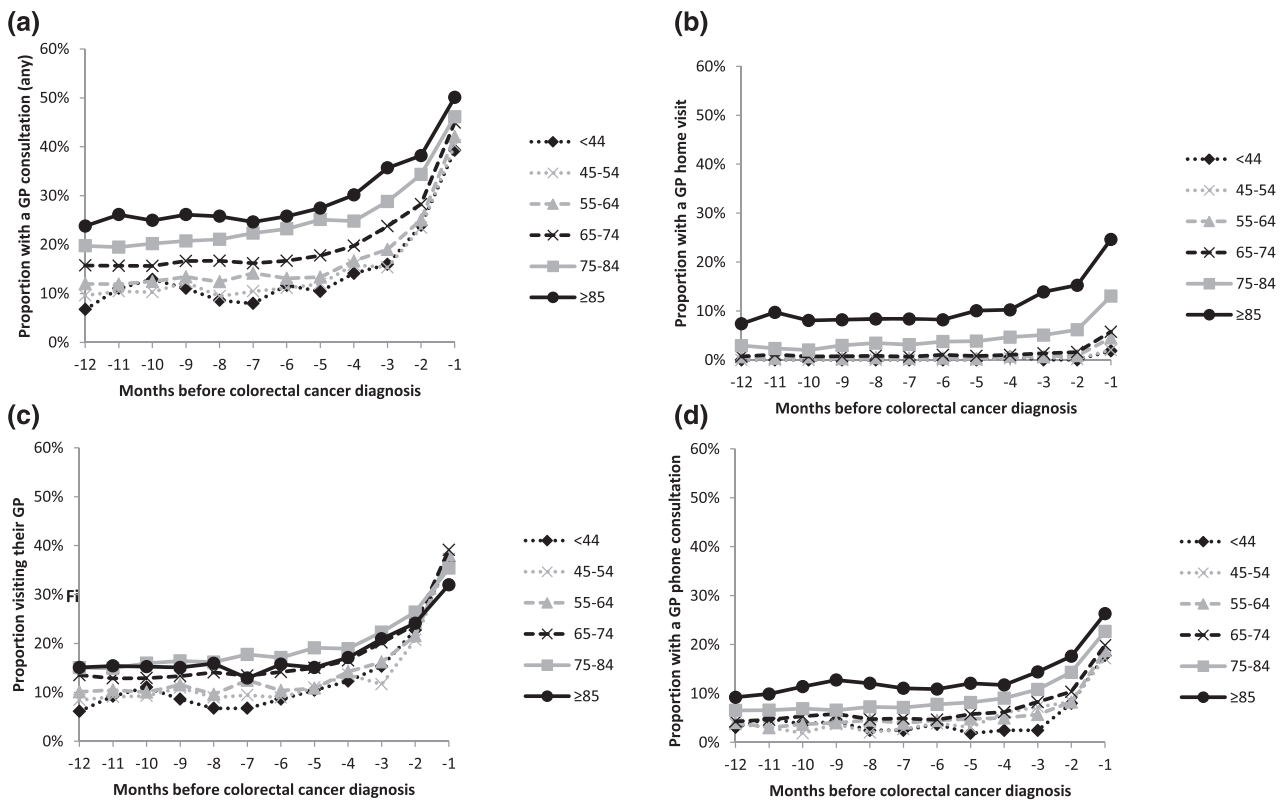
FIGURE 1 Proportion of colorectal cancer patients in the NCR-GP-CRC cohort with a (a) declared GP consultation (any), (b) GP home visit, (c) GP office visit, (d) GP phone consultation in the year prior to colorectal cancer diagnosis, stratified by age at diagnosis
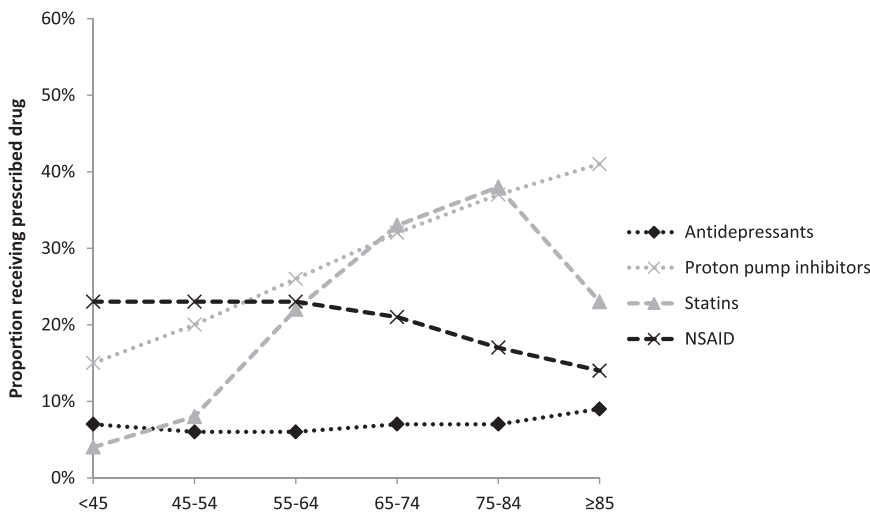
FIGURE 2 Proportion of colorectal cancer patients receiving a prescribed drug in the year prior to colorectal cancer diagnosis, stratified by age at diagnosis

also records demographics, diagnoses, symptoms, signs, prescriptions, referrals, and tests. The free text notes by the GP is not part of the standard CPRD database available to researchers as this may contain identifiable information. The NCR-PHARMO GP cohort contains both the diagnosis codes and free text which are available for research purposes (taking into account the privacy regulations). This allows a more precise estimation of less severe symptoms and comorbidities not accompanied with a code.

Previous studies that investigated the management of cancer in the primary care setting have established project specific linkages between cancer registries and primary care data to perform their research. The database resulting from this linkage often only

includes the cancer type of interest, is not accessible for other researchers, and is not updated annually. The NCR-PHARMO GP cohort is unique in its way that data are not limited to one type of cancer but includes all types of cancers registered in the NCR. Furthermore, access to data can be granted to broaden the use of the data for scientific research on the management of cancer in the primary care setting. This will provide more knowledge on this topic and eventually improvement in the quality of cancer care in the primary setting. Thereby, data are updated annually with more individual follow-up periods and new cancer patients making it a dynamic cohort.

It is crucial to have information from cancer registries when studying the management of cancer in the primary care setting as studies on cancer solely relying on primary care data may lead to uncertainties and potential misclassification bias of cancer recorded by the GP. In a previous study performed in the Netherlands, the quality of cancer patients recorded in primary care was assessed and showed that over 60% of the cancer patient are coded in concordance with the NCR (Sollie et al., 2016). Almost 40% of cancer cases registered in the NCR were missing in the electronic health care records of the GP. Including patients without cancer (false positives) and missing real cancer cases (false negatives) will introduce bias in research. It is therefore crucial to use linked primary care and cancer registry data to verify the cancer diagnosis. Thereby, details on, for example, tumour characteristics are often missing in primary care data as these details are hardly recorded by general practitioners, but these details are available in cancer registries.

There is a limitation that needs to be mentioned. Identification of medical events is limited to data that are captured as part of medical records, as with any database study. These data are not primarily collected for research purposes and rely on appropriate diagnostic coding procedures to detect these events. However, through the collaboration with STIZON, in its role as a trusted third party, it is possible to go back to the health care provider or patients for validation or extension of the retrospectively collected data by chart review or additional data collection.

In conclusion, this unique research platform makes it possible to study the current and potential position of the GP in cancer management and eventually allowing improvement in the quality of cancer care in the primary setting through a more effective allocation of care between the primary and secondary health care services.

CONFLICT OF INTEREST
Josephina G. Kuiper and Ron C.M. Herings are employees of the PHARMO Institute for Drug Outcomes Research. This independent research institute performs financially supported studies for government and related health care authorities and several pharmaceutical companies. Myrthe P.P. van Herk-Sukel, Valery E.P.P. Lemmens, and Ernst J. Kuipers declare that they have no conflict of interest.

ORCID
Josephina G. Kuiper https://orcid.org/0000-0002-4487-3546

REFERENCES

Furu, K. (2009). Establishment of the nationwide Norwegian Prescription Database (NorPD)—New opportunities for research in pharmacoepidemiology in Norway. *Norsk Epidemiologi*, 18(2), 129–36.

Global Burden of Disease Cancer Collaboration. (2017). Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: A systematic analysis for the global burden of disease study. *JAMA Oncology*, 3(4), 524–548. https://doi.org/10.1001/jamaoncol.2016.5688

Helsper, C. C. W., van Erp, N. N. F., Peeters, P. P. H. M., & de Wit, N. N. J. (2017). Time to diagnosis and treatment for cancer patients in the Netherlands: Room for improvement?. *European Journal of Cancer*, 87, 113–121. https://doi.org/10.1016/j.ejca.2017.10.003. Epub 2017 Nov 13. PMID: 29145037.

Henson, K. E., Brock, R., Shand, B., Coupland, V. H., Elliss-Brookes, L., Lyratzopoulos, G., Godfrey, P., Haigh, A., Hunter, K., McCabe, M. G., Mitchell, G., Monckton, N., Robson, R., Round, T., Wong, K., & Rashbass, J. (2018). Cohort profile: Prescriptions dispensed in the community linked to the national cancer registry in England. *BMJ Open*, 8(7), e020980. https://doi.org/10.1136/bmjopen-2017-020980

Herrett, E., Gallagher, A. M., Bhaskaran, K., Forbes, H., Mathur, R., van Staa, T., & Smeeth, L. (2015). Data resource profile: Clinical Practice Research Datalink (CPRD). *International Journal of Epidemiology*, 44(3), 827–836. https://doi.org/10.1093/ije/dyv098

Knottnerus, J. A., & ten Velden, G. H. (2007). Dutch doctors and their patients—Effects of health care reform in the Netherlands. *The New England Journal of Medicine*, 357(24), 2424–2426. https://doi.org/10.1056/NEJMp0707383

Pottegard, A., Schmidt, S. A. J., Wallach-Kildemoes, H., Sorensen, H. T., Hallas, J., & Schmidt, M. (2017). Data resource profile: The Danish National Prescription Registry. *International Journal of Epidemiology*, 46(3), 798–798f. https://doi.org/10.1093/ije/dyw213

Sollie, A., Roskam, J., Sijmons, R. H., Numans, M. E., & Helsper, C. W. (2016). Do GPs know their patients with cancer? Assessing the quality of cancer registration in Dutch primary care: A cross-sectional validation study. *BMJ Open*, 6(9), e012669. https://doi.org/10.1136/bmjopen-2016-012669

Teppo, L., Pukkala, E., & Lehtonen, M. (1994). Data quality and quality control of a population-based cancer registry. Experience in Finland. *Acta Oncologica*, 33(4), 365–369. https://doi.org/10.3109/02841869409098430

van Herk-Sukel, M. P., van de Poll-Franse, L. V., Lemmens, V. E., Vreugdenhil, G., Pruijt, J. F., Coebergh, J. W., & Herings, R. M. (2010). New opportunities for drug outcomes research in cancer patients: The linkage of the Eindhoven Cancer Registry and the PHARMO Record Linkage System. *European Journal of Cancer*, 46(2), 395–404. https://doi.org/10.1016/j.ejca.2009.09.010

Wettermark, B., Hammar, N., Fored, C. M., Leimanis, A., Otterblad Olausson, P., Bergman, U., Persson, I., Sundström, A., Westerholm, B.,

& Rosén, M. (2007). The new Swedish prescribed drug register—Opportunities for pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiology and Drug Safety*, *16*(7), 726–735. https://doi.org/10.1002/pds.1294

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.